שיעור מס' 13 - ~~מבחן~~ בדיקת השערות מורכבות:

* מבחני t עונים על שאלה אחת בנפרד (האם השפעה של משתנה מובהקת? האם שונה מאפס מובהק?)

* מבחני F מסוגלים לענות על כל השאלות יחד / קומבינציות של השפעות.

* הרעיון הוא שמשווים סטטיסטית מודלים שמורכבים עם מגבלות לעומת מודלים ללא מגבלות.

למשל:

$$Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\beta_3X_{3i}+\dots+\beta_kX_{ki}+\varepsilon_i$$

$H_0: \beta_1=\beta_2=0$ → זו בעצם האם מובהקים בו"ז ש...

וזאת לעומת:

$H_0: \beta_1=0$; $H_0: \beta_2=0$

עוד דוגמא: קומב' לינארית של משתנים: $H_0: \beta_1+\beta_2+\beta_2=1$

אפשר גם לשאול שוויון השפעה של שלושה:

$H_0: \beta_1=\beta_2=\beta_3$ → האם ההשפעה של ה משתנים X_1,X_2,X_3 היא זהה (כדי לומר למה היא שונה)

נניח שיש לנו מודל עם $K=4$

רוצים לשאול האם בו"ז → גם ההשפעה של X_3 וגם של X_4 שווה לאפס.

2 תנאים ("מגבלות")

$H_0: \beta_3=\beta_4=0$

אם ההשערה שלנו נכונה:

בעצם במקרה הזה זהו המודל: $Y_i=\beta_0+\beta_1X_{1i}+\beta_2X_{2i}+\varepsilon_i$ → המודל המוגבל $\boxed{R}$ - המודל תחת השערת האפס.

$Y_i=\beta_0+\beta_1X_{1i}+\dots+\beta_4X_{4i}+\varepsilon_i$ → המודל הלא מוגבל $\boxed{UR}$

את הבדיקה עושים באמצעות מבחן F:

<u>האינטואיציה מאחורי המבחן:</u> לוקחים את המודל המלא (הלא מוגבל) וכופים עליו את H_0. מריצים רגרסיה.

ברגע שכפינו משהו (מגבלה) ← משמעות מחיר מסוים שמשלמים אותו ב $\sum \hat{\varepsilon}_i^2$ ← מגבלה ← סכום השגיאות שלי גדל.
(פחות משתנים מסבירים)

$\sum \hat{\varepsilon}_i^{2\,R} >$ .. UR

מה שמבחן F בודק בעצם הוא האם המגבלה הזו מקפיצה את גודל השגיאות בצורה "משמעותית סטטיסטית".

אזי

$$F_{\text{סטטיסטי}} = \frac{\overbrace{\left(\sum \hat{\varepsilon}_i^{2\,R} - \sum \hat{\varepsilon}_i^{2\,UR}\right)}^{\text{חיובי בהכרח}} / q}{\sum \hat{\varepsilon}_i^{2\,UR} / (N-K-1)}$$

q – מספר המגבלות שכפינו (מספר השוויונים) ב H_0

"4" $N-K-1$ – N − מספר הפרמטרים

$$\sim F(q, (N-K-1))$$

אם $F_{\text{סטטיסטי}} > F_{crit}$ ← דוחים את השערת האפס ← המודל המוגבל בודאות
ב-95% אינו נכון

בעצם זה אומר שסכום השגיאות
שלו <u>ממש</u> גדל יותר משל
המודל הלא מוגבל.

← מבחן F ← תמיד חד״צ.
טבלה באתר ← 95%

③

אינטואיציה סטטיסטית שעומדת מאחורי מבחן F:

אם קיים משתנה: $Z \sim N(0,1)$

אזי $Z^2 \sim \chi^2_{(1)}$

$\sum_{i=1}^{n} Z_i^2 \sim \chi^2_{(n)}$ כאשר Z_i, Z_j בת"ל.

