

การวิเคราะห์ถดถอย

สถิติสำหรับจิตวิทยา 1

สันทัด พรประเสริฐมานิต

โครงร่างการนำเสนอ

- มโนทัศน์พื้นฐาน
- สมบัติวิธีการทำนาย
- สมบัติการถดถอยมาตรฐาน
- การทดสอบสมมติฐาน
- การเขียนรายงาน
- ช่วงเชื่อมั่น
- การวิเคราะห์สหสัมพันธ์และการวิเคราะห์ถดถอย

แนะนำ

- เมื่อตัวแปรสองตัวมีความสัมพันธ์กัน เรามักจะใช้ข้อมูลของตัวแปรหนึ่งในการทำนายตัวแปรอีกตัวแปรหนึ่ง เช่น
 - ส่วนสูงของคน น่าจะใช้ในการทำนายน้ำหนักของคนได้ คนที่สูงมากน่าจะทำนายว่ามีน้ำหนักมาก คนที่สูงน้อย น่าจะทำนายว่ามีน้ำหนักน้อย
 - ในการคัดเลือกบุคลากร เซวน์ปัญญาและบุคลิกภาพแบบการมีความรับผิดชอบสามารถทำนายผลการปฏิบัติงานในอนาคต
- ในบทนี้จะกล่าวถึง การทำนายเป็นความสัมพันธ์เชิงเส้นจากตัวแปรหนึ่ง (หรือหลายตัว) ไปยังอีกตัวแปรหนึ่ง
- เรียกว่า การถดถอยเชิงเส้น (Linear Regression)

มโนทัศน์พื้นฐาน

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

มโนทัศน์พื้นฐาน

- สมมติว่า ตัวแปร Y มีค่าเท่ากับ 1, 2, 3, 4, 5
- หากลองสุ่มข้อมูลตัวหนึ่งออกมา และเราไม่ทราบว่าคุณสมบัตินั้น และเราทำนายค่านั้นว่าค่าเท่ากับเท่าไร เช่น 4 เราจะเรียกว่า ค่าทำนาย (Predicted Y : $\hat{Y}$)
- โดยส่วนใหญ่ ค่าที่ทำนายได้นั้น จะไม่ตรงกับค่าจริง ความแตกต่างจะเรียกว่า ความผิดพลาดในการทำนาย (Error: e)

$$e = Y - \hat{Y}$$

- เช่น ความผิดพลาดของข้อมูลนี้เท่ากับ -3, -2, -1, 0, 1

มโนทัศน์พื้นฐาน

- ในการทำนาย ต้องการให้ความผิดพลาดในการทำนายน้อยที่สุด
- เช่น ในการทำนายข้อมูลของตัวแปรใดตัวแปรหนึ่ง หากไม่มีข้อมูลอื่นเลย ค่าที่ทำนายที่ดีที่สุด คือ ค่าเฉลี่ย

$$Y = 1, 2, 3, 4, 5 \quad \longrightarrow \quad \hat{Y} = M = 3$$



$$e = Y - \hat{Y} = -2, -1, 0, 1, 2 \quad \longrightarrow \quad M_e = 0$$

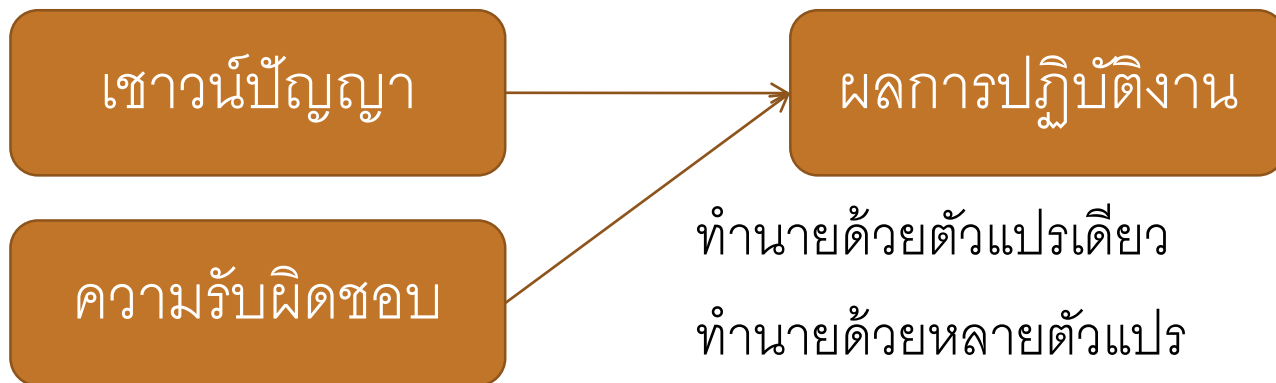
- ค่าเฉลี่ย จะทำให้ความผิดพลาดในการทำนายน้อยที่สุด เพราะมีค่าเฉลี่ยของความผิดพลาดเท่ากับ 0

มโนทัศน์พื้นฐาน

- ในกรณีมีข้อมูลของอีกตัวแปรหนึ่ง เราจะแปลงค่าข้อมูลของตัวแปรที่ใช้ทำนาย (Predictor) เพื่อให้ใกล้เคียงกับตัวแปรที่ถูกทำนาย (Criterion or Predicted Variable) มากที่สุด



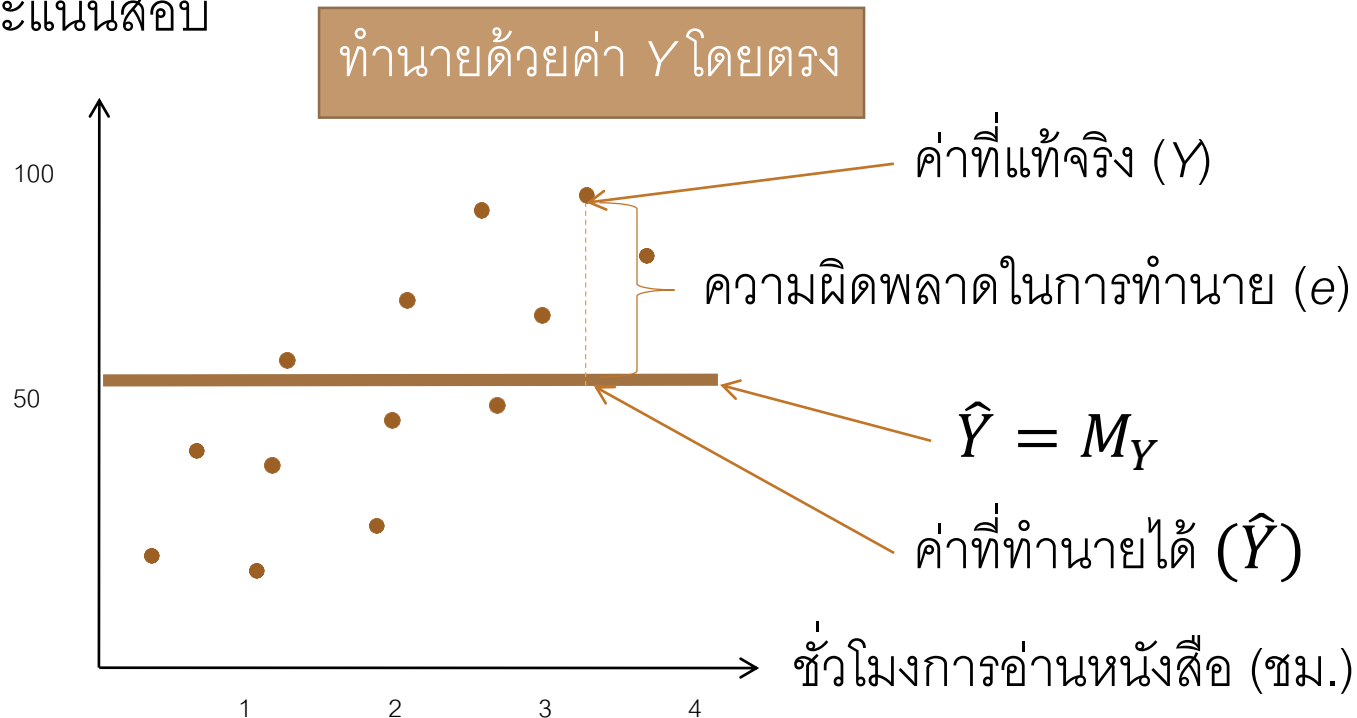
- เช่น



มโนทัศน์พื้นฐาน

- กล่าวคือ แปลงค่า X ให้เป็นค่า $\hat{Y}$ เพื่อให้ e มีค่าน้อยที่สุด
- พิจารณาแผนภาพกระจายกระจาย (Scatterplot)

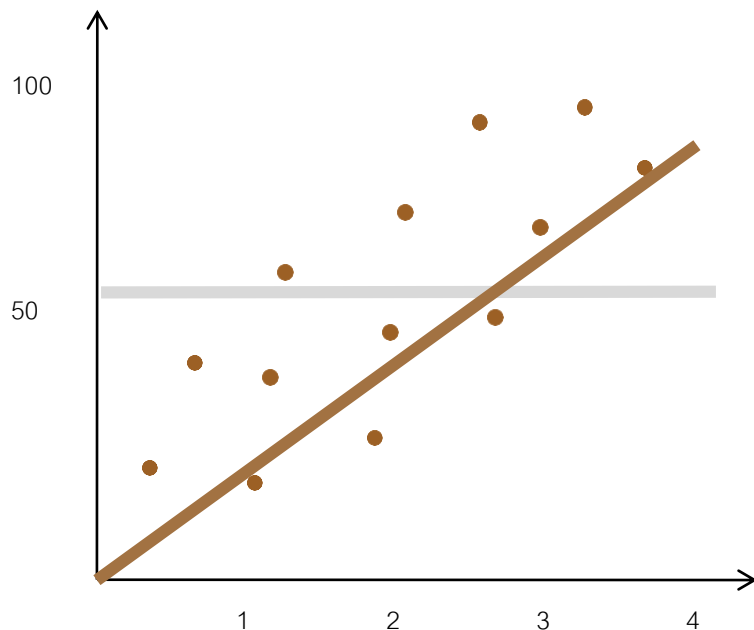
คะแนนสอบ



มโนทัศน์พื้นฐาน

- กล่าวคือ แปลงค่า X ให้เป็นค่า $\hat{Y}$ เพื่อให้ e มีค่าน้อยที่สุด
- พิจารณาแผนภาพกระจาย (Scatterplot)

คะแนนสอบ



ทำนายด้วยค่า $20X$

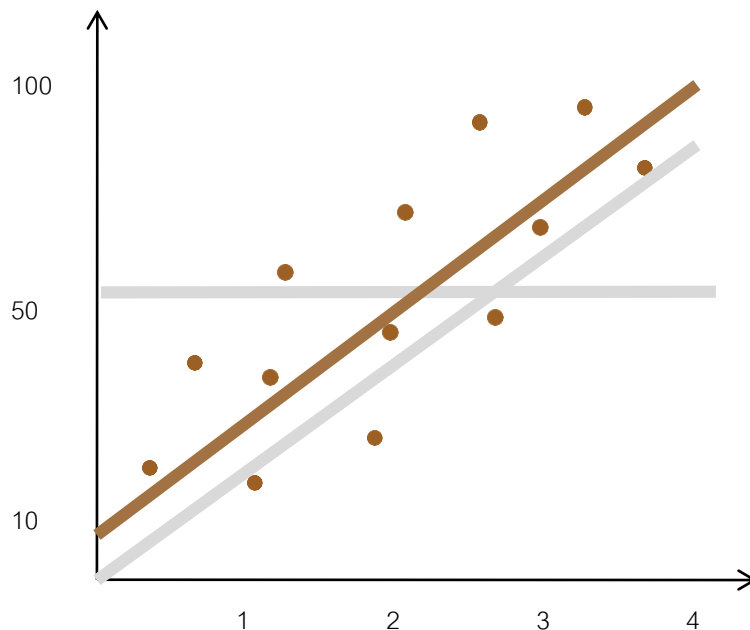
ความผิดพลาดน้อยกว่าทำนายด้วย Y เพียงอย่างเดียว

ชั่วโมงการอ่านหนังสือ (ชม.)

มโนทัศน์พื้นฐาน

- กล่าวคือ แปลงค่า X ให้เป็นค่า $\hat{Y}$ เพื่อให้ e มีค่าน้อยที่สุด
- พิจารณาแผนภาพกระจาย (Scatterplot)

คะแนนสอบ



ทำนายด้วยค่า $20X + 10$

ความผิดพลาดน้อยลงกว่าเดิม

ชั่วโมงการอ่านหนังสือ (ชม.)

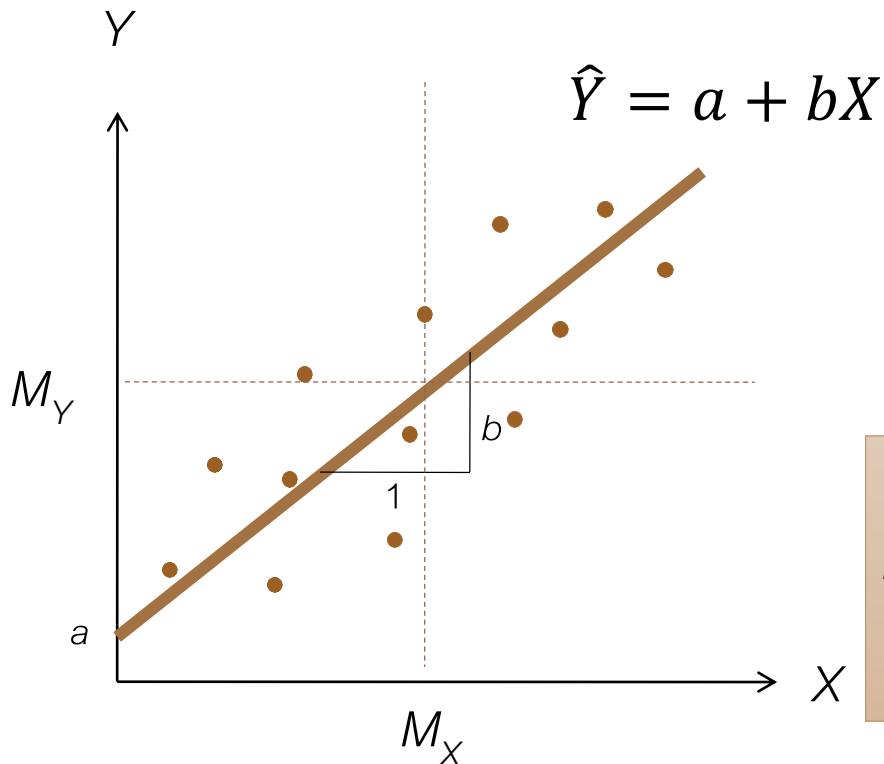
มโนทัศน์พื้นฐาน

- กล่าวคือ เราต้องการหาค่า a และ b ใน $\hat{Y} = a + bX$ ซึ่งเป็นการแปลงค่า X ด้วยสมการเชิงเส้น
- ค่า a และ b ต้องทำให้ค่าความผิดพลาดในการทำนายมีค่าน้อยที่สุด
- หรือ ต้องการเส้นที่ผ่านกลุ่มของจุดมากที่สุดนั่นเอง
- เกณฑ์ที่ใช้ในการตัดสินใจ คือ เส้นที่ทำให้ผลรวมกำลังสองของความผิดพลาดมีค่าน้อยที่สุด (Least sum of squared error)

$$\sum e^2 = \sum (Y - \hat{Y})^2$$

มโนทัศน์พื้นฐาน

- สมการเชิงเส้น ที่ทำให้ความผิดพลาดมีค่าน้อยที่สุด จะเรียกว่า สมการถดถอย (Regression Line)



