

# การวิเคราะห์ความสัมพันธ์และ วิเคราะห์ถดถอย

สันศักดิ์ พรประเสริฐมานิต

คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย

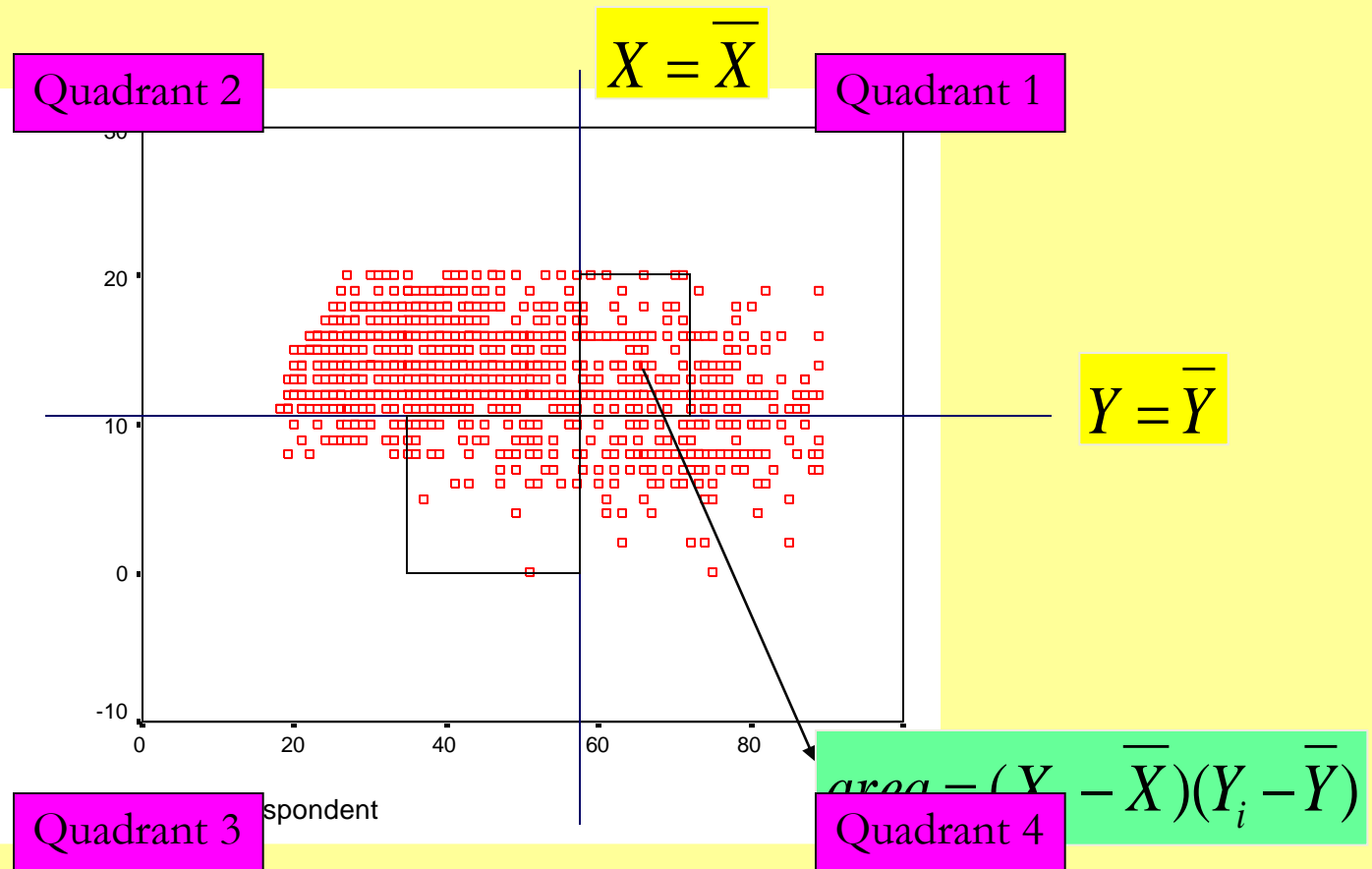
# หัวข้อสำคัญ

- การวิเคราะห์สหสัมพันธ์ (Correlation)
- การเปรียบเทียบสหสัมพันธ์ (Comparing Correlation Coefficient)
- การวิเคราะห์ถดถอยอย่างง่าย (Simple Regression)
- การวิเคราะห์ถดถอยแบบพหุ (Multiple Regression)

# การวิเคราะห์สหสัมพันธ์

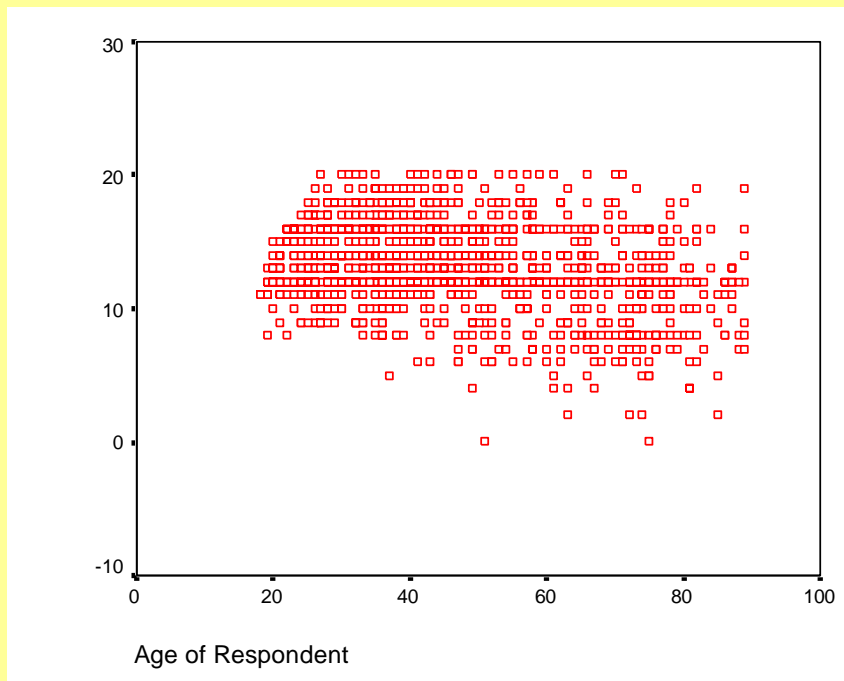
- ความสัมพันธ์ในเชิงสถิติ คือ เมื่อตัวแปรหนึ่งเปลี่ยนแปลงไป อีกตัวแปรหนึ่งเปลี่ยนแปลงด้วย
- การวิเคราะห์สหสัมพันธ์เป็นตัวชี้วัดหนึ่งที่ใช้ทดสอบว่าตัวแปรสองตัวหรือหลายตัวสัมพันธ์กันเท่าไร

# การวิเคราะห์ห้สหสัมพันธ์



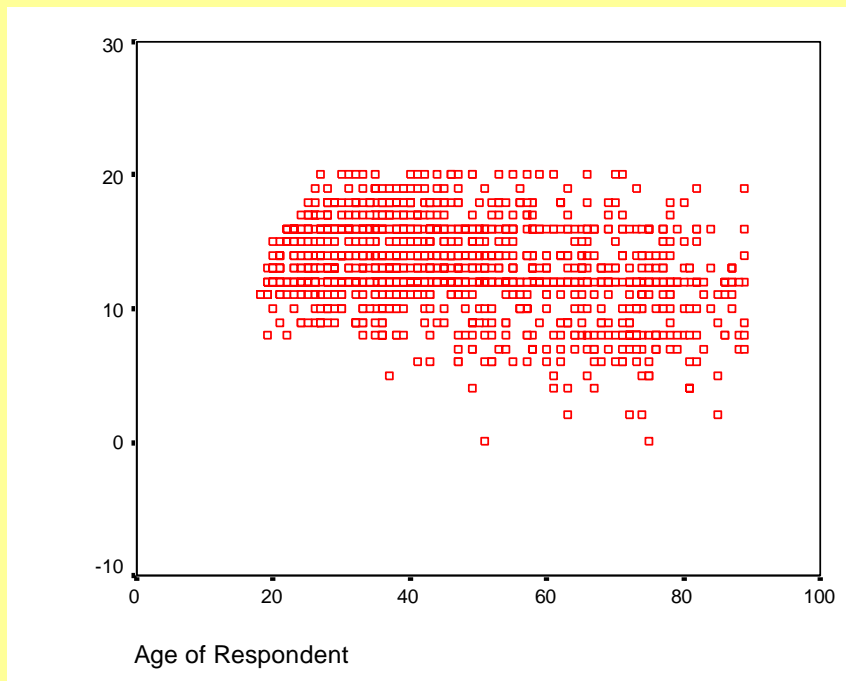
# การวิเคราะห์สหสัมพันธ์

- ผลคูณไขว้ (Cross product)  
คือ ผลรวมของค่าเบี่ยงเบนจาก  
ค่าเฉลี่ยร่วมของตัวแปรทั้งสอง



$$CP = \sum (X - \bar{X})(Y - \bar{Y})$$

# การวิเคราะห์สหสัมพันธ์



- ถ้านำผลคูณไขว้ มาเฉลี่ย (คล้ายๆ กับค่าความแปรปรวนที่นำผลรวมของค่าเบี่ยงเบน [Sum of square] มาเฉลี่ย) จะได้ความแปรปรวนร่วมเกี่ยว (Covariance)

$$\text{COV}_{XY} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n - 1}$$

# การวิเคราะห์สหสัมพันธ์

- อย่างไรก็ตาม ความแปรปรวนร่วมเกี่ยวระหว่างคู่ตัวแปรหนึ่ง อาจไม่สามารถเปรียบเทียบกับความแปรปรวนร่วมเกี่ยวระหว่างอีกคู่ตัวแปรหนึ่งได้ เพราะว่าหน่วยการวัดคนละหน่วยกัน
- ด้วยเหตุนี้ จึงต้องทำให้ค่านี้เป็นมาตรฐานก่อนที่จะเปรียบเทียบได้

