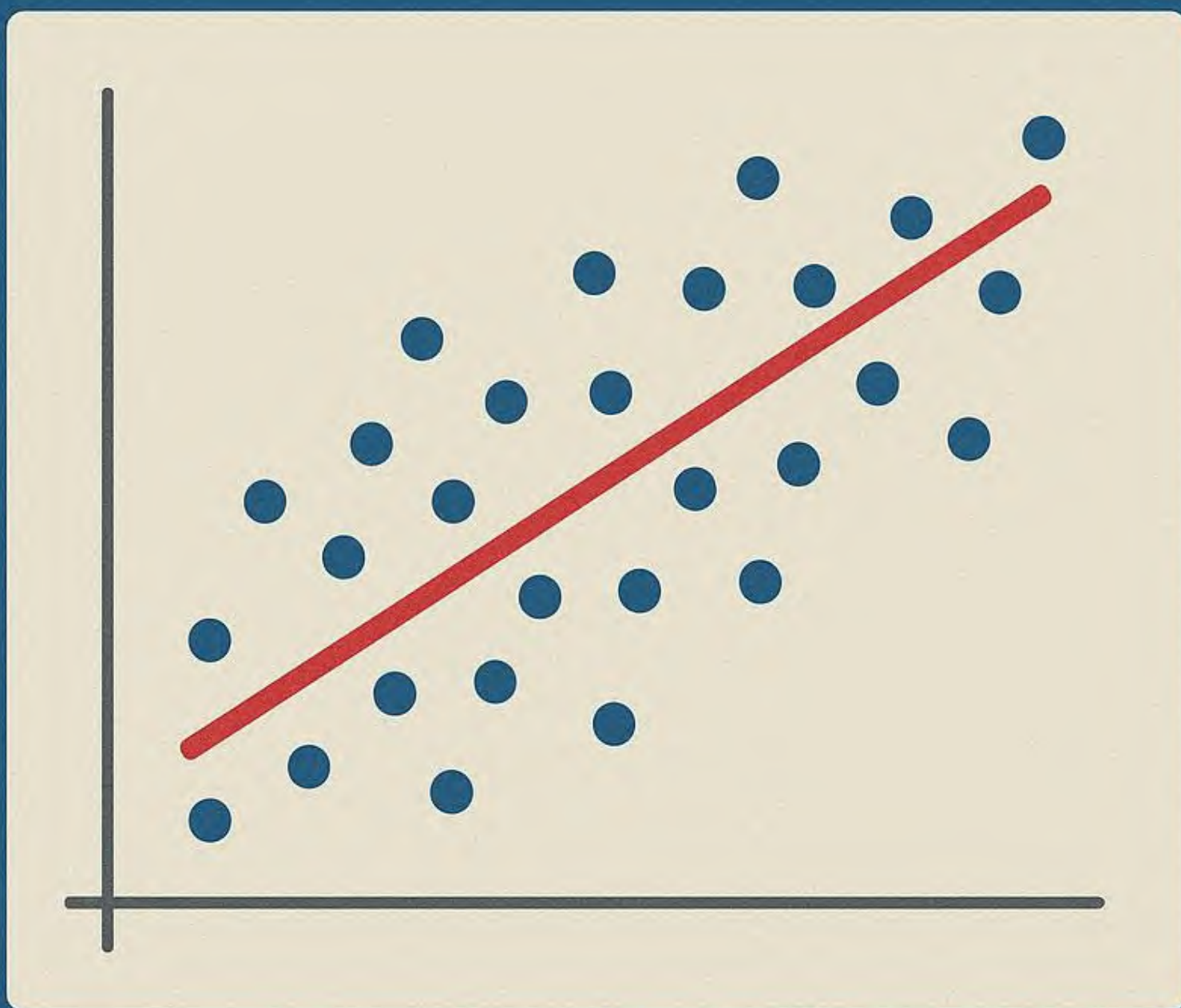


การวิเคราะห์ REGRESSION ฉบับเริ่มต้น



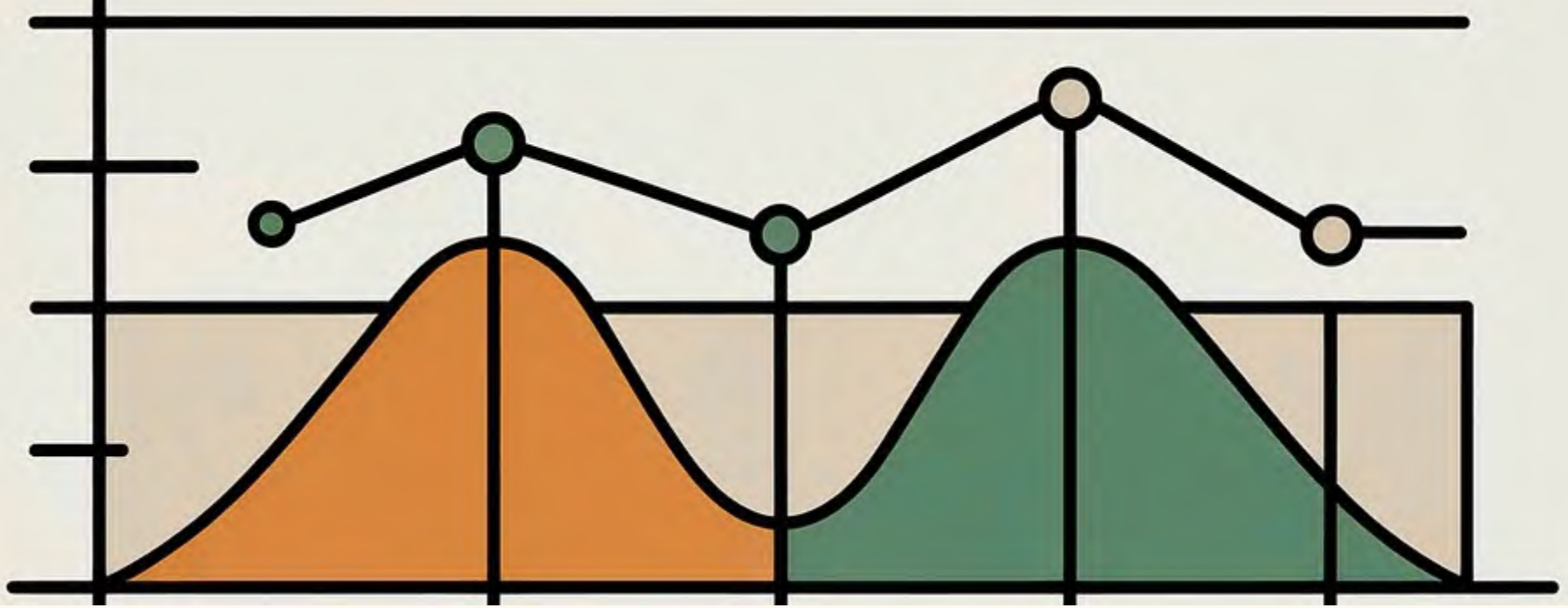
ดร.ขวัญใจ วงศ์ช่วย



บทที่ 1: Linear Regression (การถดถอยเชิงเส้นอย่างง่าย)

เรียนรู้เทคนิคการวิเคราะห์ความสัมพันธ์เชิงเส้นที่ทรงพลัง สำหรับการทำนายและเข้าใจข้อมูลในยุคดิจิทัล

การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) เป็นแบบจำลองทางสถิติพื้นฐานที่ใช้เพื่อศึกษาและแสดงความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระหนึ่งตัว (Independent Variable) และตัวแปรตามหนึ่งตัว (Dependent Variable) โดยมีเป้าหมายหลักในการสร้างสมการเส้นตรงที่ดีที่สุด เพื่อใช้ทำนายค่าของตัวแปรตามจากค่าของตัวแปรอิสระที่กำหนดให้ ตัวอย่างเช่น เราสามารถใช้เทคนิคนี้เพื่อคาดการณ์ราคาบ้านจากขนาดพื้นที่ หรือทำนายยอดขายสินค้าจากงบประมาณการตลาดที่ลงทุนไป การทำความเข้าใจหลักการนี้จะช่วยให้คุณสามารถตีความและใช้ข้อมูลเพื่อการตัดสินใจทางธุรกิจได้อย่างมีประสิทธิภาพในโลกที่ขับเคลื่อนด้วยข้อมูล



การถดถอยเชิงเส้นคืออะไร?



