שיטות רגרסיה

פרק 6 - רגרסיה מרובה ומולטיקוליניאריות

תוכן העניינים

רגרסיה מרובה ומולטיקולינאריות:

רקע:

מודל הרגרסיה המרובה:

$.Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_J X_{Ji} + U_i$

כאשר:

Y_i = משתנה תלוי.

$X_{1i} \ldots X_{Ji}$ = משתנים ב"ת.

U_i = טעות מקרית המקיימת את כל ההנחות הקלאסיות.

α = חותך אחד שמשמעותו: הציון המנובא כאשר כל המשתנים הב"ת = 0.

$\beta_1 \ldots \beta_J$ = מקדמי השיפוע. (מס' הבטות = למספר המשתנים הב"ת במודל).

משמעות מקדם השיפוע β_j: ההשפעה <u>הייחודית</u> של המשתנה הב"ת המסוים לניבוי המשתנה התלוי, בניכוי השפעתם של כל יתר המשנים הב"ת האחרים המצויים במשוואת הרגרסיה.

אמידת מודל הרגרסיה המרובה:

1. שיטת הריבועים הפחותים:

$$.Min\sum e_i^2 = Min\sum\left(Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_J X_{Ji}\right)^2$$

מפתרון פונקציית הריבועים הפחותים נקבל את אומדי הרגרסיה: $\hat{\alpha}, \hat{\beta}_1 \ldots \hat{\beta}_J$.

2. המשוואות הנורמאליות:

$\sum_{i=1}^{n} e_i = 0$ בגלל שיש חותך.

$\sum_{i=1}^{n} e_i X_{1i} = 0$ בגלל שיש את X_{1i}.

$\sum_{i=1}^{n} e_i X_{2i} = 0$ בגלל שיש את X_{2i}.

עד $\sum_{i=1}^{n} e_i X_{Ji} = 0$ בגלל שיש את X_{Ji}.

דוגמא:

מקרה פרטי, מודל עם שני משתנים מסבירים:

$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$.

הנוסחאות הנורמאליות הן:

$$\sum_{i=1}^{n} e_i = 0$$

$$\sum_{i=1}^{n} e_i X_{1i} = 0$$

$$\sum_{i=1}^{n} e_i X_{2i} = 0$$

מפתרון מערכת המשוואות נקבל את הנוסחאות הבאות לחישוב האומדים:

$$\hat{\alpha} = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$$

$$\hat{\beta}_1 = \frac{\left(r_{y1} - r_{y2} * r_{12}\right)}{1 - r_{12}^2} \cdot \frac{\hat{s}_y}{\hat{s}_{x1}}$$

$$\hat{\beta}_2 = \frac{\left(r_{y2} - r_{y1} * r_{12}\right)}{1 - r_{12}^2} \cdot \frac{\hat{s}_y}{\hat{s}_{x2}}$$

הערה:

ניתן לראות כי אם לא קיים מתאם בין המשתנים הב"ת: $r_{12} = 0$,

שיפועי הרגרסיה המרובה זהים לשיפועי הרגרסיה הפשוטה:

$$\hat{\beta}_1 = r_{y1} \cdot \frac{\hat{s}_y}{\hat{s}_{x1}} = \frac{\operatorname{cov}(y, x_1)}{\operatorname{var}(x_1)}$$

$$\hat{\beta}_2 = r_{y2} \cdot \frac{\hat{s}_y}{\hat{s}_{x2}} = \frac{\operatorname{cov}(y, x_2)}{\operatorname{var}(x_2)}$$

מולטיקולינאריות:

מולטיקולינאריות מתייחסת למתאם בין המשתנים המסבירים במודל.

מולטיקולינאריות מלאה:
מתאם מלא בין המשתנים המסבירים במודל.
הדבר קורה כאשר משתנה מסביר אחד הוא קומבינציה ליניארית מלאה של המשתנה המסביר השני: $x_1 = a + bx_2$ (x_1 הוא קומבינציה ליניארית מלאה של x_2)
מכאן ש: $r_{12} = 1$.

- שימו לב כי מדובר בטרנספורמציה ליניארית ולא בטרנספורמציה אחרת (למשל $x_1 = x_2^2$), אז בהכרח $r_{12} \neq 1$.
- מולטיקולינאריות מלאה יכולה להיווצר גם כאשר קבוצה של משתנים מסבירים מהווה קומבינציה ליניארית מלאה של אחד המשתנים המסבירים: $x_1 + x_2 = a + bx_3$.

במצב של מולטיקולינאריות מלאה אין כל השפעה של המשתנה האחד מעבר לשני ולא ניתן לאמוד את המודל שכן אר״פ אינם מוגדרים.
פתרון: הורדת אחד המשתנים ואמידת המשוואה מחדש בלעדיו.

מולטיקולינאריות חלקית:
כאשר יש מתאם גבוה מאוד (אך לא מושלם) בין 2 משתנים מסבירים במודל או בין קבוצה של משתנים מסבירים:

$$x_1 = a + bx_2 + u_i$$
$$x_1 + x_2 = a + bx_3 + u_i$$

מכיוון שיש מתאם גבוה בין המשתנים הב״ת לא נוכל לבודד באופן מלא את ההשפעה המדויקת של כל אחד מהם על ציוני המשתנה התלוי.
כל אחד מהמשתנים הב״ת ״יגזול״ מן ההשפעה הייחודית שיש למשתנה הב״ת השני על המשתנה התלוי, כך שבסופו של דבר, למרות שהמודל עם שני המשתנים הב״ת יהיה מובהק, התרומה הייחודית של כל משתנה ב״ת לניבוי התלוי לא תהיה מובהקת.