# การวิเคราะห์สหสัมพันธ์

- การทำให้เป็นมาตรฐานทำได้โดยนำความแปรปรวนร่วม  
เกี่ยวข้องด้วยส่วนเบี่ยงเบนมาตรฐานของตัวแปรทั้งสอง  
จะได้ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (Pearson  
Product Moment Correlation)

$$r_{XY} = \frac{COV_{XY}}{S_X \cdot S_Y}$$



# การวิเคราะห์สหสัมพันธ์

- สูตรอื่นของค่าสัมประสิทธิ์สหสัมพันธ์

$$r_{XY} = 1 - \left( \frac{\sum (z_X - z_Y)^2}{2(n-1)} \right)$$

$$r_{XY} = \frac{\sum z_X z_Y}{n-1}$$

# การวิเคราะห์สหสัมพันธ์

- การตีความหมายค่าสหสัมพันธ์
  - สหสัมพันธ์นี้แตกต่างจากศูนย์หรือไม่
  - สหสัมพันธ์นี้มีทิศทางใด
  - สหสัมพันธ์นี้มีขนาดความสัมพันธ์เท่าไร
  - ค่าสัมประสิทธิ์การกำหนด (Coefficient of determination) บอกว่าตัวแปรหนึ่งสามารถอธิบายความแปรปรวนของอีกตัวแปรหนึ่งได้เท่าใด

# การวิเคราะห์สหสัมพันธ์

- การตีความหมายค่าสหสัมพันธ์
  - ค่าสหสัมพันธ์ที่ได้สามารถนำไปอ้างอิงถึงประชากร ด้วยวิธีการทดสอบสมมติฐาน (Hypothesis Testing) หรือการประมาณค่าพารามิเตอร์ (Parameter Estimation)
  - การทดสอบสมมติฐาน จะตั้งสมมติฐานหลักดังนี้

$$H_0 : \rho_{XY} = 0$$

# การวิเคราะห์สหสัมพันธ์

- การตีความหมายค่าสหสัมพันธ์

— การทดสอบจะใช้สูตรทั่วไปของ t-test ดังนี้

จาก

$$t = \frac{\text{statistics} - \text{parameter}}{SE_{\text{statistics}}}$$

จะได้

$$t = \frac{r_{XY}}{\sqrt{\frac{1 - r_{XY}^2}{n - 2}}}$$

# การวิเคราะห์สหสัมพันธ์

- การตีความหมายค่าสหสัมพันธ์
  - ค่าองศาอิสระจะเท่ากับ  $n-2$  แล้วสามารถเปิดตารางเพื่อหาค่า  $p$  ได้
  - หนังสือสถิติทั่วไปจะเสนอค่า  $r$  และจะบอกค่า critical value ตามค่านัยสำคัญระดับต่างๆ ให้เลย
  - ในโปรแกรมคอมพิวเตอร์จะนำเสนอค่า  $r$ ,  $df$  และ  $p$  ให้เลย

# การวิเคราะห์สหสัมพันธ์

- การตีความหมายค่าสหสัมพันธ์
  - การประมาณค่าพารามิเตอร์ (Parameter Estimation)  
เนื่องจากการ Sampling distribution ของค่าสหสัมพันธ์ไม่ได้เป็นโค้งสมมาตร (ยกเว้นแต่กรณี  $\rho = 0$  ที่การกระจายเป็นสมมาตร)
  - ด้วยเหตุนี้ Fisher จึงคิดวิธีการแปลงค่าสหสัมพันธ์ให้มีการกระจายของ Sampling distribution เป็นรูปโค้งปกติ มีสูตรการแปลงดังนี้

# การวิเคราะห์สหสัมพันธ์

- การตีความหมายค่าสหสัมพันธ์
  - การประมาณค่าพารามิเตอร์

$$z' = \frac{1}{2} \ln \left[ \frac{1+r}{1-r} \right]$$

- การกระจายสามารถทำได้เหมือนกับการทำ Confidence Interval ในค่า  $z$  ทั่วไปคือ

$$CI_{100(1-\alpha)\%} = z' \pm z_{\alpha/2} SE_{z'} \quad \text{โดยที่} \quad SE_{z'} = 1/\sqrt{n-3}$$

# การวิเคราะห์สหสัมพันธ์

- ประเด็นที่สำคัญของค่าสหสัมพันธ์
  - พิสัยจำกัด (Range Restriction) ในกรณีที่พิสัยของข้อมูลถูกจำกัด จาก  $X$  เหลือ  $X_C$  ผู้วิจัยหาความสัมพันธ์ระหว่าง  $Y$  กับ  $X_C$  ได้เท่านั้น จะมีสูตรประมาณค่าสหสัมพันธ์ระหว่าง  $Y$  กับ  $X$  ได้ดังนี้

$$\tilde{r}_{YX} = \frac{r_{YX_C} (sd_X / sd_{X_C})}{\sqrt{1 + r_{YX_C}^2 ((sd_X^2 / sd_{X_C}^2) - 1)}}$$



# การวิเคราะห์สหสัมพันธ์

- ประเด็นที่สำคัญของค่าสหสัมพันธ์
  - ความเที่ยง (Reliability) ถ้ามาตรที่ใช้หาสหสัมพันธ์มีความเที่ยงต่ำ ส่งผลให้ค่าความคาดเคลื่อนแบบสุ่มสูง ทำให้ค่าสหสัมพันธ์ที่ได้น้อยกว่าความเป็นจริง สูตรการปรับแก้ค่าความเที่ยง (Correction of Attenuation) เป็นดังนี้

$$r_{Y_t X_t} = \frac{r_{XY}}{\sqrt{r_{XX} r_{YY}}}$$

# การวิเคราะห์สหสัมพันธ์

- ประเด็นที่สำคัญของค่าสหสัมพันธ์
  - กลุ่มหาความสัมพันธ์แตกต่างกัน (Heterogeneity of samples)
  - ค่าสุดโต่ง (Outlier)
  - ความสัมพันธ์ที่ไม่ใช่เส้นตรง (Nonlinearity)
  - ความเป็นเหตุเป็นผล (Causation)

# การวิเคราะห์สหสัมพันธ์

- การลดรูปของค่าสหสัมพันธ์ของเพียร์สัน
  - Point-biserial correlation

$$r_{pb} = \frac{(M_{Y_1} - M_{Y_0})\sqrt{pq}}{sd_Y}$$

# การวิเคราะห์สหสัมพันธ์

- การลดรูปของค่าสหสัมพันธ์ของเพียร์สัน

— Phi coefficient

A	B
C	D

$$r_{\phi} = \frac{BC - AD}{\sqrt{(A + B)(C + D)(A + C)(B + D)}}$$

— Spearman rank correlation

# การวิเคราะห์สหสัมพันธ์

- การลดรูปของค่าสหสัมพันธ์ของเพียร์สัน
  - Spearman rank correlation

$$r_s = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}$$

# การวิเคราะห์สหสัมพันธ์

- Biserial Correlation

— Point biserial correlation จะเป็นความสัมพันธ์ระหว่าง true dichotomous กับ interval แต่ biserial correlation จะเป็นระหว่าง forced dichotomous (assumed that the underlying continuous variable is normal distributed) กับ interval

$$r_b = \frac{(M_{Y_P} - M_{Y_Q})pq}{h(sd_Y)} = r_{pb} \frac{\sqrt{pq}}{h}$$

# การวิเคราะห์สหสัมพันธ์

- การเปรียบเทียบค่าสหสัมพันธ์ระหว่างกลุ่ม (Independent testing for difference between correlation)

$$z = \frac{z'_V - z'_W}{\sqrt{\frac{1}{n_V - 3} + \frac{1}{n_W - 3}}}$$

# การวิเคราะห์สหสัมพันธ์

- การเปรียบเทียบค่าสหสัมพันธ์ระหว่างตัวแปรตามและตัวแปรต้นสองตัว (Comparing two correlated correlation)

$$z = (z_{r_1} - z_{r_2}) \sqrt{\frac{n-3}{2(1-r_x)h}}$$

$$h = \frac{1 - f\bar{r}^2}{1 - r^2} = 1 + \frac{\bar{r}^2}{1 - r^2}(1 - f)$$

$$f = \frac{1 - r_x}{2(1 - r^2)}$$

$$\bar{r}^2 = \frac{r_1^2 + r_2^2}{2}$$

ค่า  $f$  ควรมีค่าน้อยกว่า 1 ถ้ามากกว่า 1 ให้ปัดเป็น 1



# การวิเคราะห์ถดถอยอย่างง่าย

- ในการวิเคราะห์สหสัมพันธ์จะหาเพียงว่าตัวแปรสองตัวสัมพันธ์ด้วยกันเท่าใด
- แต่ในการวิเคราะห์ถดถอยจะสนใจที่นำข้อมูลของตัวหนึ่งไปอธิบายค่าของอีกตัวแปรหนึ่ง ตัวแปรแรกเรียกว่าตัวแปรทำนาย (Predictor Variable) และตัวแปรหลังเรียกว่าตัวแปรเกณฑ์ (Criterion Variable)

# การวิเคราะห์ถดถอยอย่างง่าย

- การนำข้อมูลของตัวแปรหนึ่งไปอธิบายอีกตัวแปรหนึ่ง  
อาจพูดได้สองนัย
  - การอธิบาย (Explanation)
  - การทำนาย (Prediction)

# การวิเคราะห์ถดถอยอย่างง่าย

- สมมติว่ามีข้อมูลของตัวแปรหนึ่ง ค่าที่ใช้ในการทำนายได้ดีที่สุดคือค่าเฉลี่ย
- ทำไมถึงดีที่สุด เพราะว่าทำให้ผลรวมของค่าผิดพลาดในการทำนาย (Sum of error) น้อยที่สุด และยังส่งผลให้ค่าผลรวมของค่าผิดพลาดในการทำนายยกกำลังน้อยที่สุดด้วย (Sum of squared error)

# การวิเคราะห์ห้ถดถอยอย่างง่าย

ID	NCCU	Baseline Prediction	Error	Squared error
1	4	7	-3	9
2	6	7	-1	1
3	6	7	-1	1
4	7	7	0	0
5	8	7	+1	1
6	7	7	0	0
7	8	7	+1	1
8	10	7	+3	9
Total	56		0	22

NCCU = Number of Credit Cards Used

# การวิเคราะห์ถดถอยอย่างง่าย

- แต่อย่างไรก็ตาม ถ้าเรามีข้อมูลของตัวแปรอีกตัวแปรหนึ่ง ก็จะทำให้อธิบายหรือทำนายตัวแปรนั้นได้ดีขึ้น
- ยิ่งตัวแปรไหนสัมพันธ์กับตัวแปรที่ต้องการทำนายมาก จะยิ่งทำนายได้แม่นยำมากขึ้น เพราะยิ่งสัมพันธ์ก็ยิ่งอธิบายความแปรปรวนได้มาก
- หมายเหตุ การอธิบายความแปรปรวน ก็คือการอธิบาย การเปลี่ยนแปลงของอีกตัวแปรหนึ่งได้

