# מבוא לאקונומטריקה א

## פרק 13 - רגרסיה מרובה תוך שימוש בפלטים של SAS

### תוכן העניינים

1. רגרסיה מרובה ..... 1

# רגרסיה מרובה:

## רקע:

### מבחן T ו-F:

כאשר יש יותר ממשתנה מסביר אחד, מדובר ברגרסיה מרובה.

המודל הקלאסי: $Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + ... + \beta_k X_{kt} + u_t$.

- קבוע $\alpha$ יש אחד.
- מספר ה-$\beta$ טות כמספר המשתנים הב"ת במודל.

מבחן $F$ למובהקות המודל:

השערות: $\begin{matrix} H_0 = \beta_1 = \beta_2 = ... = \beta_K = 0 \\ H_1 : OTHERWISE \end{matrix}$.

סטטיסטי המבחן $F$ וכלל ההכרעה:

$$F = \frac{\frac{RSS}{k}}{\frac{ESS}{T-k-1}} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{T-k-1}} > F\left(k, T-k-1; 1-\alpha\right).$$

מבחן $t$ למובהקות ה-$\beta$ טות:

מבחן לבדיקת מובהקות $\beta$ ספציפית:

השערות: $\begin{matrix} H_0 = \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{matrix}$.

סטטיסטי המבחן $t$ וכלל ההכרעה:

$$\left| t_{\hat{\beta}_i} \right| = \left| \frac{\hat{\beta}_i}{S_{\hat{\beta}_i}} \right| > t_{\left(T-k-1; 1-\frac{\alpha}{2}\right)}.$$

**השוואה בין מודלים – $\bar{R}^2$ וחוק חיטובסקי:**

בכדי להחליט האם כדאי לנו להוסיף למודל משתנה ב"ת מסוים:
נשווה את פרופורציית השונות המוסברת המתוקנת $\bar{R}^2$ בין המודל ללא המשתנה המסביר לבין המודל עם המשתנה המסביר שהוספנו.

- ניתן להשתמש גם באומד המוטה - $R^2$ להשוואה בין מודלים אם מתקיימים שני התנאים הבאים:
  1. מספר המשתנים זהה.
  2. המשתנה המוסבר זהה.

לפי חוק חיטובסקי – בהוספת משתנה מסביר אחד בלבד למודל ה-$\bar{R}^2$ יעלה אך ורק אם: $\left|t_{\hat{\beta}}\right|>1$.

כאשר: $\left|t_{\hat{\beta}}\right|<1$ אז $\bar{R}^2$ ירד בהוספת המשתנה והוא גם לא יהיה רלוונטי למודל (מובהק).

כאשר: $\left|t_{\hat{\beta}}\right|>2$ אז $\bar{R}^2$ יעלה והמשתנה שהוסף יהיה גם מובהק.

כאשר: $1<\left|t_{\hat{\beta}}\right|<2$ אז ה-$\bar{R}^2$ יעלה אך יש לבדוק את רלוונטיות המשתנה שהוסף למודל על פי מבחן $t$.

## שאלות:

### מבחן T ו-F:

**1)** נאמד המודל: $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ והתקבלו התוצאות הבאות:

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | ------ | 646169.84 | ---------- | ---------- | 0.0000 |
| Error | ------ | ---------- | ---------- | | |
| C Total | 203 | 646790.01 | | | |

| | | | |
|---|---|---|---|
| Root MSE | ---------- | R-square | ---------- |
| Dep Mean | 178.6645 | Adj R-sq | 0.999022 |
| C.V. | 0.988075 | | |

**Parameter Estimates**

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob>\|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | 5.067731 | 0.456604 | 11.09874 | 0.0000 |
| X | 1 | ---------- | 0.042711 | 22.84485 | 0.0000 |
| Z | 1 | 3.005385 | 0.008679 | 346.2721 | 0.0000 |
| W | 1 | -5.029101 | 0.073149 | ---------- | 0.0000 |
| S | 1 | 8.974106 | 0.029075 | 308.6485 | 0.0000 |

א. השלם את הנתונים החסרים בפלט.

ב. האם המודל מובהק? בדקו ברמת מובהקות של 0.05.

ג. האם משתנה W רלוונטי למודל? בדקו ברמת מובהקות של 0.01.

**השוואה בין מודלים:**

**2)** במודל לניבוי ההכנסה על פי שנות לימוד וותק במקום העבודה, התקבל: $\bar{R}^2 = 0.266$. הוסף המשתנה היקף המשרה. במבחן למובהקות המשתנה הנוסף התקבל: $t_{\hat{\beta}} = 0.456$. האם ערך $\bar{R}^2$ יעלה/ירד/לא ישתנה בהוספת המשתנה הנוסף למודל?

