מבוא לאקונומטריקה

פרק 7 - מבחן F ו R בריבוע

תוכן העניינים

מבחן F ו-R ברגרסיה:

רקע:

מדד R^2 לטיב הרגרסיה:

מדד לפרופורציית השונות המוסברת:

$$R^2=\frac{RSS}{TSS}=1-\frac{ESS}{TSS}$$

מתבסס על הנוסחה לפירוק השונות של קו הרגרסיה:

$$TSS=RSS+ESS$$

$$\sum\left(Y_i-\bar{Y}\right)^2=\sum\left(\hat{Y}_i-\bar{Y}\right)^2+\sum e_i^2$$

תכונות R^2:

- נע בין 0 ל-1: $0\le R^2\le 1$.
  כאשר $R^2=1$ ההתאמה מושלמת ואין שום טעויות בניבוי במודל ואילו
  כאשר $R^2=0$ הכל טעות ואין שום הסבר במודל.
- אר״פ מביא למקסימום את R^2.
- לא ניתן להשוות במדד בין מודלים שבהם אין את אותו משתנה מוסבר.
- בהוספת משתנים מסבירים נוספים למודל, R^2 יכול רק לעלות או להישאר ללא שינוי. זהו למעשה החיסרון הגדול של המדד.
  כדי להתגבר על חיסרון זה קיים מדד נוסף והוא R^2_{adj} (R^2 מתוקנן):
  $$\bar{R}^2 = 1-\left(1-R^2\right)\frac{n-1}{n-k}$$
  K = מס׳ הפרמטרים במודל (כולל החותך).
- המדד המתוקנן לוקח בחשבון את מספר המשתנים הב״ת שיש במודל ויכול לרדת בהוספת משתנים למודל לכן מתקיים תמיד ש: $\bar{R}^2<R^2$.
- המדד המתוקנן – $\bar{R}^2$ עדיף על המדד – R^2 בכדי לבחון האם כדאי לנו להוסיף משתנים ב״ת למודל.

זהויות שכדאי לדעת לגבי R^2:

במודל רגרסיה פשוטה: $y_i = \alpha + \beta x_i + u_i$ מתקיים:

1. $R^2 = r_{yx}^2$.
2. $r_{yx} = \hat{\beta}\frac{S_x}{S_y}$.
3. $R^2 = \hat{\beta}^2\frac{S_x^2}{S_y^2}$.
4. במודלים: $\begin{matrix} y_i = \alpha_1 + \beta_1 x_i + u_i \\ x_i = \alpha_2 + \beta_2 y_i + \varepsilon_i \end{matrix}$ מתקיים:
   i. הם בעלי אותו R^2.
   ii. $R^2 = \beta_1 \cdot \beta_2$.

שימו לב:

1. במודל ללא שיפוע: $y_i = \alpha + u_i$, ה- R^2 שווה ל-0 כי אין מקדם הסבר לרגרסיה.
2. במודל ללא חותך: $y_i = \beta x_i + u_i$ אין משמעות ל- R^2 כיוון שלא מתקיימת המשוואה הנורמלית הראשונה: $\sum \hat{u}_i = 0$ ולכן גם $\bar{\hat{y}} \neq \bar{y}$ ולכן גם לא מתקיים: $SST = SSR + SSE$.

מבחן F:

משמש לבדיקת:

1. הגבלות שונות המתקיימות במודל (מבחן WALD).
2. מובהקות מודל הרגרסיה כולו.

מבחן WALD:

לבדיקת השערת אפס שיש בה מספר שוויונים (במבחן t היה רק שוויון אחד בהשערת האפס).

1. אומדים את המודל המקורי – הלא-מוגבל (Unrestricted) ומקבלים את סכום ריבועי הסטיות של הטעויות $\left(\sum e_{i\ UR}^2\right)$.
2. מגדירים את כל השוויונים של השערת האפס.
3. מציבים את השוויונים של השערת האפס במודל המקורי לקבלת המודל המוגבל (Restricted).
4. אומדים את המודל המוגבל ומקבלים את סכום ריבועי הסטיות של הטעויות $\left(\sum e_{i\ R}^2\right)$.
5. חישוב הסטטיסטי: $\dfrac{\left(\sum e_{\text{מ}}^2 - {}_{\text{מל}}\sum e^2\right)/m}{\sum e_{\text{מ}}^2 /_{\text{ל}}(n-k)} \sim F_{(m,n-k,1-\alpha)}$.