# การวิเคราะห์ถดถอยอย่างง่าย

- การทำนาย จะต้องเปลี่ยนแปลงตัวแปรทำนายให้ใกล้เคียงกับตัวแปรเกณฑ์ให้มากที่สุด (ให้  $SS_{\text{error}}$  น้อยที่สุด) โดยการเปลี่ยนแปลงเชิงเส้น (Linear transformation)
- การเปลี่ยนแปลงเชิงเส้น จะคูณตัวแปรนั้นด้วยค่าคงที่ และบวกด้วยค่าคงที่ (เป็นการแปลงตัวแปรระดับอันตรภาคให้อยู่ระดับอันตรภาคเหมือนเดิม) ดังสมการ

$$X' = a + bX$$

# การวิเคราะห์ถดถอยอย่างง่าย

- ดังนั้นต้องหาค่า  $a$  และค่า  $b$  ที่ทำให้  $SS_{\text{error}}$  น้อยที่สุด
- ค่า  $a$  และ  $b$  ที่ทำให้ความแตกต่างระหว่างค่าตัวแปรทำนายที่แปลงเชิงเส้นไป และค่าตัวแปรเกณฑ์น้อยที่สุด  
เป็นดังสูตร

$$b = \frac{\text{COV}_{XY}}{s_X^2}$$

$$a = \bar{Y} - b\bar{X}$$

# การวิเคราะห์ถดถอยอย่างง่าย

- จากตัวอย่างที่แล้ว ในการทำนายจำนวนบัตรเครดิตที่ใช้ มีตัวแปรที่อาจใช้ทำนายได้ 3 ตัว

	1	2	3	4
1. NCCU				
2. Family Size	.87			
3. Family Income	.83	.67		
4. Number of Automobiles	.34	.19	.30	

- ตัวแปรที่ดีที่สุดในการทำนาย คือ ขนาดของครอบครัว



# การวิเคราะห์หาคัดถอยอย่างง่าย

Prediction Equation:  $\hat{Y} = 2.87 + .97X$

ID	NCCU	Family Size	Prediction Score	Error of Prediction	Error Squared
1	4	2	4.81	-.81	.66
2	6	2	4.81	1.19	1.42
3	6	4	6.75	-.75	.56
4	7	4	6.75	.25	.06
5	8	5	7.72	.28	.08
6	7	5	7.72	-.72	.52
7	8	6	8.69	-.69	.48
8	10	6	8.69	1.31	1.72
Total	56			0	5.50

NCCU = Number of Credit Cards Used

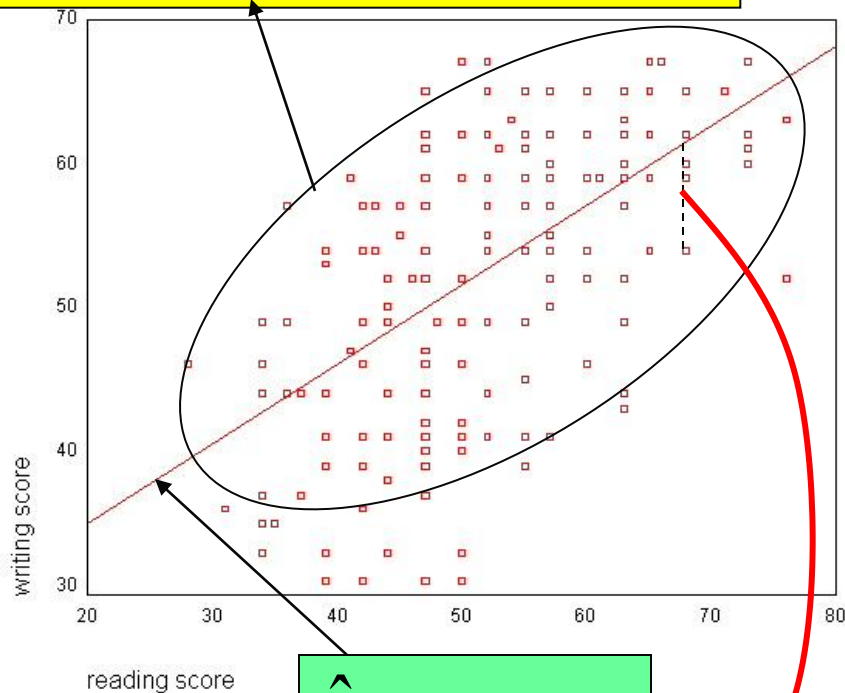
# การวิเคราะห์ถดถอยอย่างง่าย

- จะเห็นว่าค่า  $SS_{error}$  ลดลงจาก 22 เมื่อใช้ค่าเฉลี่ยในการทำนาย เป็น 5.5 เมื่อใช้ขนาดของครอบครัวในการทำนาย
- ค่าสัมประสิทธิ์การกำหนด (Coefficient of determination) คือ ค่าที่แสดงให้เห็นว่าตัวแปรทำนายสามารถอธิบายตัวแปรเกณฑ์ได้กี่เปอร์เซ็นต์ (เท่ากับค่าสหสัมพันธ์ยกกำลังสอง)

$$r^2 = \frac{SS_{total} - SS_{error}}{SS_{total}} = \frac{SS_{regression}}{SS_{total}} = \frac{22 - 5.5}{22} = .75$$

# การวิเคราะห์ห้ถดถอยอย่างง่าย

การเกาะกลุ่มคือขนาดความสัมพันธ์  
หรือความแม่นยำในการทำนาย



$$\hat{Y} = a + bX$$

ระยะห่างระหว่างจุดและเส้น คือ  
ค่าความผิดพลาดในการทำนาย

- รูปของการเปลี่ยนแปลงเชิงเส้นในตัวแปรทำนายจะสามารถวาดเป็นเส้นได้ตามการเปลี่ยนแปลงเชิงเส้นที่ได้จัดกระทำไป
- $a$  คือจุดตัดแกน  $Y$
- $b$  คือค่าความชันของเส้นหรือค่าทำนายของ  $Y$  ที่เปลี่ยนแปลงไปเมื่อ  $X$  เปลี่ยนแปลงไป 1 หน่วย
- เส้นนี้จะผ่านจุดค่าเฉลี่ยของ  $X$  และ  $Y$  เสมอ

# การวิเคราะห์ถดถอยอย่างง่าย

- การวิเคราะห์ห้ อาจทำได้โดยการแปลงตัวแปร  $X$  และ  $Y$  เป็นค่ามาตรฐานก่อน แล้วจึงนำมาวิเคราะห์ถดถอย
- ผลการวิเคราะห์ เส้นถดถอยจะผ่านจุดกำเนิด (Origin) เสมอ เนื่องจากเส้นนี้ผ่านจุดค่าเฉลี่ย  $X$  และ  $Y$  ซึ่งเป็นค่า  $(0, 0)$  ใน  $zX$  และ  $zY$  ทำให้ค่า  $a = 0$
- ส่วนค่าความชันจะเท่ากับค่าสหสัมพันธ์ (ลองแทนสูตรดู)
- ได้สมการ  $\hat{Z}_Y = \beta Z_X$

# การวิเคราะห์ห้ถดถอยอย่างง่าย

- ค่า  $\beta$  หมายถึง ถ้า  $X$  เปลี่ยนแปลงไป 1 SD จะทำให้  $Y$  เปลี่ยนแปลงไป  $\beta$  SD

$$\hat{Z}_Y = \beta Z_X$$

# การวิเคราะห์ถดถอยอย่างง่าย

- ค่าส่วนเบี่ยงเบนมาตรฐานของการทำนาย (Standard error of estimate) คือค่าส่วนเบี่ยงเบนมาตรฐานของความผิดพลาดในการทำนาย (error or residual) มีสูตรดังนี้

$$SE_{Y-\hat{Y}} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{(1 - r_{XY}^2) \sum (Y - \bar{Y})^2}{n-2}}$$

- ค่านี้จะมากหรือน้อยขึ้นกับส่วนเบี่ยงเบนมาตรฐานของ Y และค่าสหสัมพันธ์ของ X และ Y

# การวิเคราะห์ห้ถดถอยอย่างง่าย

- ค่านี้จะเป็นตัวบ่งชี้ถึงความแม่นยำในการทำนายได้
- นอกจากนี้ยังสามารถนำค่านี้ไปใช้ในการทำนายค่า  $Y$  จากค่า  $X$  ให้เป็นช่วงได้ ดังนี้

$$CI_{\alpha, \hat{Y}_i} = \hat{Y}_i \pm t_{\alpha/2, n-2} SE_{\hat{Y}_i}$$

# การวิเคราะห์ถดถอยอย่างง่าย

- สามารถหาการกระจายของค่าที่ทำนายได้แต่ละตัวดัง  
สูตร

$$SE_{\hat{Y}_i} = SE_{Y-\hat{Y}} \sqrt{\frac{1}{n} + \frac{(X_i - \bar{X})^2}{(n-1)sd_X^2}}$$

- ถ้า  $X$  เป็นค่าเฉลี่ย จะทำให้ค่าความคาดเคลื่อนน้อยที่สุด (เพราะว่าค่า  $b$  เป็นค่าประมาณพารามิเตอร์ มีโอกาสผิดพลาดอยู่ ยิ่งคุณมาก (ห่างจากค่าเฉลี่ยมาก) จะได้รับผลกระทบจากค่านี้มาก



