

## גרסיה ותוכנו ניסויים

### פרק 8 - מבחן F ו R בربוע

תוכן העניינים

1. כללי .....

## מבחן F ו- R בריבוע:

**רקע:**

**מדד  $R^2$  לטיב הרוגסיה:**

מדד לפרופורציצית השונות המוסברת :

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

מתבסס על הנוסחה לפירוק השונות של קו הרוגסיה :

$$TSS = RSS + ESS$$

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

תכונות  $R^2$  :

- נع בין 0 ל-1 :  $0 \leq R^2 \leq 1$ .
- כאשר  $R^2 = 1$  ההתאמה מושלמת ואין שום טעויות בניבוי במודל ואיילו.
- כאשר  $R^2 = 0$  הכל טעות ואין שום הסבר במודל.
- אר"פ מביא למקסימום את  $R^2$ .
- לא ניתן להשוות במידד בין מודלים שבהם אין את אותו משתנה מוסף.
- בஹוספת משתנים מסבירים נוספים למודל,  $R^2$  יכול רק לעלות או להישאר ללא שינוי. זהו למעשה החיסרונו הגדול של המדי.
- כדי להתגבר על חיסרונו זה קיים ממד נוסף והוא  $R_{adj}^2$  (מדתוקן) :

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$$

$K$  = מס' הפרמטרים במודל (כולל החותך).

- המדד המתוקן לוקח בחשבון את מספר המשתנים הב"י שיש במודל ויכול לרזרת בהhosפת משתנים למודל لكن מתקיים תמיד ש :  $\bar{R}^2 < R^2$ .
- המדד המתוקן –  $\bar{R}^2$  עדיף על המדי –  $R^2$  בכך לבחון האם כדאי לנו להוסיף משתנים ב"י למודל.

זהויות שכדי לדעת לגבי  $R^2$  :  
במודל רגרסיה פשוטה :  $y_i = \alpha + \beta x_i + u_i$  מתקיים :

$$\cdot R^2 = r_{yx}^2 \quad .1$$

$$\cdot r_{yx} = \hat{\beta} \frac{S_x}{S_y} \quad .2$$

$$\cdot R^2 = \hat{\beta}^2 \frac{S_x^2}{S_y^2} \quad .3$$

4. במודלים :  $y_i = \alpha_1 + \beta_1 x_i + u_i$  מתקיים :  
 $x_i = \alpha_2 + \beta_2 y_i + \varepsilon_i$

.ii. הם בעלי אותו  $R^2$ .

$$\cdot R^2 = \beta_1 \cdot \beta_2$$

שימוש 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 המשווה הנורמלית הראשונה :  $\sum \hat{u}_i = 0$  ולכן גם  $\bar{y} \neq \bar{\hat{y}}$  ולכן גם לא מתקיים :  $SST = SSR + SSE$

### מבחן F :

משמש לבדיקת :

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

**מבחן F: WALD**

לבדיקה השערת אפס שיש בה מספר שווויוניים (במבחן  $t$  היה רק שוויון אחד בהשערת האפס).

1. אומדים את המודל המקורי – הלא-מוגבל (Unrestricted) ומקבלים את סכום

$$\text{ריבועי הסטיות של הטעויות} = \left( \sum e_{UR}^2 \right)$$

2. מגדירים את כל השווויוניים של השערת האפס.

3. מציבים את השווויוניים של השערת האפס במודל המקורי לקבלת המודל המוגבל (Restricted).

4. אומדים את המודל המוגבל ומקבלים את סכום ריבועי הסטיות של

$$\text{הטעויות} = \left( \sum e_{R}^2 \right)$$

5. חישוב הסטטיסטי:  $\frac{\left( \sum e_{R}^2 - \sum e_{UR}^2 \right) / m}{\sum e_{UR}^2 / (n-k)} \sim F_{(m,n-k,1-\alpha)}$

(כש-  $m$  מספר המגבליות ו-  $k$  מס' הפרמטרים במודל הלא מוגבל).