กราฟผ่านจุด M_X และ M_Y เสมอ

a คือ ค่าของ $\hat{Y}$ เมื่อ $X = 0$ หรือ
จุดตัดแกน Y ของสมการถดถอย

b คือ ค่าของ $\hat{Y}$ ที่เปลี่ยนแปลงเมื่อ X
เปลี่ยนแปลงไป 1 หน่วย

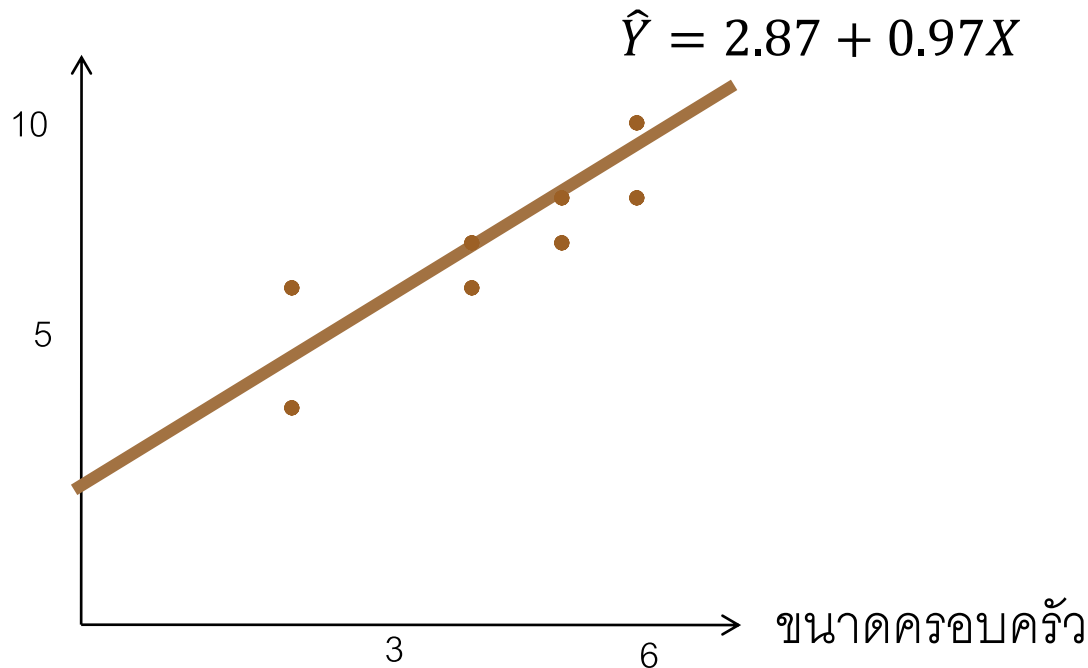
$$b = r \cdot \frac{SD_Y}{SD_X}$$

$$a = M_Y - bM_X$$

มโนทัศน์พื้นฐาน

- เช่น ทำนายจำนวนบัตรเครดิตที่ใช้ในครอบครัว ด้วยขนาดครอบครัว
- เก็บข้อมูลจาก 8 ครอบครัว

จำนวนบัตรเครดิตที่ใช้

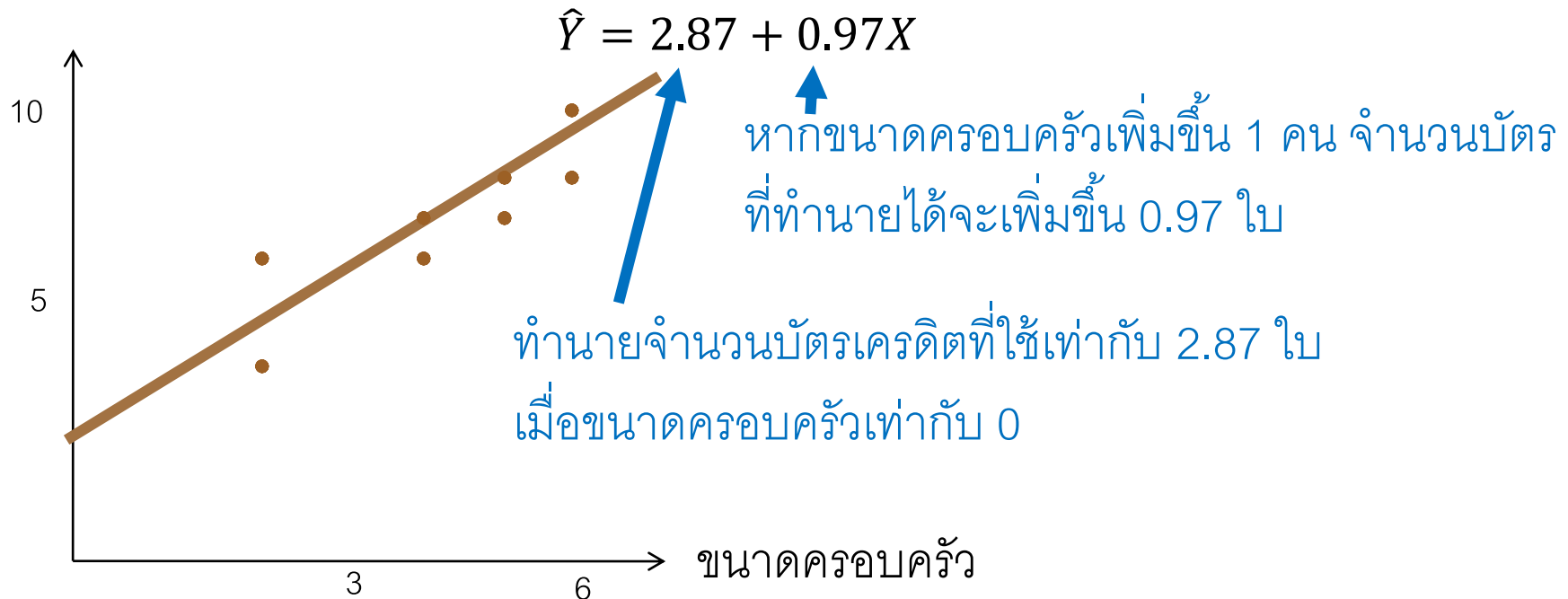


X	Y
2	4
2	6
4	6
4	7
5	8
5	7
6	8
6	10

มโนทัศน์พื้นฐาน

- เช่น ทำนายจำนวนบัตรเครดิตที่ใช้ในครอบครัว ด้วยขนาดครอบครัว
- เก็บข้อมูลจาก 8 ครอบครัว

จำนวนบัตรเครดิตที่ใช้



มโนทัศน์พื้นฐาน

- หากมีตัวแปรอิสระมากกว่า 1 ตัว สมการเชิงเส้นจะเป็นดังต่อไปนี้

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \cdots + b_pX_p$$

- เราต้องการหาค่า $a, b_1, b_2, \dots, b_p$ ให้ค่าความผิดพลาดในการทำนายมีค่าน้อยที่สุด
- กล่าวคือ ให้มีความผิดพลาดมีค่าน้อยที่สุด (Least sum of squared error)

$$\sum e^2 = \sum (Y - \hat{Y})^2$$

มโนทัศน์พื้นฐาน

$$\hat{Y} = a + b_1X_1 + b_2X_2 + \cdots + b_pX_p$$

- a คือ ค่าของ $\hat{Y}$ เมื่อ $X_1, X_2, \dots, X_p$ เท่ากับ 0 หรือจุดตัดแกน Y ของสมการถดถอย
- b_p คือ ค่าของ $\hat{Y}$ เมื่อ X_p เพิ่มขึ้น 1 หน่วย และควบคุมให้ค่าของ $X_1, X_2, \dots, X_{p-1}$ คงที่

มโนทัศน์พื้นฐาน

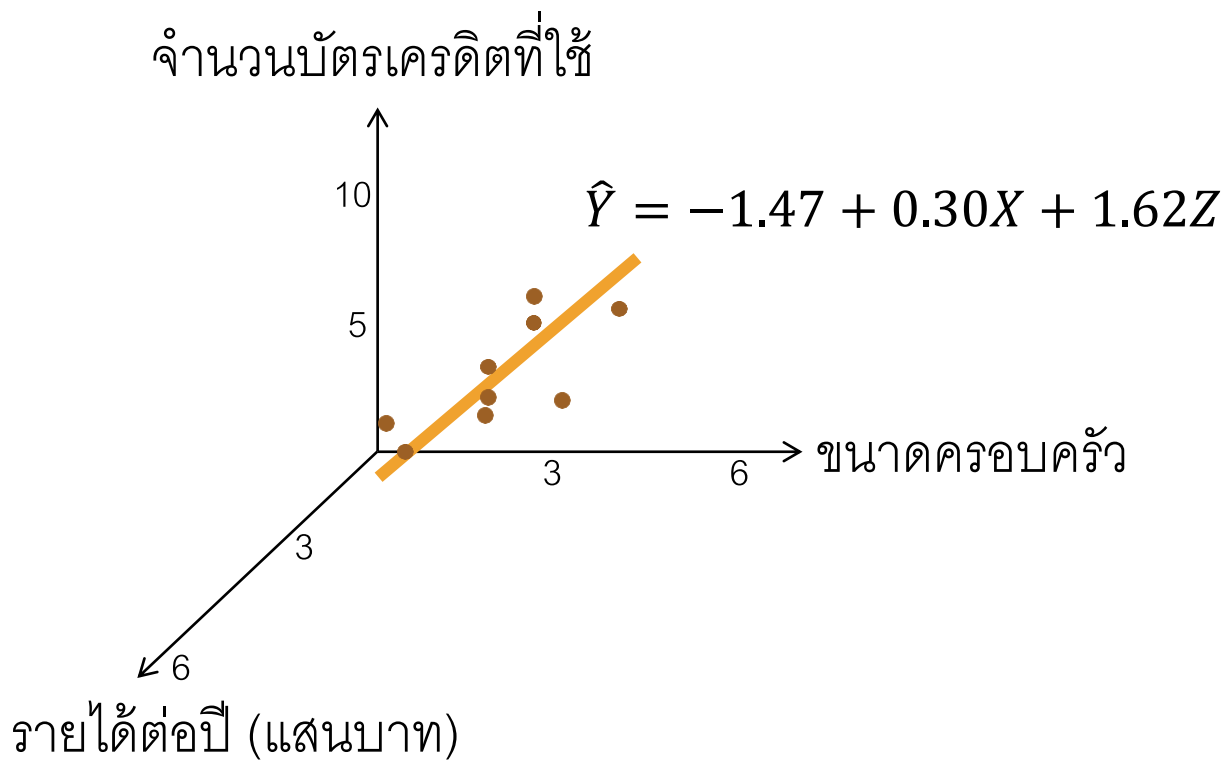
- การคำนวณจะต้องใช้การคำนวณด้วยเมทริกซ์

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

- $\mathbf{b}$ คือเมทริกซ์ขนาด $p + 1$ และ 1 ซึ่งมี $a, b_1, b_2, \dots, b_p$ เป็นสมาชิก
- $\mathbf{X}$ คือเมทริกซ์ขนาด N และ $p + 1$ ซึ่งใส่ข้อมูลตัวแปรอิสระทั้งหมด โดยเพิ่มคอลัมน์แรกให้มีสมาชิกเป็น 1 ทั้งหมด
- $\mathbf{y}$ คือเมทริกซ์ขนาด N และ 1 ซึ่งใส่ข้อมูลตัวแปรตาม

มโนทัศน์พื้นฐาน

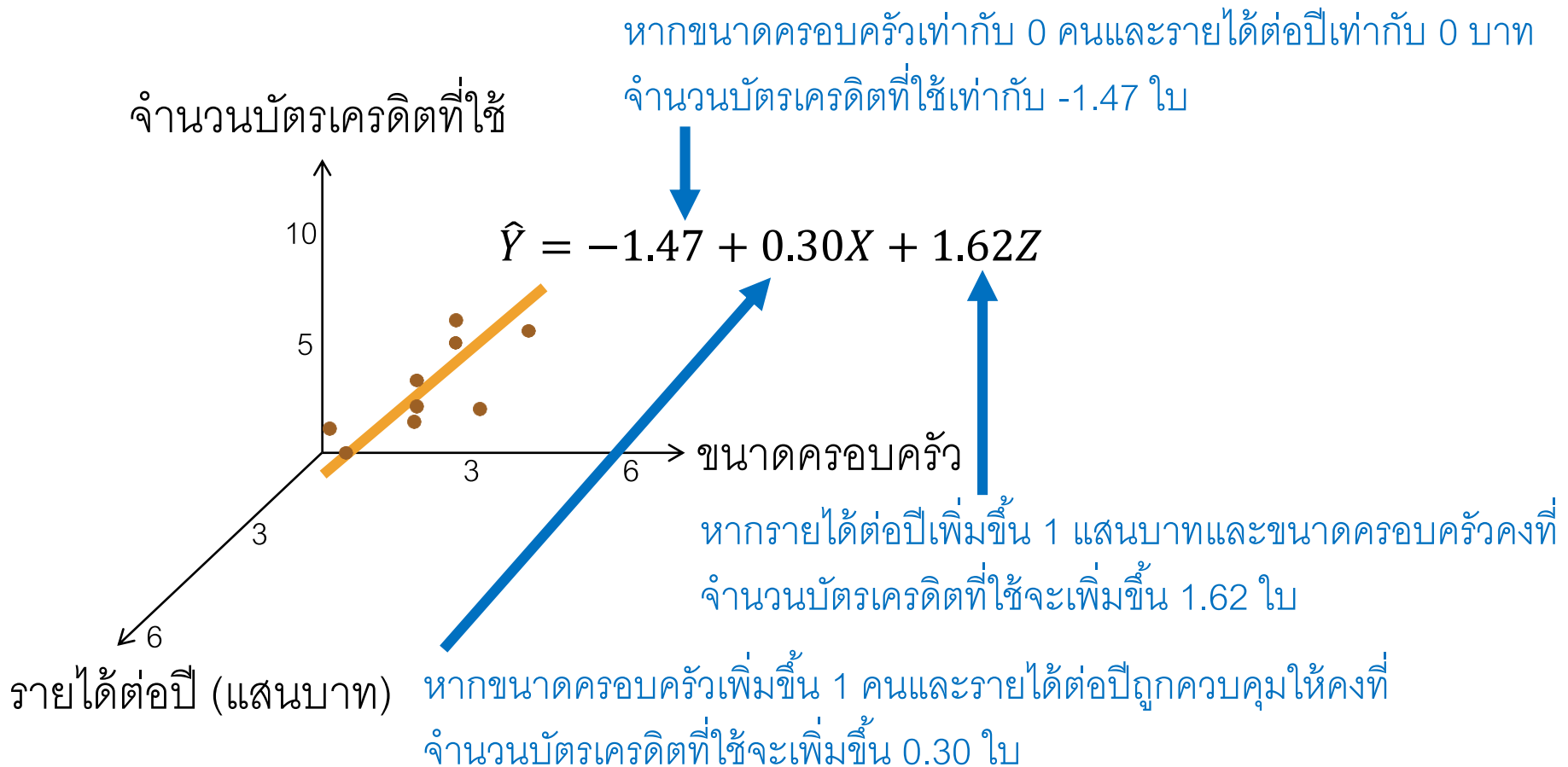
- เพิ่มตัวแปรรายได้ครอบคลุมต่อปี (หน่วยแสนบาท)