# การวิเคราะห์หัตถถอยอย่างง่าย

- การอ้างอิงถึงประชากร ในการวิเคราะห์หัตถถอย สามารถทดสอบนัยสำคัญทางสถิติได้ 3 รูปแบบ

$$H_0 : \rho = 0$$

ทดสอบว่า X มีความสัมพันธ์กับ Y หรือไม่

$$H_0 : \beta_0 = 0$$

ทดสอบว่าจุดตัดแกน Y แตกต่างจากศูนย์หรือไม่

$$H_0 : \beta_{YX} = 0$$

ทดสอบว่าการเปลี่ยนแปลงใน X สามารถทำให้ค่าของ Y ที่ทำนายได้เปลี่ยนแปลงไปหรือไม่

# การวิเคราะห์ถดถอยอย่างง่าย

- การอ้างอิงถึงประชากรอาจใช้สูตรทั่วไปของการทดสอบค่าทีคือ

$$t = \frac{\text{statistics} - \text{parameter}}{SE_{\text{statistics}}}$$

ค่าความเคลื่อนมีสูตรดังนี้

$$SE_{\beta_0} = SE_{Y-\hat{Y}} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{(n-1)s_X^2}}$$

$$SE_{\beta_{YX}} = \frac{s_Y}{s_X} \sqrt{\frac{1-r_{YX}^2}{n-2}}$$

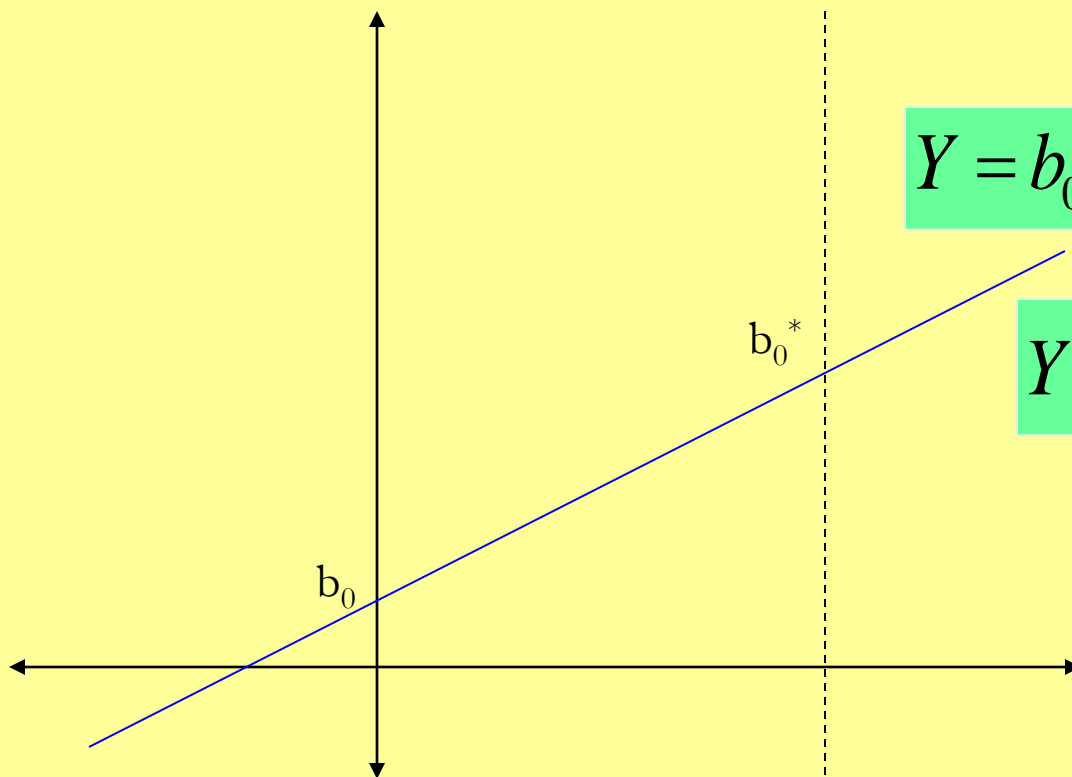
# การวิเคราะห์ถดถอยอย่างง่าย

- การกำหนดค่าศูนย์กลาง (Centering)

ในการสร้างสมการทำนายบางครั้งจุดตัดแกน  $Y$  ไม่มีความหมายในความเป็นจริง นักวิจัยจึงอาจย้ายแกน  $Y$  เพื่อให้จุดตัดแกน  $Y$  มีความหมาย เช่น ย้ายจากเส้น  $X = 0$  เป็น  $X = M(X)$  แทน ซึ่งจะทำให้จุดตัดแกน  $Y$  ใหม่ มาตัดที่จุด  $M(Y)$  แทน

# การวิเคราะห์ห้ถดถอยอย่างง่าย

- การกำหนดค่าศูนย์กลาง (Centering)



$$Y = b_0 + b_1 X$$

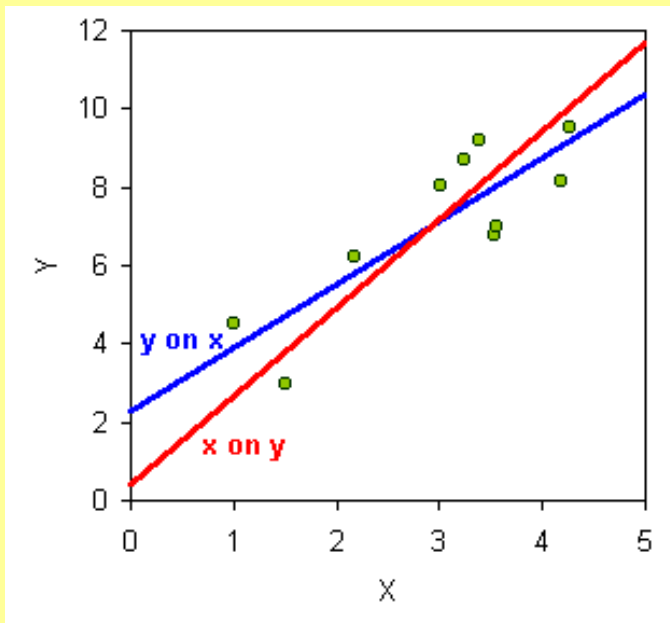
$$Y = b_0^* + b_1 (X - \bar{X})$$

เนื่องจากเส้นถดถอยจะผ่านจุด  
 $M(X)$  และ  $M(Y)$  เสมอทำ  
ให้

$$b_0^* = \bar{Y}$$

# การวิเคราะห์ถดถอยอย่างง่าย

- การสลับที่ ให้  $X$  เป็นตัวแปรเกณฑ์ และ  $Y$  เป็นตัวแปรทำนาย จะทำให้ได้สมการทำนายคนละเส้น แต่อ่านาจทำนายเท่าเดิม ( $r$  เท่าเดิม) ดังนี้



**Regress Y on X:**

**Y = Criterion, X = Predictor**

**Regress X on Y:**

**X = Criterion, Y = Predictor**

# การวิเคราะห์ถดถอยแบบพหุ

- จากการวิเคราะห์ถดถอยอย่างง่าย ถ้าเรามีตัวแปรทำนายมากกว่า 1 ตัว น่าจะช่วยให้ทำนายค่าตัวแปรเกณฑ์ได้ดีขึ้น
- ตัวแปรทำนายตัวไหนที่สามารถอธิบายความแปรปรวนที่เหลือ (Residual Variance) ได้ดีที่สุด ตัวแปรนั้นน่าจะเป็นตัวแปรที่ใส่เพิ่มเติมลงในสมการ (ทำให้  $SS_{\text{error}}$  เหลือน้อยที่สุด)

# การวิเคราะห์ถดถอยแบบพหุ

- การเปลี่ยนแปลงเชิงเส้นคราวนี้ จะต้องอยู่ในรูปการแปลงตัวแปร 2 ตัวแปรให้เกิดค่าที่ทำนายได้ค่าหนึ่ง การแปลงรูปแบบนี้จะเรียกว่า การรวมเชิงเส้น (Linear Combination หรือ Variate) ซึ่งจะส่งผลให้ระดับอันตรภาคของตัวแปรทำนายแต่ละตัวอยู่เหมือนเดิม ดังสมการ

$$\hat{Y} = b_0 + b_{Y1.2}X_1 + b_{Y2.1}X_2$$

# การวิเคราะห์ถดถอยแบบพหุ

- ดังนั้นต้องหาค่าสัมประสิทธิ์ถดถอยแต่ละตัวที่ทำให้  $SS_{\text{error}}$  น้อยที่สุด
- ค่าที่ทำให้ความแตกต่างระหว่างค่าตัวแปรทำนายที่เปลี่ยนแปลงเส้นไป และค่าตัวแปรเกณฑ์น้อยที่สุดเป็นดังสูตร

$$b_{Y1.2} = \left( \frac{r_{Y1} - r_{Y2}r_{12}}{1 - r_{12}^2} \right) \frac{s_Y}{s_1}$$

$$b_{Y2.1} = \left( \frac{r_{Y2} - r_{Y1}r_{12}}{1 - r_{12}^2} \right) \frac{s_Y}{s_2}$$

$$b_0 = \bar{Y} - b_{Y1.2} \bar{X}_1 - b_{Y2.1} \bar{X}_2$$



# การวิเคราะห์ถดถอยแบบพหุ

- จากตัวอย่างที่แล้ว ในการทำนายจำนวนบัตรเครดิตที่ใช้ มีตัวแปรที่อาจใช้ทำนายได้เพิ่มเติม 2 ตัว

	1	2	3	4
1. NCCU				
2. Family Size	.87			
3. Family Income	.83	.67		
4. Number of Automobiles	.34	.19	.30	

- ตัวแปรรายได้ครอบครัวและจำนวนรถยนต์ทำนายความแปรปรวนส่วนที่เหลือได้ 45 % และ 13 % ตามลำดับ จึงเลือกรายได้ครอบครัวเป็นตัวแปรทำนายตัวที่สอง

# การวิเคราะห์ถดถอยแบบพหุ

Prediction Equation:  $\hat{Y} = .482 + .63X_1 + .216X_2$

ID	NCCU	Family Size	Family Income	Prediction Score	Error of Prediction	Error Squared
1	4	2	14	4.76	-.76	.58
2	6	2	16	5.20	.80	.64
3	6	4	14	6.03	-.03	.00
4	7	4	17	6.68	.32	.10
5	8	5	18	7.53	.47	.22
6	7	5	21	8.18	-1.18	1.39
7	8	6	17	7.95	.05	.00
8	10	6	25	9.67	.33	.11
Total	56				0	3.04

NCCU = Number of Credit Cards Used

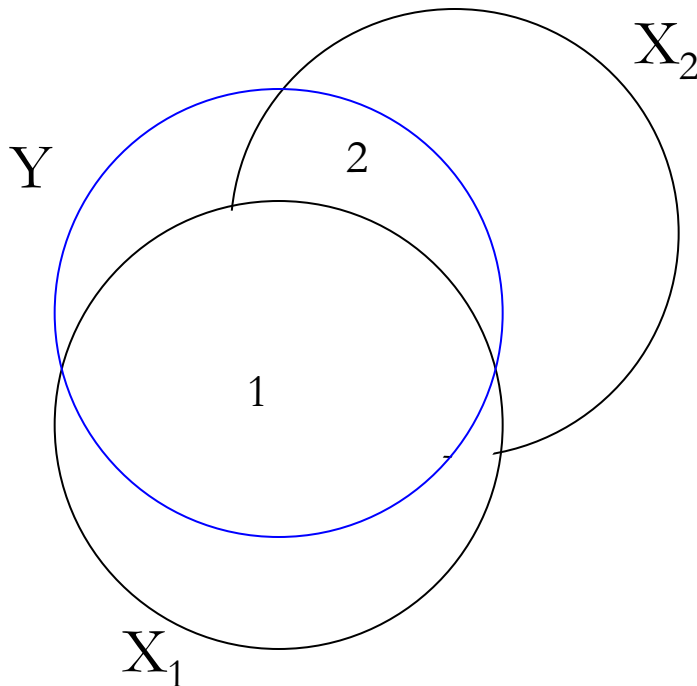
# การวิเคราะห์ถดถอยแบบพหุ

- จะเห็นว่าค่า  $SS_{error}$  ลดลงจาก 5.5 เมื่อใช้ขนาดครบคร้วในการทำนาย เป็น 3.04 เมื่อใช้ขนาดของครบคร้วและรายได้ครบคร้วร่วมกันในการทำนาย
- ค่าสัมประสิทธิ์การกำหนด (Coefficient of determination) มีค่าดังนี้

$$r^2 = \frac{SS_{total} - SS_{error}}{SS_{total}} = \frac{SS_{regression}}{SS_{total}} = \frac{22 - 3.04}{22} = .86$$

