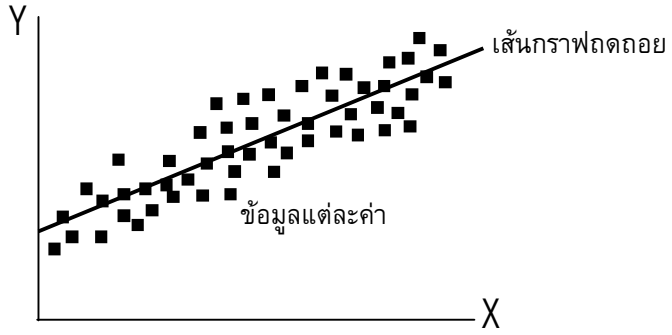
14 การวิเคราะห์การถดถอย Regression Analysis

■ บทนำ

การศึกษาความสัมพันธ์อย่างง่ายที่ผ่านมา ซึ่งเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัวว่ามีความสัมพันธ์กันหรือไม่และมีความสัมพันธ์กันในระดับใด ผลการวิเคราะห์ความสัมพันธ์อย่างง่าย จะทำให้ทราบถึงความสัมพันธ์ดังกล่าว เช่น สมมติว่าผลการเรียนวิชาโครงสร้างข้อมูล (X) กับผลการเรียนวิชาการโปรแกรม (Y) ของผู้เรียน มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ .80 แสดงว่าผลการเรียนวิชาทั้งสองมีความสัมพันธ์กันในเชิงบวกค่อนข้างสูง แสดงว่าผลการเรียนของวิชาทั้งสองมีความสัมพันธ์คล้อยตามกันไปในทิศทางเดียวกัน ซึ่งอาจจะสรุปผลได้ว่า ผู้เรียนคนใดคนหนึ่งที่ได้คะแนนวิชาโครงสร้างข้อมูลสูง ก็น่าจะได้คะแนนวิชาการโปรแกรมสูงตามไปด้วย (ซึ่งไม่ได้หมายความว่าผลการเรียนวิชาโครงสร้างข้อมูลที่มีคะแนนสูง มีผลทำให้ผลการเรียนวิชาการโปรแกรมของผู้เรียนมีค่าสูงขึ้นตามไปด้วย) แต่ถ้าหากผลการวิเคราะห์แล้วพบว่า ตัวแปรทั้งสองมีความสัมพันธ์กันเท่ากับ .30 เนื่องจากมีค่าไม่สูงมากนัก ผู้วิจัยอาจจะเกิดความไม่มั่นใจในการพยากรณ์ของผลการเรียนอีกวิชาหนึ่ง ดังนั้น จึงอาจจะต้องทำการวิเคราะห์โดยอาศัยค่าที่ทราบจากตัวแปรหนึ่งแล้วนำไปพยากรณ์ค่าของอีกตัวแปรหนึ่ง ว่ามีความแปรผันในสัดส่วนเท่าใดหรือในระดับใด การวิเคราะห์ในลักษณะของตัวเกณฑ์เพื่อใช้ในการพยากรณ์นี้ เรียกว่า การวิเคราะห์การถดถอย (Regression Analysis) ซึ่งถ้าตัวแปรทั้งสองมีความสัมพันธ์กันอย่างสมบูรณ์ โดยมีค่า r_{XY} ใกล้เคียงกับ 1.00 หรือ -1.00 ย่อมหมายถึงว่า การพยากรณ์จะยิ่งมีความถูกต้องมากขึ้น แต่ถ้าค่า r_{XY} มีค่าต่ำ ๆ การพยากรณ์ในลักษณะของการวิเคราะห์การถดถอยก็อาจจะเกิดความผิดพลาดขึ้นได้

■ การวิเคราะห์การถดถอยอย่างง่าย (Simple Regression Analysis)

การวิเคราะห์การถดถอยอย่างง่าย (Simple Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปร 2 ตัว (ในที่นี้คือตัวแปร X และ Y) ที่มีความสัมพันธ์กันในลักษณะเชิงเส้น (Linear) โดยมีสมการถดถอยก็คือ $Y = \alpha + \beta X$ ในที่นี้ Y ก็คือค่าเฉลี่ยของ Y (ไม่ใช่ค่า Y แต่ละค่า) เนื่องจากในการวิเคราะห์การถดถอยอย่างง่ายนั้น ตัวแปร X จะถูกกำหนดค่าไว้ก่อน และค่า Y จะเปลี่ยนแปลงไปตามตัวแปร X เนื่องจากค่า X ค่าหนึ่งจะมีค่า Y ที่เป็นคู่ของค่า X หลาย ๆ ค่า และเมื่อนำค่า X และ Y ทั้งหมดไปพล็อตบนแกน X, Y แล้วลากเส้นเชื่อมระหว่างจุดที่ปรากฏ เส้นกราฟที่ได้จะเป็นเส้นตรงแสดงความสัมพันธ์ระหว่างค่าเฉลี่ยของตัวแปร X กับตัวแปร Y ซึ่งก็คือ เส้นกราฟถดถอย (Regression Line) นั่นเอง



ภาพที่ 14-1 การกระจายของข้อมูลและเส้นกราฟถดถอย

จากสมการเส้นตรง $Y = \alpha + \beta X$ ซึ่ง α และ β เป็นพารามิเตอร์ที่ไม่ทราบค่า จึงจะต้องประมาณค่าโดยใช้ข้อมูลจากตัวอย่าง โดยที่วิธีที่นิยมใช้ในการประมาณค่าของ α และ β ก็คือวิธีกำลังสองน้อยที่สุด (The Least Squares Method) ซึ่งจะแทนค่าของ α และ β ด้วยค่า a และ b โดยที่ a ก็คือค่าคงที่ (Constant) เป็นค่าที่เส้นกราฟถดถอยตัดกับแกน Y ส่วน b เป็นความชัน (Slope) ของเส้นกราฟ ซึ่งแสดงถึงอัตราการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนแปลง เรียกส่วนนี้ว่า สัมประสิทธิ์การถดถอย (Regression Coefficient) หรือสัมประสิทธิ์การพยากรณ์

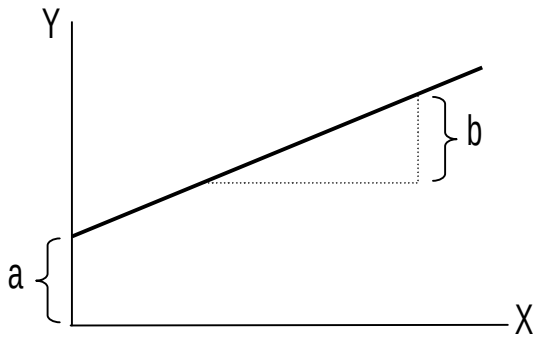
สมการถดถอยอย่างง่าย เขียนได้ดังนี้

$$Y = a + bX$$

เมื่อ

Y = ตัวแปรตาม (เนื่องจากค่าของ Y ขึ้นอยู่กับค่าของ X)
X = ตัวแปรอิสระหรือตัวแปรต้น
a = ค่าคงที่ (Constant) เป็นค่าที่ตัดกันแกน Y
b = ความชัน (Slope) ของเส้นกราฟ

สามารถเขียนเป็นกราฟสมการเส้นตรงได้ดังนี้



ภาพที่ 14-2 สมการเส้นตรงของการถดถอย

สัมประสิทธิ์การถดถอย (Regression Coefficient) หรือสัมประสิทธิ์การพยากรณ์ เป็นค่าของ b ที่เป็นความชันของกราฟเส้นตรง ที่เกิดจากสมการเชิงเส้น ถ้าทราบค่าของ b และค่าของ a แล้ว ก็จะสามารถพยากรณ์ค่าของตัวแปร Y ได้ ซึ่งสรุปได้ดังนี้

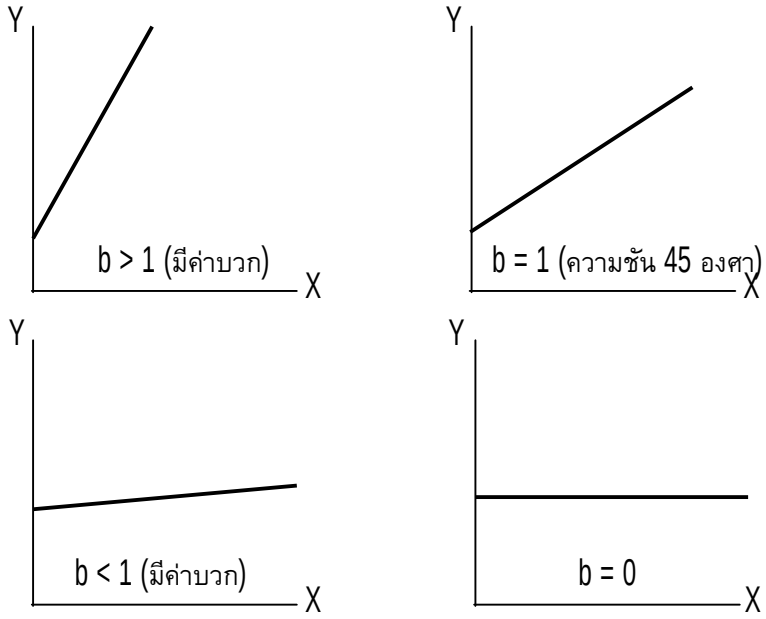
1. ถ้า b > 0 แสดงว่า X และ Y มีความสัมพันธ์กันในทิศทางเดียวกัน กล่าวคือ ถ้า X มีค่าสูงขึ้น ค่าของ Y ก็จะมีค่าสูงขึ้นตามไปด้วย