X	Z	Y
2	3.5	4
2	3.8	6
4	4.0	6
4	4.2	7
5	4.4	8
5	4.6	7
6	4.8	8
6	6.0	10

มโนทัศน์พื้นฐาน

- เพิ่มตัวแปรรายได้ครอบคลุมทั่วประเทศ (หน่วยแสนบาท)



มโนทัศน์พื้นฐาน

- สมการ เป็นเพียงเส้นที่บอกว่า ควรแปลงค่า X อย่างไรที่จะทำนายค่า Y ได้ดีที่สุด แต่ไม่ได้บอกความแม่นยำในการทำนาย
- ความแม่นยำในการทำนาย อาจมองได้ว่าเป็นการเกาะกลุ่มของข้อมูลรอบๆ เส้นถดถอย
 - หากข้อมูลเกาะกลุ่มมาก แสดงว่าแม่นยำมาก
- สามารถวัดได้ง่ายๆ ด้วยการหาค่าสหสัมพันธ์ระหว่าง Y และ $\hat{Y}$ ซึ่งจะใช้สัญลักษณ์ R

มโนทัศน์พื้นฐาน

- ในกรณีที่มีตัวแปรอิสระตัวเดียว ค่า R จะเท่ากับค่าสัมบูรณ์ของค่าสหสัมพันธ์ระหว่าง 2 ตัวแปร, $|r|$
- สัมประสิทธิ์การทำนาย (Coefficient of Determination, R^2) เป็นค่าที่บอกว่าตัวแปร X ทั้งหมดสามารถอธิบายตัวแปร Y ได้เท่าไร
 - เช่น จำนวนสมาชิกในครอบครัว สามารถทำนายความแปรปรวนของจำนวนบัตรเครดิตที่ใช้เท่ากับ 75%
 - จำนวนสมาชิกในครอบครัว และรายได้ต่อปี สามารถทำนายความแปรปรวนของจำนวนบัตรเครดิตที่ใช้เท่ากับ 89%

สัมประสิทธิ์การทำงาน

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

สัมประสิทธิ์การทำนาย

- จากที่กล่าวมาข้างต้น หากนำค่าสหสัมพันธ์มายกกำลังสองแล้ว จะเท่ากับสัมประสิทธิ์การทำนาย (Coefficient of Determination)
- หากมองวิธีการคำนวณอีกรูปแบบหนึ่ง จะเข้าใจมากขึ้นว่าทำไมถึงเรียกว่าสัมประสิทธิ์การทำนาย
- เริ่มจากการหาผลรวมความผิดพลาดกำลังสอง (Sum of squared error: SS_{Error} หรือ SS_E)

$$\sum e^2 = \sum (Y - \hat{Y})^2$$

สัมประสิทธิ์การทำนาย

- ในการบอกว่า ตัวแปร X ที่มาทำนายตัวแปร Y นั้น สามารถทำนายได้มากน้อยเพียงใด จะต้องหาก่อนว่า หากไม่มีตัวแปร X แล้ว ความผิดพลาดในการทำนายค่า Y เท่ากับเท่าไร
- ดังที่กล่าวข้างต้น หากไม่มีตัวแปรอื่นแล้ว ค่าเฉลี่ย (M_Y) จะเป็นตัวทำนายค่า Y ได้ดีที่สุด เพราะทำให้ค่าเฉลี่ยของความผิดพลาดเท่ากับศูนย์

สัมประสิทธิ์การทำนาย

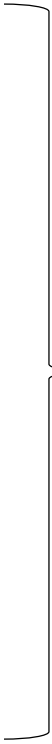
- จะใช้ ผลรวมของค่าความผิดพลาดกำลังสอง เมื่อใช้ค่าเฉลี่ยในการทำนายแทน คือ

$$SS_{\text{Error}} = \sum e^2 = \sum (Y - M_Y)^2 = SS_Y$$

สัมประสิทธิ์การทำนาย

- เช่น จำนวนบัตรเครดิตที่ใช้

y	$\hat{y}$	e	e^2
4			
6			
6			
7			
8			
7			
8			
10			


$$SS_{\text{Error}} = SS_Y = 22$$

สัมประสิทธิ์การทำนาย

- เมื่อมีตัวแปรทำนายค่า ตัวแปรนั้นจะถูกแปลงด้วยสมการถดถอย เพื่อให้ใกล้เคียงกับตัวแปร Y ให้มากที่สุด
- ความสามารถในการทำนาย สามารถบอกได้โดยผลรวมของความผิดพลาดในการทำนายเช่นกัน

$$SS_{\text{Error}} = \sum e^2 = \sum (Y - \hat{Y})^2$$

สัมประสิทธิ์การทำนาย

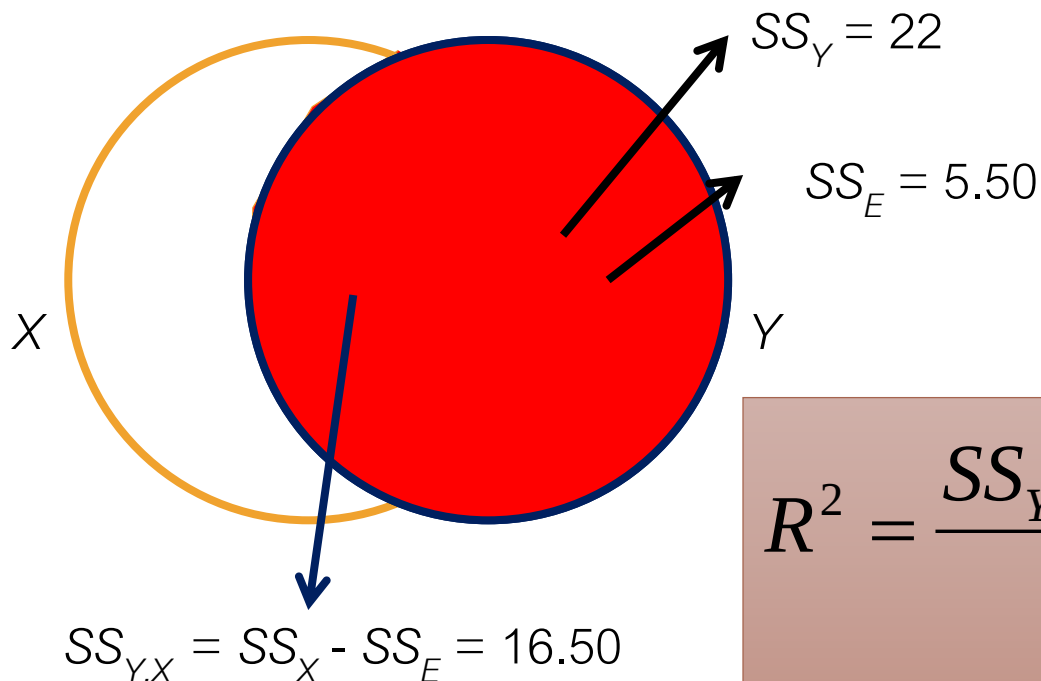
- เช่น จำนวนบัตรเครดิตที่ใช้ ถูกทำนายด้วยขนาดครอบครัว

x	y	$\hat{y}$	e	e^2
2	4			
2	6			
4	6			
4	7			
5	8			
5	7			
6	8			
6	10			

$SS_{\text{Error}} = 5.50$

สัมประสิทธิ์การทำนาย

- เมื่อไม่มีตัวแปรทำนาย และค่าเฉลี่ยทำนาย เกิด $SS_E = 22$
- เมื่อมีตัวแปรทำนาย เกิด $SS_E = 5.50$



R^2 คือ สัดส่วนของพื้นที่
ที่ตัวแปร X ทำนายได้ ต่อ
พื้นที่ตัวแปร Y ทั้งหมด

$$R^2 = \frac{SS_Y - SS_E}{SS_Y} = \frac{22 - 5.5}{22} = .75$$

ส่วนของตัวแปร X ที่สามารถอธิบาย Y ได้

สัมประสิทธิ์การทำนาย

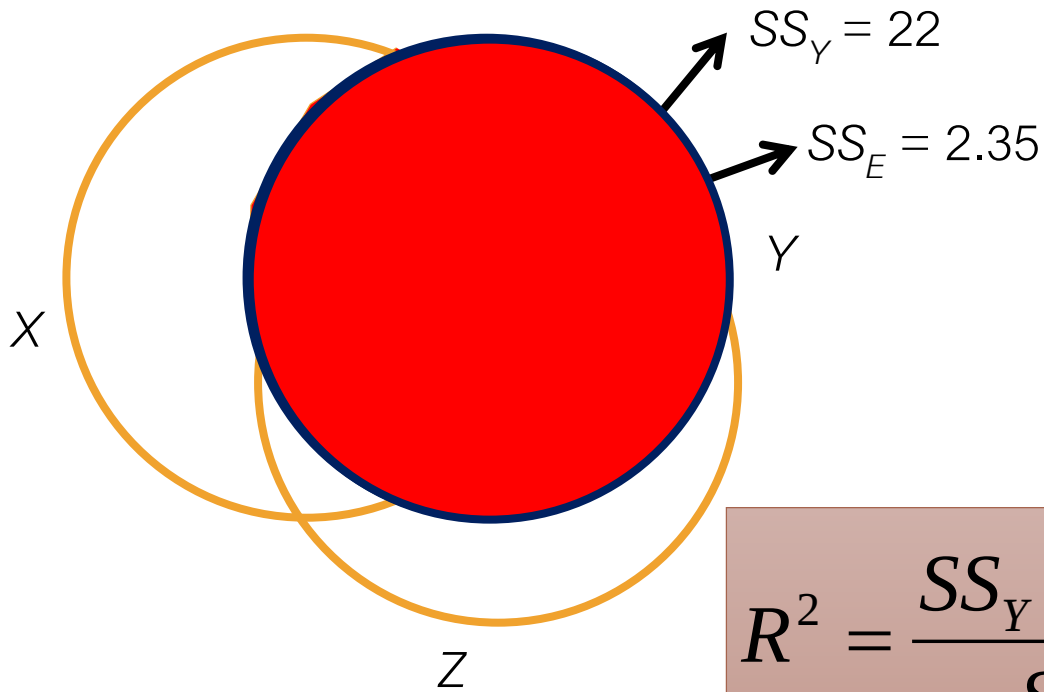
- เช่น จำนวนบัตรเครดิตที่ใช้ ถูกทำนายด้วยขนาดครอบครัวและรายได้

x	z	y	$\hat{y}$	e	e^2
2	3.5	4			
2	3.8	6			
4	4.0	6			
4	4.2	7			
5	4.4	8			
5	4.6	7			
6	4.8	8			
6	6.0	10			

$SS_{\text{Error}} = 2.35$

สัมประสิทธิ์การทำนาย

- เมื่อใส่ตัวแปรทำนาย 2 ตัว เหลือ $SS_E = 2.35$



$$R^2 = \frac{SS_Y - SS_E}{SS_Y} = \frac{22 - 2.35}{22} = .89$$

สัมประสิทธิ์การทำนาย

- แต่ทว่า R^2 เหมือนกับค่า Eta-squared ในการวิเคราะห์ความแปรปรวน ที่ไม่สามารถทำนายค่าของประชากรได้อย่างแม่นยำ
- หากตัวแปรอิสระไม่มีความสัมพันธ์กับตัวแปรตามในประชากรเลย ค่า R^2 ยังมีค่ามากกว่า 0 เกือบทุกครั้ง
- ค่า R^2 ที่ได้จากกลุ่มตัวอย่างยิ่งสูงเมื่อ
 - จำนวนกลุ่มตัวอย่างน้อย
 - จำนวนตัวแปรอิสระมาก

สัมประสิทธิ์การทำนาย

- Adjusted R^2 จึงถูกพัฒนาขึ้นมา เพื่อแก้ไขจุดบกพร่องของ R^2 (คล้ายกับ Omega-squared)

$$\tilde{R}^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - p - 1}$$

- เช่น ในตัวอย่างจำนวนบัตรเครดิตทำนายด้วยสมาชิกครอบครัวและรายได้

$$\tilde{R}^2 = 1 - \frac{(1 - .89)(8 - 1)}{8 - 2 - 1} = .85$$

สัมประสิทธิ์การทำนาย

- ค่า R^2 หรือ Adjusted R^2 สามารถใช้ตีความหมายขนาดอิทธิพลได้ โดย Cohen ได้แปลงค่า R^2 เป็น f^2 ที่สามารถตีความหมายได้ดังนี้

$$f^2 = \frac{R^2}{1 - R^2}$$

- น้อย: $f^2 = .02$, ปานกลาง: $f^2 = .15$, มาก: $f^2 = .35$
- สามารถแปลงเป็นระดับของ R^2 ได้ดังนี้
 - น้อย: $R^2 = .02$, ปานกลาง: $R^2 = .13$, มาก: $R^2 = .26$

สัมประสิทธิ์การถอดอามาตรฐาน

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันหัด พรประเสริฐมานิต

สัมประสิทธิ์การถดถอยมาตรฐาน

- โดยนิยามแล้ว สัมประสิทธิ์การถดถอย (Regression Coefficient) ในที่นี้ คือ ค่าที่อยู่ในสมการถดถอย ในที่นี้ คือ จุดตัดแกน Y (a) และความชัน ($b_1, b_2, \dots, b_p$)
- แต่โดยทั่วไป สัมประสิทธิ์ถดถอย จะหมายถึงความชัน

สัมประสิทธิ์การถดถอยมาตรฐาน

- ในการวิเคราะห์ถดถอยแบบง่าย หากนำตัวแปร X และ Y มาแปลงให้เป็นค่ามาตรฐาน z_X และ z_Y
- จากนั้น นำมาสร้างสมการถดถอยอีกครั้งหนึ่ง

$$z'_Y = \cancel{\alpha} + \beta z_X$$

$$z'_Y = \beta z_X$$

จุดตัดแกน Y เท่ากับศูนย์ เพราะผ่านจุดค่าเฉลี่ยของ X และ Y เสมอ หรือ z_X และ z_Y เท่ากับ $(0, 0)$