# การวิเคราะห์ห้ถดถอยแบบพหุ

- สังเกตภาพการอธิบายความแปรปรวน เมื่อใส่ตัวแปร  
รายได้ครบถ้วนเพิ่ม



$$X_1 \rightarrow Y : 75\%$$

$$X_1, X_2 \rightarrow Y : 86\%$$

# การวิเคราะห์ถดถอยแบบพหุ

- การวิเคราะห์ถดถอยแบบพหุ เป็นการใช้ข้อมูลจากตัวแปรทำนายหลายตัวมาสร้างผลรวมเชิงเส้นเพื่อให้ทำนายค่าตัวแปรเกณฑ์ได้ดีที่สุด ( $SS_{\text{error}}$  น้อยที่สุด) มีสมการทั่วไปดังนี้

$$\hat{Y} = b_0 + b_{Y1}X_1 + b_{Y2}X_2 + \dots + b_{Yn}X_n$$

# การวิเคราะห์ถดถอยแบบพหุ

- ค่า  $b_0$  หมายถึง ค่าของ  $Y$  ที่ทำนายได้ ในกรณีที่  $X_1, X_2, \dots, X_n = 0$  (เหมือนจุดตัดแกน  $Y$ )
- ค่า  $b_{Y_n}$  หมายถึง ค่าของ  $Y$  ที่ทำนายได้ ที่เปลี่ยนแปลงไป เมื่อ  $X_n$  เปลี่ยนแปลงไป 1 หน่วย และควบคุมให้ค่าตัวแปรทำนายตัวอื่นคงที่

# การวิเคราะห์ถดถอยแบบพหุ

- ค่าสัมประสิทธิ์การทำนาย ( $R^2$ ) สามารถหาได้โดยนำค่า  $Y$  จริง และค่า  $Y$  ที่ทำนายได้มาหาค่าสหสัมพันธ์แบบเพียร์สัน แล้วยกกำลังสอง เช่นเดียวกับสูตรที่กล่าวไปข้างต้น คือ

$$R^2 = \frac{SS_{regression}}{SS_{total}} = \frac{SS_{total} - SS_{error}}{SS_{total}} = r_{Y\hat{Y}}^2$$

# การวิเคราะห์ถดถอยแบบพหุ

- บางครั้งอาจแปลงค่าตัวแปรทุกตัว ให้เป็นค่ามาตรฐานก่อน แล้วจึงนำมาสร้างสมการถดถอย ค่า  $b_0$  จะหายไป (เนื่องจากสมการตัดจุดค่าเฉลี่ยของทุกตัว) แล้วค่าความชันของแต่ละตัวจะเปลี่ยนแปลงไป

$$\hat{z}_Y = \beta_{Y1}\hat{z}_1 + \beta_{Y2}\hat{z}_2 + \dots + \beta_{Yn}\hat{z}_n$$



# การวิเคราะห์ถดถอยแบบพหุ

- ค่า  $\beta$  (สัมประสิทธิ์ถดถอยมาตรฐาน: Standardized regression coefficient) หมายความว่า ถ้าตัวแปร  $X$  ตัวหนึ่งเปลี่ยนแปลงไป 1 ส่วนเบี่ยงเบนมาตรฐาน จะส่งผลให้ค่า  $Y$  ที่ทำนายเปลี่ยนแปลงไปก็ส่วนเบี่ยงเบนมาตรฐาน เมื่อควบคุมให้ตัวแปรทำนายอื่นคงที่
- ด้วยเหตุนี้จึงสามารถเปรียบเทียบอำนาจของแต่ละตัวแปรที่ส่งผลต่อตัวแปรตามได้

# การวิเคราะห์ห้ถดถอยแบบพหุ

- ค่าสัมประสิทธิ์ถดถอยมาตรฐานด้วย Multiple regression โดยปกติจะน้อยกว่าสัมประสิทธิ์ถดถอยมาตรฐานเมื่อวิเคราะห์ด้วย Simple regression เพราะว่าตัวแปรต้นสัมพันธ์กัน
- ถ้าตัวแปรต้นไม่สัมพันธ์กันเลย (ข้อมูลที่อธิบายได้ไม่ซ้อนกัน) จะทำให้อำนาจการทำนายของตัวแปรต้นทั้งสองรวมกันสูงสุด

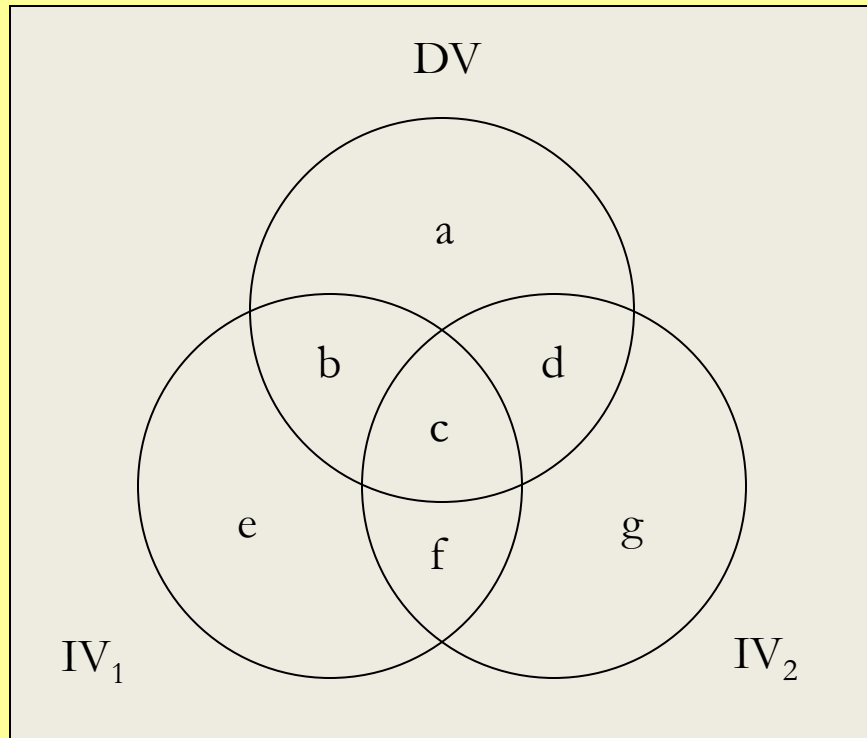
# การวิเคราะห์ถดถอยแบบพหุ

- แต่ถ้าตัวแปรต้นสัมพันธ์กันมาก จะส่งผลให้ตัวแปรที่ใส่เข้าไป ไม่สามารถอธิบายได้เพิ่มเติมมาก ค่าสัมประสิทธิ์ถดถอยมาตรฐานจะลดลงมาก
- อย่างไรก็ตาม ไม่ควรนำสัมประสิทธิ์นี้ไปเปรียบเทียบข้ามสมการ เพราะว่าถ้าเปลี่ยนชุดตัวแปรต้นแล้ว จะทำให้ค่าสัมประสิทธิ์นี้เปลี่ยนแปลงไปมาก

# การวิเคราะห์ถดถอยแบบพหุ

- การเปรียบเทียบความสำคัญของแต่ละตัวแปรต้น ใช้เพียงสัมประสิทธิ์ถดถอยมาตรฐานเพียงอย่างเดียว อาจบิดเบือนความเป็นจริง การดูอาจใช้ข้อมูลอื่นประกอบ คือ
  - ค่าสัมประสิทธิ์สหสัมพันธ์อย่างง่าย (Simple or Zero-order correlation:  $r$ )
  - Semipartial or Part Correlation ( $s_r$ )
  - สหสัมพันธ์บางส่วน (Partial Correlation:  $p_r$ )
  - สหสัมพันธ์ระหว่างตัวแปรทำนาย และค่าที่ทำนายได้ (Loading for Regression Variate)

