דוגמאות בנושא R^2

שאלה מס. 1
חוקר רצה לבדוק את הקשר בין הוצאות המשפחה על פירות וירקות ובין רמת ההכנסה.
נאספו נתונים על המשתנים הבאים:

- גודל המשפחה
- רמת ההכנסה
- הוצאות על פירות וירקות

הוא קיבל את הרגרסיה הבאה:
VEG= 40 + 8N + 0.01INC + e

- VEG- הוצאה על ירקות ופירות.
- N- מס' הנפשות במשק הבית.
- INC- הכנסה.
- Σyi^2= 2,030.
- Σei^2= 1,950.

נדרש:
1. מהו מקדם ההסבר? ועד כמה הרגרסיה מסבירה את ההוצאות על פירות וירקות?
2. אם הרגרסיה לא מסבירה היטב את ההוצאה על פירות וירקות הצע מודל חלופי?

פתרון:
1. $R^2= 1 - \Sigma ei^2 / \Sigma yi^2 = 1 – (1950/2030) = 0.039$ הרגרסיה מסבירה 3.9% מהגורמים שמשפיעים על ההוצאות על פירות וירקות.
2. המודל הבא הגיע ל-R^2=0.44 הסבר- הרגרסיה היתה:
VEG= a + b1*N + b2*INC + b3*PriceV + b4*PriceM + b5*Age + b6*M + b7*Level.

- N- מס' הנפשות במשק הבית. *השפעה חיובית.*
- INC- הכנסה. *השפעה חיובית.*
- PriceV- מחיר ירקות ופירות. *השפעה שלילית.*
- PriceM- מחיר בשר. *השפעה חיובית (מוצר תחליפי).*
- Age- ממוצע גילאי בעלי משק הבית (עורכי הקניות). *השפעה חיובית.*
- M- משקל ראש משק הבית. *לא ברורה ההשפעה (מי ששוקל הרבה- הסיבה היא שהוא לא אוכל פירות וירקות, מצד שני אם אני שוקלת הרבה אני עושה דיאטה ואז אני אוכלת יותר ירקות ופירות).*
- Level- רמת מודעות לבריאות. *השפעה חיובית.*

במודל המקדם b הוא חיובי אבל כאשר ההשפעה היא שלילית נצפה שהמקדם יהיה שלילי.

שאלה מס. 2

לפניך 2 טענות על R^2 חווה דעתך:

1. ככל שנוספות תצפיות R^2 עולה.
2. ככל שנוספים משתנים מסבירים R^2 עולה.
3. R^2 אינו יכול לקטון כאשר מוסיפים תצפיות לרגרסיה.

פתרון:

1. לא נכון בהכרח (מספיק דוגמא אחת שהיא לא נכונה כדי שזה לא יהיה נכון), כאשר התצפית היא על קו הרגרסיה R^2 גדל (היחס בין סכום השונות לסכום התצפיות קטן- מוסיפים 0 למונה כי אין סטייה ומספר מסוים למכנה). אבל כאשר התצפית מאוד רחוקה היא מורידה את R^2.
2. לא נכון בהכרח, אם המשתנה המסביר לא מסביר כלל את המשתנה המוסבר מקדם המשתנה יהיה 0 ולא תהיה לו השפעה על R^2.
3. כבר בטענה 1, ראינו שזה לא נכון.

ההנחות הקלאסיות:

- ההנחות הקלאסיות הם על הסטייה (ההפרעה המקרית), U.
- תזכורת- U הוא הסטיה של התצפית מקו הרגרסיה האמיתי.
- e – הוא הסטיה של התצפית מקו הרגרסיה הנאמד.

הנחה מספר 1

- תוחלת ההפרעה המקרית (הסטיה) היא אפס, כלומר:

$$E(u_i)=0$$

$$i=1,\ldots,N.$$

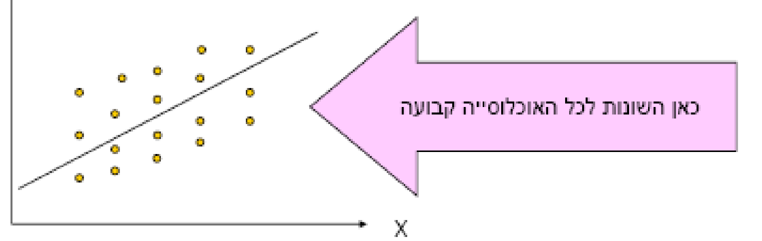
בגלל זה בקו הרגרסיה האמיתי לא יהיה לנו משתנה U.