วิเคราะห์ความสัมพันธ์

วิธีวิเคราะห์ความสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระ (X) กับตัวแปรตาม (Y)

เป้าหมายคือการหาความสัมพันธ์ในรูปแบบเส้นตรงที่แสดงให้เห็นว่าการเปลี่ยนแปลงของตัวแปรอิสระ (X) มีผลต่อตัวแปรตาม (Y) อย่างไร เพื่อทำความเข้าใจแนวโน้มของข้อมูล.



ทำนายค่าได้แม่นยำ

ใช้ทำนายค่าของ Y จากค่า X ด้วยสมการเส้นตรงที่หาได้จากข้อมูล

สมการนี้ (มักอยู่ในรูป $y = ax + b$) จะถูกปรับให้เข้ากับชุดข้อมูลที่มีอยู่ได้ดีที่สุด เพื่อใช้ในการคาดการณ์ผลลัพธ์ของตัวแปรตามได้อย่างแม่นยำและมีเหตุผลจากข้อมูลใหม่.

สมการพื้นฐานของ Linear Regression

$$Y = b_0 + b_1X$$

สมการนี้เป็นหัวใจสำคัญของการถดถอยเชิงเส้น ซึ่งช่วยให้เราสามารถสร้างแบบจำลองความสัมพันธ์ระหว่างตัวแปรต่างๆ ได้อย่างง่ายดาย โดยมีองค์ประกอบหลักสองส่วนคือจุดตัดแกน Y และความชันของเส้นถดถอย



จุดตัดแกน Y

b_0 คือจุดตัดแกน Y (intercept) - ค่าเริ่มต้นของเส้นถดถอย

ความหมาย: แสดงถึงค่าทำนายของตัวแปรตาม Y เมื่อตัวแปรอิสระ X มีค่าเป็นศูนย์ ซึ่งเป็นค่าพื้นฐานที่ไม่ขึ้นอยู่กับการเปลี่ยนแปลงของ X

ตัวอย่าง: หากทำนายยอดขายไอศกรีม (Y) จากอุณหภูมิ (X), b_0 อาจเป็นยอดขายพื้นฐานเมื่ออุณหภูมิ 0 องศาเซลเซียส หรือยอดขายที่ไม่ขึ้นกับอุณหภูมิ

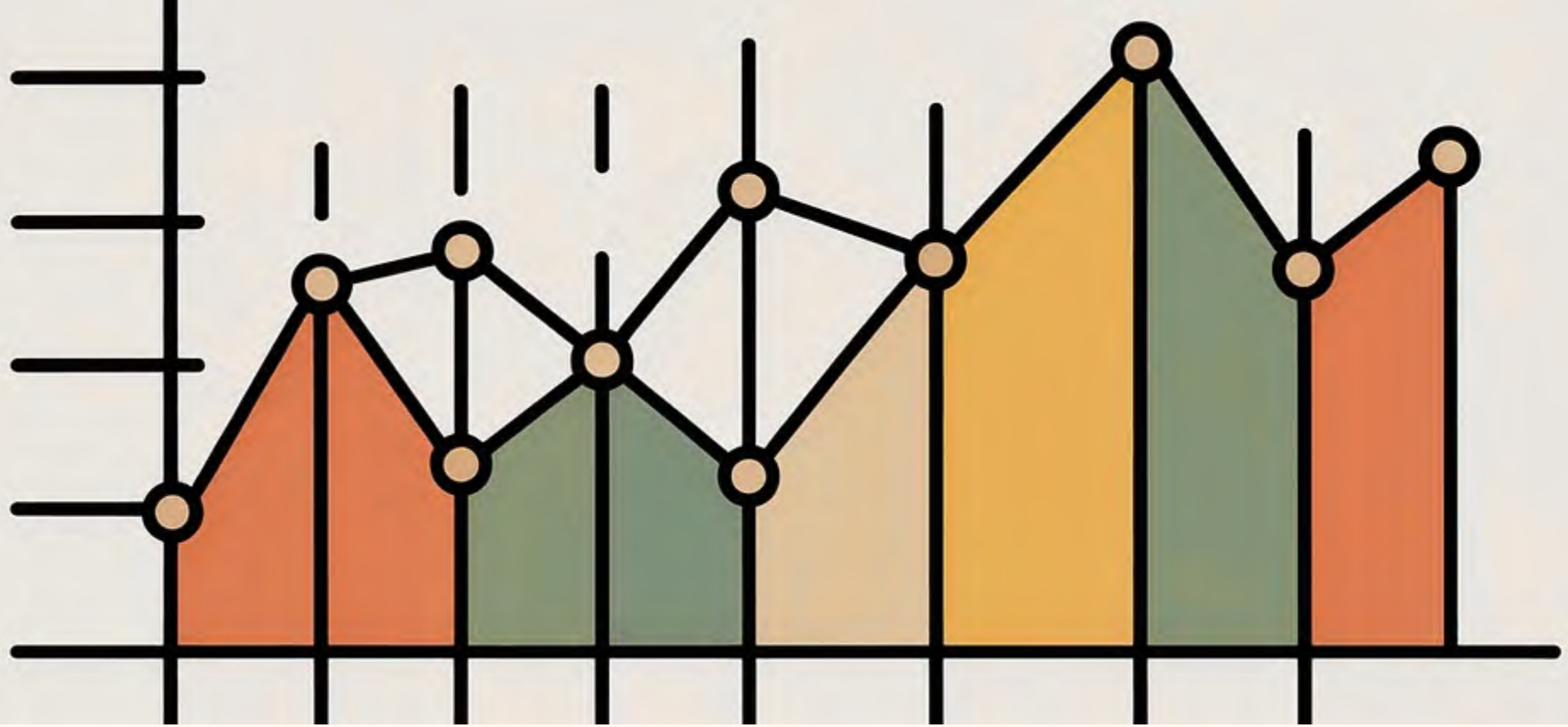


ความชันของเส้น

b_1 คือความชันของเส้น (slope) - แสดงอัตราการเปลี่ยนแปลง

ความหมาย: บ่งบอกว่าตัวแปรตาม Y จะเปลี่ยนแปลงไปเท่าใด เมื่อตัวแปรอิสระ X เพิ่มขึ้นหนึ่งหน่วย