אם קיימים 2 משתנים V עם v דה"ח $\sim \chi^2$

W עם w דה"ח $\sim \chi^2$

$$\Rightarrow \frac{V/v}{W/w} \sim F_{(v,w)}$$
דה"ח דה"ח.

למה עושים את זה?

① בדיקה של תיאוריה (האם תוספת השפיעה בזה לתשובה לנסיון)

② רוצים להסביר כמה שיותר עם כמה שפחות.

איך עושים בתוכנה - דוגמאות:

דוגמא 1:

נניח יש פונק' ייצור קוב דאגלס: $Y = A \cdot K_i^{\alpha} L_i^{\beta} \cdot e^{\varepsilon_i}$

ln

$\ln Y = \ln A + \alpha \ln K_i + \beta \ln L_i + \varepsilon_i \rightarrow$ מודל לא מוגבל

$H_0: \alpha + \beta = 1 \rightarrow$ תקין $\rightarrow$ נניח רוצים לבדוק האם פונק' הייצור מקיימת תקין

$H_1: \alpha + \beta \neq 1 \rightarrow$ לא תקין

(אם פונק' ביקוש ← הומוגנית מדרגה 0 ← הכפלת המחירים והשכר בקבוע לא משנה את הביקוש)

~~ראשית מבצעים את המודל שבודק שאלפא ובטא לפרמטר~~

מודל מוגבל:

שלב א': מריצים מודל לא מוגבל ומקבלים

$\sum \hat{\varepsilon}_i^2$ UR

שלב ב': פותחים את המודל המוגבל ומגיעים לצורה שאפשר להריץ רגרסיה:

$$\ln Y_i = \ln A + \alpha \ln K_i + (1-\alpha)\ln L_i + \varepsilon_i$$

$$\ln Y_i = \ln A + \alpha(\ln K_i - \ln L_i) + \ln L_i + \varepsilon_i$$

$$\underbrace{(\ln Y_i - \ln L_i)}_{\text{יוצר משתנה תלוי חדש}} = \underbrace{\ln A}_{\text{חותך}} + \alpha\underbrace{(\ln K_i - \ln L_i)}_{\text{משתנה מסביר חדש}} + \varepsilon_i$$

שלב ג': אחרי כן מריצים רגרסיה של המסביר על המוסבר.
מקבלים:

$\sum \hat{\varepsilon}_i^2$ R

טי' אלגב': ① משתנים ללא מקדם ← אפשר להעביר שמאלה.

② מכנה משותף לבסוף.

* הערה: את זה ↑ היה אפשר לעשות גם במבחן t.

דוגמא (2):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

$H_0: \beta_1 = 3 \, , \, \beta_2 = 7$ → 2 מגבלות

H_1: אחרת

מודל מוגבל:

$$Y_i = \beta_0 + 3X_{1i} + 7X_{2i} + \varepsilon_i$$

$(Y_i - 3X_{1i} - 7X_{2i}) = \beta_0 + \varepsilon_i$ ← מודל רק עם חותך.

(5)

דוגמא (3): מבחן מובהקות הרגרסיה:

$H_0: \beta_1 = \dots = \beta_K = 0$

H_1: לפחות אחת מהבטאות $\neq 0$.

המודל המוגבל: $Y_i = \beta_0 + \varepsilon_i$

← F_{stat} שמופיע בפלט.

זהו מבחן חשוב. שימושים:

נניח יש 2 משתנים מסבירים ← מאוד מתואמים (מולטי-קוליניאריות)

* נניח שההשפעה של שניהם לא מובהקת (לא נדחה ← $H_0: \beta_1=0$) $H_0: \beta_2=0$

** יש לנו בעיה של חוסר דיוק באמידה של האפקט הנקי, אבל ביחד הם כן משפיעים על Y ואז נדחה $H_0: \beta_1=\beta_2=0$

הבעיה היא שלא יודעים את ההשפעה הנקית.

הערה נוספת: אפשר להראות ש $t^2 = F_{(1, n-k-1)}$

הערך של t^2 שווה לערך של F כאשר יש מגבלה אחת.