2. ถ้า b < 0 แสดงว่า X และ Y มีความสัมพันธ์กันในทิศทางตรงกันข้าม กล่าวคือ ถ้า X มีค่าสูงขึ้น ค่าของ Y จะต่ำลง

3. ถ้า b มีค่าใกล้ 0 แสดงว่า X และ Y มีความสัมพันธ์กันน้อย

4. ถ้า b = 0 แสดงว่า X และ Y ไม่มีความสัมพันธ์กันเลย เส้นกราฟที่ได้จะเป็นเส้นตรงค่าของ Y จะมีค่าเท่ากับค่าคงที่ (a)

5. ถ้า b = 1 แสดงว่าความชันของเส้นกราฟมีค่าเท่ากับ 45 องศา ค่า X และ Y จะมีค่าเท่ากัน ในกรณีที่ค่าคงที่ a เท่ากับศูนย์



ภาพที่ 14-3 สมการเส้นตรงของการถดถอยเมื่อ b มีค่าแตกต่างกัน

ลักษณะของเส้นกราฟถดถอยอย่างง่าย มีดังนี้

1. ค่า a เป็นค่าคงที่ จะมีค่าเป็นบวก เมื่อเส้นกราฟตัดกับแกน Y เหนือเส้นแกน X ขึ้นไป หากเส้นกราฟตัดที่จุดกำเนิดหรือจุดกำเนิด (0,0) ค่า a จะมีค่าเป็นศูนย์ ณ จุดนี้ค่า Y จะขึ้นอยู่กับผลของค่า X กับสัมประสิทธิ์การถดถอยเท่านั้น แต่ถ้าเส้นกราฟตัดกับแกน Y ต่ำกว่าเส้นแกน X ค่า a จะมีค่าเป็นลบ

2. ค่า b ที่เป็นค่าสัมประสิทธิ์การถดถอยเป็นความชันของเส้นกราฟ เป็นค่าที่แสดงถึงการเปลี่ยนแปลงของเส้นกราฟ เมื่อตัวแปรอิสระหรือตัวแปรต้น (X) เปลี่ยนแปลงไปหนึ่งหน่วย จะ

ทำให้ตัวแปร Y เปลี่ยนแปลงไป b หน่วย ถ้าเส้นกราฟมีความชันมาก การเปลี่ยนแปลงของตัวแปร X จะทำให้ค่าของ Y เปลี่ยนแปลงไปเป็นจำนวนมาก แต่ถ้าความชันมีค่าเท่ากับ 1 การเปลี่ยนแปลงของตัวแปร X จะส่งผลให้ค่าของ Y เปลี่ยนแปลงไปเป็นจำนวนที่เป็นสัดส่วนกับค่า X และถ้าความชันมีค่าต่ำ ๆ (b < 1) จะทำให้ค่าของ Y เปลี่ยนแปลงเป็นจำนวนน้อยกว่าค่าของ X

3. ในกรณีที่ a มีค่าเป็นศูนย์ และ b มีค่าเท่ากับ 1 เส้นกราฟจะผ่านจุดกำเนิด และความชันเป็น 45 องศา ซึ่งทำให้ค่าของ X และ Y มีค่าเท่ากัน

4. ถ้าเป็นกราฟชี้ไปทางควอแดรนท์ที่ 1 (Q1) ค่า b จะมีค่าเป็นบวก แต่ถ้าเส้นกราฟชี้ไปทางควอแดรนท์ที่ 2 (Q2) ค่า b จะมีค่าเป็นลบ

■ สมการถดถอยอย่างง่าย

สมการถดถอยอย่างง่าย สามารถหาค่าของ a และ b ได้จากสูตรดังนี้

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

$$b = \frac{\sum xy}{\sum x^2}$$

เมื่อ

a	=	ค่าคงที่ของสมการถดถอยอย่างง่าย
b	=	สัมประสิทธิ์การถดถอยหรือสัมประสิทธิ์การพยากรณ์
$\overline{X}$	=	มัชฌิมเลขคณิตของตัวแปรอิสระหรือตัวแปรต้น
$\overline{Y}$	=	มัชฌิมเลขคณิตของตัวแปรตาม
x	=	ค่าเบี่ยงเบนจากมัชฌิมเลขคณิตของตัวแปรอิสระหรือตัวแปรต้น มีค่าเท่ากับ $X - \overline{X}$
y	=	ค่าเบี่ยงเบนจากมัชฌิมเลขคณิตของตัวแปรตาม มีค่าเท่ากับ $Y - \overline{Y}$

ข้อตกลงเบื้องต้นในการวิเคราะห์การถดถอยอย่างง่าย มีดังนี้

1. ตัวแปรอิสระหรือตัวแปรต้น (X) และตัวแปรตาม (Y) มีความสัมพันธ์กันในแบบเชิงเส้นตามสมการเส้นตรงก็คือ $Y = \alpha + \beta X$

2. ตัวแปรตามต้องเป็นข้อมูลชนิดต่อเนื่อง ในขณะที่ตัวแปรอิสระหรือตัวแปรต้นจะต้องเป็นตัวแปรที่ผู้วิจัยสามารถกำหนดค่าได้ ซึ่งแสดงให้เห็นว่าการเปลี่ยนแปลงของค่า X จะทำให้ค่าของ Y ที่เกิดจากการสุ่ม เปลี่ยนแปลงไป

3. ความแปรปรวนของ Y สำหรับแต่ละค่าของ X ที่กำหนด จะต้องมีค่าเท่ากัน คุณสมบัติข้อนี้ของการวิเคราะห์การถดถอย เรียกว่า Homoscedasticity

ตัวอย่างที่ 14-1

ผลการทดสอบวิชาโครงสร้างข้อมูลของผู้เรียนกลุ่มหนึ่ง พบว่ามีค่าเฉลี่ยเท่ากับ 18 และค่าเฉลี่ยวิชาการโปรแกรมมีค่าเท่ากับ 25 สมมติให้สัมประสิทธิ์การถดถอยมีค่าเท่ากับ .6 จงหาค่าคงที่ a และเขียนสมการถดถอย เพื่อหาความสัมพันธ์ของผลการสอบในวิชาทั้งสอง

จากสูตร $a = \overline{Y} - b\overline{X}$

แทนค่าในสูตร = 25 -.6 (18)

= 25 – 10.8 = 14.2

ค่าคงที่มีค่าเท่ากับ 14.2 และมีค่าเป็นบวก แสดงว่าเส้นกราฟถดถอยตัดกับแกน Y ที่ระดับ 14.2 เหนือเส้นแกนในควอแดรนท์ที่ 1 สามารถเขียนสมการถดถอยอย่างง่ายได้ว่า

Y = 14.2 + .6X

จากตัวอย่างสรุปได้ว่า หากทราบค่า X ซึ่งในที่นี้คือผลการสอบวิชาโครงสร้างข้อมูล ก็จะสามารถพยากรณ์ค่าของ Y ซึ่งเป็นผลการสอบวิชาการโปรแกรมได้ โดยการแทนค่าในสมการ Y = 14.2 + .6X เพื่อศึกษาว่าถ้าตัวแปรอิสระหรือตัวแปรต้น X มีค่าเท่ากับค่าต่าง ๆ ตัวแปรตาม Y จะมีค่าเท่าใด ดังที่ปรากฏในตารางต่อไปนี้

X	1	2	3	4	5	6	7	8	9	10
Y	14.8	15.4	16	16.6	17.2	17.8	18.4	19	19.6	20.2

การพยากรณ์การเปลี่ยนแปลงของตัวแปรตาม ที่มีความสัมพันธ์กับตัวแปรอิสระหรือตัวแปรต้นในลักษณะเชิงเส้น ค่าที่ได้จากตารางสามารถนำไปเขียนเป็นเส้นกราฟเพื่อแสดงความสัมพันธ์ของตัวแปรทั้งสองได้ เส้นกราฟที่ได้ก็จะเป็นเส้นกราฟถดถอยนั่นเอง

■ การวิเคราะห์การถดถอยจากสัมประสิทธิ์สหสัมพันธ์

ในการพยากรณ์ค่าของตัวแปรตาม Y จากการเปลี่ยนแปลงของตัวแปรอิสระหรือตัวแปรต้น X ถ้าทราบค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรทั้งสอง (r = Correlation Coefficient) จะสามารถวิเคราะห์การถดถอยเพื่อการพยากรณ์ได้ โดยไม่ต้องอาศัยค่าคงที่ a และสัมประสิทธิ์การถดถอย b แต่อย่างใด แต่จำเป็นต้องทราบส่วนเบี่ยงเบนมาตรฐานของตัวแปรทั้งสอง จึงจะสามารถหาค่าคะแนนพยากรณ์ของตัวแปรตามได้ จากสูตรดังต่อไปนี้

$$Y = \overline{Y} + r\frac{S_Y}{S_X}(X - \overline{X})$$

เมื่อ

Y = คะแนนพยากรณ์ของตัวแปรตาม

$\overline{X}$ = มัชฌิมเลขคณิตของตัวแปรอิสระหรือตัวแปรต้น

$\overline{Y}$ = มัชฌิมเลขคณิตของตัวแปรตาม

r = สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y

S_X = ส่วนเบี่ยงเบนมาตรฐานของตัวแปรอิสระหรือตัวแปรต้น

S_Y = ส่วนเบี่ยงเบนมาตรฐานของตัวแปรตาม

ตัวอย่างที่ 14-2

ผลการทดสอบของผู้เรียนห้องหนึ่งในวิชาไมโครโพรเซสเซอร์ ซึ่งมีทั้งภาคทฤษฎีและภาคประลอง ปรากฏว่าได้ข้อมูลดังนี้