ตัวอย่าง: ในการทำนายยอดขายไอศกรีม (Y) จากอุณหภูมิ (X), b_1 จะบอกว่ายอดขายไอศกรีมจะเพิ่มขึ้น (หรือลดลง) เท่าใดต่อการเพิ่มขึ้นของอุณหภูมิ 1 องศาเซลเซียส ซึ่งสะท้อนถึงผลกระทบของอุณหภูมิที่มีต่อยอดขาย



กราฟแสดงความสัมพันธ์เชิงเส้น

จุดข้อมูลแต่ละจุดบนกราฟจะแสดงค่าจริงของตัวแปรอิสระ (X) และตัวแปรตาม (Y) ที่สังเกตได้ เส้นถดถอย (Regression Line) คือเส้นที่ถูกคำนวณขึ้นมาให้เหมาะสมที่สุดกับข้อมูลเหล่านั้น โดยมีจุดประสงค์เพื่อลดระยะห่างระหว่างจุดข้อมูลกับเส้นให้เหลือน้อยที่สุด เส้นนี้ทำหน้าที่เป็นตัวแทนของแนวโน้มโดยรวมของความสัมพันธ์เชิงเส้นระหว่างตัวแปรทั้งสอง ทำให้เราสามารถมองเห็นทิศทางและความแข็งแกร่งของความสัมพันธ์ได้อย่างชัดเจน และใช้ในการคาดการณ์ค่า Y เมื่อทราบค่า X ใหม่ได้อย่างแม่นยำ เช่น การทำนายยอดขายจากการใช้จ่ายทางการตลาด

ตัวแปรใน Linear Regression

● ตัวแปรตาม

Dependent variable (Y)

ตัวแปรที่ต้องการทำนาย เป็นผลลัพธ์ที่เราสนใจศึกษา

เป็นตัวแปรที่ได้รับผลกระทบหรือเปลี่ยนแปลงไปตามตัวแปรอิสระ มักถูกเรียกว่า '**response variable**' หรือ '**outcome variable**' การเปลี่ยนแปลงของตัวแปรตามเป็นสิ่งที่เราต้องการอธิบายหรือทำความเข้าใจผ่านโมเดลการถดถอย

ตัวอย่าง: ในการทำนายยอดขายของร้านค้า (Y), ตัวแปรตามคือยอดขายที่เกิดขึ้นจริงที่ต้องการจะทำนาย

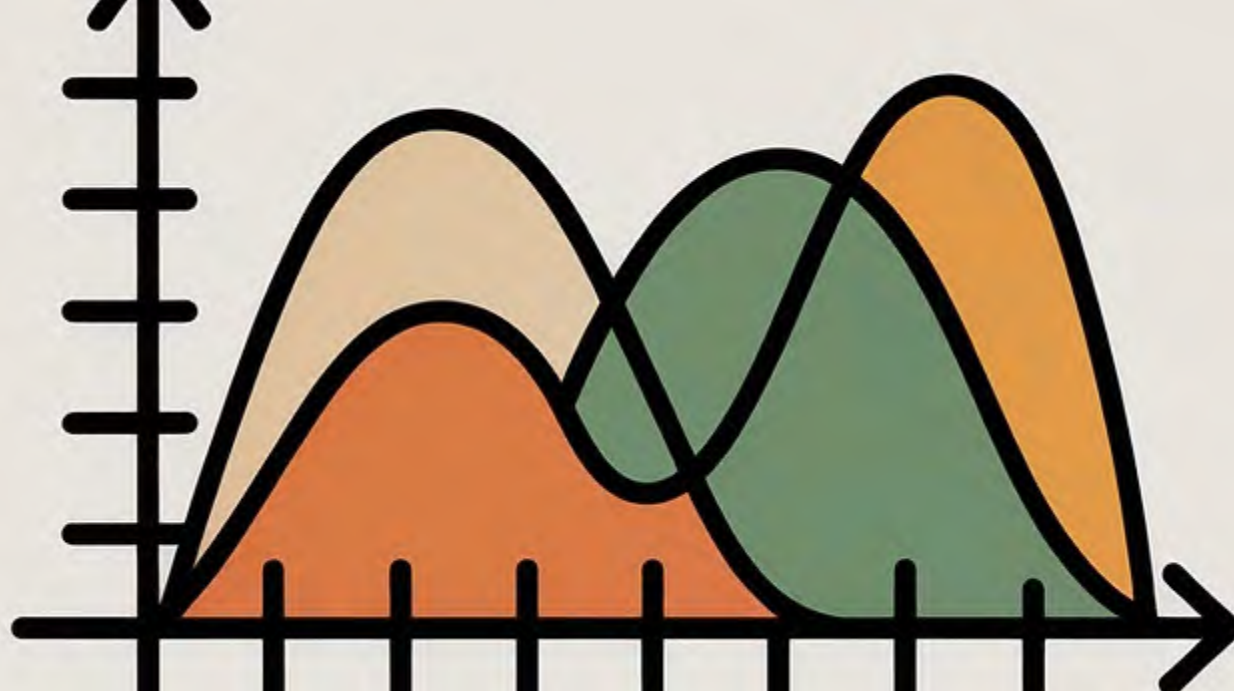
● ตัวแปรอิสระ

Independent variable (X)

ตัวแปรที่ใช้ทำนาย เป็นปัจจัยที่มีอิทธิพลต่อผลลัพธ์

เป็นตัวแปรที่ถูกควบคุม สังเกต หรือวัดค่าเพื่ออธิบายหรือทำนายการเปลี่ยนแปลงของตัวแปรตาม บางครั้งเรียกว่า '**predictor variable**' หรือ '**explanatory variable**' ในโมเดลเชิงเส้น ตัวแปรอิสระสามารถมีได้หลายตัวเพื่ออธิบายความสัมพันธ์กับตัวแปรตาม

ตัวอย่าง: ตัวแปรอิสระที่อาจส่งผลต่อยอดขาย (X) อาจเป็นจำนวนโฆษณาที่ใช้, ราคาของสินค้า หรือช่วงฤดูกาล



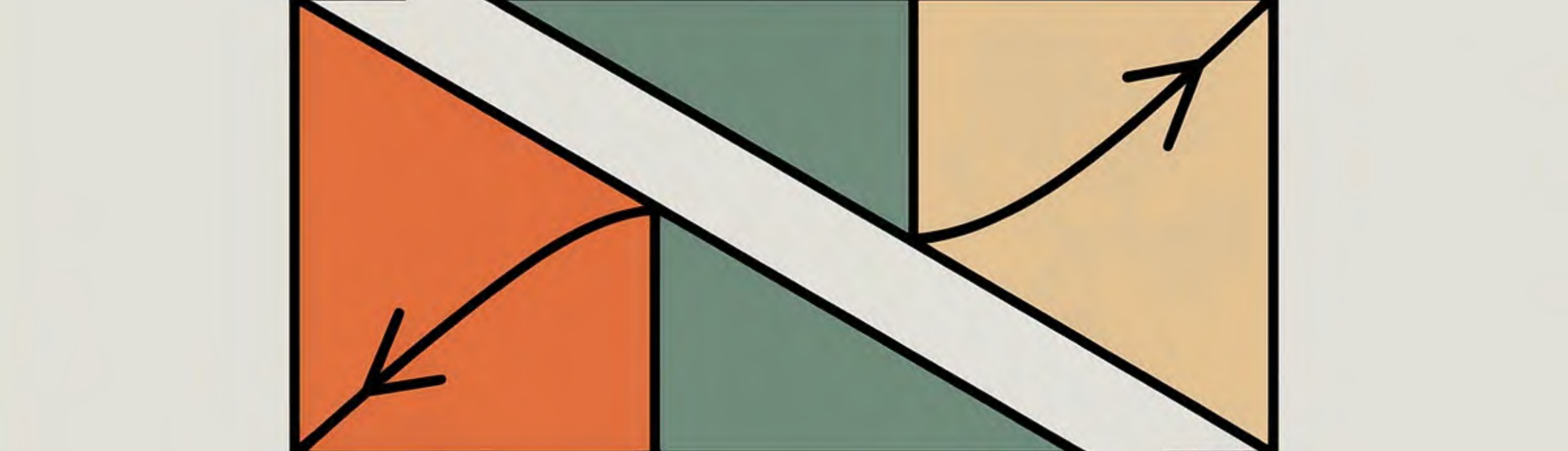
จุดตัดแกน Y (Intercept) คืออะไร?