דוגמא (4): אמידת פונק' ביקוש:

כבר הוכח לא פשוטה שהגמישויות הבודדות.

$$\log Y = \beta_0 + \beta_1 \log I + \beta_2 \log P_y + \beta_3 \log P_x$$

מחירי המוצרים האחרים

(התייחסות למשתנים אחרים וכו'.)

$H_0: \beta_1+\beta_2+\beta_3=0 \Rightarrow \beta_3 = -(\beta_1+\beta_2)$

מציבים:

$$\log y = \beta_0 + \beta_1 \log \frac{I}{P_x} + \beta_2 \frac{P_y}{P_x} + \varepsilon_i$$

משוואת רגרסיה במונחים ריאליים.

תרגול (5):

השוואה בין רגרסיות שונות:

נניח משוואת פונ' תצרוכת בתק' מיתון לעומת תקופה רגילה:

נניח שיש תצפיות משתי תקופות - ~~[illegible]~~

המודל הלא מוגבל מורכב מ-2 רג' (נפרדות)

$$\boxed{ESS_u = ESS_1 + ESS_2}$$

$$\begin{cases} Y_i^1 = \beta_0^1 + \beta_1^1 X_{1i} + \beta_2^1 X_{2i} + \varepsilon_i^1 & (n_1 \text{ obs.}) \\ Y_i^2 = \beta_0^2 + \beta_1^2 X_{1i} + \beta_2^2 X_{2i} + \varepsilon_i^2 & (n_2 \text{ obs.}) \end{cases}$$

H_0: האם שתי הרגרסיות מתנהגות באותו אופן $\rightarrow \beta_0^1 = \beta_0^2,\ \beta_1^1 = \beta_1^2,\ \beta_2^1 = \beta_2^2$

H_1: אחרת

המודל המוגבל (pooled) ← שתי הרגרסיות ביחד

עכשיו עורכים מבחן F (קרוי גם צ'או פישר):

$$F_{stat} = \frac{\dfrac{(ESS_1 + ESS_2)^{UR} - ESS^{R}}{k+1}}{\dfrac{(ESS_1 + ESS_2)}{(n_1+n_2-2k-2)}}$$

$k+1$ ← מס' המגבלות

(n_1+n_2-2k-2) ← סה"כ דה"ח במודל הלא מוגבל

יש עוד דרך ← עם משתני דמי.

דרך אחרת להראות את מבחן F:
(שימוש ב-R^2):

$$\text{Geo } F = \frac{\left[\dfrac{\sum \hat{\varepsilon}_i^{2\,R}}{TSS} - \dfrac{\sum \hat{\varepsilon}_i^{2\,UR}}{TSS}\right] / q}{\dfrac{\sum \hat{\varepsilon}_i^{2\,UR}}{TSS} / (n-k-1)} = \frac{\left[1-R^2_R - (1-R^2_{UR})\right] / q}{1-R^2_{UR} / n-k-1}$$

$$= \frac{(R^2_{UR} - R^2_R) / q}{1-R^2_{UR} / n-k-1}$$

$\rightarrow$ תמיד $R^2_{UR} \geq R^2_R$

מגבלה תשובה: הנוסחא הזו נכונה רק כאשר המשתנה התלוי במודל ללא שינוי ← כלומר כאשר TSS זהה.

← אחרת משתמשים בנוסחא הסטנדרטית.

אפשר להראות לכן שמבחן מובהקות הרגרסיה:

$$F_{stat} = \frac{R^2_{UR} / k}{(1-R^2_{UR}) / (n-k-1)}$$

~~[illegible]~~

נושאי הסף נוסף:

adjusted R^2:

* למה צריך אותו (R^2 תמיד גדל עם כמות משתנים מסבירים)

* הנוסחא:

$$\text{adjusted } \bar{R}^2 = 1 - \frac{\sum \hat{\varepsilon}_i^2}{TSS} \cdot \frac{(n-1)}{(n-k-1)}$$

נושא נוסף בעיות ספציפיקציה ← טעיות מדידה

(בוומלות CRS ,

(β