Week 2: Regression

กิตติรักษ์ ม่วงมิ่งสุข (กิ๊)

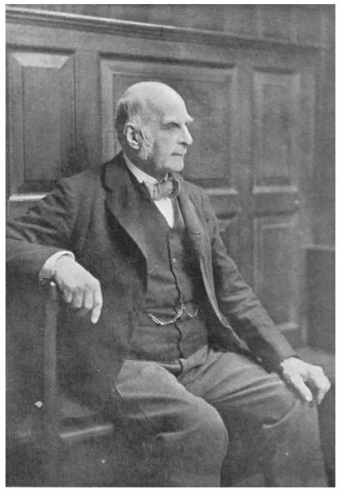
Last Modify: NOV 26, 2025

Regression

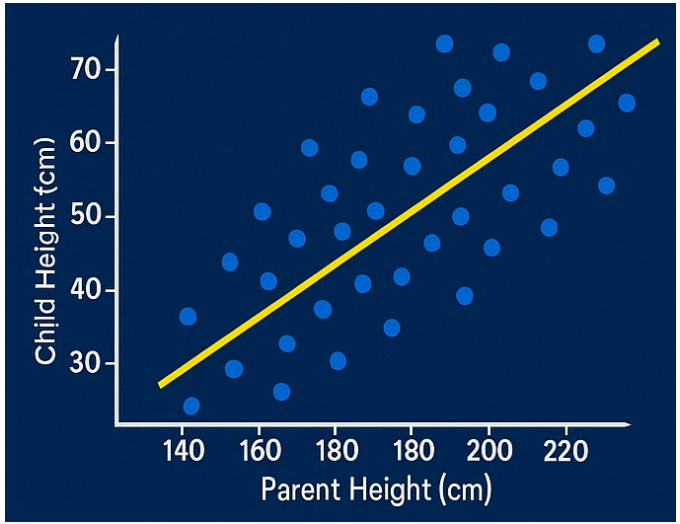
- คำว่า Regression (การถดถอย) ถูกนำมาใช้ครั้งแรกโดยนักสถิติชาวอังกฤษชื่อ **เซอร์ ฟรานซิส กาลตัน (Sir Francis Galton)** ในช่วงปลายศตวรรษที่ 19
- กาลตันศึกษา**ความสัมพันธ์ระหว่าง ส่วนสูงของพ่อแม่ กับส่วนสูงของลูก** ๆ เขาพบปรากฏการณ์ที่เขาเรียกว่า "Regression towards Mediocrity in Hereditary Stature" ซึ่งหมายถึง "**การถดถอยเข้าสู่ค่าเฉลี่ยในเรื่องส่วนสูงที่ถ่ายทอดทางพันธุกรรม**"

Galton

- Obsessed with measurement
- Tried to measure everything from the weather to female beauty
- Invented correlation and regression



รูปภาพจาก https://the4thparadigm.medium.com/milestone-0-the-origin-of-the-term-regression-4b6e9804ac6s0



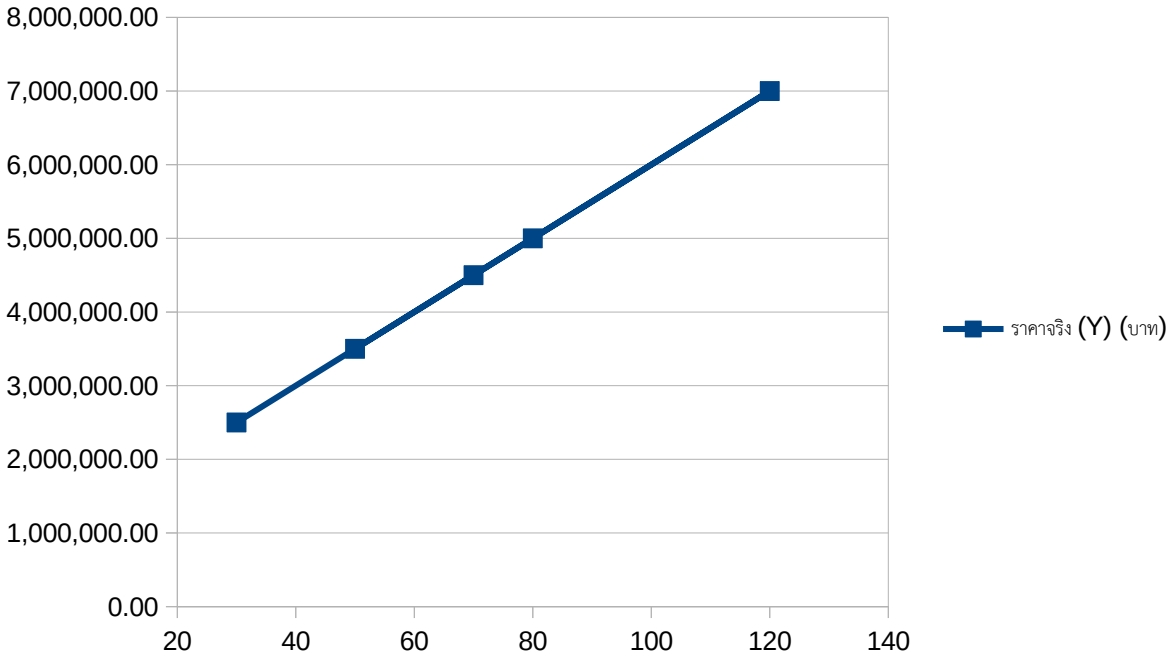
รูปภาพจาก https://www.postnetwork.co/the-regression-revolution-how-sir-francis-galtons-work-laid-the-groundwork-for-neural-networkssi/