(כש- m מספר המגבלות ו- k מסי הפרמטרים במודל הלא מוגבל).

- כאשר לשתי הרגרסיות (המוגבלת והלא מוגבלת) אותו משתנה מוסבר ניתן להשתמש גם בנוסחה הבאה: $\dfrac{\left(R_{\text{מ}}^2 - {}_{\text{מ}}R^2\right)_{\text{ל}}/m}{\left(1-R_{\text{מ}}^2\right)_{\text{ל}}/(n-k)} \sim F_{(m,n-k,1-\alpha)}$.

כלל הכרעה לדחיית H_0: $F_{stat} > F_{(m,n-k;1-\alpha)}$.

אם דוחים את H_0 המסקנה היא שהמודל המקורי (הלא-מוגבל) הוא הרלוונטי ולהיפך.

מבחן F למובהקות המודל:

משמש לבדיקה האם מודל הרגרסיה שלנו לניבוי משתנה תלוי מסוים על ידי המשתנים הב״ת, מובהק באוכלוסייה.

השערות:
$$\begin{aligned}&H_0: \beta_1 = \beta_2 = \ldots = \beta_K = 0\\&H_1: OTHERWISE\end{aligned}$$

המודל הלא מוגבל יהיה: $Y_t = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u_t$ U:

המודל המוגבל יהיה: $Y_t = \alpha + u_t$ R:

מאחר ו- $R_{מ}^2 = 0$ ו- $m = k-1$: $F = \dfrac{\dfrac{R_U^2 - R_R^2}{m}}{\dfrac{1-R_U^2}{n-k}} = \dfrac{\dfrac{R_U^2}{k-1}}{\dfrac{1-R_U^2}{n-k}}$.

הערה:
בדיקת מובהקות המודל ברגרסיה מרובה ניתנת לביצוע רק על ידי מבחן F מאחר ויש יותר ממגבלה אחת בהשערת האפס.
לעומת זאת בדיקת מובהקות המודל ברגרסיה חד משתנית ניתנת לביצוע גם על ידי מבחן t שכן יש רק מגבלה אחת בהשערת האפס: $F = t^2$.

לסיכום:

1. מתי נשתמש במבחן t ומתי במבחן F?
   - רק t: השערות חד צדדיות (סימן אי שוויון בהשערות).
   - t או F (כאשר: $F = t^2$): מגבלה אחת (שוויון אחד בלבד) בהשערת האפס.
   - רק F: כאשר יש כמה מגבלות (שוויונים) בהשערת האפס.
2. מצב של סתירה בין מבחן F למבחני t:
   כאשר המודל מובהק אולם אף אחד מהשיפועים לא יוצא מובהק – בעיה של מולטיקוליניאריות חלקית במודל (מתאמים גבוהים בין המשתנים הב״ת).

שאלות:

R בריבוע:

1) דרגו את המודלים הבאים (לפי קריטריון R^2):

1. $y_i = \alpha + \beta x_{1i} + u_i$ $R^2 = 0.15$.
2. $y_i = \alpha + u_i$.
3. $y_i = \beta x_{1i} + u_i$.
4. $y_i = \alpha + \beta\, x_{1i} + \beta_2 x_{2i} + \varepsilon_i$.
5. $y_i = \alpha + \beta_2 x_{2i} + u_i$ $R^2 = 0.20$.

2) על סמך מדגם של 100 תצפיות נאמדו המודלים הבאים:

1. $\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}$.
2. $\hat{y}_i = \hat{\delta}_0 + \hat{\delta}_1 x_{1i} + \hat{\delta}_2 x_{2i}$ $R^2 = 0.70$.
3. $\hat{y}_i = \hat{\gamma}_0 + \hat{\gamma}_2 x_{2i}$ $R^2 = 0.65$.