สัมประสิทธิ์การถดถอยมาตรฐาน

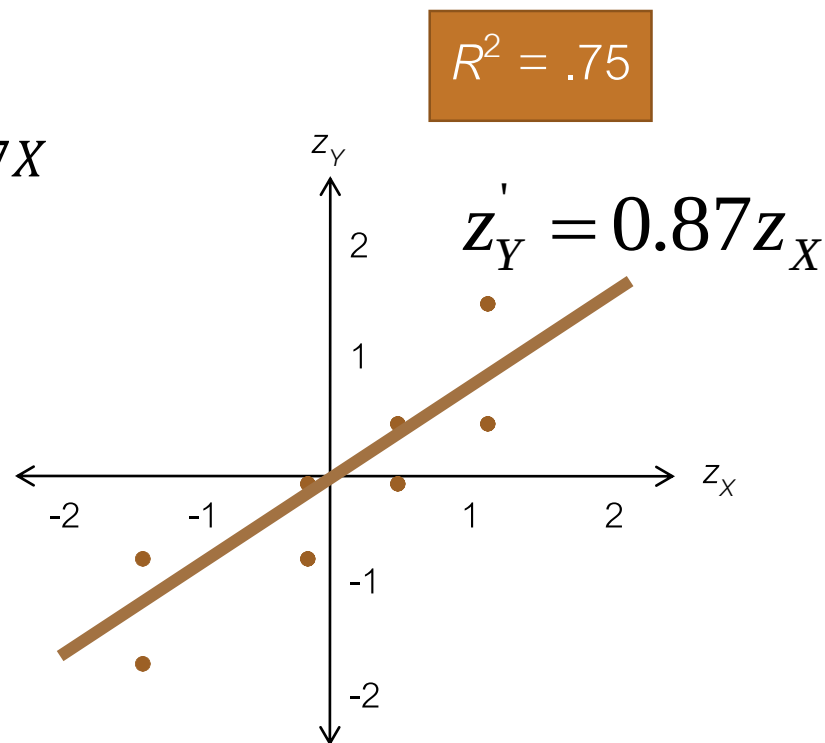
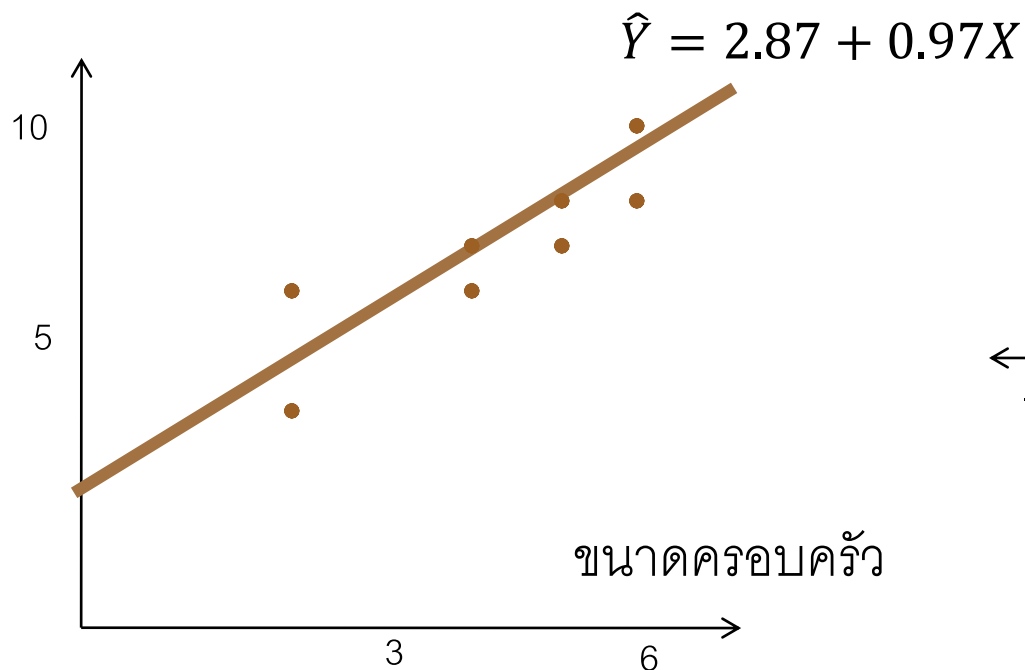
- ค่าความชันนี้ จะเรียกว่า สัมประสิทธิ์การถดถอยมาตรฐาน (Standardized Regression Coefficient, β)
- ค่าสัมประสิทธิ์ถดถอยมาตรฐานในกรณีนี้ จะเท่ากับค่าสหสัมพันธ์ มีค่าที่เป็นไปได้ตั้งแต่ -1 ถึง 1

$$\hat{z}_Y = \beta z_X$$

สัมประสิทธิ์การถดถอยมาตรฐาน

- เช่น ทำนายจำนวนบัตรเครดิตที่ใช้ในครอบครัว ด้วยขนาดครอบครัว
- เก็บข้อมูลจาก 8 ครอบครัว

จำนวนบัตรเครดิตที่ใช้



สัมประสิทธิ์การถดถอยมาตรฐาน

- ในการวิเคราะห์ถดถอยแบบพหุ ให้แปลงตัวแปรทุกตัวเป็นคะแนนมาตรฐาน และสร้างสมการถดถอย

$$\hat{z}_Y = \beta_1 z_{X_1} + \beta_2 z_{X_2} + \cdots + \beta_p z_{X_p}$$

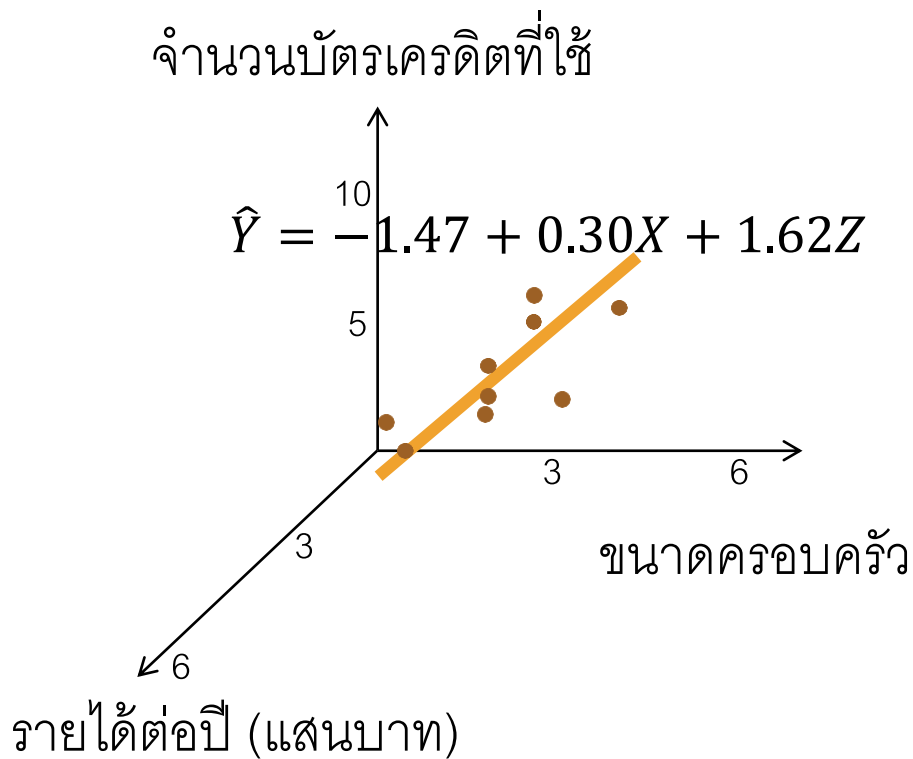
จุดตัดแกน Y เท่ากับศูนย์เช่นกันเพราะผ่านจุดค่าเฉลี่ยของตัวแปรทุกตัวเสมอ
กล่าวคือ $\hat{z}_Y = 0$ เมื่อ $z_{X_1} = z_{X_2} = \cdots = z_{X_p} = 0$

สัมประสิทธิ์การถดถอยมาตรฐาน

- ค่าสัมประสิทธิ์การถดถอยมาตรฐานนี้ จะมีประโยชน์ในการเปรียบเทียบอำนาจในการทำนายของตัวแปรภายในสมการ
- กล่าวคือ หากค่าสัมบูรณ์ของสัมประสิทธิ์การถดถอยมาตรฐานของตัวแปรใด สูงกว่า แสดงว่าตัวแปรนั้นมีอิทธิพลมากกว่า

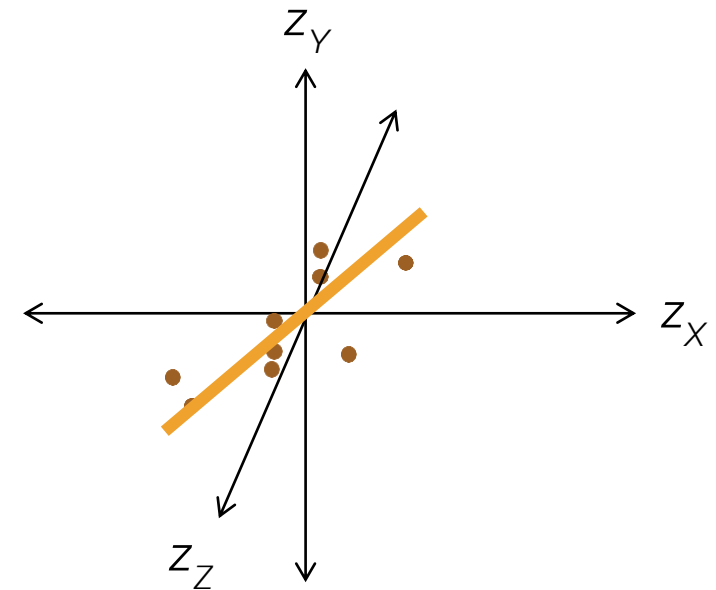
มโนทัศน์พื้นฐาน

- เช่น ตัวอย่างจำนวนบัตรเครดิต



$$R^2 = .89$$

$$\hat{z}_Y = 0.27z_X + 0.71z_Z$$



สัมประสิทธิ์การถดถอยมาตรฐาน

- (ข้อควรระวัง) คนส่วนใหญ่มักเข้าใจว่า ค่าสัมประสิทธิ์การถดถอยมาตรฐานจะมีค่าระหว่าง -1 ถึง 1 เท่านั้น แท้จริงแล้วไม่จริง
- หากมีตัวแปรอิสระมากกว่า 2 ตัว และความสัมพันธ์ระหว่างตัวแปรอิสระไม่เท่ากับ 0 แล้ว ค่าสัมประสิทธิ์การถดถอยมาตรฐานอาจสูงกว่า 1 หรือต่ำกว่า -1 ได้
- โดยเฉพาะอย่างยิ่ง ความสัมพันธ์ระหว่างตัวแปรอิสระสูงมากๆ

$r(X_1, X_2) = .5$

ในเซลล์แสดงถึง β_1, β_2

$r(Y, X_1) \setminus r(Y, X_2)$	.1	.2	.3	.4	.5	.6
.1	.067, .067	.000, .200	-.067, .333	-.133, .467	-.200, .600	-.267, .733
.2	.200, .000	.133, .133	.067, .267	.000, .400	-.067, .533	-.133, .667
.3	.333, -.067	.267, .067	.200, .200	.133, .333	.067, .467	.000, .600
.4	.467, -.133	.400, .000	.333, .133	.267, .267	.200, .400	.133, .533
.5	.600, -.200	.533, -.067	.467, .067	.400, .200	.333, .333	.267, .467
.6	.733, -.267	.667, -.133	.600, .000	.533, .133	.467, .267	.400, .400

$r(X_1, X_2) = .9$

ในเซลล์แสดงถึง β_1, β_2

$r(Y, X_1) \setminus r(Y, X_2)$	.1	.2	.3	.4	.5	.6
.1	.052, .052	.421, .579	-.895, 1.105	-1.368, 1.631	-1.842, 2.158	
.2	.579, .421	.105, .105	-.368, .632	-.842, 1.158	-1.316, 1.684	-1.789, 2.211
.3	1.105, -.895	.632, -.368	.158, .158	-.316, .684	-.789, 1.211	-1.263, 1.737
.4	1.631, -1.368	1.158, -.842	.684, -.316	.211, .211	-.263, .737	-.737, 1.263
.5	2.158, -1.842	1.684, -1.316	1.211, -.789	.737, -.263	.263, .263	-.211, .789
.6		2.211, -1.789	1.737, -1.263	1.263, -.737	.789, -.211	.316, .316

สัมประสิทธิ์การถดถอยมาตรฐาน

- (ข้อควรระวัง) สัมประสิทธิ์ถดถอยมาตรฐาน สามารถใช้เปรียบเทียบอิทธิพลของตัวแปรในสมการเดียวกันเท่านั้น ไม่สามารถเปรียบเทียบค่ากับสมการอื่นได้

$$\hat{z}_Y = \beta_1 z_{X_1} + \beta_2 z_{X_2} + \cdots + \beta_p z_{X_p}$$

เปรียบเทียบกันได้

กลุ่มที่ 1

$$\hat{z}_Y = \beta_1 z_{X_1} + \beta_2 z_{X_2}$$

เปรียบเทียบกันไม่ได้

กลุ่มที่ 2

$$\hat{z}_Y = \beta_1 z_{X_1} + \beta_2 z_{X_2}$$

สัมประสิทธิ์การถดถอยมาตรฐาน

- (ข้อควรระวัง) คนส่วนใหญ่มักเข้าใจว่า ค่าสัมประสิทธิ์ถดถอยมาตรฐาน และค่าสหสัมพันธ์ต้องเป็นไปในทิศทางเดียวกัน
- จริงๆ แล้วไม่ใช่ ถึงแม้ว่าค่าสหสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามในการวิเคราะห์ถดถอยแบบพหุเป็นไปในทางบวก ความสัมพันธ์แบบถดถอยอาจเป็นไปในทางลบ

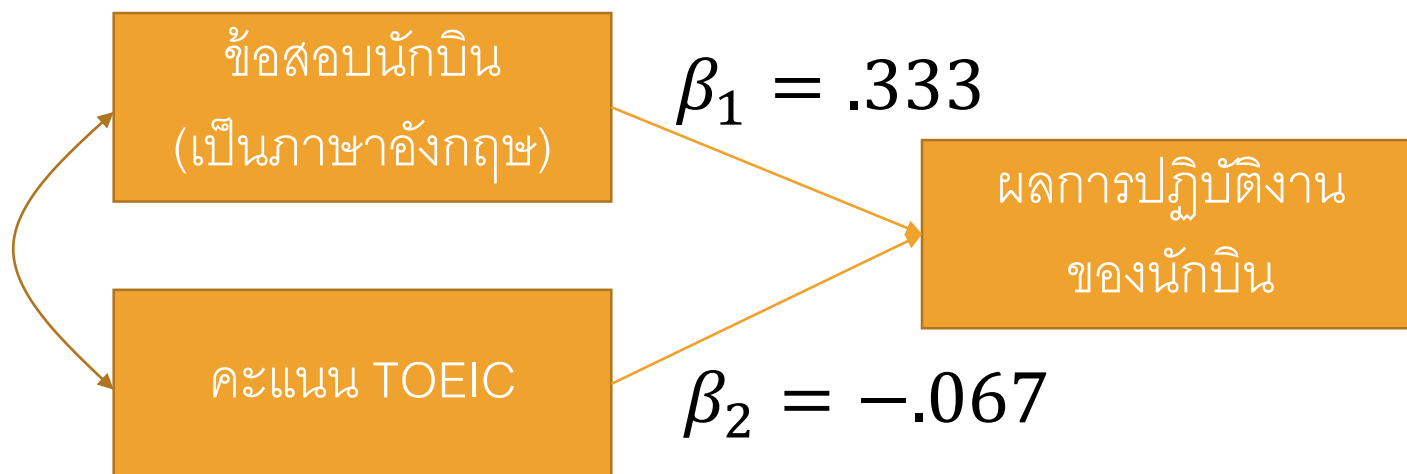
$$r(X_1, X_2) = .5$$

ในเซลล์แสดงถึง β_1, β_2

$r(Y, X_1) \setminus r(Y, X_2)$	.1	.2	.3	.4	.5	.6
.1	.067, .067	.000, .200	-.067, .333	-.133, .467	-.200, .600	-.267, .733