วิชา	Mean	SD
ภาคทฤษฎี	58	8
ภาคประลอง	65	12

จงพยากรณ์คะแนนภาคประลองของผู้เรียนคนหนึ่งที่สอบภาคทฤษฎีได้ 62 คะแนน สมมติให้สัมประสิทธิ์สหสัมพันธ์ของวิชาไมโครโพรเซสเซอร์ทั้งภาคทฤษฎีและภาคประลองเท่ากับ .5

จากสูตร $$Y = \overline{Y} + r\frac{S_Y}{S_X}(X - \overline{X})$$

$$Y = 65 + .5\frac{12}{8}(62 - 58) = 65 + .5(\frac{3}{2})(4) = 68$$

คะแนนภาคประลองของผู้เรียนคนนั้น มีค่าเท่ากับ 68 คะแนน

■ ความคลาดเคลื่อนมาตรฐานในการพยากรณ์ (Standard Error of Estimate)

ในการพยากรณ์ค่าของตัวแปรตาม Y ที่เกิดจากการกำหนดค่าของตัวแปรอิสระหรือตัวแปรต้น X จากสมการถดถอยที่ได้ จะมีประโยชน์ต่อการพยากรณ์มากเพียงใดนั้น ขึ้นอยู่กับข้อมูลที่รวบรวมมาว่ามีการกระจัดกระจายไปจากเส้นกราฟหรือสมการถดถอยมากน้อยเพียงใด ถ้าค่าของตัวแปรตาม Y แต่ละค่าแตกต่างไปจากเส้นกราฟถดถอยที่ประมาณขึ้นมามาก การพยากรณ์ค่าของตัวแปรตาม Y โดยใช้เส้นกราฟถดถอยก็จะเกิดความคลาดเคลื่อนมาก แต่ถ้าค่าที่รวบรวมมา

แต่ละค่าอยู่บนเส้นกราฟถดถอยหรืออยู่ใกล้ ๆ เส้นกราฟถดถอยมากเท่าใด การพยากรณ์ค่าของตัวแปรตาม Y โดยใช้เส้นกราฟถดถอยก็จะถูกต้องมากขึ้นเท่านั้น

ความแตกต่างระหว่างค่าของตัวแปรตาม Y ที่เกิดจากการพยากรณ์กับค่าที่ได้จากการรวบรวม เรียกว่า ความคลาดเคลื่อน (Error หรือ Residual) ใช้ตัวย่อว่า e โดยที่ e = Y-Y' ซึ่งค่าความคลาดเคลื่อนนี้จะใช้เป็นตัวพยากรณ์ค่าการกระจายของข้อมูลรอบ ๆ เส้นกราฟถดถอย สำหรับสถิติที่ใช้วัดการกระจายของข้อมูลรอบ ๆ เส้นกราฟถดถอย เรียกว่า ความคลาดเคลื่อนมาตรฐานในการพยากรณ์ (Standard Error of Estimate) ใช้ตัวย่อว่า SEE หรือ $SE_{est.Y}$ ซึ่งเป็นส่วนเบี่ยงเบนมาตรฐานชนิดหนึ่งที่เป็นการเบี่ยงเบนของคะแนนพยากรณ์จากคะแนนที่รวบรวมมาได้

ความคลาดเคลื่อนมาตรฐาน สามารถหาได้จากสมการดังต่อไปนี้

$$SEE = \sqrt{\frac{\sum (Y - Y')^2}{N - k - 1}}$$

เมื่อ

SEE = ความคลาดเคลื่อนมาตรฐานในการพยากรณ์

Y‘ = คะแนนพยากรณ์ของตัวแปรตาม

N = จำนวนสมาชิก

k = จำนวนตัวแปรอิสระหรือตัวแปรต้นในสมการถดถอย

$\sum (Y - Y')^2$ = ผลรวมกำลังสองของส่วนที่เหลือ (Sum Squares of Residuals) ซึ่งใช้ตัวย่อว่า SSR หรือ SS_{res}

ในกรณีของสมการถดถอยเชิงเส้นซึ่งมีจำนวนตัวแปรอิสระหรือตัวแปรต้นเพียงตัวเดียว

$$SEE = \sqrt{\frac{\sum (Y - Y')^2}{N - 2}}$$

ถ้าแทนค่าผลรวมกำลังสองของส่วนที่เหลือ (SSR) ลงในสูตรที่ผ่านมา จะได้สูตรสำหรับหาความคลาดเคลื่อนมาตรฐานในการพยากรณ์ได้ว่า

$$SEE = \sqrt{\frac{SSR}{N - 2}}$$

ในทางปฏิบัติ ข้อมูลที่เกี่ยวข้องเพื่อแทนค่าในสูตรจะมีเป็นจำนวนมาก สูตรดังกล่าวข้างต้นจึงไม่เหมาะกับการคำนวณ จึงได้มีการเปลี่ยนแปลงสูตรสำหรับหาความคลาดเคลื่อนมาตรฐานในการพยากรณ์ให้ง่ายขึ้น ดังนี้

$$SEE = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{N - 2}}$$

นอกจากนี้ ถ้าทราบค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรอิสระหรือตัวแปรต้น X กับตัวแปรตาม Y ก็สามารถหาความคลาดเคลื่อนมาตรฐานในการพยากรณ์ได้จากสูตรดังนี้

$$SEE = S_Y\sqrt{1-r^2}$$

เมื่อ

S_Y = ส่วนเบี่ยงเบนมาตรฐานของตัวแปร Y

r = สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y

ตัวอย่างที่ 14-3

ข้อมูลจากการเข้าค่ายฝึกทักษะการเขียนเกมคอมพิวเตอร์ออนไลน์ ของเยาวชนที่มีน้ำหนักตัว (กิโลกรัม) แตกต่างกัน จำนวน 12 คน เมื่อเปรียบเทียบกับอัตราการใช้พลังงานในแต่ละวัน (กิโลแคลอรี) ซึ่งพบว่ามีความสัมพันธ์ของข้อมูลทั้งสองในลักษณะเชิงเส้น ปรากฏผลดังตาราง จงหาค่าต่าง ๆ ต่อไปนี้

1. สมการแสดงความสัมพันธ์ระหว่างน้ำหนักตัวของเยาวชนกับอัตราการใช้พลังงาน
2. อัตราการใช้พลังงานในแต่ละวันของเยาวชนคนหนึ่งที่มีน้ำหนักตัว 52.5 กิโลกรัม
3. ความคลาดเคลื่อนมาตรฐานในการพยากรณ์

คนที่	น้ำหนัก (X)	พลังงาน (Y)
1	36.1	995
2	33.1	913
3	40.3	1,189
4	42.4	1,124
5	41.2	1,204
6	34.5	1,052
7	42	1,418
8	54.6	1,425
9	48.5	1,396
10	42	1,256
11	51.1	1,347
12	50.6	1,502
รวม	516.4	14,821

จากข้อมูลในตารางที่กำหนดให้ จะเห็นว่าแนวโน้มของความสัมพันธ์ระหว่างตัวแปร X คือ น้ำหนักตัว กับตัวแปร Y คืออัตราการใช้พลังงาน จะมีลักษณะเป็นสมการเส้นตรง ดังนั้น สมการเส้นตรงที่ใช้ในการหาความสัมพันธ์ระหว่างตัวแปรทั้งสอง ก็คือ Y = a + bX

โดยที่ $$b = \frac{\sum xy}{\sum x^2}$$

$$b = \frac{N\sum XY - \sum X.\sum Y}{N\sum X^2 - (\sum X)^2}$$

เมื่อ $\sum X = 516.4, \sum X^2 = 22{,}741.34, \sum XY = 650{,}264.8, \sum Y = 14{,}821$

$$b = \frac{12(650{,}264.8) - (516.4)(14{,}821)}{12(22{,}741) - (516.4)^2} = 24.03$$

และ $$a = \overline{Y} - b\overline{X}$$

$$a = \frac{14{,}821}{12} - (24.03)\frac{516.4}{12} = 200.99$$

ดังนั้น สมการที่แสดงความสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y ซึ่งเป็นสมการแสดงความสัมพันธ์ระหว่างน้ำหนักตัวของเยาวชนกับอัตราการใช้พลังงาน ก็คือ

Y = 200.99 + 24.03X

อัตราการใช้พลังงานในแต่ละวัน (Y) ของเยาวชนคนหนึ่งที่มีน้ำหนักตัว 52.5 กิโลกรัม (X) สามารถหาได้จากการแทนค่าในสมการแสดงความสัมพันธ์ ดังนี้