**מבחן Wald ו-T מורכב:**

**3)** נאמד המודל: $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ והתקבלו התוצאות הבאות:

## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 4 | 646169.84 | 161542.46 | 51835.84 | 0.0000 |
| Error | 199 | 620.1683 | 3.1164236 | | |
| C Total | 203 | 646790.01 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 1.7653395 | R-square | 0.999041 |
| Dep Mean | 178.6645 | Adj R-sq | 0.999022 |
| C.V. | 0.988075 | | |

## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob>\|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | 5.067731 | 0.456604 | 11.09874 | 0.0000 |
| X | 1 | 0.975736 | 0.042711 | 22.84485 | 0.0000 |
| Z | 1 | 3.005385 | 0.008679 | 346.2721 | 0.0000 |
| W | 1 | -5.029101 | 0.073149 | -68.75141 | 0.0000 |
| S | 1 | 8.974106 | 0.029075 | 308.6485 | 0.0000 |

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?

להלן אמידת המודל המוגבל:

## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 2 | 646166.01 | 323083.01 | | |
| Error | 201 | 623.9983 | 3.104469 | | |
| C Total | 203 | 646790.01 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 1.7619504 | R-square | 0.999035 |
| Dep Mean | 173.6645 | Adj R-sq | 0.999026 |
| C.V. | 1.0145714 | | |

## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob>\|T\| |
|---|---|---|---|---|---|
| X | 1 | 0.978491 | 0.036399 | 26.88240 | 0.0000 |
| Z+3S | 1 | 2.999995 | 0.003669 | 817.6080 | 0.0000 |
| W | 1 | -5.043109 | 0.071218 | -70.81249 | 0.0000 |

ג. חשב את הסטטיסטי של WALD.

ד. כמה דרגות חופש יש במונה וכמה במכנה?

ה. האם דוחים או מקבלים את השערת האפס?

4) על מנת לאמוד את פונקציית התצרוכת נאספו נתונים על 42 משקי בית בשנת 2007 ונאמדה המשוואה הבאה: $C_t = \alpha + \beta_1 \cdot W_t + \beta_2 \cdot P_t + u_t$.
להלן תוצאות האמידה של המשוואה הנ"ל:

## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | --- | --------- | -------- | ------- | -------- |
| Error | --- | --------- | 52968 | | |
| C Total | --- | --------- | | | |

## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob>\|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | -107.226 | ------ | ---------- | ------- |
| W | 1 | 0.743 | ------ | | |
| P | 1 | 0.561 | ------- | | |

## Covariance of Estimates

| COV | INTERCEP | W | P |
|---|---|---|---|
| INTERCEP | ---------- | ---------- | -------- |
| W | ---------- | 0.0046 | -0.0090 |
| P | ---------- | -0.0090 | 0.016 |

על מנת לבדוק את ההשערה שהנטייה השולית לצרוך מתוך ההכנסה זהה לנטייה השולית לצרוך מתוך ההון, נאמדה גם המשוואה הבאה:
$C_t = \alpha + \beta_1 \cdot Y_t + u_t$, כאשר: $Y_t$ = סה"כ ההכנסה של משק בית t.
התקבל: $ESS = 0.4566$.
בדקו את ההשערה בשתי דרכים.

**תרגיל מסכם:**

**5)** חוקר אמד את התצרוכת של 500 משקי בית כפונקציה של הכנסה שלהן לפי המשוואה: $EXPENSE_t = \alpha + \beta \cdot INCOME_t + u_t$.

$EXPENSE_t$ - התצרוכת של משק הבית ה-t-י באלפי שקלים.

$INCOME_t$ - ההכנסה של משק הבית ה-t-י באלפי שקלים.

ההפרעות האקראיות מקיימות את כל ההנחות הקלאסיות התקבל הפלט הבא:

## Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 1 | 2013.105 | 2013.105 | 6495.745 | 0.0000 |
| Error | 498 | 154.3358 | 0.3099112 | | |
| C Total | 499 | 2167.441 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 0.556697 | R-square | 0.928794 |
| Dep Mean | 3.990208 | Adj R-sq | 0.928651 |
| C.V. | 13.95157 | | |

## Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob>\|T\| |
|---|---|---|---|---|---|
| INTERCEP | 1 | 0.041995 | 0.054951 | 0.764236 | 0.4451 |
| INCOME | 1 | 0.713503 | 0.008853 | 80.59618 | 0.0000 |

א. מהו Pvalue לבדיקת מובהקות המודל ע"י מבחן F?

ב. מהו אחוז השונות בתצרוכת המוסבר ע"י ההכנסה?

ג. מהו אומדן לתצרוכת ההתחלתית של משק בית?

ד. האם אומדן זה מובהק?