สัมประสิทธิ์การถดถอยมาตรฐาน

- ตัวอย่างของกรณีนี้ เช่น



	ผลงาน	ข้อสอบ	TOEIC
ผลงาน			
ข้อสอบ	.3		
TOEIC	.1	.5	

การทดสอบสมมติฐาน

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

การทดสอบสมมติฐาน

- แท้จริงแล้ว ผู้วิจัยต้องการทดสอบว่า ตัวแปร X ทั้งหมด สามารถทำนาย Y ได้จริงหรือไม่
- กล่าวคือ การใส่ตัวแปร X ทั้งหมดนั้นสามารถช่วยทำนายตัวแปร Y ได้ดีกว่าค่าเฉลี่ยของตัวแปร Y หรือไม่
- การทดสอบนี้ คือ การทดสอบว่าสัมประสิทธิ์ทำนายสูงกว่า 0 หรือไม่ จะทดสอบด้วย F -test

$$H_0: P^2 = 0$$

$$H_1: P^2 > 0$$

P หมายถึงตัว Rho ขนาดใหญ่
ซึ่งเป็นสัญลักษณ์ค่าพารามิเตอร์ของ R^2

การทดสอบสมมติฐาน

- ในการวิเคราะห์การถดถอย จะมีรูปแบบการแสดงผลค่า F ดังต่อไปนี้

ความแตกต่างที่สามารถอธิบายได้ด้วยตัวแปรอิสระ

Effects	SS	df	MS	F	p
Regression	SS_{Reg}	p	MS_{Reg}	F	p
Error	SS_{Error}	$n - p - 1$	MS_{Error}		
Total	SS_{Total}				

ความแตกต่างที่ไม่สามารถวัดได้ด้วยตัวแปรอิสระ

การทดสอบสมมติฐาน

- ในการวิเคราะห์การถดถอย จะมีรูปแบบการแสดงผลค่า F ดังต่อไปนี้

$$SS_{\text{Total}} = \sum (Y - \bar{Y})^2$$

$$SS_{\text{Error}} = \sum (Y - \hat{Y})^2$$

Effects	SS	df	MS	F	p
Regression	SS_{Reg}	p	MS_{Reg}	F	p
Error	SS_{Error}	$n - p - 1$	MS_{Error}		
Total	SS_{Total}				

$$SS_{\text{Total}} = SS_{\text{Reg}} + SS_{\text{Error}}$$

การทดสอบสมมติฐาน

- ในการวิเคราะห์การถดถอย จะมีรูปแบบการแสดงผลค่า F ดังต่อไปนี้

$$MS_{\text{Reg}} = SS_{\text{Reg}}/df_{\text{Reg}}$$
$$MS_{\text{Error}} = SS_{\text{Error}}/df_{\text{Error}}$$
$$F = MS_{\text{Reg}}/MS_{\text{Error}}$$

Effects	SS	df	MS	F	p
Regression	SS_{Reg}	p	MS_{Reg}	F	p
Error	SS_{Error}	$n - p - 1$	MS_{Error}		
Total	SS_{Total}				

ทดสอบอทิทธิพลของตัวแปรอิสระ ถ้าน้อยกว่า $\alpha \rightarrow \text{Reject } H_0$

การทดสอบสมมติฐาน

- ในการวิเคราะห์การถดถอย จะมีรูปแบบการแสดงผลค่า F ดังต่อไปนี้

ส่วนเบี่ยงเบนมาตรฐานของ e หรือ $Y - \hat{Y}$

(Standard error of the estimate)

$$s_e = \sqrt{MS_{\text{Error}}}$$

ใน ANOVA เหมือนกับส่วนเบี่ยงเบนมาตรฐาน
ของตัวแปรตามภายในกลุ่ม (ใช้คำนวณ d)

Effects	SS	df	MS	F	p
Regression	SS_{Reg}	p	MS_{Reg}	F	p
Error	SS_{Error}	$n - p - 1$	MS_{Error}		
Total	SS_{Total}				

การทดสอบสมมติฐาน

- ผลการวิเคราะห์การถดถอย ในตัวอย่างเรื่องจำนวนบัตรเครดิต
 - $SS_{\text{Reg}} = 19.652$, $SS_{\text{Error}} = 2.348$, $SS_{\text{Total}} = 22.000$
 - $df_{\text{Reg}} = 2$, $df_{\text{Error}} = 5$, $df_{\text{Total}} = 7$
 - $MS_{\text{Reg}} = 9.826$, $MS_{\text{Error}} = 0.470$, $s_e = 0.686$
 - $F(2, 5) = 20.928$, $p = .004$

Effects	SS	df	MS	F	p
Regression	19.652	2	9.826	20.928	.004
Error	2.348	5	0.470		
Total	22.000	7			

ตัวแปรอิสระมีผลต่อตัวแปรตาม ทดสอบพบว่าตัวใดบ้าง

$p < \alpha \rightarrow \text{Reject } H_0$

การทดสอบสมมติฐาน

- หลังจากนี้ จะทดสอบต่อว่าตัวแปรอิสระแต่ละตัว มีผลต่อตัวแปรตามหรือไม่
- กล่าวคือ ทดสอบว่าสัมประสิทธิ์ถดถอยของตัวแปร X แต่ละตัวแตกต่างจาก 0 หรือไม่ ด้วย t -test

$$H_0: b_{X_p} = 0$$

$$H_1: b_{X_p} \neq 0$$

b_{X_p} ในที่นี้หมายถึงค่าสัมประสิทธิ์ถดถอยในประชากรในการทำนาย Y ด้วย X_p โดยควบคุมตัวแปรอื่นให้คงที่

การทดสอบสมมติฐาน

- นอกจากนี้ ยังมีการทดสอบว่า จุดตัดแกน Y แตกต่างจาก 0 หรือไม่ ด้วย t -test

$$H_0: a = 0$$

$$H_1: a \neq 0$$

- หากค่า X เท่ากับ 0 ไม่มีความหมาย การทดสอบนี้ก็ไร้ความหมายเช่นกัน
 - เช่น เป็นไปไม่ได้ ที่จำนวนครอบครัวเท่ากับ 0
 - การทดสอบนี้จะมีประโยชน์ หากมีการเปลี่ยนค่าศูนย์กลาง (Centering) ซึ่งกล่าวในบทถัดไป

การทดสอบสมมติฐาน

- ผลการวิเคราะห์การถดถอย ในตัวอย่างเรื่องจำนวนบัตรเครดิต

ตัวแปร	b	SE_b	β	t	p
จุดตัดแกน Y	-1.47	1.83		-0.801	.459
ขนาดครอบครัว (X)	0.31	0.31	.27	0.998	.364
รายได้ (Z)	1.63	0.63	.70	2.585	.049

$$\hat{Y} = -1.47 + 0.31X + 1.63Z$$

$$\hat{z}_Y = 0.27z_X + 0.70z_Z$$

การทดสอบสมมติฐาน

- ผลการวิเคราะห์การถดถอย ในตัวอย่างเรื่องจำนวนบัตรเครดิต

ตัวแปร	b	SE_b	β	t	p
จุดตัดแกน Y	-1.47	1.83		-0.801	.459
ขนาดครอบครัว (X)	0.31	0.31	.27	0.998	.364
รายได้ (Z)	1.63	0.63	.70	2.585	.049

$a = -1.47$, $t(5) = -0.801$, $p = .46 \rightarrow$ Fail to reject H_0

จำนวนบัตรเครดิตเท่ากับ -1.47 ไบ เมื่อขนาดครอบครัวเท่ากับ 0 และรายได้เท่ากับ 0 ซึ่งจำนวนบัตรนั้นไม่แตกต่างจาก 0 อย่างมีนัยสำคัญ

การทดสอบสมมติฐาน

- ผลการวิเคราะห์การถดถอย ในตัวอย่างเรื่องจำนวนบัตรเครดิต

ตัวแปร	b	SE_b	β	t	p
จุดตัดแกน Y	-1.47	1.83		-0.801	.459
ขนาดครอบครัว (X)	0.31	0.31	.27	0.998	.364
รายได้ (Z)	1.63	0.63	.70	2.585	.049

$$b_x = 0.31, t(5) = 0.998, p = .36 \rightarrow \text{Fail to reject } H_0$$

จำนวนบัตรเครดิตเพิ่มขึ้น 0.31 ใบหากขนาดครอบครัวเพิ่มขึ้น 1 คน เมื่อควบคุมรายได้ให้คงที่ อิทธิพลของขนาดครอบครัวต่อจำนวนบัตรเมื่อกontrolรายได้นั้นไม่ถึงระดับนัยสำคัญ

การทดสอบสมมติฐาน

- ผลการวิเคราะห์การถดถอย ในตัวอย่างเรื่องจำนวนบัตรเครดิต

ตัวแปร	b	SE_b	β	t	p
จุดตัดแกน Y	-1.47	1.83		-0.801	.459
ขนาดครอบครัว (X)	0.31	0.31	.27	0.998	.364
รายได้ (Z)	1.63	0.63	.70	2.585	.049

$$b_Z = 1.63, t(5) = 2.585, p = .049 \rightarrow \text{Reject } H_0$$

จำนวนบัตรเครดิตเพิ่มขึ้น 1.63 ใบหากรายได้ต่อปีเพิ่มขึ้น 1 แสนบาทเมื่อควบคุมขนาดครอบครัวให้คงที่ กล่าวคือ อิทธิพลของรายได้ต่อจำนวนบัตรเครดิตถึงระดับนัยสำคัญเมื่อควบคุมขนาดครอบครัว

การทดสอบสมมติฐาน

- ในกรณี การวิเคราะห์ถดถอยอย่างง่าย ตัวแปรต้นจะมีแค่ตัวแปรเดียว จะทำให้ผลการวิเคราะห์ง่ายขึ้นดังต่อไปนี้
 - การทดสอบ R^2 ที่ได้ค่า F และการทดสอบค่า b ที่ได้ค่า t จะมีค่า p เหมือนกัน โดย $t^2 = F$
 - ค่า R^2 จะเท่ากับค่าสหสัมพันธ์ของตัวแปรทั้งสองยกกำลังสอง หรือ r^2
 - การทดสอบค่าสหสัมพันธ์ จะมีค่า p เท่ากับการทดสอบ F และ t ด้วย
 - ด้วยเหตุนี้ การรายงานผลการทดสอบทางสถิติ สามารถเลือกรายงานผลเพียงแค่ตัวใดตัวหนึ่ง ซึ่งมักนิยมรายงานการทดสอบสัมประสิทธิ์ถดถอย

การทดสอบสมมติฐาน

อายุและจำนวนการเลือกของเล่นถูกเพศ

จำนวน 10 คน



สมมติฐาน คือ เมื่ออายุมากขึ้นมี
แนวโน้มจะเลือกของเล่นถูกเพศมากขึ้น

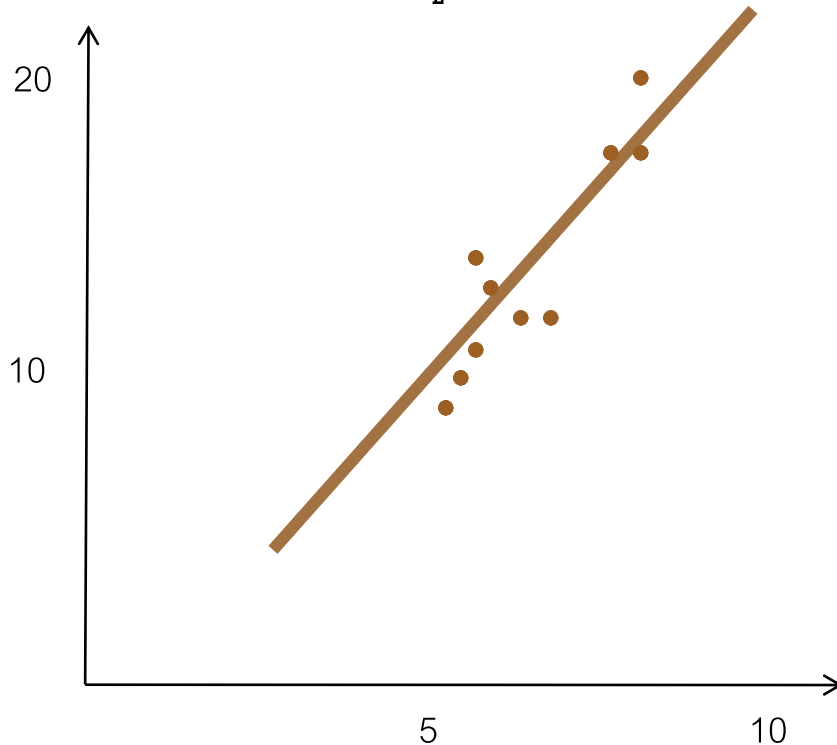
$$H_0: \boldsymbol{\rho}_{XY} = 0$$

$$H_1: \rho_{XY} \neq 0$$

การทดสอบสมมติฐาน

อายุและจำนวนการเลือกของเล่นถูกเพศ
จำนวน 10 คน

จำนวนการเลือกของเล่นถูกเพศ



X	Y
7.5	18
7.0	13
5.5	11
8.0	20
6.5	13
6.0	14
5.0	9
8.0	18
6.5	14
6.0	10

$$M_X = 6.6$$

$$M_Y = 14$$

$$s_{XY} = 3.44$$

$$SD_X = 1.02$$

$$SD_Y = 3.65$$

$$r_{XY} = .923$$

$$p = .0002$$

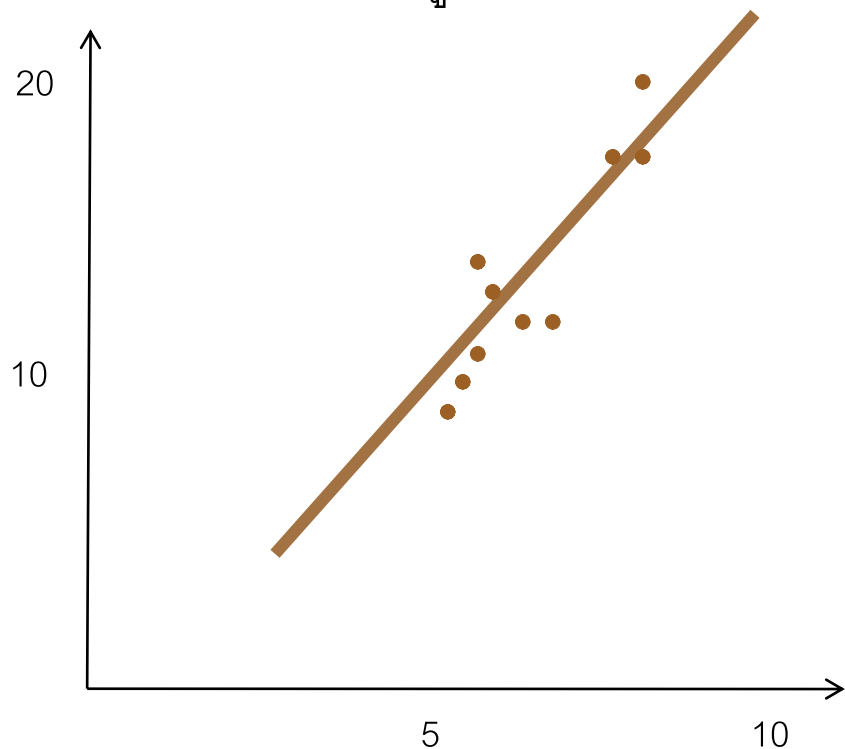
อายุ

การทดสอบสมมติฐาน

อายุและจำนวนการเลือกของเล่นถูกเพศ

จำนวน 10 คน

จำนวนการเลือกของเล่นถูกเพศ



$$\begin{aligned} b &= r \cdot \frac{SD_Y}{SD_X} \\ &= .923 \cdot \frac{3.65}{1.02} \\ &= 3.30 \end{aligned}$$

$$\begin{aligned} a &= M_Y - bM_X \\ &= 14 - 3.30 \times 6.6 \\ &= -7.78 \end{aligned}$$