Y = 200.99 + 24.03(52.5)
= 200.99 + 1,261.575

ดังนั้น Y = 1,462.565 กิโลแคลอรีต่อวัน

หลังจากที่ได้สมการเส้นตรง Y = 200.99 + 24.03X ซึ่งแสดงความสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y แล้ว จะพบว่า b มีค่า 24.03 ซึ่ง b ก็คือค่าสัมประสิทธิ์การถดถอย โดยมีค่าเป็นบวก แสดงว่าเมื่อเยาวชนมีน้ำหนักตัวเพิ่มขึ้น อัตราการใช้พลังงานแต่ละวันก็จะมีค่าเพิ่มขึ้นโดยเฉลี่ย 24.03 กิโลแคลอรี อย่างไรก็ตามเมื่อพิจารณาจากสมการดังกล่าวนี้จะพบว่า เมื่อ X มีค่าเป็นศูนย์ (ในที่นี้ก็คือเยาวชนไม่มีน้ำหนักตัวเลย) แต่ก็ยังต้องใช้พลังงาน 200.99 กิโลแคลอรีต่อวัน ซึ่งเป็นไปไม่ได้ ดังนั้น ค่าของ a บางครั้งก็ไม่สามารถตีความในลักษณะดังกล่าวได้ ทั้งนี้เนื่องจาก

สมการถดถอยจะใช้ได้ก็ต่อเมื่อ ค่าของตัวแปรอิสระหรือตัวแปรต้นอยู่ในช่วงค่าที่กำหนดไว้เท่านั้น ไม่ได้หมายความว่าสมการถดถอยผิดพลาด

จากสมการถดถอย Y = 200.99 + 24.03X สามารถนำไปคำนวณหาค่าความคลาดเคลื่อนของตัวแปร Y ได้ โดยแทนค่าลงในสมการและเขียนลงในตารางเพื่อเปรียบเทียบค่าที่ได้จากการพยากรณ์กับค่าที่แท้จริง (Y) ดังนี้

คนที่	น้ำหนัก (X)	พลังงาน (Y)	Y = 200.99 + 24.03X	ความคลาดเคลื่อน
1	36.1	995	1,068.5	-73.5
2	33.1	913	996.4	-83.4
3	40.3	1,189	1,169.4	19.6
4	42.4	1,124	1,219.9	-95.9
5	41.2	1,204	1,191	13
6	34.5	1,052	1,030	22
7	42	1,418	1,210.2	207.8
8	54.6	1,425	1,513	-88
9	48.5	1,396	1,366.4	29.6
10	42	1,256	1,210.2	45.8
11	51.1	1,347	1,428.9	-81.9
12	50.6	1,502	1,416.9	85.1

เมื่อพิจารณาจากตาราง จากการพยากรณ์โดยใช้สมการถดถอย Y = 200.99 + 24.03X จะพบว่าเกิดความคลาดเคลื่อนแตกต่างกันไป ความคลาดเคลื่อนที่เกิดขึ้นน้อยที่สุดก็คือคนที่ 5 ซึ่งมีน้ำหนักตัว 41.2 กิโลกรัมและพลังงานที่ใช้ในแต่ละวัน 1,204 กิโลแคลอรี ซึ่งค่าที่ได้จากการพยากรณ์พลังงานที่ใช้ในแต่ละวันโดยใช้สมการถดถอยก็คือ 1,191 กิโลแคลอรี แสดงว่าเกิดความคลาดเคลื่อนไปจากข้อมูลจริง 13 กิโลแคลอรี ในขณะที่คนที่ 7 เกิดความคลาดเคลื่อนมากที่สุดถึง 207.8 กิโลแคลอรี จากการใช้สมการถดถอยดังกล่าวพยากรณ์

ค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์ หาได้จากสูตร

$$SEE = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{N-2}}$$

เมื่อ $a = 200.99, b = 24.03, N = 12$

$\sum Y = 14{,}821$

$$\sum Y^2 = 18{,}695{,}125$$

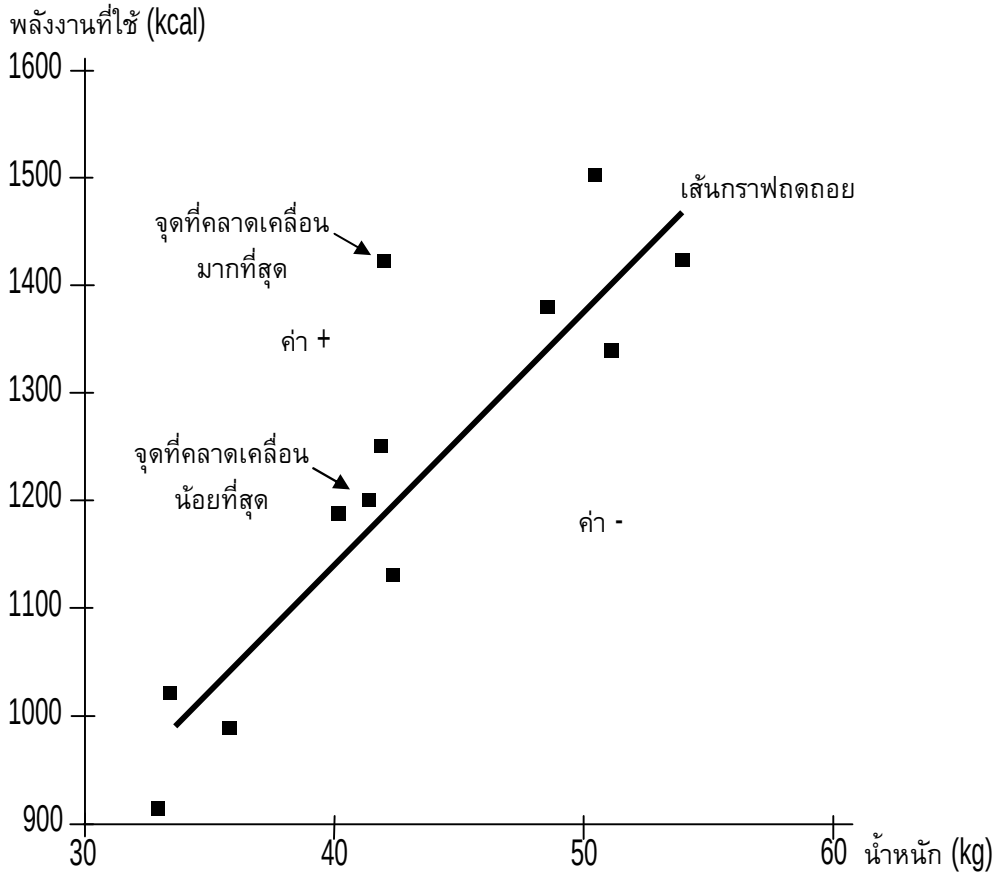
$$\sum XY = 650{,}264.8$$

แทนค่า $$SEE = \sqrt{\frac{18{,}695{,}125 - (200.99)(14{,}821) - 24.03(650{,}264.8)}{12-2}}$$

$$SEE = 95.07$$

แสดงว่าค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์ ที่เกิดจากสมการถดถอยดังกล่าวนี้ ก็คือ 95.07 ซึ่งก็คือส่วนเบี่ยงเบนมาตรฐานของข้อมูลรอบ ๆ เส้นกราฟถดถอย ถ้าส่วนเบี่ยงเบนมาตรฐานรอบ ๆ เส้นกราฟถดถอยมีค่าเป็นศูนย์ ก็จะแสดงว่าไม่มีการถดถอยเกิดขึ้นเลย หมายถึงค่าทุกค่าจะอยู่บนเส้นกราฟถดถอยที่ได้จากการพยากรณ์ แต่ถ้าส่วนเบี่ยงเบนมาตรฐานมีค่ามากก็จะแสดงว่าข้อมูลรอบ ๆ เส้นกราฟถดถอยเกิดการกระจายมาก การพยากรณ์ค่าของตัวแปรตามจากการใช้เส้นกราฟถดถอยก็จะผิดพลาดมากขึ้น

เมื่อนำค่าของ X และ Y ที่ได้จากตารางไปพล๊อตบนกราฟ จะได้ดังภาพ



ตัวอย่างที่ 14-4

บริษัทตัวแทนจำหน่ายคอมพิวเตอร์ยี่ห้อหนึ่ง ต้องการหาความสัมพันธ์ระหว่างค่าโฆษณาลงในสื่อต่าง ๆ กับยอดขายคอมพิวเตอร์ประจำเดือน โดยในรอบปีที่ผ่านมาได้มีข้อสรุปรายงานความสัมพันธ์ระหว่างตัวแปรทั้งสอง ปรากฏตามตารางดังนี้

เดือน	มค.	กพ.	มีค.	เมย.	พค.	มิย.	กค.	สค.	กย.	ตค.	พย.	ธค.
โฆษณา (แสนบาท)	1.0	1.6	1.8	2.0	2.6	2.2	3.0	3.0	4.5	4.0	4.0	4.0
ยอดขาย (ล้านบาท)	28	22	22	26	18	35	30	38	45	30	40	45