א. שלושה חוקרים העלו טענה לגבי מקדם R^2 של משוואה מסי (1):

1. אי אפשר לדעת מהנתונים המובאים לעיל אם R^2 של משוואה (1) הוא גדול או קטן מ-0.70.
2. אי אפשר לדעת מהנתונים המובאים לעיל אם R^2 של משוואה (1) הוא גדול או קטן מ-0.65.
3. ניתן לצפות כי R^2 של משוואה (1) יהיה גדול מ-0.70.

בהתייחס לטענות החוקרים ניתן לומר:

i. רק הטענה של חוקר 1 נכונה.
ii. רק הטענה של חוקר 2 נכונה.
iii. רק הטענה של חוקר 3 נכונה.
iv. כל הטענות שגויות.

ב. חוו דעתכם על הטענות הבאות המתייחסות ל- $\bar{R}^2$:

i. ניתן לצפות ש- $\bar{R}^2$ של משוואה (1) יהיה גדול מ-0.7.   נכון/לא נכון/ לא ניתן לדעת
ii. ניתן לצפות כי $\bar{R}^2$ של משוואה (2) יהיה קטן מ-0.7.   נכון/לא נכון/ לא ניתן לדעת
iii. ניתן לצפות כי $\bar{R}^2$ של משוואה (3) יהיה קטן מ-0.7.   נכון/לא נכון/לא ניתן לדעת

3) על סמך מדגם של 80 משפחות המונות כל אחת 4 ילדים, נאמדו המשוואות הבאות:

1. $\hat{y}_i = 5+2x_{1i}+2x_{2i}$ $R^2=0.77$.
2. $\hat{y}_i = 24+0.8x_{1i}$ $R^2=0.62$.
3. $\hat{y}_i = 14+0.7x_{2i}$ $R^2=0.25$.
4. $\hat{y}_i = 4+0.5w_i$ $R^2=0.30$.
5. $\hat{\ln}(y)_i = 7+0.9x_{1i}+0.6x_{2i}$ $R^2=0.45$.
6. $\hat{\ln}(y)_i = 11+0.7x_{1i}+0.9x_{2i}+0.6x_{3i}$.
7. $\hat{y}_i = 13+8x_{1i}+7x_{2i}+2x_{3i}+9x_{4i}$.

כאשר y_i הינו סה"כ הוצאות משק בית i, x_{ji} הינו גילו של הילד j, ונתון כי: $w_i = 2x_{3i}+x_{1i}-x_{2i}$.
דרגו את הרגרסיות לפי קריטריון R^2 (הימני עדיף על השמאלי).

4) נתונות שתי המשוואות הבאות: $y_i = 58+b_1x_i+e_{1i}$ ו- $x_i = a_2-0.2y_i+e_{2i}$,
כאשר: $\bar{y}=\bar{x}=40$. למה שווה מקדם המתאם של פירסון בין X ל- Y?

א. 0.09.
ב. 0.69.
ג. 0.3.
ד. 0.72.
ה. אף תשובה לא נכונה.

5) נתון מודל רגרסיה: $y_i = \alpha+\beta x_i+u_i$.
הוכיחו כי: $SST = SSR+SSE$.

מבחן F:

6) נאמד המודל: $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ והתקבל כי: $\sum e^2 = 620.1683$ וכי: $R^2 = 0.99$.
הועלתה ההשערה כי ההשפעה על Y של משתנה S היא פי 3 מזו של משתנה Z, וכן כי החותך הוא 5.
א. מהי השערת האפס?
ב. מהו המודל המוגבל שאותו צריך לאמוד?

מאמידת המודל המוגבל התקבל כי: $\sum e^2 = 623.99$ וכי: $R^2 = 0.99$.
ג. חשב את הסטטיסטי של WALD.
ד. כמה דרגות חופש יש במונה וכמה במכנה?
ה. האם דוחים או מקבלים את השערת האפס?