- כאשר לשתי הרגression (המוגבלת והלא מוגבלת) אותו משתנה מושבר ניתן

$$\frac{\left( R_z^2 - R^2 \right) / m}{\left( 1 - R_z^2 \right) / (n-k)} \sim F_{(m,n-k,1-\alpha)} \quad \text{להשתמש גם בנוסחה הבאה:}$$

כל הכרעה לדחיה  $H_0$ :  $F_{stat} > F_{(m,n-k,1-\alpha)}$   
 אם דוחים את  $H_0$  המסקנה היא שהמודל המקורי (הלא-מוגבל) הוא הרלוונטי ולהיפך.

**מבחן F לモביהקות המודל:**

משמש לבדיקה האם מודל הרgression שלנו לניבוי משתנה תלוי מסוים על ידי המשתנים הב"ית, מובהק באוכולוסייה.

השערות:  
 $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$   
 $H_1: \text{OTHERWISE}$

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

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

$$. F = \frac{\frac{R_U^2 - R_R^2}{m}}{\frac{1 - R_U^2}{n-k}} = \frac{\frac{R_U^2}{k-1}}{\frac{1 - R_U^2}{n-k}} : m = k-1 \text{ ו } R_U^2 = 0$$

הערה :

בדיקת מובהקות המודל ברגرسיה מרובה ניתנת לביצוע רק על ידי מבחן F לאחר  
ויש יותר מגבלה 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 :  $F = t^2$ .

### לטיכום :

1. متى השתמש במבחן t ומתי במבחן F?

- רק t : השערות חד צדיות (סימן אי שווין בהשערות).
- t או F (כאשר :  $F = t^2$ ) : מגבלה אחת (שוויון אחד בלבד) בהשערת האפס.
- רק F : כאשר יש כמה מגבלות (שוויוניים) בהשערת האפס.

2. מצב של סטייה בין מבחן F למבחן t :

כאשר המודל מובhawk אולם אף אחד מהSHIPועים לא יוצא מובhawk – בעיה של  
מולטיקוליניריות חלקית במודל (מתאימים גבויים בין המשתנים הב"ת).

**שאלות:****R בריבוע:**

**1)** דרגו את המודלים הבאים (לפי קритריון  $R^2$ ) :

$$\cdot R^2 = 0.15 \quad y_i = \alpha + \beta x_{1i} + u_i \quad .1$$

$$\cdot y_i = \alpha + u_i \quad .2$$

$$\cdot y_i = \beta x_{1i} + u_i \quad .3$$

$$\cdot y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \quad .4$$

$$\cdot R^2 = 0.20 \quad y_i = \alpha + \beta_2 x_{2i} + u_i \quad .5$$

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

$$\cdot \hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \quad .1$$

$$\cdot R^2 = 0.70 \quad \hat{y}_i = \hat{\delta}_0 + \hat{\delta}_1 x_{1i} + \hat{\delta}_2 x_{2i} \quad .2$$

$$\cdot R^2 = 0.65 \quad \hat{y}_i = \hat{\gamma}_0 + \hat{\gamma}_2 x_{2i} \quad .3$$

א. שלושה חוקרים העלו טענה לגבי מקדם  $R^2$  של משווה מס' (1) :

1. אי אפשר לדעת מהנתונים המובאים לעיל אם  $R^2$  של משווה מס' (1) הוא גדול או קטן מ-0.70.

2. אי אפשר לדעת מהנתונים המובאים לעיל אם  $R^2$  של משווה מס' (1) הוא גדול או קטן מ-0.65.

3. ניתן לצפות כי  $R^2$  של משווה מס' (1) יהיה גדול מ-0.70.

בהתיחס 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 המתיחסות ל- $\bar{R}^2$  :

- i. ניתן לצפות ש- $\bar{R}^2$  של משווה מס' (1)

יכוון/לא נכון/ לא ניתן לדעת יהיה גדול מ-0.7.

- ii. ניתן לצפות כי  $\bar{R}^2$  של משווה מס' (2)

יכוון/לא נכון/ לא ניתן לדעת יהיה קטן מ-0.7.