ה. על עוזר מחקר הטיל החוקר לבדוק את ההשערה כי על כל 1000 ₪ נוספים בהכנסה צורך הפרט 700 ₪, כנגד ההשערה כי הוא צורך יותר מ-700 ₪. נסח את השערת האפס ואת ההשערה האלטרנטיבית.

ו. מהו הסטטיסטי t לבדיקת ההשערה?

ז. מהו הסטטיסטי WALD לבדיקת ההשערה?

ח. התברר כי הייתה טעות בנתונים, וכי יש להוסיף 1000 ₪ לתצרוכת של כל משק בית:

i. ההוספה תגדיל את האומד ל- $\alpha$: נכון / לא נכון / לא ניתן לדעת.

ii. בעקבות ההוספה האומד ל- $\alpha$ יהיה מובהק: נכון / לא נכון / לא ניתן לדעת.

iii. ההוספה תשנה את האומד ל- $\beta$ : נכון / לא נכון / לא ניתן לדעת.

iv. ההוספה תשנה את $R^2$ : נכון / לא נכון / לא ניתן לדעת.

החוקר טען כי יש להוסיף לפונקציית התצרוכת גם את השפעת העושר.
העושר של משק בית מורכב מתוכניות החסכון שלו ($SAVINGS$) ומניירות הערך שיש לו ($NE$). שתי סדרות הנתונים הן באלפי שקלים.
החוקר אמד את המשוואה:

$$EXPENSE_t = \alpha + \beta_1 \cdot INCOME_t + \beta_2 \cdot SAVINGS_t + \beta_3 \cdot NE_t + u_t ,$$

וקיבל כי סכום ריבועי הסטיות של הטעויות הוא 121.

ט. מהי השערת האפס לבדיקת הטענה של החוקר (שהמודל החדש נכון ולא המקורי)?

י. מהו הסטטיסטי WALD לבדיקת ההשערה?

החוקר רצה לבדוק את ההשערה כי הנש"צ מתוך ההכנסה שווה ל-0.6 וכי השפעת ניירות הערך על התצרוכת היא פי 2 מהשפעת תוכניות החסכון.

יא. מהי השערת האפס לבדיקה זו?

יב. המודל המוגבל לבדיקת ההשערה יהיה מהצורה: $Z_t = \gamma_0 + \gamma_1 W_t + v_t$ .

בטא את $Z_t$ , ו- $W_t$ באמצעות המשתנים המקוריים.

## תשובות סופיות:

**1)** א.

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Value | Prob>F |
|---|---|---|---|---|---|
| Model | 4 | 646169.84 | **161542.46** | **51835.84** | 0.0000 |
| Error | 199 | **620.1683** | **3.1164236** | | |
| C Total | 203 | 646790.01 | | | |

| | | | |
|---|---|---|---|
| **Root MSE** | **1.7653395** | **R-square** | **0.999041** |
| **Dep Mean** | 178.6645 | **Adj R-sq** | 0.999022 |
| **C.V.** | 0.988075 | | |

**Parameter Estimates**

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob>\|T\| |
|---|---|---|---|---|---|
| **INTERCEP** | **1** | **5.067731** | 0.456604 | 11.09874 | 0.0000 |
| **X** | **1** | 0.975736 | 0.042711 | 22.84485 | 0.0000 |
| **Z** | **1** | 3.005385 | 0.008679 | 346.2721 | 0.0000 |
| **W** | **1** | -5.029101 | 0.073149 | **-68.75141** | 0.0000 |
| **S** | **1** | 8.974106 | 0.029075 | 308.6485 | 0.0000 |

ב. יש עדות לכך. ג. יש עדות לכך.

**2)** ירד.

**3)** 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. $WALD_{stat}=0.6145$. ד. מונה: 2 , מכנה: 199-.

ה. מקבלים.

**4)** בדיקה ע"י מבחן WALD ו-t: אין עדות לכך.

**5)** א. $PF=0.000$. ב. 92%. ג. $\hat{\alpha}=0.04195$. ד. לא.

ה. $H_0:\beta=0.70\ ,\ H_1:\beta>0.70$. ו. $t_{\hat{\beta}}=1.583$. ז. $WALD_{stat}=2.505$.

ח.i. נכון. ii. נכון. iii. לא נכון. iv. לא נכון.

ט. $H_0:\beta_2=\beta_3=0\ ,\ H_1:OTHERWISE$. י. $WALD_{stat}=68.32$.

יא. $\beta_1=0.6\ ,\ H_0:\beta_3=2\cdot\beta_2$.

יב. $Z_t=EXPENCE_t-0.6\cdot INCOME_t\ ,\ W_t=SAVINGS_t+2\cdot NE_t$.