จงใช้ข้อมูลความสัมพันธ์ระหว่างตัวแปรทั้งสองในรอบปีที่ผ่านมา เพื่อพยากรณ์ยอดขายที่บริษัทจะได้รับจากการจำหน่ายคอมพิวเตอร์ในเดือนมกราคมปีถัดไป โดยที่บริษัทได้กำหนดงบประมาณค่าโฆษณาไว้เป็นจำนวนเงิน 520,000 บาท พร้อมทั้งหาค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์

จากข้อมูลในตารางที่กำหนดให้ จะพบว่าแนวโน้มของความสัมพันธ์ระหว่างตัวแปรค่าโฆษณากับยอดขายประจำเดือนจะมีลักษณะเป็นสมการเส้นตรง ดังนั้น สมการเส้นตรงที่ใช้ในการหาความสัมพันธ์ระหว่างตัวแปรทั้งสองก็คือ $Y = a + bX$ ซึ่งสามารถคำนวณหาค่าของ a และ b ได้ดังนี้

เดือน	X	Y	X^2	Y^2	XY
มค.	1.0	28	1	784	28
กพ.	1.6	22	2.56	484	35.2
มีค.	1.8	22	3.24	484	39.6
เมย.	2.0	26	4	676	52
พค.	2.6	18	6.76	324	46.8
มิย.	2.0	35	4	1,225	70
กค.	3.0	30	9	900	90
สค.	3.0	38	9	1,444	114
กย.	4.5	45	20.25	2,025	202.5
ตค.	4.0	30	16	900	120
พย.	4.0	40	16	1,600	160
ธค.	4.0	45	16	2,025	180
รวม	33.5	379	107.81	12,871	1,138.1

จากตาราง $\sum X = 33.5, \sum Y = 379$

$\sum X^2 = 107.81, \sum XY = 1{,}138.1$

หาค่า a และ b $b = \dfrac{\sum xy}{\sum x^2}$

$$b = \frac{N\sum XY - \sum X.\sum Y}{N\sum X^2 - (\sum X)^2}$$

$$b = \frac{12(1{,}138.1) - (33.5)(379)}{12(107.81) - (33.5)^2} = 5.60$$

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

$$a = \frac{379}{12} - (5.60)\frac{33.5}{12} = 15.95$$

ดังนั้น สมการที่แสดงความสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y ซึ่งเป็นสมการแสดงความสัมพันธ์ระหว่างค่าโฆษณาลงในสื่อต่าง ๆ กับยอดขายคอมพิวเตอร์ประจำเดือน ก็คือ

$$Y = 15.95 + 5.60X$$

ยอดขายคอมพิวเตอร์ที่บริษัทจะได้รับในเดือนมกราคมปีนี้ จากการพยากรณ์โดยใช้สมการถดถอยที่ได้ ซึ่งบริษัทได้กำหนดงบประมาณค่าโฆษณาไว้เป็นจำนวนเงิน 520,000 บาท (หรือ 5.2 แสนบาท) ก็คือ

$$Y = 15.95 + 5.60(5.2)$$
$$= 15.95 + 29.12$$

ดังนั้น $Y = 45.07$ ล้านบาท

แสดงว่าในเดือนมกราคมปีหน้า ยอดขายคอมพิวเตอร์ที่เกิดจากการพยากรณ์โดยใช้สมการถดถอยก็คือ 45.07 ล้านบาท ในที่นี้ b มีค่าเท่ากับ 5.60 ซึ่งหมายถึง ถ้าบริษัทใช้จ่ายค่าโฆษณาลงสื่อต่าง ๆ เพิ่มขึ้นทุก 100,000 บาท (1 หน่วยของ X) ก็จะทำให้ยอดขายคอมพิวเตอร์เพิ่มขึ้น 5,600,000 บาท หรือถ้าลดค่าโฆษณาลงทุก 100,000 บาท ยอดขายคอมพิวเตอร์ในแต่ละเดือนก็จะลดลง 5,600,000 บาท และถ้าไม่มีการโฆษณาเลย บริษัทก็ยังคงมียอดขายคอมพิวเตอร์อยู่ที่ 15.95 ล้านบาทต่อเดือน

จากสมการถดถอย $Y = 15.95 + 5.60X$ สามารถนำไปคำนวณหาค่าความคลาดเคลื่อนของการพยากรณ์ยอดขาย (ตัวแปร Y) ได้ดังนี้

ค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์ หาได้จากสูตร

$$SEE = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{N-2}}$$

เมื่อ $a = 15.95, b = 5.60, N = 12$

แทนค่า $$SEE = \sqrt{\frac{12{,}871 - (15.95)(379) - (5.60)(1{,}138.1)}{12-2}}$$

$$SEE = 6.727$$

ค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์ยอดขายคอมพิวเตอร์ ที่เกิดจากสมการถดถอย Y = 15.95 + 5.60X ก็คือ 6.727 ซึ่งมีค่าน้อยกว่าค่าความคลาดเคลื่อนมาตรฐานในตัวอย่างที่ 14-3 ที่ผ่านมา แสดงว่าส่วนเบี่ยงเบนมาตรฐานของข้อมูลรอบ ๆ เส้นกราฟถดถอยมีค่าน้อย ทำให้ผลการพยากรณ์มีความถูกต้องมากขึ้นกว่าตัวอย่างที่ผ่านมา

■ การอ้างอิงโดยใช้สัมประสิทธิ์การถดถอย

จากสมการถดถอย Y = a + bX เมื่อพิจารณาค่าของ b หรือค่าสัมประสิทธิ์การถดถอยหรือสัมประสิทธิ์การพยากรณ์ (Regression Coefficient) จะพบว่าความสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y จะขึ้นอยู่กับค่าสัมประสิทธิ์การถดถอยกับค่าคงที่ a ซึ่งสามารถนำค่าสัมประสิทธิ์การถดถอยไปใช้อ้างอิงในการวิจัยเกี่ยวกับการวิเคราะห์ความถดถอยได้ สรุปได้ดังนี้

1. การเปลี่ยนแปลงของตัวแปร Y จะขึ้นอยู่กับอัตราการเปลี่ยนแปลงของตัวแปร X ที่เปลี่ยนแปลงไปแต่ละหน่วย นั่นคือ เมื่อตัวแปร X เปลี่ยนแปลงไปหนึ่งหน่วย ทำให้ตัวแปร Y เปลี่ยนแปลงไปเท่ากับ b หน่วย เช่น จากสมการถดถอยในตัวอย่างที่ผ่านมา Y = 15.95 + 5.60X เมื่อตัวแปร X เปลี่ยนไป 1 หน่วย จะทำให้ค่าของตัวแปร Y เปลี่ยนไปเท่ากับผลคูณของตัวแปร X กับสัมประสิทธิ์การถดถอย (5.60) บวกกับค่าคงที่ 15.95

2. ถ้าสัมประสิทธิ์การถดถอยมีค่าเท่ากับศูนย์ (b = 0) แสดงว่าตัวแปร X กับตัวแปร Y ไม่มีความสัมพันธ์กัน แต่ถ้าสัมประสิทธิ์การถดถอยมีค่ามากกว่าศูนย์ (b > 0) แสดงว่าตัวแปร X กับตัวแปร Y มีความสัมพันธ์กัน สามารถหาความสัมพันธ์ระหว่างตัวแปรทั้งสองโดยการทดสอบสมมติฐานจากค่าของสัมประสิทธิ์การถดถอยได้ โดยกำหนดให้

$H_0 : \beta_0 = 0$ (หมายความว่า ตัวแปร X กับตัวแปร Y ไม่มีความสัมพันธ์กัน)

$H_1 : \beta_0 \neq 0$ (หมายความว่า ตัวแปร X กับตัวแปร Y มีความสัมพันธ์กัน)

หลังจากนั้น จึงทำการทดสอบสมมติฐานโดยนำค่าที่ได้จากตัวอย่างมาเปรียบเทียบกับค่าที่ต้องการพยากรณ์โดยใช้สถิติ t-test หรือ z-test ซึ่งการตัดสินใจว่าจะเชื่อตามสมมติฐานหลัก (H_0)

หรือไม่นั้น ก็ใช้วิธีการทดสอบค่าเฉลี่ยเหมือนกับที่ได้กล่าวมาแล้วในบทที่ 10 ถ้าค่าที่คำนวณได้มีค่าเกินกว่าค่าที่ได้จากการเปิดตาราง แสดงว่าจะต้องปฏิเสธสมมติฐานเป็นกลาง (H_0) และยอมรับสมมติฐานตรงข้าม (H_1) ซึ่งก็คือ ตัวแปร X กับตัวแปร Y มีความสัมพันธ์กันนั่นเอง ในที่นี้ไม่ได้หมายความว่าตัวแปรทั้งสองเป็นเหตุเป็นผลต่อกัน ความสัมพันธ์กันที่เกิดจากค่าสัมประสิทธิ์การถดถอยมีค่ามากกว่า 0 หมายถึง การที่ตัวแปรทั้งสองมีความสัมพันธ์กันในลักษณะใดลักษณะหนึ่งเท่านั้น ไม่เพียงพอที่จะสามารถสรุปถึงความเป็นเหตุและผลระหว่างตัวแปรทั้งสองแต่อย่างใด