הנחה מספר 2

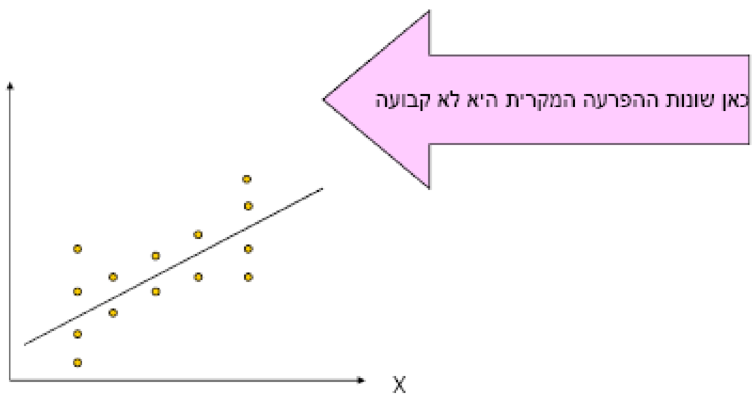
- שונות ההפרעה המקרית היא קבועה. כלומר:

$$\operatorname{var}(u_i) = \sigma^2$$

- הנחה זו מכונה הנחת ההומוסקדסטיות)



- השונות מבטאת את הפיזור של ה- e לכל X והוא קבוע.

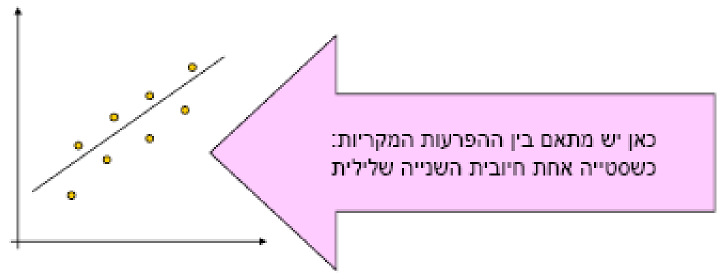


כאן הפיזור הוא לא קבוע- בXים גבוהים השונות גבוה ובאמצע נמוכה.

הנחה מספר 3

- ההפרעות המקריות בכל שתי תצפיות הן בלתי מתואמות (הנחת חוסר מתאם סדרתי) כלומר,

$$\operatorname{cov}(u_i, u_j) = 0, \quad i \neq j, \quad i, j = 1, \ldots, N$$



הנחה מספר 4

- ההפרעות המקריות אינן מתואמות עם המשתנה המסביר. כלומר,

$$\operatorname{cov}(u_i, x_i) = 0$$

- לדוגמא, אם ב-X-ים נמוכים הסטיות קטנות וב-X-ים גבוהים הסטיות גבוהות-ההנחה לא מתקיימת

הנחה זאת גם משתמעת מכך שהפיזור הוא קבוע (הנחה מס. 2) כיוון שהסטיות לא יכולות להשתנות הן גם לא יכולות להיות תלויות בגודלו של X.

הנחה מספר 5

- המשתנים האקסוגנים, ה-X-ים, הם קבועים ואינם משתנה מיקרי.
- השונות של ה-X-ים היא אפס.
- הנחה זו חשובה להמשך.

משפט גאוס-מרקוב

- אומדני הריבועים הפחותים (OLS) במודל הרגרסיה

$$Y_i = \alpha + \beta X_i + u_i$$

כאשר u מקיים את ההנחות הקלאסיות הם האומדים היעילים ביותר והם BLUE.

BLUE

- Best Linear Unbiased Estimator

- כלומר האומדנים הם ליניאריים, חסרי הטיה והטובים ביותר.

- תוחלת של קבוע שווה לעצמו!

- אומדנים חסרי הטיה הם אומדנים שהתוחלת שלהם שווה להם!

אומדנים חסרי הטיה:

- נחשב את התוחלת של Y :

$$E(Y_i)=E(\alpha+\beta X_i+u_i)=\alpha+\beta X_i+\underbrace{E(u_i)}_{\text{שווה ל-0 ע"פ ההנחות הקלאסיות}}=\alpha+\beta X_i$$

- נחשב את השונות של Y :

$$\mathrm{var}(Y_i)=\underbrace{\mathrm{var}(\alpha+\beta X_i+u_i)}_{\text{שונות של קבועים היא אפס}}=\mathrm{var}(u_i)=\sigma^2$$

12

$\alpha + \beta Xi$ ← שניהם קבועים

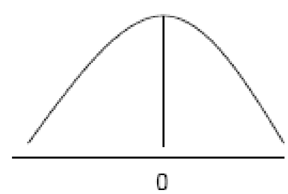
ניתן לסכם:

- U מתפלג נורמלית, כאשר: $u_i = N\left(0,\sigma^2\right)$

התפלגות נורמלית

תוחלת

שונות



- Yi מתפלג נורמלית, כאשר:

$$Y_i = N(\alpha+\beta X,\sigma^2)$$