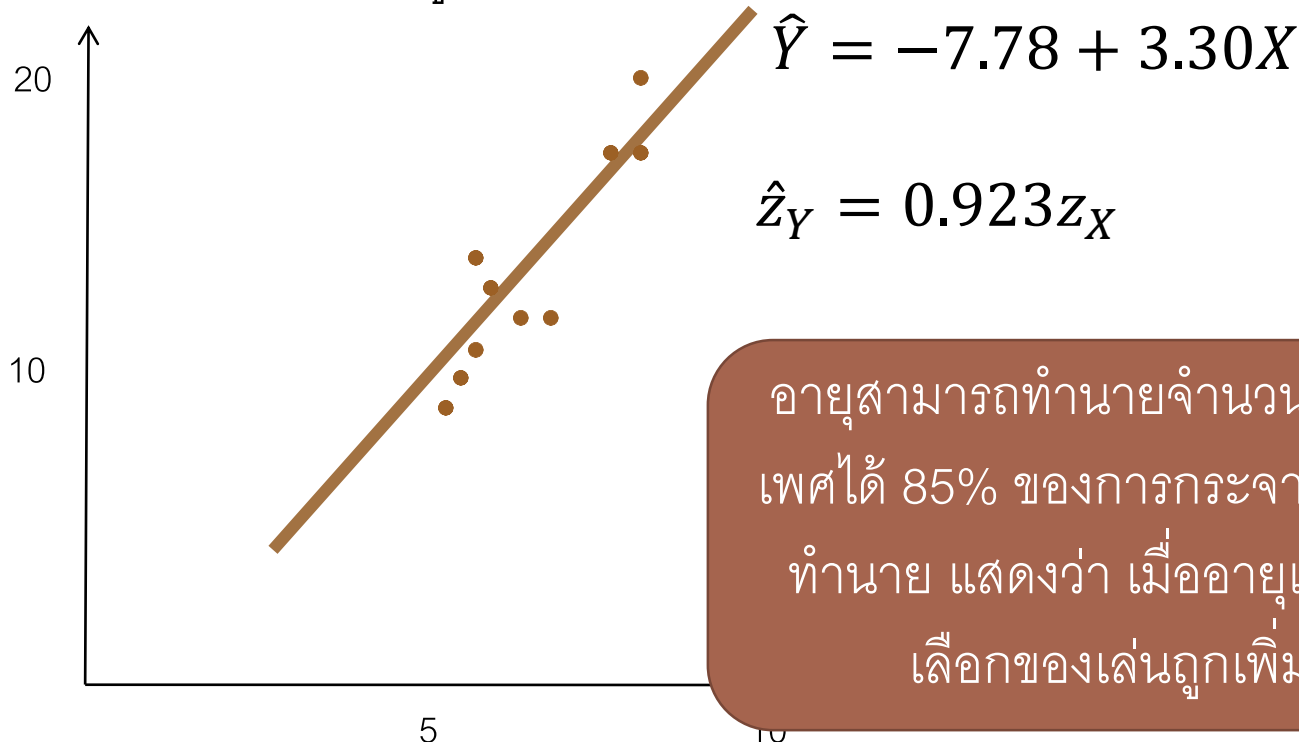
การทดสอบสมมติฐาน

อายุและจำนวนการเลือกของเล่นถูกเพศ
จำนวน 10 คน

$$r_{XY} = .923, p = .0002$$

ตัวแปรทั้งสองมีความสัมพันธ์กันจริง

จำนวนการเลือกของเล่นถูกเพศ



$$R^2 = .85$$

อายุสามารถทำนายจำนวนการเลือกของเล่นถูกเพศได้ 85% ของการกระจายทั้งหมด จากสมการทำนาย แสดงว่า เมื่ออายุเพิ่ม 1 ปี จำนวนการเลือกของเล่นถูกเพิ่มขึ้น 3.30 ครั้ง

การทดสอบสมมติฐาน

ผู้วิจัยต้องการทราบว่าลักษณะผู้นำ
แบบปฏิรูปของหัวหน้า จะสามารถทำนาย
ผลการปฏิบัติงานของลูกน้องได้อย่างไร



เก็บข้อมูลจากพนักงาน 200 คน



สมมติฐาน คือ ลักษณะผู้นำแบบ
ปฏิรูปของหัวหน้าสามารถทำนาย
ผลงานได้



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



$$H_1: \rho_{XY} \neq 0$$

การทดสอบสมมติฐาน

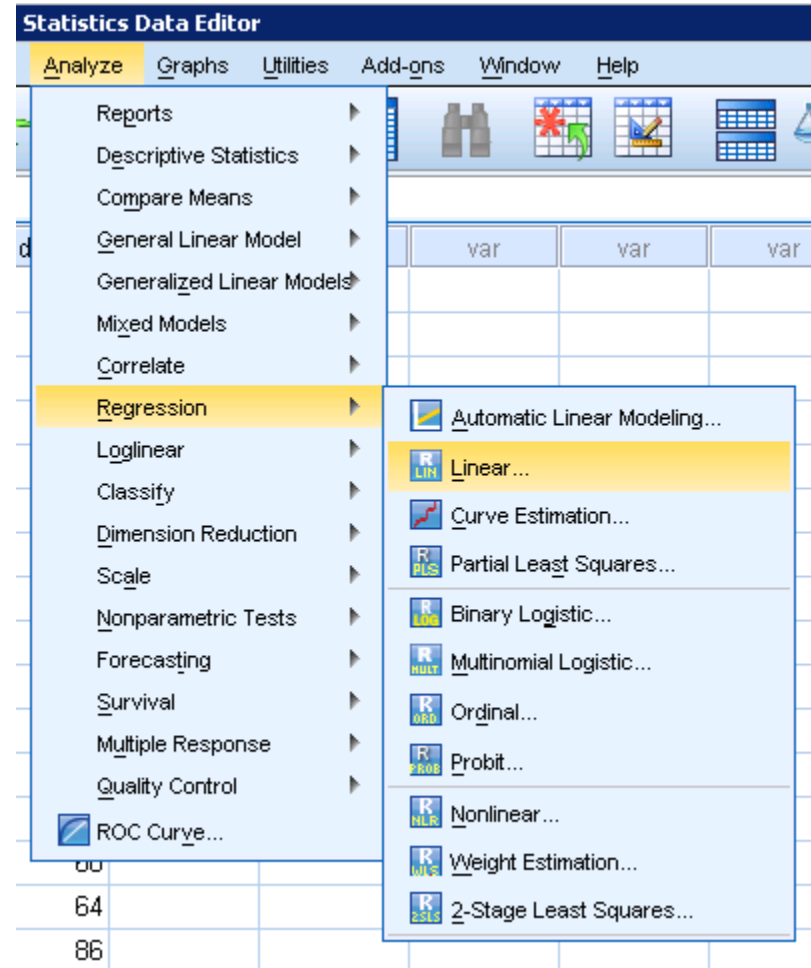
ผู้วิจัยต้องการทราบว่าลักษณะผู้นำ
แบบปฏิรูปของหัวหน้า จะสามารถทำนาย
ผลการปฏิบัติงานของลูกน้องได้อย่างไร



*Untitled2 [DataSet2] - IBM SPSS Statistics Data Editor						
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help						
	tfleader	producti	var	var	var	var
1	3.90	76				
2	3.98	64				
3	3.77	81				
4	4.14	82				
5	2.86	67				
6	3.68	84				
7	3.56	88				
8	3.69	66				
9	3.21	66				
10	3.28	73				
11	4.54	65				
12	2.70	72				
13	3.15	79				
14	2.86	68				
15	3.46	64				
16	4.02	86				
17	3.63	79				
18	4.25	85				
19	3.71	84				
20	3.39	76				
21	2.34	67				

การทดสอบสมมติฐาน

ผู้วิจัยต้องการทราบว่าลักษณะผู้นำ
แบบปฏิรูปของหัวหน้า จะสามารถทำนาย
ผลการปฏิบัติงานของลูกน้องได้อย่างไร



การทดสอบสมมติฐาน

ผู้วิจัยต้องการทราบว่าลักษณะ
แบบปฏิรูปของหัวหน้า จะสามารถ
ผลการปฏิบัติงานของลูกน้องได้



Linear Regression

Dependent: producti

Independent(s): tleader

Method: Enter

Selection Variable:

Case Labels:

WLS Weight:

OK Paste Reset Cancel Help

Statistics... Plots... Save... Options...

Previous Next

Block 1 of 1

ใส่ตัวแปร Y

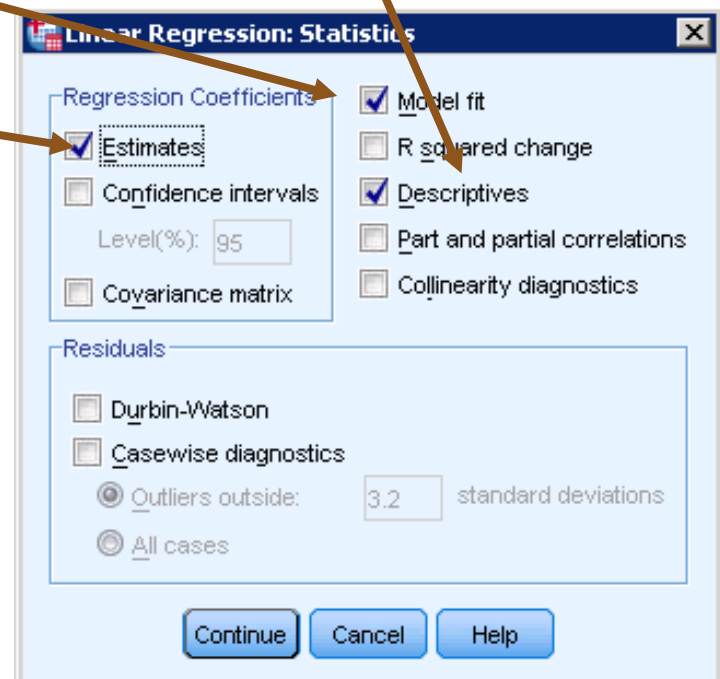
ใส่ตัวแปร X
ทั้งหมด

การทดสอบสมมติฐาน

ค่าสถิติพื้นฐานและค่าสหสัมพันธ์ของตัวแปรแต่ละตัว

ค่าสัมประสิทธิ์การทำนาย

ค่าสัมประสิทธิ์ถดถอย



กด Statistics... เพื่อเรียกข้อมูลเพิ่มเติม

การทดสอบสมมติฐาน

ผู้วิจัยต้องการทราบว่าลักษณะผู้นำ
แบบปฏิรูปของหัวหน้า จะสามารถทำนาย
ผลการปฏิบัติงานของลูกน้องได้อย่างไร



$$r = .458, p < .001$$

Descriptive Statistics

	Mean	Std. Deviation	N
producti	75.32	7.758	200
tfleader	3.5390	.52580	200

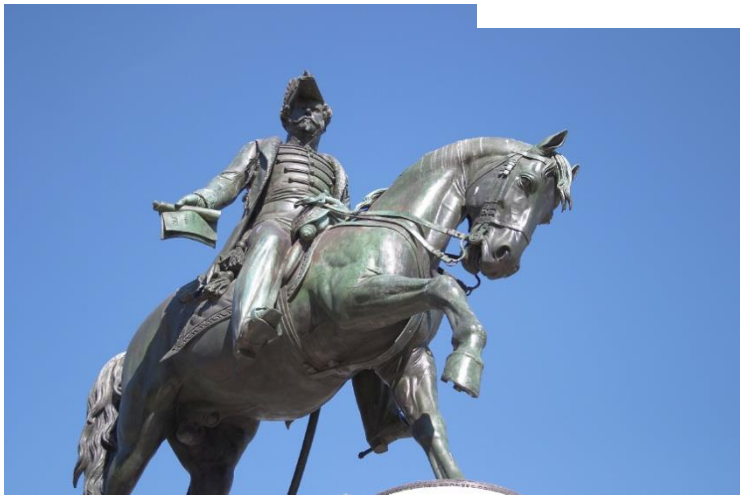
Correlations

		producti	tfleader
Pearson Correlation	producti	1.000	.458
	tfleader	.458	1.000
Sig. (1-tailed)	producti	.	.000
	tfleader	.000	.
N	producti	200	200
	tfleader	200	200

ค่าสถิติพรรณนาและค่าสหสัมพันธ์
ควรรายงานในผลงานวิจัยเสมอ

การทดสอบสมมติฐาน

ผู้วิจัยต้องการทราบว่า
แบบปฏิรูปของหัวหน้า จะ
ผลการปฏิบัติงานของลูก



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.458 ^a	.209	.205	6.915

a. Predictors: (Constant), ffileader

$R = |r| = .458$

$R^2 = .209$

Adjusted $R^2 = .205$

$s_e = 6.915$

การทดสอบสมมติฐาน

ผู้วิจัยตั้ง
แบบปฏิรูป
ผลการปฏิ

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2508.523	1	2508.523	52.456	.000 ^b
	Residual	9468.632	198	47.821		
	Total	11977.155	199			

a. Dependent Variable: producti

b. Predictors: (Constant), ttleader

$$R^2 = .215, F(1, 198) = 52.46, p < .001$$

ตัวแปรอิสระสามารถทำนายตัวแปรตามได้อย่างมีนัยสำคัญ

$$p < \alpha$$

ทดสอบต่อว่าตัวแปรอิสระตัวใด ที่สามารถทำนายตัวแปรตามได้ ในตัวอย่างนี้มีตัวแปรเดียว จึงแน่นอนว่าต้องเป็นตัวแปรการเป็นผู้นำแบบนักปฏิรูป

การทดสอบสมมติฐาน

ผู้วิจัยต้องการทราบว่าลักษณะผู้นำ
แบบปฏิรูปของหัวหน้า จะสามารถทำนาย
ผลการปฏิบัติงานของลูกน้องได้อย่างไร

$$Y = a + bX$$

$\beta = .458$

$a = 51.418$

$b = 6.752$

Coefficient

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	51.418	3.336		15.415	.000
	tfleader	6.752	.932	.458	7.243	.000

a. Dependent Variable: productl

Y-intercept แตกต่าง
จากศูนย์อย่างมีนัยสำคัญ

ทดสอบสมมติฐาน
ของ Y-intercept = 0

slope แตกต่าง
จากศูนย์อย่างมีนัยสำคัญ

ทดสอบสมมติฐาน
ของ slope = 0

การทดสอบสมมติฐาน

ผู้วิจัยต้องการทราบว่าลักษณะผู้นำ
แบบปฏิรูปของหัวหน้า จะสามารถทำนาย
ผลการปฏิบัติงานของลูกน้องได้อย่างไร

สรุป



ลักษณะผู้นำแบบปฏิรูปของ
หัวหน้า สามารถทำนายผล
การปฏิบัติงานของลูกน้องได้
21% ของการกระจาย

เมื่อลักษณะผู้นำแบบปฏิรูปของ
หัวหน้า เพิ่มขึ้น 1 คะแนน จะ
ทำนายผลการปฏิบัติงานของ
ลูกน้องเพิ่มขึ้น 7.25 คะแนน

