# Dummy Variable

ตัวอย่าง 1) $Y_i$ = เงินเดือนครูรัฐ $i$

$D_{2i}$ = 1 ถ้ารัฐนั้นอยู่ทางเหนือ, 0 = อื่นๆ

$D_{3i}$ = 1 ถ้ารัฐนั้นอยู่ทางใต้, 0 = อื่นๆ

$D_{1i}$ = 1 ถ้ารัฐนั้นอยู่ทางตะวันตก 0 = อื่นๆ

สมมติว่าประมาณสมการถดถอยได้ดังนี้

$$\hat{Y}_i = 26{,}158.62 - 1734.473\,D_{2i} - 3{,}264.615\,D_{3i} \quad (1)$$
$$(1128.523) \quad (1435.953) \quad (1{,}499.615)$$
$$t \quad (23.1759)^{*} \quad (-1.2078) \quad (-2.176)^{*}$$

(* หมายถึงมีนัยสำคัญทางสถิติ)

จากสมการข้างต้น กล่าวได้ว่า เงินเดือนโดยเฉลี่ยของครูในรัฐทางตะวันตก เท่ากับ 26,158 $ โดยเงินเดือนครูรัฐทางเหนือต่ำกว่าอยู่จำนวน 1,734.473 และเงินเดือนครูรัฐทางใต้ต่ำกว่าอยู่จำนวน 3,264.615

ค่าเฉลี่ยรายได้ของครูรัฐอื่น เป็นดังนี้

ทางเหนือ 26,158 − 1734.473 = 24,424 $

ทางใต้ 26,158 − 3265 = 22,894 $

พิจารณา slope $D_{2i}$ ไม่ sig. แสดงว่า ค่าเฉลี่ยเงินเดือนของครูรัฐทางตะวันตก ไม่ต่างจากเงินเดือนครูทางเหนือ แต่ค่าเฉลี่ยเงินเดือนครูทางใต้ต่ำกว่า ทางตะวันตกอย่างมีนัยสำคัญ

## ข้อพึงระวัง

(1) อาจเกิด Perfect collinearity

(2) Dummy variables ที่ถูกละไปจากสมการนั้นจะถือว่าเป็น base สำหรับการเปรียบเทียบ

(3) Intercept คือค่าเฉลี่ยของตัวเปรียบเทียบ

(4) ค่าสัมประสิทธิ์หรือ slope จะแสดงถึงว่าตัว Dummy นั้นต่างจากค่า Intercept มากน้อยเพียงใด

(5) การเลือก base ขึ้นอยู่กับผู้วิจัย

(6) การใส่ Dummy โดยไม่มี Intercept ค่าสัมประสิทธิ์ก็คือค่าเฉลี่ยสำหรับตัวแปร Dummy นั้นๆ

(7) โมเดลควรจะมี intercept หรือไม่มี intercept

**ตัวอย่าง 2)** $$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 X_i + u_i \qquad (2)$$

โดย $X_i$ คือ รัฐอุดหนุนค่าใช้จ่ายต่อหัวเด็กนักเรียน

ประมาณสมการเป็นดังนี้

$$\hat{Y}_i = 13269.11^{*} - 1673.514 D_{2i}^{*} - 1144.157 D_3 + 3.2889 X_i^{*}$$

$$(1395.056) \quad (801.1703) \quad (861.1182) \quad (0.3176)$$

$$t \quad (9.5115) \quad (-2.0889) \quad (-1.3286) \quad (10.3539)$$

(* = หมายถึงมีนัยสำคัญทางสถิติ) $R^2 = 0.7266$

**การตีความ** ถ้ารัฐอุดหนุนเพิ่ม 1 \$ ทำให้ $Y_i$ เพิ่ม 3.28 \$

ตัวอย่าง 3) $Y_t = \alpha_1 + \alpha_2 D_{2t} + \beta_1 X_t + \beta_2 (D_{2t} X_t) + u_t$ —(3)

Y คือ การออม

X คือ รายได้

t คือ เวลา

$D_1 = 1$ หมายถึงข้อมูล 1982-1995

$= 0$ ช่วงเวลาอื่น (1970-1981)

ประมาณสมการได้ดังนี้

$$\hat{Y}_t = 1.0161 + 152.4786 D_t^{*} + 0.0803 X_t^{*} - 0.0655 D_t X_t^{*} \quad —(4)$$

se (20.1648) (33.0824) (0.0144) (0.0159)

t (0.0504) (4.6090) (5.5413) (-4.0963) $R^2 = 0.8819$

ค่าสัมประสิทธิ์ $D_t X_t$ มีนัยสำคัญ แสดงว่า สมการถดถอย 2 ช่วงเวลานั้นแตกต่างกัน ดังนั้น ตัวประมาณสมการจะได้