ค่าของ Y เมื่อ $X = 0$

เป็นจุดที่เส้นถดถอยตัดกับแกน Y ซึ่งแสดงถึงค่าพื้นฐานหรือค่าเริ่มต้นของตัวแปรตาม (Y) เมื่อตัวแปรอิสระ (X) มีค่าเป็นศูนย์ ในทางคณิตศาสตร์ อินเตอร์เซปต์ (b_0 หรือ β_0) คือค่าคงที่ในสมการถดถอยเชิงเส้น $Y = b_0 + b_1X + \epsilon$ การทำความเข้าใจค่านี้ช่วยให้เราทราบว่าตัวแปรตามมีค่าเท่าใดในกรณีที่ไม่มีอิทธิพลจากตัวแปรอิสระ

ค่าพื้นฐานของโมเดล

อินเตอร์เซปต์ทำหน้าที่เป็นค่าเริ่มต้นหรือค่าพื้นฐานของตัวแปรตามในโมเดล ก่อนที่ตัวแปรอิสระจะถูกนำมาพิจารณาและมีผลกระทบต่อการเปลี่ยนแปลงค่า Y ตัวอย่างเช่น หากเราทำนายยอดขาย (Y) จากงบประมาณโฆษณา (X) ค่าอินเตอร์เซปต์อาจหมายถึงยอดขายขั้นต่ำที่เกิดขึ้นได้แม้จะไม่มีงบโฆษณาเลย สิ่งนี้มีความสำคัญในการตีความโมเดลในบริบททางธุรกิจและวิทยาศาสตร์

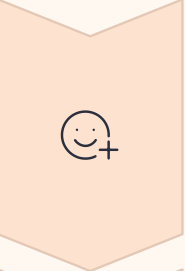


ความชัน (Slope) คืออะไร?



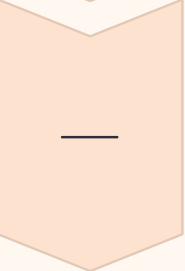
การเปลี่ยนแปลง

บอกถึงอัตราการเปลี่ยนแปลงของตัวแปรตาม (Y) เมื่อตัวแปรอิสระ (X) เพิ่มขึ้น 1 หน่วย โดยสมมติว่าปัจจัยอื่น ๆ คงที่ ทำให้เราเข้าใจขนาดและทิศทางของผลกระทบที่ X มีต่อ Y ได้อย่างชัดเจน



ความสัมพันธ์เป็นบวก

ถ้า $b_1 > 0$ ความสัมพันธ์เป็นบวก หมายถึงเมื่อค่าของ X เพิ่มขึ้น ค่าของ Y ก็มีแนวโน้มที่จะเพิ่มขึ้นตามไปด้วย เช่น ยิ่งออกกำลังกายมาก สุขภาพก็จะดีขึ้น ซึ่งกราฟจะมีลักษณะชันขึ้นไปทางขวา แสดงถึงความสัมพันธ์ในทิศทางเดียวกัน



ความสัมพันธ์เป็นลบ

ถ้า $b_1 < 0$ ความสัมพันธ์เป็นลบ หมายถึงเมื่อค่าของ X เพิ่มขึ้น ค่าของ Y กลับมีแนวโน้มที่จะลดลง เช่น ยิ่งอุณหภูมิสูงขึ้น ยอดขายเสื้อกันหนาวก็จะลดลง ซึ่งกราฟจะมีลักษณะชันลงไปทางขวา แสดงถึงความสัมพันธ์ในทิศทางตรงกันข้าม

ตัวอย่างสมการถดถอย

สมการถดถอยเชิงเส้นอย่างง่ายช่วยให้เราสามารถประมาณค่าของตัวแปรตาม (Y) โดยอาศัยค่าของตัวแปรอิสระ (X) และเข้าใจความสัมพันธ์ระหว่างตัวแปรทั้งสองได้อย่างชัดเจน



$$Y = 5 + 2X$$

ในสมการนี้:

- ค่าคงที่ 5 คือจุดตัดแกน Y (Intercept) ซึ่งเป็นค่าประมาณของ Y เมื่อ $X = 0$
- ค่าสัมประสิทธิ์ 2 คือความชัน (Slope) หมายความว่าทุกครั้งที่ X เพิ่มขึ้น 1 หน่วย Y ที่คาดการณ์จะเพิ่มขึ้น 2 หน่วย

01

กำหนดค่า X

สมมติว่าเรากำลังศึกษาความสัมพันธ์ระหว่างชั่วโมงการอ่านหนังสือ (X) กับคะแนนสอบที่คาดการณ์ (Y) เราต้องการทราบว่าถ้าอ่านหนังสือ 3 ชั่วโมง จะได้คะแนนสอบประมาณเท่าไร