# การวิเคราะห์ถดถอยแบบพหุ



$$R_{Y.12}^2 = \frac{b+c+d}{a+b+c+d}$$

$$r_{Y1}^2 = \frac{b+c}{a+b+c+d}$$

$$sr_{Y1}^2 = \frac{b}{a+b+c+d}$$

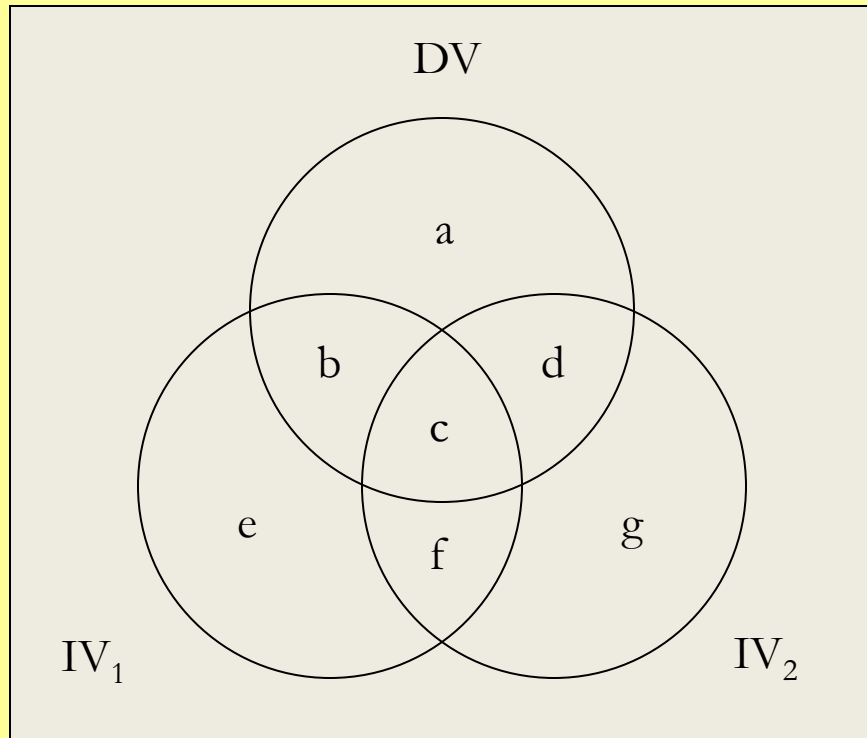
$$pr_{Y1}^2 = \frac{b}{a+b}$$

$$r_{Y2}^2 = \frac{c+d}{a+b+c+d}$$

$$sr_{Y2}^2 = \frac{d}{a+b+c+d}$$

$$pr_{Y2}^2 = \frac{d}{a+d}$$

# การวิเคราะห์หัดถดถอยแบบพหุ



$$R_{Y.12}^2 = \frac{b+c+d}{a+b+c+d}$$

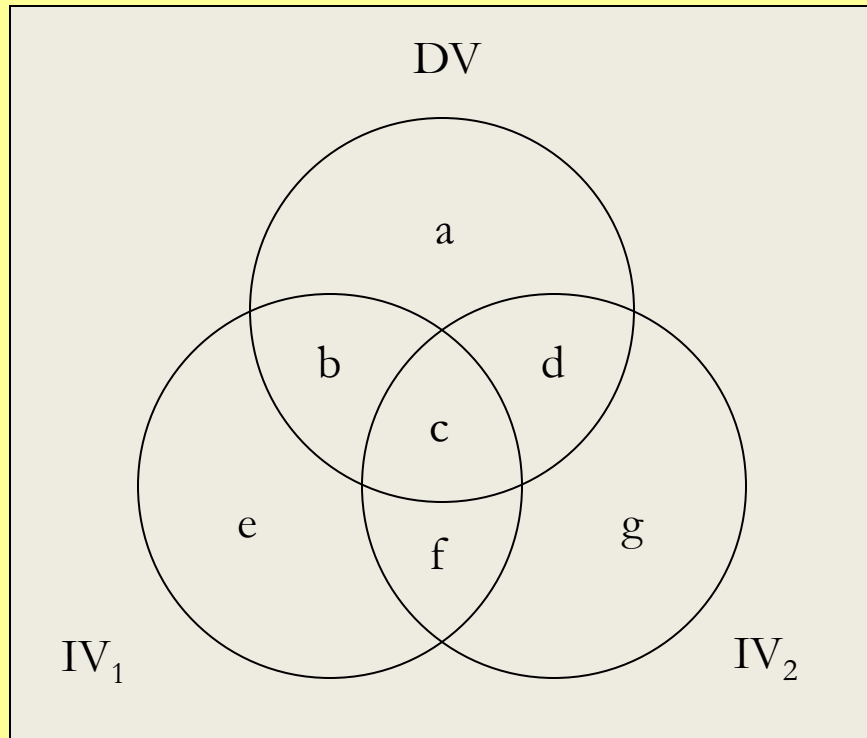
$$sr_{Y1}^2 = R_{Y.12}^2 - r_{Y2}^2$$

$$sr_{Y2}^2 = R_{Y.12}^2 - r_{Y1}^2$$

$$pr_{Y1}^2 = \frac{R_{Y.12}^2 - r_{Y2}^2}{1 - r_{Y2}^2}$$

$$pr_{Y2}^2 = \frac{R_{Y.12}^2 - r_{Y1}^2}{1 - r_{Y1}^2}$$

# การวิเคราะห์ห้ถดถอยแบบพหุ



$$R_{Y.12}^2 = \frac{b+c+d}{a+b+c+d}$$

$$sr_{Y1} = r_{Y(X_1 - \hat{X}_{1.2})}$$

$$sr_{Y2} = r_{Y(X_2 - \hat{X}_{2.1})}$$

$$pr_{Y1} = r_{(Y - \hat{Y}_2)(X_1 - \hat{X}_{1.2})}$$

$$pr_{Y2} = r_{(Y - \hat{Y}_1)(X_2 - \hat{X}_{2.1})}$$

# การวิเคราะห์ถดถอยแบบพหุ

- ค่า partial correlation เหมือนกับการหาความสัมพันธ์ของตัวแปรสองตัว เมื่อควบคุมให้อีกตัวแปรหนึ่งคงที่ เช่น แรงบีบมือ กับพลังหมัดอาจจะไม่สัมพันธ์กันเลย ถ้าควบคุมด้วยน้ำหนัก
- ค่า part correlation เป็นตัวบอกกว่าความแปรปรวนเฉพาะตัวของมัน (Unique variance) ส่งผลต่อตัวแปรตามเท่าใด



# การวิเคราะห์ถดถอยแบบพหุ

- การวิเคราะห์ถดถอยแบบพหุ อาจพบได้ว่าค่าสหสัมพันธ์อย่างง่ายระหว่างตัวแปรต้นตัวหนึ่งและตัวแปรตามมีค่ามาก แต่เมื่อมาวิเคราะห์ถดถอยกับตัวแปรต้นตัวอื่นพบว่าค่าสัมประสิทธิ์ถดถอยมีค่าน้อย
- แสดงว่าตัวแปรนั้นน่าจะมีผลทำนายตัวแปรตามได้ แต่ว่าตัวแปรตัวนี้ไปสัมพันธ์กับตัวแปรต้นตัวอื่นสูง จึงทำให้ค่าสัมประสิทธิ์ถดถอยน้อยลง

# การวิเคราะห์ถดถอยแบบพหุ

- ดังนั้น การที่ค่าสัมประสิทธิ์ถดถอยมีค่าน้อยไม่ได้หมายความว่าตัวแปรนั้นไม่สามารถอธิบายความแปรปรวนของตัวแปรตามได้ อาจเป็นเพราะตัวแปรตัวนี้ไปสัมพันธ์กับตัวแปรต้นอื่นสูง
- ด้วยเหตุนี้ การตีความต้องใช้สัมประสิทธิ์ถดถอยสัมประสิทธิ์สหสัมพันธ์อย่างง่าย part และ partial correlation มาช่วยอธิบาย

# การวิเคราะห์ถดถอยแบบพหุ

- ส่วนค่าสหสัมพันธ์ระหว่างตัวแปรต้น และค่าตัวแปรตามที่ทำนายได้ อาจใช้ช่วยอธิบายว่าตัวแปรใด สามารถอธิบายตัวแปรตามได้ แต่นักวิจัยบางท่านบอกว่าไม่จำเป็นต้องใช้ เพราะค่านี้ใกล้เคียงกับค่าสหสัมพันธ์อย่างง่ายระหว่างตัวแปรต้นแต่ละตัวกับตัวแปรตาม

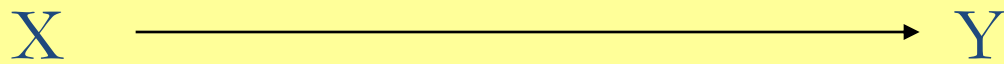
# การวิเคราะห์หัตถดถอยแบบพหุ

- การตีความความสัมพันธ์ระหว่างตัวแปรตาม และตัวแปรต้นสองตัว
  - ผลทางตรง (Direct Effect) และผลทางอ้อม (Indirect Effect)
  - ความซ้ำซ้อนบางส่วน (Partial Redundancy) หรือความซ้ำซ้อนทั้งหมด (Full Redundancy)
  - ความสัมพันธ์ปลอม (Spurious Relationship)
  - ผลกด (Suppression Effect)

# การวิเคราะห์ห้ถดถอยแบบพหุ

- ผลทางตรง (Direct Effect) และผลทางอ้อม (Indirect Effect)

Direct Effect



Indirect Effect



$$r_{YX} < sr_{YX.Z}$$

Independent Variable

Mediator  
Intervening Variable

Dependent Variable

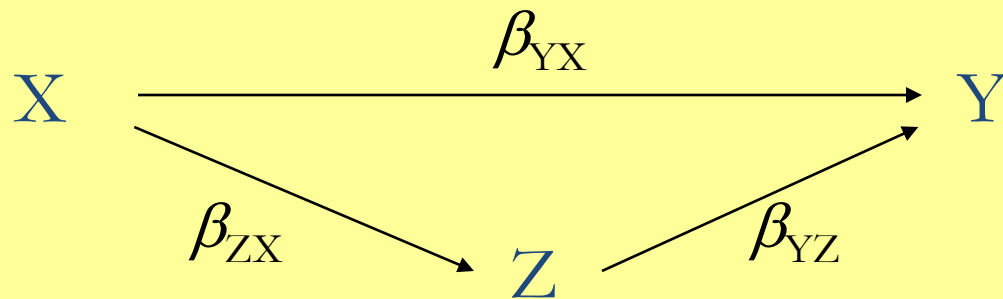
เช่น ระดับเสียงสูงส่งผลให้ทำงานได้ไม่มีประสิทธิภาพ อาจผ่านตัวแปรสื่อ คือ ความเครียด

# การวิเคราะห์ห้ถดถอยแบบพหุ

- ผลทางตรง (Direct Effect) และผลทางอ้อม (Indirect Effect)

Direct Effect

Indirect Effect



$$r_{YX} = \beta_{YX.Z} + \beta_{ZX} \beta_{YZ}$$

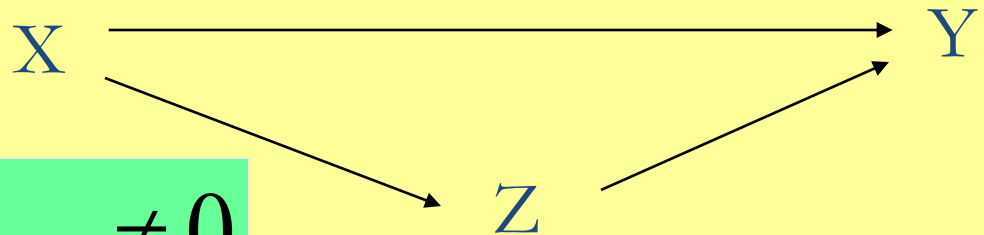
DE

IE

# การวิเคราะห์หัตถดถอยแบบพหุ

- ความซ้ำซ้อนบางส่วน (Partial Redundancy) หรือความซ้ำซ้อนทั้งหมด (Full Redundancy)