■ การทดสอบสมมติฐานสำหรับการวิเคราะห์การถดถอยอย่างง่าย

สำหรับการทดสอบสมมติฐานสำหรับการวิเคราะห์การถดถอยอย่างง่าย จะมีการทดสอบสมมติฐานอยู่ 2 ลักษณะ ได้แก่ การทดสอบสมมติฐานค่าคงที่ เพื่อทดสอบว่าตัวแปรอิสระหรือตัวแปรต้นกับตัวแปรตามมีความสัมพันธ์กันหรือไม่ และ การทดสอบสมมติฐานค่าสัมประสิทธิ์ของตัวแปรอิสระหรือตัวแปรต้น เพื่อทดสอบว่าตัวแปรทั้งสองมีความสัมพันธ์ในทิศทางเดียวกันหรือตรงกันข้ามกัน หรือไม่

การทดสอบสมมติฐานค่าคงที่ มีขั้นตอนดังนี้

1. กำหนดสมมติฐาน
   $H_0 : \beta_0 = 0$ (ค่าคงที่เท่ากับศูนย์ แสดงว่า ไม่มีความสัมพันธ์ระหว่าง X และ Y)
   $H_1 : \beta_0 \neq 0$ (ค่าคงที่ไม่เป็นศูนย์ แสดงว่า มีความสัมพันธ์ระหว่าง X และ Y)
2. คำนวณหาค่า t จากสูตร

$$t = \frac{a}{S_a}$$

เมื่อ

a = ค่าโดยประมาณของ β_0
S_a = ส่วนเบี่ยงเบนมาตรฐานของ a
SEE = ความคลาดเคลื่อนมาตรฐานในการพยากรณ์

โดยที่ $S_a = SEE\sqrt{\frac{\sum X^2}{N\sum (X-\overline{X})^2}}$ หรือ $S_a = SEE\sqrt{\frac{\sum X^2}{N(\sum X^2 - N\overline{X}^2)}}$

3. เปิดตาราง t เพื่อหาค่าวิกฤติที่ df_1 ที่ $\alpha/2$ และ df_2 ที่ n-k-1 (เมื่อ n = จำนวนกลุ่มตัวอย่าง และ k = จำนวนตัวแปรอิสระหรือตัวแปรต้น)

4. เปรียบเทียบค่าที่คำนวณได้กับค่าจากตาราง ถ้าพบว่าค่าที่ได้จากการคำนวณมีค่ามากกว่าค่าจากตาราง ก็จะปฏิเสธสมมติฐานเป็นกลาง (H_0) และยอมรับสมมติฐานตรงข้าม (H_1) สรุปผลตาม H_1 ได้ว่า มีความสัมพันธ์ระหว่าง X และ Y

การทดสอบสมมติฐานค่าสัมประสิทธิ์ของตัวแปรอิสระหรือตัวแปรต้น มีขั้นตอนดังนี้

1. กำหนดสมมติฐาน มีอยู่ 2 กรณี ได้แก่
   1.1 กรณีที่ความชันมากกว่าศูนย์ ($b > 0$)
   $H_0 : \beta_1 \leq 0$ (X และ Y ไม่มีความสัมพันธ์ในทิศทางเดียวกัน)
   $H_1 : \beta_1 > 0$ (X และ Y มีความสัมพันธ์ในทิศทางเดียวกัน)
   1.2 กรณีที่ความชันน้อยกว่าศูนย์ ($b < 0$)
   $H_0 : \beta_1 \geq 0$ (X และ Y ไม่มีความสัมพันธ์ในทิศทางตรงกันข้าม)
   $H_1 : \beta_1 < 0$ (X และ Y มีความสัมพันธ์ในทิศทางตรงกันข้าม)
2. คำนวณหาค่า t จากสูตร

$$t = \frac{b}{S_b}$$

เมื่อ

b = ค่าโดยประมาณของ β_1
S_b = ส่วนเบี่ยงเบนมาตรฐานของ b
SEE = ความคลาดเคลื่อนมาตรฐานในการพยากรณ์

โดยที่ $S_b = \frac{SEE}{\sqrt{\sum (X - \overline{X})^2}}$ หรือ $S_b = \frac{SEE}{\sqrt{(\sum X^2 - N\overline{X}^2)}}$

3. เปิดตาราง t เพื่อหาค่าวิกฤติที่ df_1 ที่ α และ df ที่ n-k-1 (เมื่อ n = จำนวนกลุ่มตัวอย่าง และ k = จำนวนตัวแปรอิสระหรือตัวแปรต้น)
4. เปรียบเทียบค่าที่คำนวณได้กับค่าจากตาราง ซึ่งมีอยู่ 2 กรณีดังนี้
   4.1 กรณีที่ความชันมากกว่าศูนย์ ($b > 0$)

ถ้าค่าที่ได้จากการคำนวณมีค่ามากกว่าค่าจากตาราง ($t > t_{(\alpha,\ n-k-1)}$) ก็จะปฏิเสธสมมติฐานเป็นกลาง (H_0) และยอมรับสมมติฐานตรงข้าม (H_1) สรุปผลตาม H_1 ได้ว่า X และ Y มีความสัมพันธ์ในทิศทางเดียวกัน

   4.2 กรณีที่ความชันน้อยกว่าศูนย์ ($b < 0$)

ถ้าค่าที่ได้จากการคำนวณมีค่าน้อยกว่าค่าจากตาราง ($-t > -t_{(\alpha,\ n-k-1)}$) ก็จะปฏิเสธสมมติฐานเป็นกลาง (H_0) และยอมรับสมมติฐานตรงข้าม (H_1) สรุปผลตาม H_1 ได้ว่า X และ Y มีความสัมพันธ์ในทิศทางตรงกันข้าม

ตัวอย่างที่ 14-5

จากการทดสอบฮาร์ดดิสค์จำนวน 10 ตัว เกี่ยวกับเวลาเฉลี่ยในการค้นหาข้อมูล (Average Seek Time) เป็น ms กับอัตราการขนย้ายข้อมูล (Data Transfer Rate) เป็น MB/s ปรากฏดังนี้

ฮาร์ดดิสค์	1	2	3	4	5	6	7	8	9	10
Av.Seek time.	1.4	1.5	1.5	1.6	2.0	2.0	2.2	2.4	2.5	2.5
Data Transfer.	11.5	11.7	11.7	10.1	13.1	13.2	14.2	14.0	13.7	14.1

จงหาสมการถดถอยที่แสดงความสัมพันธ์ระหว่างเวลาเฉลี่ยในการค้นหาข้อมูล กับอัตราการขนย้ายข้อมูล และทดสอบสมมติฐานสมการถดถอยที่ได้ ที่ระดับนัยสำคัญ .05

ฮาร์ดดิสค์	X	Y	X^2	Y^2	XY
1	1.4	11.5	1.96	132.25	16.1
2	1.5	11.7	2.25	136.89	17.55
3	1.5	11.9	2.25	141.61	17.85
4	1.6	10.1	2.56	102.01	16.16
5	2	13.1	4	171.61	26.2
6	2	13.2	4	174.24	26.4
7	2.2	14.2	4.84	201.64	31.24
8	2.4	14	5.76	196	33.6
9	2.5	13.7	6.25	187.69	34.25
10	2.5	14.1	6.25	198.81	35.25
	19.6	127.3	40.12	1,642.75	254.6

จากตาราง $\overline{X} = \frac{19.6}{10} = 1.96, \overline{Y} = \frac{127.3}{10} = 12.73$

โดยที่ $$b = \frac{N\sum XY - \sum X.\sum Y}{N\sum X^2 - (\sum X)^2}$$

แทนค่า $$b = \frac{10(254.6) - (19.6)(1127.3)}{10(40.12) - (19.6)^2} = \frac{50.92}{17.04} = 2.988$$

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

$$a = 12.73 - (2.988)(1.96) = 6.873$$

ดังนั้น สมการถดถอยที่แสดงความสัมพันธ์ระหว่างเวลาเฉลี่ยในการค้นหาข้อมูล กับอัตราการขนย้ายข้อมูล ก็คือ

$$Y = 6.873 + 2.988X$$

การทดสอบสมมติฐานค่าคงที่

$H_0 : \beta_0 = 0$ (ไม่มีความสัมพันธ์ระหว่างเวลาเฉลี่ยในการค้นหาข้อมูล กับอัตราการขนย้ายข้อมูล)

$H_1 : \beta_0 \neq 0$ (มีความสัมพันธ์ระหว่างเวลาเฉลี่ยในการค้นหาข้อมูล กับอัตราการขนย้ายข้อมูล)

โดยที่ $t = \frac{a}{S_a}$ และ $S_a = SEE\sqrt{\frac{\sum X^2}{N(\sum X^2 - N\overline{X}^2)}}$

เมื่อ $SEE = \sqrt{\frac{\sum Y^2 - a\sum Y - b\sum XY}{N-2}}$

แทนค่า $SEE = \sqrt{\frac{1{,}642.75 - (6.873)(127.3) - (2.988)(254.6)}{10-2}} = \frac{7.0723}{8} = .884$