02

แทนค่าในสมการ

นำค่า $X = 3$ ไปแทนที่ในสมการถดถอยที่เราได้ เพื่อเริ่มการคำนวณหาค่า Y

$$Y = 5 + 2 \times 3$$

03

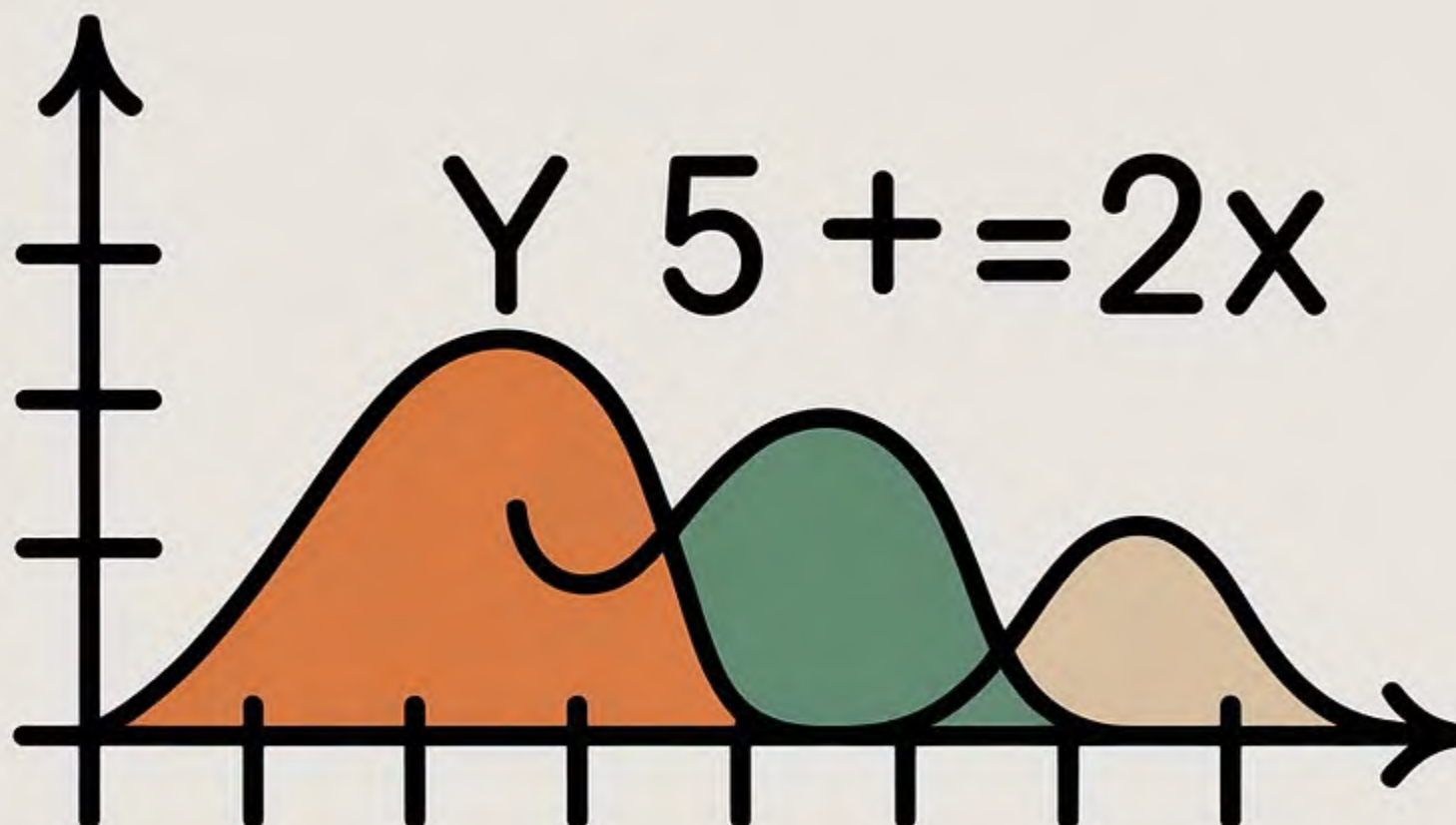
คำนวณผลลัพธ์

ทำการคำนวณตามลำดับคณิตศาสตร์ จะได้ผลลัพธ์ของ Y ที่คาดการณ์ได้

$$Y = 5 + 6 = 11$$

ดังนั้น หากอ่านหนังสือ 3 ชั่วโมง คะแนนสอบที่คาดการณ์ได้คือ 11 คะแนน

สมการถดถอยนี้มีประโยชน์มาก เพราะเมื่อ X มีค่าเท่าไร เราสามารถคำนวณหาค่า Y ที่คาดการณ์ได้ทันทีจากสมการนี้ ทำให้เราสามารถทำนายหรือประมาณค่าของตัวแปรตามได้อย่างรวดเร็วและมีหลักการ



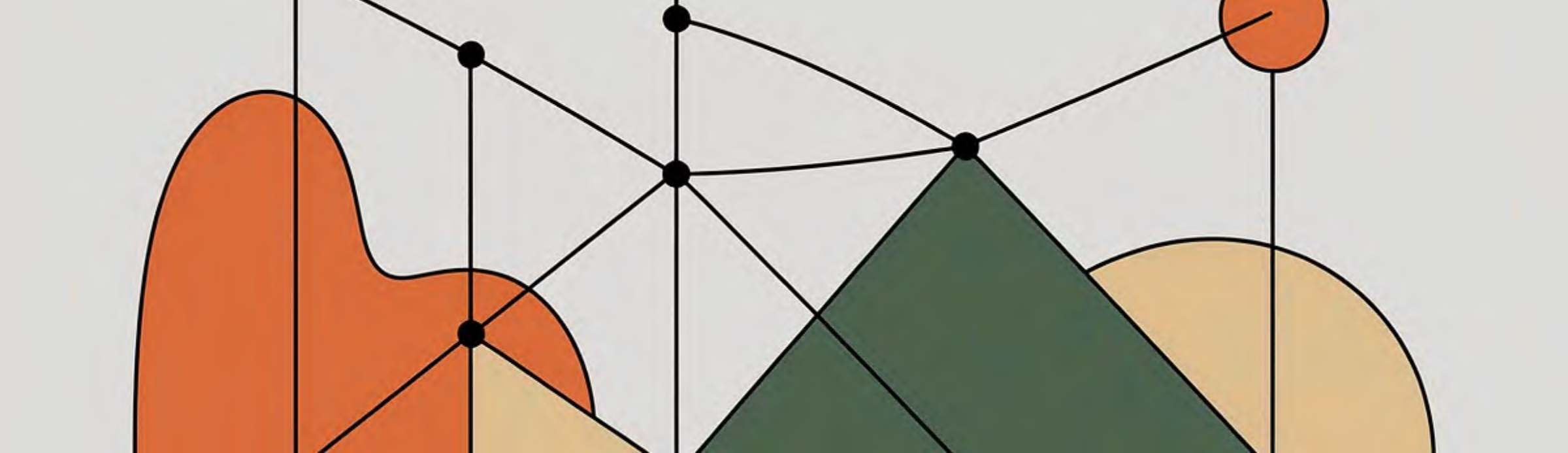
กราฟแสดงสมการ $Y = 5 + 2X$

เส้นตรงนี้แสดงความสัมพันธ์เชิงเส้นที่ชัดเจนระหว่างตัวแปร X และ Y โดยมีลักษณะสำคัญสองประการคือ จุดตัดแกน Y และความชัน

จุดตัดแกน Y อยู่ที่ 5 หมายความว่าเมื่อค่า X เป็นศูนย์ (คือตำแหน่งที่เส้นกราฟตัดกับแกน Y) ค่าของ Y จะเท่ากับ 5 ซึ่งเป็นจุดเริ่มต้นของเส้นตรงนี้บนระนาบพิกัด

ความชันเท่ากับ 2 บ่งบอกถึงอัตราการเปลี่ยนแปลงของ Y ต่อการเปลี่ยนแปลงของ X ทุกๆ 1 หน่วย นั่นคือ ทุกครั้งที่ X เพิ่มขึ้น 1 หน่วย ค่า Y จะเพิ่มขึ้น 2 หน่วยเสมอ ยกตัวอย่างเช่น หาก X มีค่าเป็น 1 ค่า Y จะเป็น 7 ($5 + 2 \times 1$) และหาก X มีค่าเป็น 2 ค่า Y จะเป็น 9 ($5 + 2 \times 2$) ซึ่งการเปลี่ยนแปลงนี้คงที่ตลอดทั้งเส้นกราฟ

กราฟนี้ช่วยให้เราเห็นภาพความสัมพันธ์โดยตรงและสามารถใช้ในการทำนายค่า Y สำหรับค่า X ใดๆ ได้อย่างง่ายดาย ทำให้การทำความเข้าใจข้อมูลเชิงปริมาณเป็นไปได้อย่างมีประสิทธิภาพ



วิธีหาค่าพารามิเตอร์

วิธี Least Squares

วิธี Least Squares (วิธีกำลังสองน้อยที่สุด) เป็นเทคนิคทางสถิติที่ใช้ในการประมาณค่าพารามิเตอร์ b_0 (จุดตัดแกน Y) และ b_1 (ความชัน) ในสมการถดถอยเชิงเส้น เพื่อสร้างแบบจำลองที่อธิบายความสัมพันธ์ระหว่างตัวแปรต้นและตัวแปรตาม