7) במדגם של 82 תצפיות התקבל: $y_i = 12 + 3x_{1i} + 4x_{2i} + e_i$ $R^2 = 0.73$.
א. בחנו את ההשערה כי: $\begin{matrix} H0: \beta_2 = 0 \\ H1: \beta_2 \neq 0 \end{matrix}$.
כאשר נתון כי לאחר אמידת המודל המוגבל התקבל כי: $R^2 = 0.6$.
ב. חשבו את $S_{\hat{\beta}_2}$.

8) על מנת לאמוד את פונקציית התצרוכת נאספו נתונים על 42 משקי בית בשנת 2007 ונאמדה המשוואה הבאה: $C_i = \alpha + \beta_1 \cdot W_i + \beta_2 \cdot P_i + u_i$.
מתוצאות האמידה של המשוואה הנ"ל התקבל כי: $\sum e^2 = 52968$.
על מנת לבדוק את ההשערה שהנטייה השולית לצרוך (נש"צ) מתוך ההכנסה זהה לנטייה השולית לצרוך מתוך ההון נאמדה גם המשוואה הבאה: $C_i = \alpha + \beta_1 \cdot Y_i + u_i$ כאשר: $Y_i =$ סה"כ ההכנסה של משק בית t $(W_i + P_i)$.
התקבל: $\sum e^2 = 54156$.
א. בדקו את ההשערה.
ב. חשבו את סטטיסטי t לבדיקת ההשערה.

מבחן F למובהקות המודל:

9) נתון המודל: $y = A\frac{x_{1i}^{\beta_1}}{x_{3i}^{\beta_3}} e^{\beta_2 x_2} e^{u_i}$.
באמידת מדגם של 58 נבדקים התקבל: $R^2 = 0.56$.
האם המודל מובהק?

תרגול מסכם:

(10 נאמדו חמשת המודלים הבאים על 70 תצפיות:

1. $ESS = 130$ $\quad I_i = 12 + 0.13 \cdot \exp_i + 0.08 \cdot scl_i + 2 \cdot workh_i + u_i$.
2. $ESS = 150$ $\quad I_i = 11 + 0.1 \cdot scl + 0.1 \cdot workh_i + u_i$.
3. $ESS = 151$ $\quad I_i = 9 + 0.22 \cdot scl + u_i$.
4. $ESS = 152$ $\quad I_i = 15 + 0.15 \cdot workh_i + u_i$.
5. $ESS = 200$ $\quad I_i = 25 + u_i$.

המשתנה המוסבר הוא הכנסה מעבודה (I) והמשתנים המסבירים שבחנו הם מספר שנות הלימוד (scl), מספר שעות עבודה ($workh$) וותק בעבודה (exp)

הערה: הניחו כי ערך F הקריטי הוא 4.

א. האם לשעות עבודה ($workh$) ישנה השפעה מובהקת על ההכנסה במשוואה 2?
ב. האם לשנות לימוד ישנה השפעה מובהקת על ההכנסה במשוואה 2?
ג. האם רגרסיה 2 מובהקת? (בחנו האם יש הסבר במודל 2), כיצד זה מסתדר עם תשובתכם ל-א' ו-ב'.
ד. האם השפעת הוותק יכול להיות 0.15?
ה. כלכלן נוסף הציע להריץ את המודל:
$I_i + \exp_i = 2 - 3(scl_i - workh_i) + u_i \ , \ ESS = 145$.
איזו השערה ניתן לבחון באמצעות מודל זה?
כמה דרגות חופש יש לסטטיסטי שנקבל? בחנו אותה.

(11 על סמך מדגם של 40 תצפיות נאמדו המשוואות הבאות:

1. $y_i = 2 + 3X_{1i} + 4X_{2i} + e_i \quad R^2 = 0.76$.
2. $y_i = 3 + 5D_i + e_i \quad R^2 = 0.60$.
3. $D_i = 0.2X_{1i} + X_{2i}$.

כאשר Y הינו הציון בתואר ראשון, X_1 ציוני הבגרות ו-X_2 ציוני הפסיכומטרי.