Partial Redundancy



$$sr_{YX.Z} < r_{YX}; sr_{YX.Z} \neq 0$$

Full Redundancy

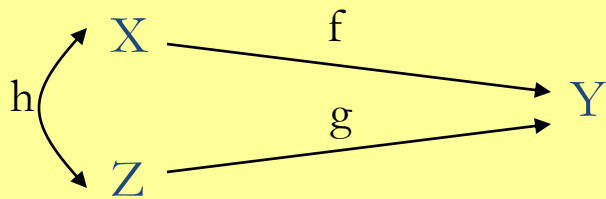


$$sr_{YX.Z} < r_{YX}; sr_{YX.Z} = 0$$

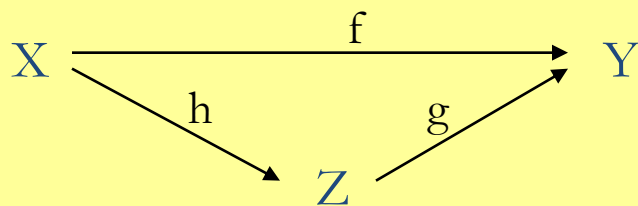
# การวิเคราะห์ห้ถดถอยแบบพหุ

- ความสัมพันธ์ปลอม (Spurious Relationship)

## Partial Redundancy

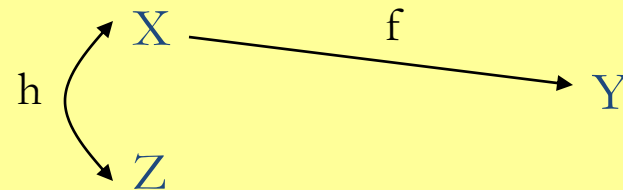


Model without mediator

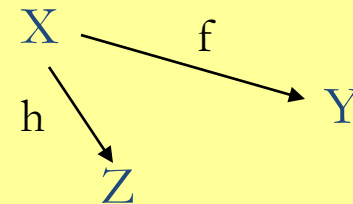


Model with mediator

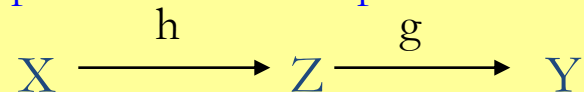
## Full Redundancy



Spurious effect from Z to Y



Spurious relationship in Y and Z



Indirect effect (Full Redundancy)

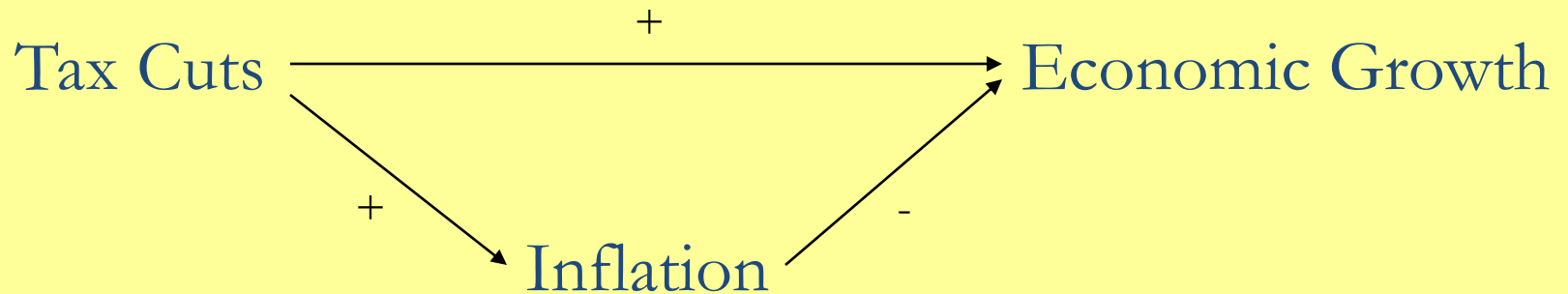


# การวิเคราะห์หัตถถอยแบบพหุ

- ผลกิด (Spurious Relationship)
  - เกิดขึ้นเมื่อใส่ตัวแปรต้นเพิ่มเติมเข้าไปในสมการ ทำให้อำนาจการทำนายของตัวแปรต้นเดิมมีค่าสูงขึ้น หรือกลับทิศทาง
  - ผลกิดก็คือการกดความสัมพันธ์ที่แท้จริงไว้
  - ตัวอย่างเช่น การทำนายความสามารถทางการบิน ด้วยทักษะทางด้านเครื่องกล พบว่าทำนายได้ระดับหนึ่ง แต่เมื่อเอาคะแนนภาษาอังกฤษไปร่วมทำนาย พบว่าทำนายได้สูงขึ้น

# การวิเคราะห์ห้ถดถอยแบบพหุ

- ผลกิด (Spurious Relationship)



# การวิเคราะห์ถดถอยแบบพหุ

- การอ้างอิงถึงประชากร ค่าสัมประสิทธิ์ที่สามารถอ้างอิงถึงประชากรมีดังนี้
  - การทดสอบว่าตัวแปรต้นทั้งหมดสามารถอธิบายตัวแปรตามได้อย่างมีนัยสำคัญหรือไม่ ทดสอบด้วยค่าสถิติ  $F$
  - การทดสอบว่าตัวแปรต้นแต่ละตัวสามารถอธิบายความแปรปรวนได้อย่างมีนัยสำคัญหรือไม่ ทดสอบด้วยค่าสถิติ  $t$
  - การทดสอบว่าจุดตัดแกน  $Y$  แตกต่างจากศูนย์อย่างมีนัยสำคัญหรือไม่ ทดสอบด้วยค่าสถิติ  $t$

# การวิเคราะห์ถดถอยแบบพหุ

- การอ้างอิงถึงประชากร ค่าสัมประสิทธิ์ที่สามารถอ้างอิงถึงประชากรมีดังนี้
  - ถ้าตัวแปรต้นตัวใดสามารถทำนายตัวแปรตามได้อย่างมีนัยสำคัญ หมายถึงค่า  $b$ ,  $\beta$ ,  $sr$ ,  $pr$  ถึงระดับนัยสำคัญทั้งสามตัว
  - ค่าความคาดเคลื่อนมาตรฐานของการทดสอบตัวแปรต้นแต่ละตัวขึ้นอยู่กับขนาดกลุ่มตัวอย่าง (ถ้ามาก จะทำให้ SE ต่ำ) และความสัมพันธ์ของตัวแปรต้นนั้นกับตัวแปรต้นตัวอื่น (ถ้ามาก จะทำให้ SE สูง)

# การวิเคราะห์หาคัดถอยแบบพหุ

- การอ้างอิงถึงประชากร ค่าสัมประสิทธิ์ที่สามารถอ้างอิงถึงประชากรมีดังนี้
  - บางครั้งอาจพบกรณีที่ค่า  $F$  ในภาพรวมถึงระดับนัยสำคัญ แต่ไม่พบตัวแปรต้นตัวใดที่ถึงระดับนัยสำคัญ ในที่นี้เป็นเพราะตัวแปรต้นสัมพันธ์กันเองสูงมาก ภาวะนี้เรียกว่าภาวะร่วมเส้นตรงแบบพหุ (Multicollinearity)

# การวิเคราะห์ห้ถดถอยแบบพหุ

- ค่าสัมประสิทธิ์การทำนายแบบปรับแก้
  - เราพบว่าค่า  $R^2$  จะมีค่าสูงขึ้นเมื่อเราใส่ตัวแปรต้นเพิ่มเข้าไป (ค่าความสามารถในการทำนายที่สูงขึ้นอาจเกิดจากผลสุ่มที่ได้จากตัวแปรต้นนั้น) ทำให้การทำนายค่า  $R^2$  ในประชากรไม่ถูกต้อง นักสถิติจึงคิดสูตรปรับแก้ค่า  $R^2$  คือ

$$\tilde{R}_Y^2 = 1 - (1 - R_Y^2) \frac{n-1}{n-k-1}$$

# การวิเคราะห์หัตถดถอยแบบพหุ

- วิธีการใส่ตัวแปรต้น
  - การใส่ตัวแปรพร้อมๆ กัน หรือ Method Enter
  - การใส่ตัวแปรโดยคอมพิวเตอร์กำหนด มี 3 วิธี คือ Forward Backward และ Stepwise
  - การใส่ตามลำดับของผู้วิจัย หรือ Method Hierarchical

# การวิเคราะห์ถดถอยแบบพหุ

- วิธีการใส่ตัวแปรต้น
  - การใส่ตัวแปรพร้อมๆ กัน หรือ Method Enter เป็นวิธีการนำตัวแปรต้นทั้งหมดมาวิเคราะห์ในสมการพร้อมๆ กัน การเขียนสมการทำนายต้องเขียนโดยใส่ตัวแปรต้นทุกตัว ไม่ว่าจะทำนายได้ถึงระดับนัยสำคัญหรือไม่ก็ตาม
  - จำนวนกลุ่มตัวอย่างที่ควรใช้คือ 10 คน ต่อ 1 ตัวแปร
  - วิธีนี้เหมาะสมในงานวิจัยที่มีสมมติฐานมาก่อน แล้วใช้วิธีนี้ทดสอบสมมติฐาน



# การวิเคราะห์ถดถอยแบบพหุ

- วิธีการใส่ตัวแปรต้น

— การใส่ตัวแปรโดยคอมพิวเตอร์กำหนด เป็นวิธีการที่ให้  
คอมพิวเตอร์คัดเลือกตัวแปรเข้าไปสมการทำนาย มี 3 วิธี คือ

- Forward ค่อยๆ ใส่ตัวแปรที่  $p$  สูงสุด เข้าไปจนกระทั่งตัวที่ไม่ถึงระดับนัยสำคัญจะตัดออก เกณฑ์มักจะเป็น  $p < .05$
- Backward ใส่ตัวแปรต้นทั้งหมดออก แล้วค่อยๆ ตัดตัวแปรที่มีค่า  $p$  สูงสุดออก จนกระทั่ง  $p$  ทั้งหมดต่ำกว่าเกณฑ์ เกณฑ์มักจะเป็น  $p > .10$
- Stepwise เป็นวิธีที่รวมกันระหว่าง Forward และ Backward

# การวิเคราะห์ทัศนคติแบบพหุ

- วิธีการใส่ตัวแปรต้น
  - วิธีการให้คอมพิวเตอร์คัดเลือกตัวแปร เป็นวิธีที่ไม่เหมาะสม นักในงานวิจัย เพราะว่าต้องทดสอบนัยสำคัญหลายครั้ง ทำให้ค่าคาดเคลื่อนเพิ่ม นอกจากนั้นค่า *pr* ที่แตกต่างกันนิดเดียว ส่งผลให้ตัวแปรบางตัวอาจถูกตัดทิ้ง บางตัวถูกนำเข้ามาในสมการ ทำให้ผลลัพธ์ไม่คงที่
  - จำนวนกลุ่มตัวอย่างที่ควรใช้คือ 40 คน ต่อ ตัวแปรต้น 1 ตัว