เป้าหมายหลักคือการหาค่าของ b_0 และ b_1 ที่ทำให้ผลรวมของความคลาดเคลื่อนกำลังสอง (Sum of Squared Errors, SSE) ระหว่างค่าที่สังเกตได้จริงกับค่าที่ทำนายโดยแบบจำลองมีค่าน้อยที่สุด วิธีนี้ช่วยให้เส้นถดถอยที่ได้มีความเหมาะสมกับข้อมูลมากที่สุด โดยเฉพาะเมื่อข้อมูลมีการกระจายตัว

การทำให้ความคลาดเคลื่อนยกกำลังสองน้อยที่สุดยังช่วยให้นักวิเคราะห์สามารถหาเส้นที่ดีที่สุดที่ผ่านจุดข้อมูลต่างๆ ได้อย่างเป็นระบบ ซึ่งเป็นวิธีมาตรฐานที่ใช้กันอย่างแพร่หลายในการวิเคราะห์การถดถอยเพื่อสร้างแบบจำลองที่แม่นยำและสามารถอธิบายความสัมพันธ์ระหว่างตัวแปรได้อย่างมีประสิทธิภาพและสามารถนำไปใช้ในการทำนายค่าต่างๆ ได้



ความหมายของ Least Squares

หลักการทํางาน

ลดความแตกต่างระหว่างค่าจริงกับค่าที่ทำนายโดยเส้นถดถอยให้เหลือน้อยที่สุด โดยการหาผลรวมกำลังสองของความคลาดเคลื่อน (residuals) ระหว่างค่าที่สังเกตได้จริงกับค่าที่คาดการณ์จากแบบจำลองให้น้อยที่สุด

วัดความคลาดเคลื่อน

ยกกำลังสองเพื่อให้ค่าเป็นบวกและเน้นความคลาดเคลื่อนที่มากขึ้น การยกกำลังสองนี้ทำให้ค่าความคลาดเคลื่อนที่เป็นบวกและลบไม่หักล้างกัน และยังช่วยให้เส้นถดถอยพยายามลดความคลาดเคลื่อนใหญ่ๆ ลงเป็นพิเศษ

เส้นที่ดีที่สุด

เส้นถดถอยที่ได้จะเป็นเส้นที่ใกล้เคียงจุดข้อมูลทั้งหมดมากที่สุด เส้นนี้ไม่เพียงแต่ผ่านกลางจุดข้อมูลเท่านั้น แต่ยังสะท้อนแนวโน้มโดยรวมของข้อมูล ทำให้สามารถใช้ในการทำนายค่าใหม่ๆ ได้อย่างแม่นยำ

สมการคำนวณความชัน

$$b_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}$$

สูตรนี้ใช้เพื่อคำนวณค่าความชัน (b_1) ของเส้นถดถอยเชิงเส้น ซึ่งแสดงถึงการเปลี่ยนแปลงโดยเฉลี่ยของตัวแปรตาม (Y) เมื่อตัวแปรอิสระ (X) เปลี่ยนไปหนึ่งหน่วย พุดง่าย ๆ คือมันบอกเราว่า Y จะเพิ่มขึ้นหรือลดลงเท่าใดในแต่ละครั้งที่ X เพิ่มขึ้นหนึ่งหน่วย การคำนวณนี้อาศัยความสัมพันธ์ร่วมกันระหว่าง X และ Y (ความแปรปรวนร่วม) และปรับด้วยความแปรปรวนของ X เพื่อให้ได้ค่าความชันที่เหมาะสมที่สุดซึ่งแสดงถึงแนวโน้มของข้อมูล

📌 $\bar{X}$ และ $\bar{Y}$ คือค่าเฉลี่ยของตัวแปร X และ Y ตามลำดับ

การทำความเข้าใจส่วนประกอบของสูตรนี้ช่วยให้เห็นภาพว่าค่าความชันเป็นตัวแทนของอัตราการเปลี่ยนแปลงของ Y ที่สัมพันธ์กับ X ได้อย่างไร

สมการคำนวณจุดตัดแกน

$$b_0 = \bar{Y} - b_1 \bar{X}$$

สมการนี้ใช้สำหรับหาค่าจุดตัดแกน Y (b_0) ซึ่งเป็นค่าที่เส้นถดถอยเชิงเส้นตัดกับแกน Y เมื่อค่าตัวแปรอิสระ (X) เป็นศูนย์ ค่านี้สำคัญในการทำนายค่า Y เมื่อไม่มีผลกระทบจาก X และเป็นองค์ประกอบสำคัญในการสร้างสมการเส้นถดถอยที่สมบูรณ์

1 คำนวณค่าเฉลี่ย

เริ่มต้นด้วยการหาค่าเฉลี่ยของข้อมูลตัวแปรอิสระ ($\bar{X}$) และตัวแปรตาม ($\bar{Y}$) จากชุดข้อมูลทั้งหมดที่คุณมี การคำนวณค่าเฉลี่ยเหล่านี้เป็นพื้นฐานสำคัญสำหรับขั้นตอนต่อไปในการหาจุดตัดแกน

2 คำนวณความชัน

ก่อนที่จะหาจุดตัดแกน Y คุณจำเป็นต้องทราบค่าความชัน (b_1) ของเส้นถดถอยก่อน ซึ่งสามารถคำนวณได้จากสูตรความชันที่ได้กล่าวถึงในส่วนก่อนหน้านี้ ค่าความชันนี้จะบอกถึงอัตราการเปลี่ยนแปลงของ Y ต่อหนึ่งหน่วยการเปลี่ยนแปลงของ X

3 แทนค่าในสมการ

เมื่อได้ค่าเฉลี่ย $\bar{X}$ และ $\bar{Y}$ รวมถึงค่าความชัน b_1 แล้วให้นำค่าเหล่านี้ไปแทนลงในสมการ $b_0 = \bar{Y} - b_1 \bar{X}$ เพื่อคำนวณหาค่า b_0 ซึ่งเป็นจุดตัดแกน Y ค่านี้จะช่วยให้คุณสามารถกำหนดตำแหน่งเริ่มต้นของเส้นถดถอยได้อย่างถูกต้อง

แสดงค่าพารามิเตอร์บนกราฟ

จุดตัดแกน (b_0)

จุดที่เส้นถดถอยตัดกับแกน Y เมื่อตัวแปรอิสระ X มีค่าเป็น 0

ในทางปฏิบัติ b_0 แสดงถึงค่า Y ที่คาดการณ์ไว้เมื่อตัวแปรอิสระ X มีค่าเป็นศูนย์ อย่างไรก็ตาม ในบางสถานการณ์ การที่ X เป็นศูนย์อาจไม่มีความหมายเชิงปฏิบัติ จึงควรตีความค่านี้อย่างระมัดระวังและพิจารณาบริบทของข้อมูล

ความชัน (b_1)