หาค่า S_a $S_a = .884\sqrt{\frac{40.12}{10(40.12 - 10(1.96)^2)}} = 1.53$

หาค่า t $t = \frac{6.873}{1.53} = 4.492$

จากการเปิดตาราง t ที่ระดับนัยสำคัญ .05 และที่ df = n-2 = 8 พบว่ามีค่าเท่ากับ 1.860 ซึ่งมีค่าน้อยกว่าค่าที่คำนวณได้ (4.492) แสดงว่าค่าที่คำนวณได้อยู่นอกเขตยอมรับ (หรืออยู่ในเขตวิกฤติ) จึงปฏิเสธ H_0 และยอมรับ H_1 สรุปตาม H_1 ได้ว่ามีความสัมพันธ์ระหว่างเวลาเฉลี่ยในการค้นหาข้อมูล กับอัตราการขนย้ายข้อมูล

การทดสอบสมมติฐานค่าสัมประสิทธิ์ของตัวแปรอิสระหรือตัวแปรต้น

$H_0 : \beta_1 \leq 0$ (เวลาเฉลี่ยในการค้นหาข้อมูล กับอัตราการขนย้ายข้อมูล ไม่มีความสัมพันธ์ในทิศทางเดียวกัน)

$H_1 : \beta_1 > 0$ (เวลาเฉลี่ยในการค้นหาข้อมูล กับอัตราการขนย้ายข้อมูล มีความสัมพันธ์ในทิศทางเดียวกัน)

โดยที่ $t = \frac{b}{S_b}$ และ $S_b = \frac{SEE}{\sqrt{(\sum X^2 - N\overline{X}^2)}}$

หาค่า S_b $$S_b = \frac{.884}{\sqrt{40.12 - 10(1.96)^2}} = .677$$

หาค่า t $$t = \frac{2.988}{.677} = 4.41$$

เมื่อเปรียบเทียบกับค่าที่ได้จากการเปิดตาราง (2.306) พบว่า มีค่าน้อยกว่าค่าที่คำนวณได้ (4.41) แสดงว่าค่าที่คำนวณได้อยู่นอกเขตยอมรับ (หรืออยู่ในเขตวิกฤติ) จึงปฏิเสธ H_0 และยอมรับ H_1 เช่นเดียวกันกับกรณีแรก สรุปได้ว่า เวลาเฉลี่ยในการค้นหาข้อมูลกับอัตราการขนย้ายข้อมูลมีความสัมพันธ์ในทิศทางเดียวกัน

การทดสอบสมมติฐานทั้ง 2 กรณี สรุปได้ว่า เวลาเฉลี่ยในการค้นหาข้อมูลมีความสัมพันธ์กับอัตราการขนย้ายข้อมูลของฮาร์ดดิสค์ และความสัมพันธ์ระหว่างตัวแปรทั้งสองมีทิศทางไปในทางเดียวกัน ที่ระดับนัยสำคัญ .05 โดยมีสมการถดถอยก็คือ Y = 6.873 + 2.988X

■ สัมประสิทธิ์ของการตัดสินใจ (Coefficient of Determination)

ในการวิเคราะห์สมการถดถอย เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหรือตัวแปรต้นกับตัวแปรตาม สมการถดถอยที่ได้จะสามารถใช้พยากรณ์ค่าของตัวแปรตามได้ดีเพียงใดนั้นย่อมขึ้นอยู่กับอิทธิพลของตัวแปรอิสระหรือตัวแปรต้นว่าจะส่งผลอย่างไรกับตัวแปรตาม หรือกล่าวอีกนัยหนึ่งได้ว่า การเปลี่ยนแปลงของตัวแปรตามเป็นผลมาจากอิทธิพลของตัวแปรอิสระหรือตัวแปรต้นมากน้อยเพียงใด สมการถดถอยที่ได้จะสามารถพยากรณ์การเปลี่ยนแปลงของตัวแปรตามได้ดีหรือไม่ ย่อมขึ้นอยู่กับค่าสัมประสิทธิ์ของการตัดสินใจ หรือ Coefficient of Determination ซึ่งเกิดจากค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ยกกำลังสอง ใช้ตัวย่อว่า R^2

ดังนั้นสูตรที่ใช้ ก็คือ

$$R^2 = \frac{(\sum XY - N\overline{X}\overline{Y})^2}{(\sum X^2 - N\overline{X}^2)(\sum Y^2 - N\overline{Y}^2)}$$

หรือ $$R^2 = \frac{(N\sum XY - (\sum X)(\sum Y))^2}{(N\sum X^2 - (\sum X)^2)(N\sum Y^2 - (\sum Y)^2)}$$

เมื่อ

R^2 = สัมประสิทธิ์ของการตัดสินใจ มีค่าระหว่าง 0 ถึง 1

สัมประสิทธิ์ของการตัดสินใจ ยิ่งมีค่าใกล้ 1.00 มากเท่าใด แสดงว่าสามารถอธิบายค่าของตัวแปรตามได้ดี เนื่องจากตัวแปรอิสระหรือตัวแปรต้นกับตัวแปรตามมีความสัมพันธ์กันมาก แต่ถ้ามีค่าใกล้ 0 แสดงว่า สมการถดถอยสามารถอธิบายค่าของตัวแปรตามได้ไม่ดี หรือกล่าวโดยสรุปได้ว่า ตัวแปรทั้งสองมีความสัมพันธ์กันน้อย

ตัวอย่างที่ 14-6

จากตัวอย่างที่ 14-5 ที่ผ่านมา จงหาสัมประสิทธิ์ของการตัดสินใจของมาการถดถอยที่ได้

จากสูตร $$R^2 = \frac{(\sum XY - N\overline{X}\overline{Y})^2}{(\sum X^2 - N\overline{X}^2)(\sum Y^2 - N\overline{Y}^2)}$$

เมื่อ $\overline{X} = 1.96, \overline{Y} = 12.73, \sum X^2 = 40.12, \sum Y = 127.3, \sum Y^2 = 1{,}642.75, \sum XY = 253.6$

แทนค่า $$R^2 = \frac{((254.6 - (10)(1.96)(12.73))^2}{(40.12 - (10)(1.96)^2)(1{,}642.75 - (10)(12.73)^2)}$$

$$R^2 = \frac{25.928}{37.864} = .6847$$

สัมประสิทธิ์ของการตัดสินใจมีค่าเท่ากับ .684 หรือเท่ากับ 68.47% ซึ่งหมายความว่า อัตราการขนย้ายข้อมูล (Y) ขึ้นอยู่เวลาเฉลี่ยในการค้นหาข้อมูล (X) ประมาณ 68.47% ส่วนอีก 31.53% จะขึ้นอยู่กับองค์ประกอบอื่น ๆ ได้แก่ ความเร็วรอบของฮาร์ดดิสค์ ความเร็วของกลไกอ่านข้อมูล และองค์ประกอบอื่น ๆ ที่ไม่ได้นำมาศึกษาในที่นี้ หรือกล่าวอีกนัยหนึ่งได้ว่า สมการถดถอย Y = 6.873 + 2.988X สามารถพยากรณ์ค่าของอัตราการขนย้ายข้อมูลของฮาร์ดิสค์ได้ถูกต้อง 68.47%

■ การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis)

การวิเคราะห์การถดถอยอย่างง่ายที่ผ่านมา เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรจำนวน 2 ตัว ประกอบด้วยตัวแปรอิสระหรือตัวแปรต้น (X) ที่ทำหน้าที่พยากรณ์ค่าของตัวแปรตาม (Y) อย่างละหนึ่งตัว ว่าจะมีค่าเท่าใดหรือมีความสัมพันธ์กันอย่างไร โดยมีสมการถดถอยอย่างง่ายก็คือ $Y = \alpha + \beta X$ แต่การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis) จะเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระหรือตัวแปรต้นที่ทำหน้าที่พยากรณ์ตั้งแต่ 2 ตัวขึ้นไป กับตัวแปรตาม 1 ตัว เช่น ต้องการพยากรณ์ผลการเรียนของนักศึกษาแผนกคอมพิวเตอร์ (Y) โดยใช้ตัวพยากรณ์ 3 ตัว ประกอบด้วย ความสนใจของผู้เรียน (X_1) ความรู้พื้นฐานของผู้เรียน (X_2) และคุณภาพการสอนของผู้สอน (X_3) เป็นต้น ในการวิเคราะห์การถดถอยพหุคูณนั้นจะต้องหาค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ (Multiple Correlation Coefficient) เพื่อให้ทราบถึงความสัมพันธ์ระหว่างตัวแปรอิสระหรือตัวแปรต้นจำนวนทั้ง 3 ตัวกับตัวแปรตามว่ามีความสัมพันธ์กันเช่นใด

สำหรับการวิเคราะห์การถดถอยพหุคูณ จะต้องหาสมการถดถอยเพื่อใช้ในการพยากรณ์ค่าของตัวแปรตาม (Y) เช่นเดียวกับการวิเคราะห์การถดถอยอย่างง่าย และหาค่าความคลาดเคลื่อนมาตรฐาน รวมทั้งหาค่าสหสัมพันธ์พหุคูณ (Multiple Correlation) เพื่อหาความสัมพันธ์เชิงเส้นตรงที่เป็นไปได้สูงสุดระหว่างตัวแปรอิสระหรือตัวแปรต้นกับตัวแปรตาม