- iii. ניתן לצפות כי  $\bar{R}^2$  של משווה מס' (3)

יכוון/לא נכון/ לא ניתן לדעת יהיה קטן מ-0.7.

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

$$R^2 = 0.77 \quad \hat{y}_i = 5 + 2x_{1i} + 2x_{2i} \quad .1$$

$$R^2 = 0.62 \quad \hat{y}_i = 24 + 0.8x_{1i} \quad .2$$

$$R^2 = 0.25 \quad \hat{y}_i = 14 + 0.7x_{2i} \quad .3$$

$$R^2 = 0.30 \quad \hat{y}_i = 4 + 0.5w_i \quad .4$$

$$R^2 = 0.45 \quad \ln(y)_i = 7 + 0.9x_{1i} + 0.6x_{2i} \quad .5$$

$$\ln(y)_i = 11 + 0.7x_{1i} + 0.9x_{2i} + 0.6x_{3i} \quad .6$$

$$\hat{y}_i = 13 + 8x_{1i} + 7x_{2i} + 2x_{3i} + 9x_{4i} \quad .7$$

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

$$w_i = 2x_{3i} + x_{1i} - x_{2i}$$

דרגו את הרגרסיות לפי קритריון  $R^2$  (הימני עדיף על השמאלי).

(4) נתונות שתי המשוואות הבאות:  $1. x_i = a_2 - 0.2y_i + e_{2i}$  ו  $y_i = 58 + b_1x_i + e_{1i}$

כאשר:  $\bar{y} = 40$ . למה שווה מקדם המתאים של פירסום 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:  $y_i = \alpha + \beta x_i + u_i$

$$SST = SSR + SSE$$

**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.
- מהי השערת האפס?
  - מהו המודל המוגבל שאותו צריך לאמוד?

מאמידת המודל המוגבל התקבל כי:  $\sum e^2 = 623.99$  וכי:  $R^2 = 0.99$ .

- חשב את הסטטיסטי של WALD.
- כמה דרגות חופש יש במונה וכמה במכנה?
- האם דוחים או מקבלים את השערת האפס?

7) במדגם של 82 תוצאות התקבל:  $R^2 = 0.73$   $y_i = 12 + 3x_{1i} + 4x_{2i} + e_i$ .

- בחנו את ההשערה כי:  $H0: \beta_2 = 0$  .  $H1: \beta_2 \neq 0$

כאשר נתון כי לאחר אמידת המודל המוגבל התקבל כי:  $R^2 = 0.6$ .

- חשבו את  $S_{\hat{\beta}_2}$ .

8) על מנת לאמוד את פונקציית הצרוכת נאספו נתונים על 42 משקי בית בשנת 2007 ונאמדה המשוואה הבאה:

$$C_i = \alpha + \beta_1 \cdot W_i + \beta_2 \cdot P_i + u_i \quad \sum e^2 = 52968$$

על מנת לבדוק את ההשערה שהנטיעיה השולית לצריך (נש"צ) מותן ההכנסה זהה לנטייה השולית לצריך מתוך ההון נאמדה גם המשוואה הבאה:  $C_i = \alpha + \beta_1 \cdot Y_i + u_i$  כאשר  $Y_i = \text{סה"כ ההכנסה של משק בית t} (W_i + P_i)$ . התקבל:  $\sum e^2 = 54156$ .

- בדקו את ההשערה.
- 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 חמישת המודלים הבאים על 70 תצפיות :

$$\cdot I_i = 12 + 0.13 \cdot \exp_i + 0.08 \cdot scl_i + 2 \cdot workh_i + u_i \quad ESS = 130 \quad .1$$

$$\cdot I_i = 11 + 0.1 \cdot scl + 0.1 \cdot workh_i + u_i \quad ESS = 150 \quad .2$$

$$\cdot I_i = 9 + 0.22 \cdot scl + u_i \quad ESS = 151 \quad .3$$

$$\cdot I_i = 15 + 0.15 \cdot workh_i + u_i \quad ESS = 152 \quad .4$$

$$\cdot I_i = 25 + u_i \quad ESS = 200 \quad .5$$

המשתנה המושבר הוא הכנסה מעובודה ( $I$ ) והמשתנים המסבירים שבחנו הם מספר שנים הלימוד ( $scl$ ), מספר שעות עבודה ( $workh$ ) וותק עבודה ( $\exp$ ).  
הערה: הניחו כי ערך  $F$  הקרייטי הוא 4.