มุมเอียงของเส้นแสดงอัตราการเปลี่ยนแปลงของ Y เมื่อ X เปลี่ยนแปลงไป

ค่า b_1 บ่งบอกว่าเมื่อตัวแปร X เปลี่ยนแปลงไป 1 หน่วย ตัวแปร Y จะเปลี่ยนแปลงไปกี่หน่วย ถ้า b_1 เป็นบวก แสดงว่า X และ Y มีความสัมพันธ์ไปในทิศทางเดียวกัน แต่ถ้า b_1 เป็นลบ แสดงถึงความสัมพันธ์ในทิศทางตรงกันข้าม ทำให้เราเข้าใจผลกระทบของ X ที่มีต่อ Y



การตีความค่าพารามิเตอร์



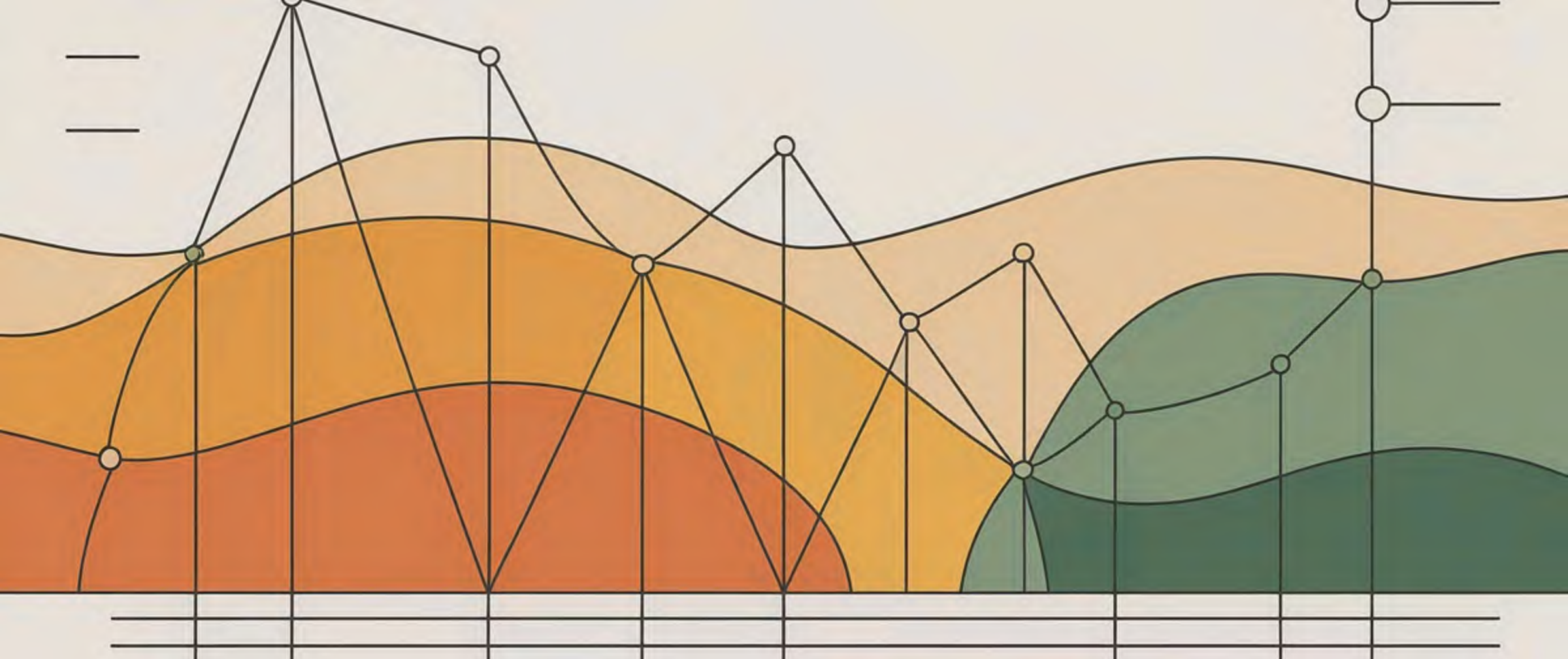
ความชัน b_1

การเปลี่ยนแปลงเฉลี่ยของ Y ต่อ 1 หน่วยของ X บอกความแข็งแกร่งและทิศทางของความสัมพันธ์ ความชันที่เป็นบวกหมายถึงเมื่อ X เพิ่มขึ้น Y ก็จะเพิ่มขึ้นด้วย ในขณะที่ความชันที่เป็นลบหมายถึงเมื่อ X เพิ่มขึ้น Y จะลดลง ขนาดของความชันยังบ่งชี้ถึงความไวของการเปลี่ยนแปลง Y ต่อ X หากความชันมีค่ามาก แสดงว่า Y เปลี่ยนแปลงอย่างรวดเร็วเมื่อ X เปลี่ยนไปเพียงเล็กน้อย ซึ่งช่วยให้เราเข้าใจว่าตัวแปรอิสระมีผลกระทบต่อตัวแปรตามมากน้อยเพียงใด



จุดตัดแกน b_0

ค่าคาดการณ์ของ Y เมื่อ $X = 0$ เป็นค่าพื้นฐานหรือค่าเริ่มต้นของตัวแปรตาม ในบางกรณีที่ $X = 0$ ไม่มีความหมายในทางปฏิบัติ ค่า b_0 อาจเป็นเพียงจุดอ้างอิงทางคณิตศาสตร์เท่านั้น ตัวอย่างเช่น หาก X คือจำนวนชั่วโมงการทำงานและ Y คือรายได้ b_0 จะแสดงถึงรายได้เมื่อไม่ได้ทำงานเลย หรือค่าคงที่อื่นๆ ที่ไม่ขึ้นกับ X ดังนั้น การตีความ b_0 จะต้องพิจารณาบริบทของข้อมูลและสถานการณ์จริงอย่างรอบคอบเสมอ



ตัวอย่างจริง: อายุกับ BMI

ความสัมพันธ์ที่พบ

จากการวิเคราะห์ข้อมูลพบว่า BMI เพิ่มขึ้น 0.049 หน่วยต่ออายุที่เพิ่มขึ้น 1 ปี

ตัวเลขนี้แสดงให้เห็นว่าเมื่อคนเรามีอายุมากขึ้น ดัชนีมวลกาย (BMI) มีแนวโน้มที่จะสูงขึ้นเล็กน้อยและค่อยเป็นค่อยไปในแต่ละปี ซึ่งอาจบ่งชี้ถึงการสะสมไขมันหรือการลดลงของมวลกล้ามเนื้อ

การเปลี่ยนแปลงนี้อาจเกิดจากปัจจัยหลายอย่าง เช่น การลดลงของการเผาผลาญพลังงาน (metabolism) กิจกรรมทางกายที่ลดลง หรือการเปลี่ยนแปลงของมวลกล้ามเนื้อเมื่อเข้าสู่วัยชรา

การศึกษานี้ไม่เพียงช่วยให้เราเข้าใจความสัมพันธ์ระหว่างอายุกับดัชนีมวลกายได้ชัดเจนขึ้นเท่านั้น แต่ยังเป็นข้อมูลพื้นฐานสำคัญสำหรับการวางแผนดูแลสุขภาพในแต่ละช่วงวัย

โดยเฉพาะในการพิจารณาแนวทางป้องกันภาวะน้ำหนักเกินหรือโรคอ้วนที่อาจเกี่ยวข้องกับอายุ และการให้คำแนะนำด้านสุขภาพที่เหมาะสมกับแต่ละกลุ่มอายุ



การทดสอบสมมติฐาน

สมมติฐานศูนย์

Null Hypothesis (H_0)