สมการแรก $E(Y_t|D_t=0, X_t) = \alpha_1 + \beta_1 X_t$ —(5)

รอง $E(Y_t|D_t=1, X_t) = (\alpha_1+\alpha_2) + (\beta_1+\beta_2) X_t$ —(6)

- กรณี $D_{2t}$ แค่ $\alpha_2 = 0$ แสดงว่า สมการ 2 ช่วงเวลามี Intercept เท่ากัน
- ถ้า $\beta_2 = 0$ สมการ 2 ช่วงเวลามี Slope เท่ากัน แต่ Intercept ต่างกัน

* กรณีสมการ (3) ข้างต้น ต้องการทดสอบว่า Intercept ค่าสัมประสิทธิ์ 2 ช่วงเวลานั้นแตกต่างกันหรือไม่ (Test of stability) นั่นคือ ทดสอบว่า

$\alpha_2 = 0, \beta_2 = 0$ หรือไม่ สามารถทดสอบโดย Restricted และ Unrestricted Model ถ้ายอมรับสมมติฐาน Ho แสดงว่า $\alpha_2 = 0, \beta_2 = 0$ นั่นคือ โครงสร้างการออมไม่เปลี่ยนแปลงใน 2 ช่วงเวลาดังกล่าว

## ต.ย. ทดสอบ Restricted กับ Unrestricted Model

เช่น $\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$ หลักทดสอบว่า

$\beta_2 + \beta_3 = 1$ หรือไม่ สามารถทดสอบโดยใช้ F-test

เราจะแทนค่า $\beta_2 = 1 - \beta_3$ จะได้

$$\ln Y_i/X_{2i} = \beta_0 + \beta_3 \ln\left(\frac{X_{3i}}{X_{2i}}\right) + u_i \longrightarrow (7) \text{ Restricted Model}$$

สถิติทดสอบ

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)} = \frac{(R^2_{UR} - R^2_R)/m}{(1-R^2_{UR})/(n-k)}$$

$m$ = จำนวน linear restriction

$k$ = จำนวน parameter ใน UR

$n$ = จำนวน observations

$RSS = \sum \hat{u}_i^2$ (Residual sum of squares)

ถ้า F ไม่ sig ยอมรับ $H_0$ นั่นคือ $\beta_2 + \beta_3 = 1$

*. ดังนั้นสมการที่เหมาะสมคือ สมการ (7)

[ดูตัวอย่างเพิ่มเติม Gujarati 4 ed. (2003) หน้า 269 – 273]

## Interaction effect ของ Dummy

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i \quad \text{— (8)}$$

$Y_i$ = ค่าจ้างรายชั่วโมง \$ $\quad D_2 = 1$ = หญิง, 0 = ชาย

$X$ = จำนวนปีที่ศึกษา $\quad D_3 = 1$, ผิว c

0, อื่นๆ

* ถ้าเราตั้งสมมติฐาน ผู้หญิงที่ทำงานเป็นคนผิว c รายได้ต่ำกว่า ผู้ชายที่ทำงานเป็นคนผิว c

ต้องใส่ตัวแปรที่ต้องเพิ่มคือ $D_2 D_3$

ดังนั้น $Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{2i} D_{3i} + \beta X_i + u_i$ —— (9)

$E(Y_i / D_{2i}=1, D_{3i}=1, X_i) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i$ —— (10)

ประมาณสมการ (9) เป็นดังนี้

$$\hat{Y}_i = -0.2610^{*} - 2.3606^{*} D_{2i} - 1.7327^{*} D_{3i} + 2.1289^{\dagger} D_{2i} D_{3i} + 0.8028^{*} X_i$$

$$t \quad (-0.2357) \quad (-5.4873) \quad (-2.1803) \quad (1.7420) \quad (9.9095)$$

$R^2 = 0.2032$, $n = 528$

ถ้ามีจำนวนปีที่ศึกษาคงที่ และรวมค่าสัมประสิทธิ์ของ Dummy ทุกตัว จะได้ $-2.3606 - 1.77327 + 2.1289 = -1.964$ แสดงว่า ค่าจ้างของผู้หญิงที่ทำงานในภาคใต้จะลดลง 1.96 \$ ซึ่งอยู่ระหว่าง ( $D_{2i}=1$, $D_{3i}=1$ )

2.3606 และ 1.7327