א. האם לשעות עבודה ( $workh$ ) ישנה השפעה מובהקת על הכנסה במשווה ?

ב. האם לשנות לימוד ישנה השפעה מובהקת על הכנסה במשווה ?

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 :

$$\cdot I_i + \exp_i = 2 - 3(scl_i - workh_i) + u_i, \quad ESS = 145$$

איזו השערה ניתן לבחון באמצעות מודל זה ?

כמה דרגות חופש יש לסטטיסטי שנקבל ? בוחנו אותה.

**11)** על סמך מדגם של 40 תצפיות נאמדו המשוואות הבאות :

$$\cdot R^2 = 0.76 \quad y_i = 2 + 3X_{1i} + 4X_{2i} + e_i \quad .1$$

$$\cdot R^2 = 0.60 \quad y_i = 3 + 5D_i + e_i \quad .2$$

$$\cdot D_i = 0.2X_{1i} + X_{2i} \quad .3$$

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

א. בדקו את ההשערה כי ציוני הבגרות וציוני הפסיכומטרי ביחיד 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:

$$\cdot R^2 = 0.6 \quad \hat{y}_i = 5 + 2X_{1i} + 2X_{2i} \quad .1$$

$$\cdot R^2 = 0.45 \quad \hat{y}_i = 11 + 0.9x_{2i} + 0.6x_{3i} \quad .2$$

$$\cdot R^2 = 0.78 \quad \hat{y}_i = 13 + 8x_{1i} + 7x_{2i} + 2x_{3i} \quad .3$$

כאשר  $y_i$  הינו סה"כ הוצאות משק בית  $i$ ,  $x_{ji}$  הינו גילו של הילד  $j$ .  
חשבו את האומדן לסתיות התקן של המקדם  $X_3$  ברגression.

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

א. מהי המשווה לאמידת המקדים של המודל?

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

ג. מהן דרגות החופש במוניה 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 התשלומיות 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. לא ניתן לדעת.      .iii. נכון.  
 (3) 3 , 4 , 2 , 1 , 7  
 (4) ג'.  
 (5) הוכחה.
- (6) א.  $H_0 : \alpha = 5$  ,  $\beta_s = 3\beta_z$   
 ג.  $F = 0.6145$  . ד. מכנה: 2 , מכנה: 199 .  
 ה. מתקבלים.
- (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.  
 ה.  $H_0: \beta_{\text{exp}} = -1$  ;  $\beta_{\text{work}} = -\beta_{\text{scl}}$   
 ז.  $\beta_1 = 0.2\beta_2$       ג. יש עדות לכך.
- (11) א. יש עדות לכך.      ב. יש עדות לכך.  
 ג. מכנה: 2 , מכנה: 4 .  
 ד.  $S.E = 0.25$  (12)
- (13)  $\ln(y_i) = \ln(A) + \beta_1 \ln(X_{1i}) + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$  . א.  
 ב.  $\ln(y_i) = \ln(A) + \beta_3 (\ln(X_{1i}) + 3X_{2i} + X_{3i}) + u_i$   
 ג. מכנה:  $m = 2$  ,  $n - k = n - 4$  :  

$$F = \frac{\frac{R_U^2 - R_R^2}{n-k}}{\frac{m}{1-R_U^2}}$$
- (14)  $\ln\left(\frac{P_i}{S_i}\right) = \alpha + \beta_J \ln\left(\frac{J_i}{S_i}\right) + \varepsilon_i$