א. בדקו את ההשערה כי ציוני הבגרות וציוני הפסיכומטרי ביחד לא משפיעים על ציוני תואר ראשון.
ב. בדקו את ההשערה כי רגרסיה 2 מובהקת.
ג. איזה השערה ניתן לבדוק באמצעות רגרסיה 1 ו-2?

12) על סמך מדגם של 80 משפחות המונות כל אחת 4 ילדים, נאמדו המשוואות הבאות:

1. $\hat{y}_i = 5 + 2X_{1i} + 2X_{2i}$ $R^2 = 0.6$.
2. $\hat{y}_i = 11 + 0.9x_{2i} + 0.6x_{3i}$ $R^2 = 0.45$.
3. $\hat{y}_i = 13 + 8x_{1i} + 7x_{2i} + 2x_{3i}$ $R^2 = 0.78$.

כאשר y_i הינו סה"כ הוצאות משק בית i, x_{ji} הינו גילו של הילד j.

חשבו את האומדן לסטיית התקן של המקדם X_3 ברגרסיה 3.

13) נתון המודל: $y_i = AX_{1i}^{\beta_1} e^{\beta_2 X_{2i} + \beta_3 X_{3i}} e^{u_i}$.

א. מהי המשוואה לאמידת המקדמים של המודל?

ב. מה המודל המוגבל עבור ההשערה: $\beta_1 = 2\beta_3$; $\beta_2 = 3\beta_3$.

ג. מהן דרגות החופש במונה ובמכנה?

ד. רשמו את הנוסחה לחישוב סטטיסטי המבחן.

14) המודל הבא מתאר את פונקציית הייצור של מוצר P:

$$\ln(P_i) = \alpha + \beta_S \ln(S_i) + \beta_J \ln(J_i) + \varepsilon_i .$$

כאשר S ו- J הן שתי התשומות בייצור (S = תשומת ההון ו- J = תשומת העבודה).

מהו המודל המוגבל המתאים לבדיקת ההשערה כי פונקציית הייצור מקיימת תק"ל (תשואה קבועה לגודל)?

תשובות סופיות:

1) $4>5>1=3>2$.

2) א. iii. ב. i. לא ניתן לדעת. ii. נכון. iii. נכון.

3) 7 , 1 , 2 , 4 , 3 ובאופן נפרד: 6 , 5.

4) ג'.

5) הוכחה.

6) א. $H_0: \alpha=5,\ \beta_s=3\beta_z$. ב. $Y_t-5=\beta_x X_t+\beta_z\left(Z_t+3S_t\right)+\beta_w W_t+u_t$.

ג. $F=0.6145$. ד. מונה: 2, מכנה: 199. ה. מקבלים.

7) א. יש עדות לכך. ב. $S_{\hat{\beta}_2}=0.645$.

8) א. אין עדות לכך. ב. $t=0.934$.

9) יש עדות לכך.

10) א. אין עדות לכך. ב. אין עדות לכך. ג. יש עדות לכך.

ד. כן. ה. $H0;\ \beta_{\exp}=-1;\ \beta_{work}=-\beta_{scl}$.

11) א. יש עדות לכך. ב. יש עדות לכך. ג. $\beta_1=0.2\beta_2$.

12) $S.E=0.25$.

13) א. $\ln\left(y_i\right)=\ln\left(A\right)+\beta_1\ln\left(X_{1i}\right)+\beta_2X_{2i}+\beta_3X_{3i}+u_i$.

ב. $\ln\left(y_i\right)=\ln\left(A\right)+\beta_3\left(\ln\left(X_{1i}\right)+3X_{2i}+X_{3i}\right)+u_i$.

ג. מונה: $m=2$, מכנה: $n-k=n-4$.

ד. $F=\dfrac{\dfrac{R_U^2-R_R^2}{m}}{\dfrac{1-R_U^2}{n-k}}$.

14) $\ln\left(\dfrac{P_i}{S_i}\right)=\alpha+\beta_J\ln\left(\dfrac{J_i}{S_i}\right)+\varepsilon_i$.