# การวิเคราะห์ถดถอยแบบพหุ

- วิธีการใส่ตัวแปรต้น
  - วิธีการให้ผู้วิจัยสร้างลำดับเอง (Hierarchical Method) เป็นวิธีที่ผู้วิจัยกำหนดลำดับความสำคัญของตัวแปรตามวัตถุประสงค์ดังต่อไปนี้
    - เพื่อควบคุมตัวแปร
    - เพื่อกำหนดภาพในการอธิบาย ว่าตัวแปรใดเป็นตัวแปรต้น หรือตัวแปรใดเป็นตัวแปรสื่อ
    - เพื่อนำตัวแปรที่ผู้วิจัยสนใจ มาใส่เพิ่มเติมในสมการ โดยไม่ไปทำลายผลการวิเคราะห์ของตัวแปรหลัก

# การวิเคราะห์ถดถอยแบบพหุ

- วิธีการใส่ตัวแปรต้น

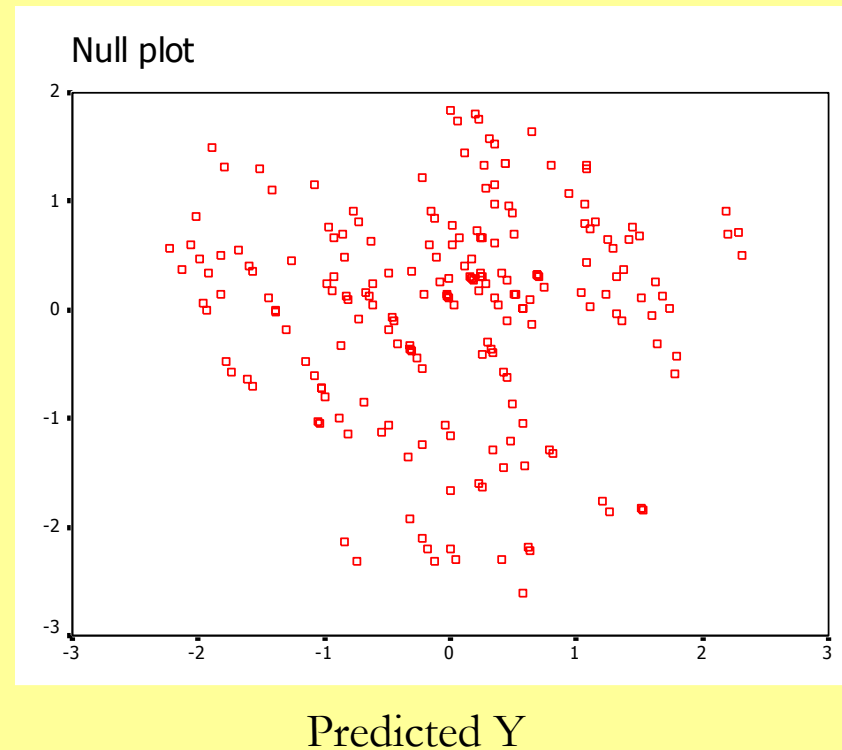
— การใช้วิธีการกำหนดลำดับตัวแปรเอง ให้สังเกตที่ค่า  $R^2$  ที่เปลี่ยนแปลง (มีการทดสอบว่า  $R^2$  ที่เปลี่ยนแปลงไป ถึงระดับนัยสำคัญหรือไม่ด้วย) และสังเกตที่ค่า  $\beta$  ที่เปลี่ยนแปลงด้วย

# การวิเคราะห์ถดถอยแบบพหุ

- ข้อตกลงเบื้องต้นของการวิเคราะห์ถดถอยแบบพหุ
  - ค่าความผิดพลาดในการทำนาย มีการกระจายเป็นโค้งปกติ (Normality)
  - การกระจายของตัวแปรตามในทุกๆ ค่าของตัวแปรต้นมีค่าเท่ากัน (Homoscedasticity)
  - ตัวแปรต้นมีความสัมพันธ์เชิงเส้นตรงกับตัวแปรตาม (Linearity)
  - การวัดตัวแปรเป็นอิสระจากกัน (Independent of error terms)
  - ไม่มีค่าความผิดพลาดในการวัด (No measurement error)
  - ไม่พบภาวะร่วมเส้นตรงแบบพหุ (Multicollinearity)

# การวิเคราะห์ห้ถดถอยแบบพหุ

- วิธีหนึ่งที่อาจตรวจสอบ  
Normality, Homoscedasticity,  
Linearity และ Independent of  
error terms ได้ คือการสร้าง  
Null plot คือการสร้างกราฟ  
แกนหนึ่งเป็นค่า  $Y$  ที่ทำนายได้  
และค่าหนึ่งเป็นค่าผิดพลาดใน  
การทำนาย
- อาจตรวจสอบรายตัวแปรด้วย  
Partial Regression Plot



# การวิเคราะห์ถดถอยแบบพหุ

- การตรวจสอบภาวะร่วมเส้นตรงแบบพหุ (Multicollinearity) มีวิธี 3 วิธี
  - ตรวจสอบความสัมพันธ์อย่างง่ายระหว่างตัวแปรต้นด้วยกัน ค่าสหสัมพันธ์ระหว่างตัวแปรต้นไม่ควรเกิน .9
  - ตรวจสอบค่า Tolerance และ Variance Inflation Factor
    - ค่า Tolerance คือ ค่า Unique Variance ที่ตัวแปรต้นนั้นเหลืออยู่เมื่อถูกอธิบายด้วยตัวแปรต้นตัวอื่น ค่าไม่ควรต่ำกว่า .1 หรือ 10 %
    - ค่า VIF คือ  $1/\text{Tolerance}$  ค่านี้จะบอกว่า  $SE_b$  เพิ่มขึ้นเป็นกี่เท่า ค่านี้ไม่ควรเกิน 10

# การวิเคราะห์หาคัดถอยแบบพหุ

- การตรวจสอบภาวะร่วมเส้นตรงแบบพหุ (Multicollinearity) มีวิธี 3 วิธี
  - กระบวนการตรวจสอบสองขั้นตอน
    - ให้ดูว่าค่า Condition index ใดเกิน 30
    - หาว่ามีตัวแปรต้นสัมพันธ์กับ Condition Index นั้นเกิน .9 มากกว่า 2 ตัวหรือไม่ ถ้ามีแสดงว่าตัวแปรต้นเหล่านั้นมี Multicollinearity



# การวิเคราะห์ถดถอยแบบพหุ

- การตรวจสอบภาวะร่วมเส้นตรงแบบพหุ (Multicollinearity)
  - อาจแก้ไขโดยการตัดตัวแปรที่ซ้ำซ้อนทิ้ง หรือใช้วิธีการรวมตัวแปรด้วยการวิเคราะห์องค์ประกอบ (Factor Analysis)

# การวิเคราะห์ถดถอยแบบพหุ

- การตรวจสอบภาวะร่วมเส้นตรงแบบพหุ (Multicollinearity)
  - อาจแก้ไขโดยการตัดตัวแปรที่ซ้ำซ้อนทิ้ง หรือใช้วิธีการรวมตัวแปรด้วยการวิเคราะห์องค์ประกอบ (Factor Analysis)

# การวิเคราะห์ถดถอยแบบพหุ

- การสร้างตัวแปรดัมมี่

จะเป็นการแยกตัวแปรที่มีลักษณะเป็นกลุ่ม มาเป็นตัวแปรแบบ metric เท่ากับจำนวนกลุ่มลบหนึ่งตัวแปร วิธีการสร้างมีทั้งหมด 3 วิธีด้วยกัน

- แบบดัมมี่ (Dummy Coding)

- แบบอิทธิพล (Effect Coding)

- แบบตั้งฉาก (Orthogonal Coding) (ยากไป ไม่อธิบาย)

# การวิเคราะห์ถดถอยแบบพหุ

- แบบดัมมี่ (Dummy Coding)  
จะตั้งกลุ่มอ้างอิงเท่ากับ 0 ใน  
ทุกตัวแปร ส่วนตัวแปรไหนชี้  
ถึงกลุ่มใด กลุ่มนั้นจะมีค่า  
เท่ากับ 1 กลุ่มอื่นนอกจาก  
กลุ่มอ้างอิงและกลุ่มที่ตัว  
แปรนั้นบ่งชี้ให้เท่ากับ 0

	<b>D<sub>1</sub></b>	<b>D<sub>2</sub></b>
<b>E<sub>1</sub></b>	<b>1</b>	<b>0</b>
<b>E<sub>2</sub></b>	<b>0</b>	<b>1</b>
<b>C</b>	<b>0</b>	<b>0</b>

# การวิเคราะห์หาคัดถอยแบบพหุ

- ถ้านำตัวแปรต้นที่มีใส่ในสมการ ค่า  $b$  จะหมายความว่ากลุ่มที่ตัวแปรต้นมีแทนถึงต่างกับกลุ่มอ้างอิงมากน้อยเท่าใด และจุดตัดแกน  $Y$  เท่ากับค่าเฉลี่ยของกลุ่มอ้างอิง

$$Y = 3 + 5X_1 + 2X_2$$

# การวิเคราะห์ห้ถดถอยแบบพหุ

- แบบอิทธิพล (Effect Coding)

จะตั้งกลุ่มอ้างอิงเท่ากับ -1

ในทุกตัวแปร ส่วนตัวแปร

ไหนชี้ถึงกลุ่มใด กลุ่มนั้นจะมี

ค่าเท่ากับ 1 กลุ่มอื่นนอกจาก

กลุ่มอ้างอิงและกลุ่มที่ตัว

แปรนั้นบ่งชี้ให้เท่ากับ 0

	<b>D<sub>1</sub></b>	<b>D<sub>2</sub></b>
<b>E<sub>1</sub></b>	<b>1</b>	<b>0</b>
<b>E<sub>2</sub></b>	<b>0</b>	<b>1</b>
<b>C</b>	<b>-1</b>	<b>-1</b>

# การวิเคราะห์ถดถอยแบบพหุ

- ถ้านำตัวแปรต้นที่มีใส่ในสมการ ค่า  $b$  จะหมายความว่ากลุ่มที่ตัวแปรต้นมีแทนถึงต่างกับค่าเฉลี่ยโดยรวม (Grand mean) มากน้อยเท่าใด และจุดตัดแกน  $Y$  เท่ากับค่าเฉลี่ยโดยรวมนั่นเอง

$$Y = 5.33 + 2.67X_1 - 0.33X_2$$