שאלות:

רגרסיה מרובה:

1) כלכלן החליט לאמוד מודל ליניארי עם שלושה משתנים מסבירים: x_1, x_2, x_3.
$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$.

א. מהי בעיית ה-OLS שעליו לפתור?

ב. מצאו את תנאי סדר ראשון של הבעיה.

2) כלכלן החליט לבחון מה משפיע על שער הדולר בישראל.
לכן אסף מדגם בין ארבע תצפיות חודשיות.
להלן טבלה מסכמת:

חודש	X_2 השקעות זרים בישראל (במיליוני דולרים)	X_1 שער הריבית	Y דולר	טעות (e^i)
אוגוסט	100	3	3.2	-5
ספטמבר	95	3.5	3.6	6
אוקטובר	90	3.5	3.8	0
נובמבר	100	3	3.5	-2

מהו המודל אשר אותו אמד הכלכלן?

3) הניחו כי הקשר באוכלוסייה בין X ל- Y נתון ע"י המשוואה הבאה: $Y_i = 2 + \beta_1 X_{1i} + 5X_{2i} + u_i$ וכל ההנחות הקלאסיות מתקיימות.

נתון האומד: $\tilde{\beta}_1 = \dfrac{\sum X_{1i}\left((Y_i - 8X_{2i} - 2) - \overline{(y - 8X_2 - 2)}\right)}{\sum X_{1i}^2}$.

א. חשבו את תוחלת האומד.

ב. חשבו את שונות האומד.

ג. מהו היחס בין שונות האומד הנ"ל, לבין שונות אומד הריבועים הפחותים?

4) הניחו כי הקשר באוכלוסייה בין X ל- Y נתון ע"י המשוואה הבאה: $y_i = \alpha + \beta_1 x_{1i} + 8x_{2i} + u_i$.

כל ההנחות הקלאסיות מתקיימות וכן: $\sum x_{1i} = 0$.

אומדים את β_1 באופן הבא: $b_1 = \dfrac{\sum (x_{1i} - \bar{x})\left(y_i - 8x_{2i} - \left(\overline{y - 8x_2}\right)\right)}{\sum (x_{1i} - \bar{x})^2}$.

א. האם האומד חסר הטיה?

ב. מהי שונות האומד?

מולטיקולינאריות:

5) נתון המודל: $Y_i = \alpha + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + U_i$.

חוו דעתכם על הטענות הבאות (כל סעיף עומד בפני עצמו):

א. בהנחה כי מתקיים: $X_{1i} - 2X_{2i} = 1$,
לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים. נכון/לא נכון/ לא ניתן לדעת.

ב. בהנחה כי מתקיים: $x_{1i} = x_{2i}^2$,
לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים. נכון/לא נכון/ לא ניתן לדעת.

ג. הוכיחו תשובותיכם לסעיפים ב' ו-ג'.

ד. בהנחה כי מתקיים: $r_{12} = 0.98$,

i. לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים. נכון/לא נכון/ לא ניתן לדעת.

ii. איזו בעיה עלולה להיווצר במודל ומהן השלכותיה.

6) כלכלן אמד את המודל: $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$.
בשל החשש ממולטיקוליניאריות בחן הכלכלן את המתאם בין כל זוג של משתנים מסבירים וקיבל: $r_{x_1,x_2} = 0.9$, $r_{x_1,x_3} = 0.99$, $r_{x_3,x_2} = 0.5$.
לכן הסיק כי אין בעיה של מולטיקוליניאריות מושלמת במודל.
האם הוא צודק?

7) כלכלן אמד את המודל הבא: $\ln(Q_i) = \alpha + \beta_1 \ln(K_i) + \beta_2 \ln(K_i^2) + \beta_3 L_i^{0.5} + u_i$.
האם קיימת בעיה של מולטיקוליניאריות במודל?

8) להלן מודל של שכר W_i, כפונקציה של שנות לימוד S_i ושל גיל A_i:

1. $W_i = \alpha + \beta_1 \cdot S_i + \beta_2 \cdot A_i + u_i$.

בנוסף למשתנים במשוואה, החליט החוקר להוסיף גם את משתנה הוותק: EXP_i.
מכיוון שלא היו בידו נתונים על הוותק, החליט החוקר להעריכו עבור כל עובד על ידי הגיל של העובד פחות 24 שנים (מתוך ההנחה שהחיים המקצועיים מתחילים בגיל זה לערך).
להלן משוואה מס' 2:

2. $W_i = \alpha + \beta_1 \cdot S_i + \beta_2 \cdot A_i + \beta_3 \cdot EXP_i + w_i$.

חווה דעתך על המשוואה השנייה.

תשובות סופיות:

1) א. $\min \sum \left(y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i} \right)^2$.

ב. $\sum_{i=1}^{n} e_i = 0$, $\sum_{i=1}^{n} e_i X_{1i} = 0$, $\sum_{i=1}^{n} e_i X_{2i} = 0$, $\sum_{i=1}^{n} e_i X_{3i} = 0$.

2) $y_i = \beta_1 x_{1i} + u_i$.

3) א. לא ניתן לחשב. ב. $Var = \frac{\sigma^2}{\sum X_{1i}^2}$. ג. לא ניתן לדעת.

4) א. כן. ב. $V(b_1) = \frac{\sigma^2}{\sum x_{1i}^2}$.

5) א. נכון. ב. לא נכון. ג. הוכחה. ד.i. לא נכון.
ii. מולטיקולינאריות חלקית.

6) הכלכלן לא יכול להיות בטוח.

7) כן.

8) ראו סרטון.