การทดสอบสมมติฐาน

นักวิจัยต้องการทดสอบว่าความกดดัน
จากที่บ้านและความกดดันจาก
ที่ทำงาน อะไรมีผลต่อความเครียด
มากกว่ากัน



นักวิจัยจึงเก็บข้อมูลจากพนักงานบริษัท
จำนวน 200 คน

สมมติฐาน คือ ความกดดันทั้ง
สองแบบมีผลต่อความเครียด

$$H_0: P^2 = 0$$

$$H_1: P^2 > 0$$

การทดสอบสมมติฐาน

นักวิจัยต้องการทดสอบว่าความกดดัน
จากที่บ้านและความกดดันจาก
ที่ทำงาน อะไรมีผลต่อความเครียด
มากกว่ากัน



ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน
ของตัวแปรแต่ละตัว

Descriptive Statistics

	Mean	Std. Deviation	N
stress	52.09897959	11.33471799	100
homepress	50.25	10.758	100
workpress	51.03	10.634	100

การทดสอบสมมติฐาน

นักวิจัยต้องการทดสอบว่าความ
จากที่บ้านและความกดดัน
ที่ทำงาน อะไรมีผลต่อความ
มากกว่ากัน



Correlations

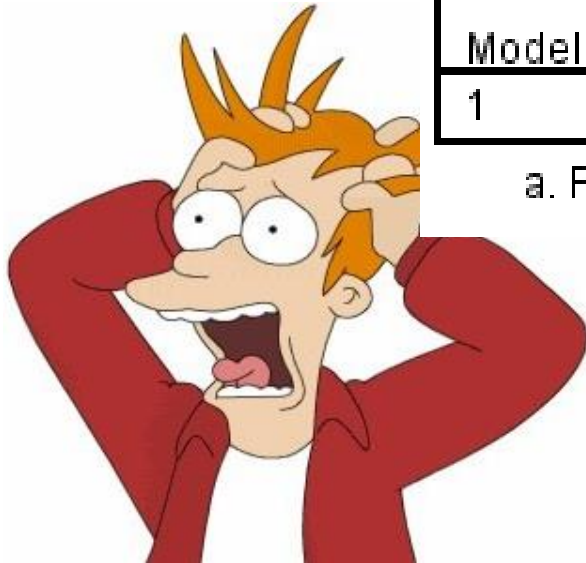
		stress	homepress	workpress
Pearson Correlation	stress	1.000	.568	.481
	homepress	.568	1.000	.554
	workpress	.481	.554	1.000
Sig. (1-tailed)	stress	.	.000	.000
	homepress	.000	.	.000
	workpress	.000	.000	.
N	stress	100	100	100
	homepress	100	100	100
	workpress	100	100	100

ความสัมพันธ์ของตัวแปรแต่ละต้นอยู่ระหว่าง
สูงหรือค่อนข้างสูง และถึงระดับนัยสำคัญ

การทดสอบสมมติฐาน

นักวิจัยต้องการทดสอบว่าความกดดัน

จากที่บ้านและความกดดันจาก
ที่ทำงาน อะไรมีผลต่อ
มากกว่ากัน



$$R = .602$$

Model Summary

$$R^2 = .363$$

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.602 ^a	.363	.350	9.141825756

a. Predictors: (Constant), workpress, homepress

$$\text{Adjusted } R^2 = .350$$

$$s_e = 9.14$$

การทดสอบสมมติฐาน

นักวิจัยต้องการทดสอบว่าความกดดัน

จากที่บ้านและ
ที่ทำงาน อะไรมีผล
มากกว่า

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4612.528	2	2306.264	27.596	.000 ^b
	Residual	8106.579	97	83.573		
	Total	12719.107	99			

a. Dependent Variable: stress

b. Predictors: (Constant), workpress, homepress

$$R^2 = .363, F(2, 97) = 27.60, p < .001$$

ตัวแปรอิสระสามารถทำนายตัวแปรตามได้อย่างมีนัยสำคัญ

$$p < \alpha$$

ทดสอบต่อว่าตัวแปรอิสระตัวใด ที่สามารถทำนายตัวแปรตามได้

การทดสอบสมมติฐาน

$$\hat{Y} = 16.011 + 0.459X_1 + 0.255X_2$$

นักวิจัยต้อง
จากทาง
ที่ทำงาน

Coefficients ^a					
		Unstandardized Coefficients		Standardized Coefficients	
Model		B	Std. Error	Beta	
1	(Constant)	16.011	5.021		3.189 .002
	homepress	.459	.103	.436	4.473 .000
	workpress	.255	.104	.240	2.461 .016

a. Dependent Variable: stress



$$a = 16.011$$

จุดตัดแกน Y ไม่มีความหมาย เพราะความกดดันทางบ้าน
และจากที่ทำงาน คะแนนเท่ากับ 0 ไม่มีความหมาย

ทดสอบสมมติฐาน
ของ Y-intercept = 0

Y-intercept แตกต่าง
จากศูนย์อย่างมีนัยสำคัญ

การทดสอบสมมติฐาน

นักวิจัยต้อง
จากทางบ้าน
ที่ทำงาน

$$b_{\text{HOME}} = 0.459, t(97) = 4.47, p < .001, \beta = .436$$

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	16.011	5.021		3.189	.002
	homepress	.459	.103	.436	4.473	.000
	workpress	.255	.104	.240	2.461	.016

a. Dependent Variable: stress



$$b_{\text{WORK}} = 0.255, t(97) = 2.46, p = .016, \beta = .240$$

ความกดดันจากทางบ้านและที่ทำงานมีผลอย่างมีนัยสำคัญทั้งคู่ แต่อิทธิพลจากทางบ้านสูงกว่าชัดเจน (การเปรียบเทียบนี้ไม่ได้ทดสอบทางสถิติ ไม่แน่ว่าจะต่างจริงในประชากรหรือไม่)

การทดสอบสมมติฐาน

นักวิจัยต้องการทดสอบว่าความกดดัน
จากที่บ้านและความกดดันจาก
ที่ทำงาน อะไรมีผลต่อความเครียด
มากกว่ากัน

สรุป



ตัวแปรอิสระทั้งสองสามารถ
อธิบายการกระจายของ
ความเครียดได้ 36%

ทั้งความกดดันจากที่บ้านและ
ความกดดันจากที่ทำงานมีผล
อย่างมีนัยสำคัญ แต่ความกดดัน
จากที่บ้านมีอิทธิพลมากกว่า

การทดสอบสมมติฐาน

ท่านต้องการทดสอบว่าอะไร
ทำให้คนอื่นมองว่าคนน่ารัก
ท่านจึงทำนายความน่ารักด้วย
(1) หน้าตาดี (2) หุ่นดี (3) นิสัยดี
(4) ความใกล้ชิดสนิทสนม



นักวิจัยจึงเก็บข้อมูลจากนิสิตระดับปริญญาตรี
จากวิชาหนึ่ง จำนวน 50 คน



สมมติฐาน คือ ตัวแปรอิสระทั้ง 4
ตัวสามารถทำนายความน่ารักได้



$$H_0: P^2 = 0$$



$$H_1: P^2 > 0$$



การทดสอบสมมติฐาน

ท่านต้องการทดสอบว่าอะไร
ทำให้คนอื่นมองว่าคนน่ารัก
ท่านจึงทำนายความน่ารักด้วย
(1) หน้าตาดี (2) หุ่นดี (3) นิสัยดี
(4) ความใกล้ชิดสนิทสนม



ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน
ของตัวแปรแต่ละตัว

Descriptive Statistics

	Mean	Std. Deviation	N
cute	52.8548	18.49537	50
face	51.64	15.265	50
shape	50.96	16.747	50
habit	48.14	15.268	50
close	50.40	16.397	50

การทดสอบสมมติฐาน

ท่านต้องการทดสอบ

ทำให้คนอื่นมองว่า

ท่านจึงทำนายว่า

(1) หน้าตาดี (2) หุ่น

(4) ความใกล้ชิด

Correlations

		cute	face	shape	habit	close
Pearson Correlation	cute	1.000	.655	.306	.116	.230
	face	.655	1.000	.447	-.019	.333
	shape	.306	.447	1.000	-.191	.130
	habit	.116	-.019	-.191	1.000	.537
	close	.230	.333	.130	.537	1.000
Sig. (1-tailed)	cute	.	.000	.015	.212	.054
	face	.000	.	.001	.449	.009
	shape	.015	.001	.	.091	.185
	habit	.212	.449	.091	.	.000
	close	.054	.009	.185	.000	.
N		50	50	50	50	50

ดูความสัมพันธ์ระหว่างตัวแปรอิสระ หน้าตาสัมพันธ์ทางบวกกับรูปร่างและ
ความใกล้ชิด นิสัยสัมพันธ์ทางบวกกับความใกล้ชิด
มีเพียงหน้าตาและรูปร่าง มีความสัมพันธ์กับความน่ารักที่เป็นตัวแปรตาม

การทดสอบสมมติฐาน

ท่านต้องการทดสอบว่าอะไร
ทำให้คนอื่นมองว่าคนน่ารัก
ท่านจึงทำนายความน่ารักด้วย
(1) หน้าตาดี (2) นุ่นดี (3)
(4) ความใกล้ชิดสนิท



Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.673 ^a	.453	.405	14.26869

a. Predictors: (Constant), close, shape, face, habit

$R = .673$

$R^2 = .453$

Adjusted $R^2 = .405$

$s_e = 14.27$

การทดสอบสมมติฐาน

ท่านต้องการทดสอบว่าอะไร

ทำให้คนอื่นมองว่าน่ารัก

ท่านจึงทำนายความ

(1) หน้าตาดี (2) น่ารัก

(4) ความใกล้ชิด

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7600.061	4	1900.015	9.332	.000 ^b
	Residual	9161.797	45	203.595		
	Total	16761.859	49			

a. Dependent Variable: cute

b. Predictors: (Constant), close, shape, face, habit

$$R^2 = .453, F(4, 45) = 9.33, p < .001$$

$$p < \alpha$$

ตัวแปรอิสระสามารถทำนายตัวแปรตามได้อย่างมีนัยสำคัญ

ทดสอบต่อว่าตัวแปรอิสระตัวใด ที่สามารถทำนายตัวแปรตามได้

การทดสอบสมมติฐาน

$$\hat{Y} = 2.368 + 0.808X_1 + 0.065X_2 + 0.237X_3 - 0.118X_4$$

ทำให้คนอื่นมองว่าคน

ท่านจึงทำนายความน่า

(1) หน้าตาดี (2) หุ่นดี (3)

(4) ความใกล้ชิดสนิท

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	Sig.
		B	Std. Error	Beta	
1	(Constant)	2.368	10.908		.829
	face	.808	.158	.667	.000
	shape	.065	.140	.059	.646
	habit	.237	.168	.195	.166
	close	-.118	.162	-.105	.472

a. Dependent Variable: cute

ไม่ถึงระดับนัยสำคัญ

$a = 2.368, t(45) = 0.217, p = .83$

จุดตัดแกน Y ไม่มีความหมาย เพราะคะแนน
ตัวแปรอิสระเท่ากับ 0 ไม่มีความหมาย

การทดสอบสมมติฐาน

$$b_{\text{SHAPE}} = 0.065, t(45) = 0.462, p = .65, \beta = .059$$

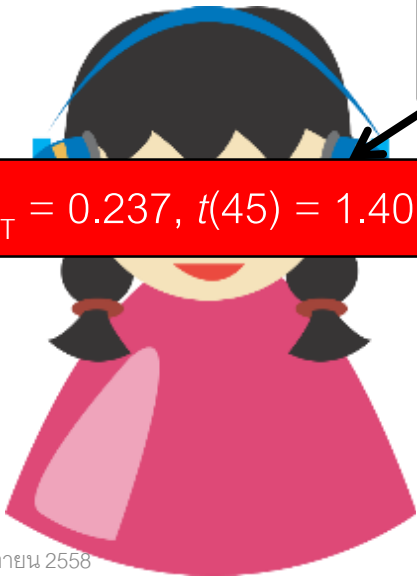
$$b_{\text{FACE}} = 0.808, t(45) = 5.104, p < .001, \beta = .667$$

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.368	10.908		.217	.829
	face	.808	.158	.667	5.104	.000
	shape	.065	.140	.059	.462	.646
	habit	.237	.168	.195	1.407	.166
	close	-.118	.162	-.105	-.726	.472

a. Dependent Variable: cute

$$b_{\text{HABIT}} = 0.237, t(45) = 1.407, p = .17, \beta = .195$$

$$b_{\text{CLOSE}} = -0.118, t(45) = -0.73, p = .47, \beta = -.105$$



หน้าตามีผลต่อความน่ารักอย่างมีนัยสำคัญ
ขณะที่ตัวแปรมีผลไม่ถึงระดับนัยสำคัญ

การทดสอบสมมติฐาน

ท่านต้องการทดสอบว่าอะไร
ทำให้คนอื่นมองว่าคนน่ารัก
ท่านจึงทำนายความน่ารักด้วย
(1) หน้าตาดี (2) หุ่นดี (3) นิสัยดี
(4) ความใกล้ชิดสนิทสนม

สรุป



ตัวแปรอิสระทั้งหมดสามารถ
อธิบายการกระจายของความ
น่ารักได้ 45%

หน้าตาเป็นตัวบ่งชี้ความน่ารัก
เพียงอย่างเดียวที่ถึงระดับ
นัยสำคัญ

การเขียนรายงาน

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

การเขียนรายงาน

- มี 2 แนวทางในการวิเคราะห์ถดถอย
 - เน้นการทำนายค่าของตัวแปรตาม การเขียนรายงานจะเน้นที่การสร้างสมการทำนายและความแม่นยำของค่าที่ทำนายได้
 - เน้นการอธิบายอิทธิพลของตัวแปรอิสระ ค้นหาว่าตัวแปรใดที่กำหนดค่าของตัวแปรตาม การเขียนรายงานจะเน้นแค่ขนาดอิทธิพลและระดับนัยสำคัญของตัวแปรตาม
- งานวิจัยในจิตวิทยาจะมีทั้ง 2 รูปแบบ
 - ในงานวิจัยเชิงวิทยาศาสตร์บริสุทธิ์จะเน้นการอธิบายอิทธิพลของตัวแปรอิสระ (เช่น ตัวกำหนดความเครียด)
 - ในงานวิจัยเชิงประยุกต์ (เช่น การคัดเลือกคนเข้าทำงาน การคัดกรองผู้ป่วย) จะเน้นการทำนายค่าตัวแปรตาม

การเขียนรายงาน (เน้นทำนาย)

- งานวิจัยต้องการทดสอบว่า ความเป็นผู้นำแบบนักปฏิรูป ($M = 3.54$, $SD = 0.53$) สามารถทำนายผลการปฏิบัติงาน ($M = 75.32$, $SD = 7.76$) ได้หรือไม่ จากการเก็บข้อมูลจากพนักงาน 500 คน พบว่าความเป็นผู้นำแบบนักปฏิรูป สามารถทำนายผลการปฏิบัติงานได้อย่างมีนัยสำคัญ ($R^2 = .22$, $p < .05$) โดยสมการทำนาย คือ $49.17 + 7.25X$ โดยส่วนเบี่ยงเบนมาตรฐานของความผิดพลาดในการทำนายเท่ากับ 6.915

การเขียนรายงาน (เน้นทำนาย)

- ผู้วิจัยต้องการทำนายความเครียดจากความกดดันจากทางบ้านและความกดดันจากที่ทำงาน ตารางที่ 1 แสดงค่าสถิติพรรณนาและค่าสหสัมพันธ์ จากการวิเคราะห์ถดถอยแบบพหุ พบว่า ตัวแปรทั้งสองสามารถทำนายความเครียดได้อย่างมีนัยสำคัญ ($R^2 = .36, p < .05$) โดยสมการทำนาย คือ $16.011 + 0.459(\text{บ้าน}) + 0.255(\text{ที่ทำงาน})$ โดยส่วนเบี่ยงเบนมาตรฐานของความผิดพลาดในการทำนายเท่ากับ 9.14 ซึ่งความชันของตัวแปรอิสระทั้งสองถึงระดับนัยสำคัญ ($p < .05$)

การเขียนรายงาน (เน้นทำนาย)

- ตารางที่ 1 แสดงสถิติพรรณนาและค่าสหสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม ($N = 100$)

ตัวแปร	1	2	3
1. ความเครียด			
2. ความกดดันจากทางบ้าน	.568**		
3. ความกดดันจากที่ทำงาน	.481**	.554**	
M	52.10	50.25	51.03
SD	11.33	10.76	10.63

หมายเหตุ * $p < .05$, ** $p < .01$

การเขียนรายงาน (เน้นอธิบาย)

- ผู้วิจัยต้องการทดสอบว่าอายุ ($M = 6.6$, $SD = 1.02$) มีอิทธิพลต่อจำนวนการเลือกของเล่นตุ๊กเพศ ($M = 14$, $SD = 3.65$) หรือไม่ จากการวิเคราะห์ถดถอยพบว่าอายุสามารถทำนายได้อย่างมีนัยสำคัญ, $b = 3.30$, $p < .001$, $r^2 = .85$.