■ บทสรุป

การวิเคราะห์การถดถอย (Regression Analysis) เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตาม ที่เรียกว่าตัวเกณฑ์ กับตัวแปรอิสระหรือตัวแปรต้นตัวอื่น ๆ ที่สัมพันธ์กัน ตั้งแต่ 2 ตัวขึ้นไป ซึ่งเรียกว่าตัวพยากรณ์ ถ้ามีตัวพยากรณ์เพียงตัวเดียวจะเรียกการศึกษาความสัมพันธ์นั้นว่า การวิเคราะห์การถดถอยอย่างง่าย (Simple Regression Analysis) หรือการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis) แต่ถ้ามีตัวพยากรณ์ตั้งแต่ 2 ตัวขึ้นไป โดยมีตัวเกณฑ์เพียงตัวเดียว จะเรียกการศึกษาความสัมพันธ์นั้นว่า การวิเคราะห์การถดถอยพหุคูณ (Multiple Regression Analysis)

การวิเคราะห์การถดถอยอย่างง่าย จะใช้วิธีพยากรณ์ค่าของตัวแปรตามจากตัวแปรอิสระหรือตัวแปรต้นที่ทำหน้าที่เป็นตัวพยากรณ์ จากสมการเส้นตรง $Y = \alpha + \beta X$ ซึ่ง α และ β เป็นพารามิเตอร์ที่ประมาณค่าโดยใช้ข้อมูลจากตัวอย่าง โดยนิยมใช้วิธีกำลังสองน้อยที่สุด และแทนค่าด้วย a และ b โดยที่ a ก็คือค่าคงที่ ส่วน b เป็นความชันของเส้นกราฟ ซึ่งแสดงถึงอัตราการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนแปลงไป ซึ่งเรียกว่า สัมประสิทธิ์การถดถอยหรือสัมประสิทธิ์การพยากรณ์ การพยากรณ์ค่าของตัวแปรตามจะถูกต้องหรือไม่ขึ้นอยู่กับสมการถดถอย ถ้าพบว่าสมการถดถอยมีส่วนเบี่ยงเบนน้อย ก็แสดงว่าค่าของตัวแปรตามจะถูกต้องมากขึ้น อย่างไรก็ตามในการพยากรณ์แต่ละครั้งจะเกิดความคลาดเคลื่อนขึ้น ซึ่งสามารถคำนวณหาค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์ได้โดยการหาค่า SEE และทดสอบสมมติฐานทางสถิติสมการถดถอยได้เช่นเดียวกันกับการหาค่า t-test ที่ผ่านมา

■ แบบฝึกหัดท้ายบท

จงตอบคำถามต่อไปนี้

1. ตัวเกณฑ์และตัวพยากรณ์ในการวิเคราะห์การถดถอย หมายถึงอะไร
2. การวิเคราะห์การถดถอยอย่างง่าย มีลักษณะเป็นอย่างไร
3. การวิเคราะห์การถดถอยพหุคูณ มีลักษณะเป็นอย่างไร
4. ให้อธิบายสมการเส้นตรง $Y = a + bX$ เกี่ยวกับความสัมพันธ์ของตัวแปร X และ Y ในลักษณะของการพยากรณ์ค่าของตัวแปร Y
5. สัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X กับตัวแปร Y หมายถึงอะไร
6. ค่าของตัวแปร X และตัวแปร Y จะมีค่าเท่ากัน ในกรณีใดบ้าง
7. ถ้าความชันของเส้นกราฟถดถอยมีค่าเท่ากับศูนย์ ตัวแปร Y จะเป็นอย่างไร
8. ค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์ หมายถึงอะไร
9. ให้อธิบายการเปลี่ยนแปลงของตัวแปร Y จากสมการถดถอย $Y = 15.25 + 5.50X$ เมื่อตัวแปร X เปลี่ยนไปแต่ละหน่วย
10. สัมประสิทธิ์ของการตัดสินใจมีค่าเท่ากับ 75% หมายความว่าอย่างไร

11. จากข้อมูลที่ปรากฏในตารางต่อไปนี้

X	1.2	1.4	1.6	1.8	1.1	1.7	2.2	2.8	3.2	4.5	2.9	1.6	2.4	4.3
Y	4	4.7	5.2	4.8	3.6	4.3	4	5.8	6.4	6.8	4.2	3.5	4.9	6.7

จงหาค่าต่าง ๆ ต่อไปนี้

11.1 สมการถดถอย

11.2 ให้หาค่าของ Y เมื่อ X มีค่าเท่ากับ 4.25 และ 1.95

11.3 สัมประสิทธิ์การถดถอย

11.4 เขียนเส้นกราฟถดถอยแสดงความสัมพันธ์ระหว่าง X และ Y

11.5 ค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์

12. จากการทดสอบการทำงานในสภาวะรับโหลด (Load Testing) ของ Stepping Motor จำนวน 10 ครั้ง เพื่อคัดเลือก Stepping Motor นำไปใช้งานกับเครื่องพิมพ์ Line Printer ยี่ห้อหนึ่ง พบว่าเมื่อเพิ่มโหลดมากขึ้น (mN) จะทำให้ความเร็วลดลง (rpm) ในลักษณะเชิงเส้น จากความเร็วปกติที่ 750 rpm ปรากฏผลดังตารางต่อไปนี้

ครั้งที่	1	2	3	4	5	6	7	8	9	10
ขนาดโหลด	2.05	2.55	2.95	3.15	3.25	3.75	4.25	4.95	8.5	12.5
ความเร็ว	749	748	747	747	745	745	742	742	740	739

จงหาค่าต่าง ๆ ต่อไปนี้

12.1 สมการถดถอย

12.2 ให้หาความเร็วของมอเตอร์ เมื่อมีโหลดอยู่ที่ 3.5 mN

12.3 เขียนเส้นกราฟถดถอย

12.4 สัมประสิทธิ์การถดถอย

12.5 ค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์

12.6 จงทดสอบสมมติฐานของการถดถอยที่ระดับนัยสำคัญ .05

13. ผลจากการทดลองใช้บทเรียน m-Learning ในวิชาเทคโนโลยีสารสนเทศเพื่อชีวิตของผู้เรียนระดับปริญญาตรี โดยใช้แบบแผนการทดลองแบบ One-Group Pretest Posttest Design ซึ่งมีจำนวน 12 คน พบว่าค่าเฉลี่ยของคะแนนแบบฝึกหัดระหว่างบทเรียน (E1) จำนวน 8 โมดูล (คะแนนเต็ม 80) มีความสัมพันธ์ในลักษณะเชิงเส้นกับค่าเฉลี่ยของคะแนนแบบทดสอบหลังบทเรียน (E2) (คะแนนเต็ม 100) ปรากฏดังตารางต่อไปนี้

คนที่	E1	E2
1	58	83
2	62	78
3	60	69
4	61	80
5	62	82
6	70	85
7	68	83
8	55	82
9	67	84
10	71	87
11	66	86
12	59	76

จงหาค่าต่าง ๆ ต่อไปนี้

13.1 ค่าของ E2 ที่ได้จากการพยากรณ์โดยใช้สมการถดถอย เมื่อผู้เรียนคนหนึ่งทำแบบฝึกหัดระหว่างบทเรียนได้ 65.5 คะแนน

13.2 สัมประสิทธิ์การถดถอย

13.3 ค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์

13.4 จงทดสอบสมมติฐานของการถดถอยที่ระดับนัยสำคัญ .01

13.5 สัมประสิทธิ์ของการตัดสินใจ และอธิบายความหมาย

14. จากการติดตามผลบัณฑิตที่จบการศึกษาสาขาวิทยาการคอมพิวเตอร์ ที่ทำงานในบริษัทแห่งหนึ่ง พบว่าบัณฑิตจำนวน 7 คน ได้รับเงินเดือนแตกต่างกันดังนี้

คนที่	1	2	3	4	5	6	7
GPA	3.47	2.60	3.35	2.86	3.82	2.21	3.26
เงินเดือน	29,400	24,800	27,900	25,300	30,300	23,000	28,200

จงหาสมการถดถอยที่แสดงความสัมพันธ์ระหว่าง GPA กับเงินเดือนที่ได้รับ และทดสอบสมมติฐานสมการถดถอยที่ได้ ที่ระดับนัยสำคัญ .05

15. การเพิ่มความจุ RAM ของคอมพิวเตอร์ เป็นผลทำให้ความเร็วในการประมวลผลของคอมพิวเตอร์เพิ่มขึ้นในลักษณะเชิงเส้น จากการทดลองเพิ่มความจุ RAM ของเครื่องคอมพิวเตอร์ (MB) ที่ใช้ CPU แตกต่างกันจำนวน 10 ยี่ห้อ สามารถวัดความเร็วในการประมวลผล (ps) ได้ดังนี้

ยี่ห้อ	A	B	C	D	E	F	G	H	I	J
RAM	1	1	2	4	4	8	8	16	16	32
ความเร็ว	9.75	8.85	11.2	12.25	15	21.5	20.5	36.8	34.5	47.8

จงหาต่าง ๆ ต่อไปนี้

15.1 สัมประสิทธิ์การถดถอย

15.2 ถ้าเพิ่มความจุ RAM เป็น 12 MB ความเร็วในการประมวลผลจะเป็นเท่าไร

15.3 ค่าความคลาดเคลื่อนมาตรฐานในการพยากรณ์

15.4 จงทดสอบสมมติฐานของการถดถอยที่ระดับนัยสำคัญ .01

15.5 สัมประสิทธิ์ของการตัดสินใจ และอธิบายความหมาย