ไม่มีความสัมพันธ์ระหว่าง X กับ Y

$$b_1 = 0$$

สมมติฐานนี้จะตั้งขึ้นมาเพื่อรอการปฏิเสธ มักจะบ่งบอกว่าไม่มีผลกระทบ ไม่มีความแตกต่าง หรือไม่มีความสัมพันธ์ที่มีนัยสำคัญทางสถิติระหว่างตัวแปรที่ศึกษา

ในการทดสอบ เราจะพยายามรวบรวมข้อมูลเพื่อดูว่ามีหลักฐานเพียงพอที่จะปฏิเสธสมมติฐานนี้หรือไม่

สมมติฐานทางเลือก

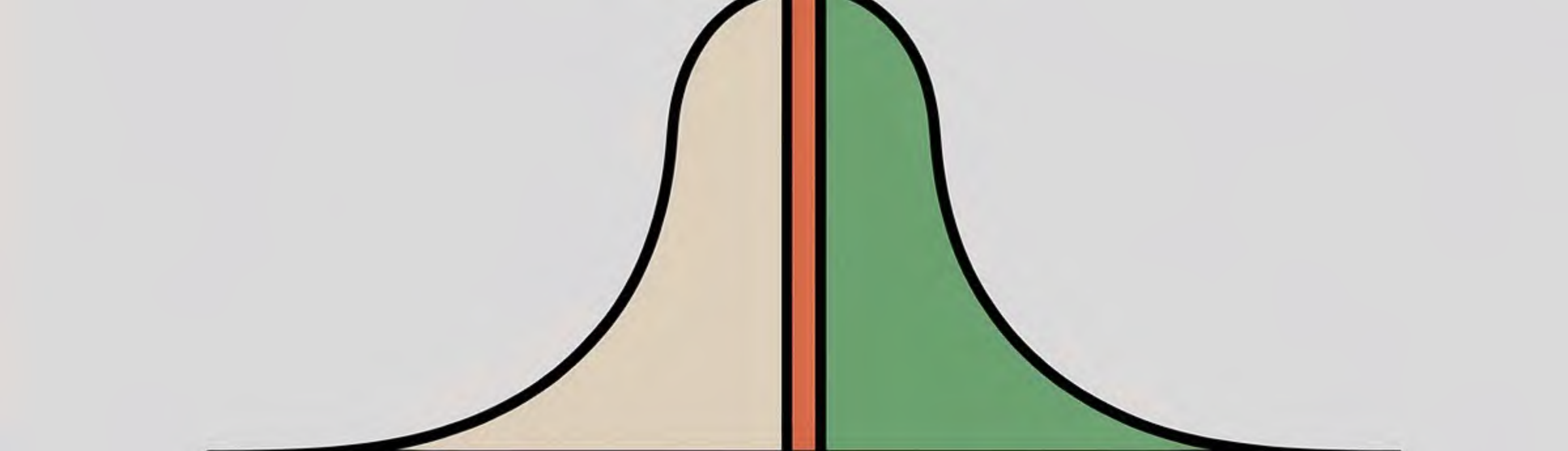
Alternative Hypothesis (H_1)

มีความสัมพันธ์ระหว่าง X กับ Y

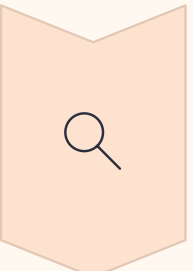
$$b_1 \neq 0$$

สมมติฐานนี้ตรงกันข้ามกับสมมติฐานศูนย์ และเป็นสิ่งที่ผู้วิจัยต้องการพิสูจน์หรือหาหลักฐานสนับสนุนว่ามีผลกระทบ ความแตกต่าง หรือความสัมพันธ์ที่มีนัยสำคัญทางสถิติ

หากหลักฐานทางสถิติมีมากพอที่จะปฏิเสธ H_0 เราก็จะยอมรับ H_1 ซึ่งหมายความว่าพบความสัมพันธ์หรือผลกระทบที่สำคัญ

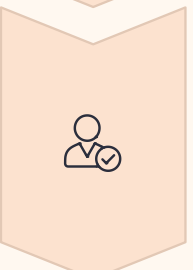


ค่า p-value คืออะไร?



ทดสอบความน่าเชื่อถือ

ใช้ทดสอบความน่าเชื่อถือของความสัมพันธ์ที่พบ หรือความแตกต่างระหว่างกลุ่มตัวอย่างต่างๆ ว่าไม่ได้เกิดขึ้นจากความบังเอิญ แต่มีนัยสำคัญทางสถิติ



เกณฑ์การตัดสินใจ

ถ้า $p\text{-value} < 0.05$ แสดงว่าความสัมพันธ์มีนัยสำคัญทางสถิติ ซึ่งหมายถึงเราจะปฏิเสธสมมติฐานศูนย์ (H_0) และยอมรับสมมติฐานทางเลือก (H_1) โดยทั่วไป ค่า 0.05 (หรือ 5%) มักถูกใช้เป็นระดับนัยสำคัญ (alpha level) มาตรฐาน

ค่า p-value ต่ำหมายถึงโอกาสที่ความสัมพันธ์ที่พบเกิดจากความบังเอิญมีน้อยมาก แสดงว่าหลักฐานที่เรามีนั้นแข็งแกร่งพอที่จะสนับสนุนข้อสรุปว่ามีผลกระทบหรือความสัมพันธ์อยู่จริงในประชากรที่เราศึกษา

P%	P%	P%	P%	P%
P%	P%	P%	P%	P%

ตารางผลลัพธ์การถดถอย

ตารางแสดงค่าสัมประสิทธิ์ (Coefficients), ค่า p-value และช่วงความเชื่อมั่น (Confidence Intervals) ซึ่งช่วยในการตีความความสำคัญทางสถิติของแต่ละตัวแปร ค่าสัมประสิทธิ์จะบอกถึงทิศทางและขนาดของความสัมพันธ์ ในขณะที่ค่า p-value จะระบุว่าความสัมพันธ์นั้นมีนัยสำคัญทางสถิติหรือไม่ โดยทั่วไป หาก p-value ต่ำกว่า 0.05 แสดงว่าตัวแปรนั้นมีอิทธิพลอย่างมีนัยสำคัญต่อตัวแปรตาม

R-squared: ความเหมาะสมของโมเดล



วัดความเหมาะสม

แสดงสัดส่วนความแปรปรวนของตัวแปรตาม (Y) ที่อธิบายได้ด้วยตัวแปรอิสระ (X) ในโมเดลถดถอย

เป็นการวัดว่าตัวแปรอิสระ (X) สามารถอธิบายความผันผวนของตัวแปรตาม (Y) ได้ดีเพียงใด ช่วยให้เราเข้าใจประสิทธิภาพของโมเดลในการทำนาย



ค่าระหว่าง 0-1

ค่า R-squared อยู่ระหว่าง 0 ถึง 1 โดยค่าใกล้ 1 หมายถึงโมเดลอธิบายข้อมูลได้ดีมากและมีความแม่นยำสูง

ในทางกลับกัน ค่าใกล้ 0 บ่งบอกว่าโมเดลไม่สามารถอธิบายความแปรปรวนของ Y ได้เลย อย่างไรก็ตาม ค่า R-squared ที่สูงมากอาจไม่ได้หมายถึงโมเดลที่ดีที่สุดเสมอไป อาจต้องพิจารณาปัจจัยอื่นร่วมด้วย เช่น การเกิด Overfitting หรือปริมาณของข้อมูล