การเขียนรายงาน (เน้นอธิบาย)

- งานวิจัยนี้ต้องการทดสอบว่าลักษณะใดของเพศหญิงที่มีผลทำให้ผู้อื่นมองว่าน่ารัก สถิติพรรณนาและค่าสหสัมพันธ์แสดงในตารางที่ 1 จากการสำรวจนิสิตระดับปริญญาตรีจำนวน 50 คน และวิเคราะห์ด้วยการวิเคราะห์ถดถอยแบบพหุ พบผลดังตารางที่ 2 ตัวแปรอิสระทั้งหมดสามารถทำนายความน่ารักได้อย่างมีนัยสำคัญ ($R^2 = .45, p < .001$) โดยหน้าตาเป็นตัวแปรเดียวที่สามารถทำนายความน่ารักได้อย่างมีนัยสำคัญ ($p < .05$)

การเขียนรายงาน (เน้นอธิบาย)

- ตารางที่ 1 แสดงสถิติพรรณนาและค่าสหสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม ($N = 50$)

ตัวแปร	1	2	3	4	5
1. ความน่ารัก					
2. หน้าตา	.655**				
3. รูปร่าง	.306*	.447**			
4. นิสัย	.116	-.019	-.191		
5. ความใกล้ชิด	.230	.333**	.130	.537**	
<i>M</i>	52.85	51.64	50.96	48.14	50.40
<i>SD</i>	18.50	15.27	16.75	15.27	16.40

หมายเหตุ * $p < .05$, ** $p < .01$

การเขียนรายงาน (เน้นอธิบาย)

- ตารางที่ 2 แสดงผลการวิเคราะห์ถดถอย ในการทำนายความน่ารัก ($N = 50$)

ตัวแปร	b	SE	β
ความน่ารัก	0.808**	0.158	.667
หน้าตา	0.065	0.140	.059
รูปร่าง	0.237	0.168	.195
นิสัย	-0.118	0.162	-.105
$R^2 = .45$, Adjusted $R^2 = .41$, $p < .001$			

หมายเหตุ * $p < .05$, ** $p < .01$

การหาช่วงเชื่อมั่น

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

การหาช่วงเชื่อมั่นของสัมประสิทธิ์ถดถอย

- สามารถคำนวณได้จาก SPSS โดยตรง โดยเลือกที่ **Statistics...** แล้วทำเครื่องหมายถูกที่ **Confidence intervals**
- ช่วงเชื่อมั่นนี้มีประโยชน์มากเมื่อใช้การวิเคราะห์ถดถอยในการทำนายค่าของตัวแปรตาม เพื่อดูว่าหากตัวแปรอิสระเพิ่มขึ้น 1 หน่วยตัวแปรตามมีแนวโน้มที่จะเพิ่มขึ้นเท่าไรบ้างในประชากร

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
1 (Constant)	51.418	3.336		15.415	.000	44.840	57.995
tflader	6.752	.932	.458	7.243	.000	4.914	8.591

a. Dependent Variable: producti

การหาช่วงเชื่อมั่นของสัมประสิทธิ์การถดถอย

- ช่วงเชื่อมั่นนี้มีประโยชน์ทั้งในการใช้การวิเคราะห์ถดถอยเพื่ออธิบายอิทธิพลของตัวแปรอิสระต่อตัวแปรตาม และเพื่อการทำนายค่าของตัวแปรตาม
- SPSS ไม่มีวิธีที่ใช้ในการหาได้ และสูตรที่สามารถคำนวณด้วยตนเองได้ของ Wishart (1931) มีข้อผิดพลาดสูงมาก เช่น อาจมีค่าช่วงเชื่อมั่นที่น้อยกว่า 0 หรือมากกว่า 0

$$CI_{1-\alpha} R^2 \pm t_{\alpha/2, N-p-1} SE_{R^2}$$

$$SE_{R^2} = \sqrt{\frac{4R^2 (1-R^2)^2 (N-p-1)^2}{(N+3)}}$$

การหาช่วงเชื่อมั่นของสัมประสิทธิ์การถดถอย

- วิธีการที่ดีที่สุดในปัจจุบัน คือ การใช้วิธีการหาช่วงเชื่อมั่นแบบ Bootstrap ซึ่งสามารถใช้โปรแกรม R คำนวณได้
- ยกตัวอย่างเช่น
 - การทำนายจำนวนบัตรเครดิตด้วยขนาดครอบครัวและรายได้, $R^2 = .89$, ช่วงเชื่อมั่นระดับ .95 = (.53, .97) → สังเกตว่าช่วงเชื่อมั่นไม่สมมาตร
 - การทำนายความเครียดจากความกดดันจากทางบ้านและครอบครัว, $R^2 = .36$, ช่วงเชื่อมั่นระดับ .95 = (.20, .52)

การหาช่วงเชื่อมั่นของค่าตัวแปรตามที่ทำนายได้

- ค่าตัวแปรตามที่ทำนายได้ (Predicted value) เป็นการทำนายค่าเป็นจุด ซึ่งมักจะไม่ใช่ค่าตัวแปรตามจริง
- อย่างไรก็ตาม การทำนายค่าเป็นช่วงจะบอกได้ว่าค่าตัวแปรตามที่เป็นไปจะอยู่ในช่วงใด

$$CI_{1-\alpha} = \hat{Y} \pm z_{\alpha/2} SE_{\hat{Y}}$$

จริงๆ ค่าวิกฤตต้องใช้จากการกระจายแบบที่ แต่เพื่อความง่ายใช้จากโค้งปกติ

$$SE_{\hat{Y}} = s_e \sqrt{1 + \frac{1}{N} + \frac{(X - M_X)^2}{(N - 1)s_X^2}}$$

การหาช่วงเชื่อมั่นของค่าตัวแปรตามที่ทำนายได้

- สังเกตว่า ยิ่งค่า X แตกต่างจากค่าเฉลี่ยของ X มากเท่าใด $SE_{\hat{y}}$ จะสูงขึ้นมากเท่านั้น
- เนื่องจากว่า ความชัน (Slope; b) ก็เป็นค่าที่ถูกประมาณค่าขึ้นมา ด้วยเหตุนี้ ถ้าข้อมูลที่ยิ่งห่างจากค่าเฉลี่ย ยิ่งต้องถูกคูณด้วยค่าความชันมาก ทำให้ความคาดเคลื่อนยิ่งสูง ทำให้ช่วงเชื่อมั่น ต้องเผื่อให้กว้างมากขึ้น
- สูตรที่ได้มีเฉพาะสำหรับตัวแปรอิสระตัวเดียว หากมีตัวแปรอิสระมากกว่า 1 ตัว ต้องใช้โปรแกรม R ในการคำนวณ

การหาช่วงเชื่อมั่นของค่าตัวแปรตามที่ทำนายได้

- เช่น จากตัวอย่างจำนวนบัตรเครดิต ลองทำนายจำนวนบัตรเครดิตที่ใช้ เมื่อขนาดครอบคลุมเท่ากับ 5 คน

$$\hat{Y} = 2.87 + 0.97(5) = 7.72$$

$$s_e = 0.956$$

$$SE_{\hat{Y}} = (0.956) \sqrt{1 + \frac{1}{8} + \frac{(5 - 4.25)^2}{(8 - 1)(2.5)}} = 1.028$$

$$CI_{1-\alpha} = 7.72 \pm (1.96)(1.028) = (5.70, 9.74)$$

ช่วงเชื่อมั่นระดับ .95 ของจำนวนบัตรเครดิตที่ใช้เมื่อ
ครอบคลุมมีขนาด 5 คน เท่ากับ 5.70 ถึง 9.74 ใบ

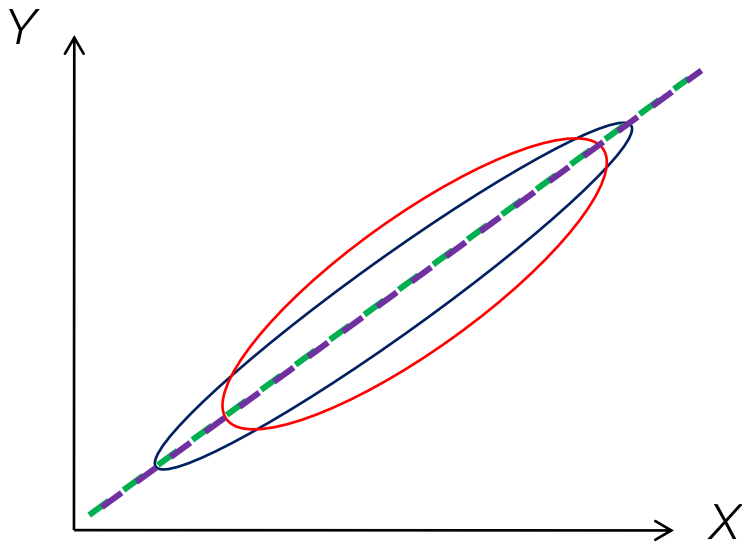
ความถดถอยและค่าสหสัมพันธ์

สถิติขั้นนำสำหรับจิตวิทยา

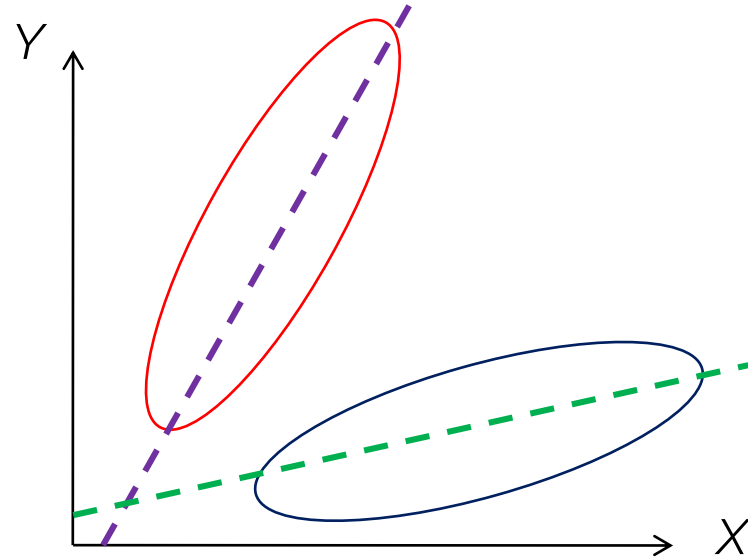
อ. สันทัด พรประเสริฐมานิต

ความถดถอยและค่าสหสัมพันธ์

- ค่าสหสัมพันธ์ คือ การเกาะกลุ่ม
- ค่าสัมประสิทธิ์ถดถอย คือ อิทธิพลของตัวแปร X ที่มีต่อ Y



$$r_1 \neq r_2; b_1 = b_2; a_1 = a_2$$



$$r_1 = r_2; b_1 \neq b_2; a_1 \neq a_2$$

ความถดถอยและค่าสหสัมพันธ์

- ค่าสหสัมพันธ์ เป็นสถิติที่สมมาตร
 - สหสัมพันธ์ระหว่าง X และ Y และสหสัมพันธ์ระหว่าง Y และ X เหมือนกัน
 - $r_{XY} = r_{YX}$
- ค่าสัมประสิทธิ์ถดถอย เป็นสถิติที่ไม่สมมาตร
 - ค่าความชันและจุดตัดแกน Y เมื่อทำนาย Y จาก X และทำนาย X จาก Y จะไม่เหมือนกัน
 - $b_{XY} \neq b_{YX}; a_{XY} \neq a_{YX}$
 - เนื่องจาก ทำนาย Y จาก X จะต้องแปลง X ไปใกล้ Y แต่เมื่อทำนาย X จาก Y จะต้องแปลง Y ไปใกล้ X

ความถดถอยและค่าสหสัมพันธ์

- มีบางคนเข้าใจว่า การสหสัมพันธ์ไม่สามารถบอกความสัมพันธ์เชิงเหตุและผลได้ แต่การวิเคราะห์ถดถอยสามารถบอกความสัมพันธ์เชิงเหตุและผลได้
- ความเชื่อนี้ไม่ถูกต้องทั้งหมด เพราะความสัมพันธ์ก็สามารถใช้บอกความสัมพันธ์เชิงเหตุผลได้ (ดังที่กล่าวไปในเรื่องความสัมพันธ์)
 - ตัวแปร X มีความสัมพันธ์กับตัวแปร Y
 - X ต้องมาก่อน Y
 - การเกิด Y จะต้องไม่ได้รับการอธิบายด้วยตัวแปรอื่น
 - มีทฤษฎีหรือกระบวนการอธิบายว่าทำไม X ถึงก่อให้เกิด Y

ความถดถอยและค่าสหสัมพันธ์

- แต่การวิเคราะห์ถดถอย เหมาะสมในการประกอบการอธิบายเชิงสาเหตุ มากกว่า
 - การวิเคราะห์ถดถอยเป็นการมองให้ละเอียดมากขึ้น ว่าหากเหตุเพิ่มขึ้น 1 หน่วยแล้ว ผลเพิ่มขึ้นกี่หน่วย
 - นอกจากนี้ การวิเคราะห์ถดถอย สามารถบอกอิทธิพลของตัวแปรเหตุ เมื่อควบคุมตัวแปรควบคุมให้คงที่

คาบต่อไป

- ส่งการบ้านที่ 12
- เรียนคาบปฏิบัติการครั้งที่ 10
- ครั้งต่อไปยังคงสอนเรื่องการวิเคราะห์ถดถอย ครอบคลุมถึง
 - การหาค่าล้าง
 - ข้อตกลงเบื้องต้นก่อนการใช้สถิติ
 - ลำดับการใส่ตัวแปรอิสระ
 - การเปลี่ยนศูนย์กลางของตัวแปรอิสระ