ปรากฏการณ์ที่กาลตันค้นพบ

- **ถ้าพ่อแม่สูงกว่าค่าเฉลี่ยมาก**: **ลูก ๆ** ของพวกเขา**มีแนวโน้มที่จะ สูง เช่นกัน** แต่ส่วนสูงของลูก ๆ โดยเฉลี่ยแล้วจะ ถดถอย (Regression) หรือ เข้าใกล้ค่าเฉลี่ย (Mediocrity/Mean) ของประชากรโดยรวมมากขึ้น
- **ถ้าพ่อแม่เตี้ยกว่าค่าเฉลี่ยมาก: ลูก ๆ** ของพวกเขา**ก็มีแนวโน้มที่จะ เตี้ย** แต่ส่วนสูงของลูก ๆ โดยเฉลี่ยแล้วจะ ถดถอย (Regression) หรือ เข้าใกล้ค่าเฉลี่ย ของประชากรโดยรวมมากขึ้นเช่นกัน
- Francis Galton - ริเริ่มแนวคิด "Regression" และสร้างแบบจำลองความสัมพันธ์เชิงเส้น ใช้ Scatter Plot อธิบาย
- Karl Pearson พัฒนาสมการทางคณิตศาสตร์ Least Squares ทำให้เกิด Linear Regression

ตัวอย่าง Linear Regression: การทำนายราคาบ้านจากขนาดพื้นที่

ขนาดพื้นที่ (X) (ตร.ม.)	ราคาจริง (Y) (บาท)
50	3,500,000.00
70	4,500,000.00
120	7,000,000.00
80	5,000,000.00
30	2,500,000.00



ตัวอย่าง Linear Regression: การทำนายราคาบ้านจากขนาดพื้นที่

- สมการเส้นตรง $\boxed{Y = \beta_0 + \beta_1 X}$
- Y (ตัวแปรตาม): สิ่งที่ต้องการทำนาย (ราคาบ้าน)
- X (ตัวแปรอิสระ): ข้อมูลที่คุณใช้ในการทำนาย (ขนาดพื้นที่)
- β_1 (อ่านว่า เบต้า) (**Slope**/Weight): คือ ความชัน หรือ อัตราการเปลี่ยนแปลง (เช่น ถ้าบ้านใหญ่ขึ้น 1 ตร.ม. ราคาจะเพิ่มขึ้นเท่าไหร่)
- β_0 (**Intercept** หรือ Bias): คือ จุดตัดแกน Y หรือ ราคาเริ่มต้น เมื่อขนาดพื้นที่เป็น 0 (มักไม่สมเหตุสมผลในทางปฏิบัติ แต่จำเป็นสำหรับสมการ)

- โมเดลคำนวณได้ดังนี้

$$\boxed{\text{ราคา}=1{,}000{,}000+(50{,}000\times\text{ขนาดพื้นที่})}$$

- การทำนาย: ถ้าบ้านขนาด 100 ตร.ม.
- $= 1{,}000{,}000 + (50{,}000 \times 100)$
- $= 6{,}000{,}000$

การคำนวณ β_0 และ β_1 ใน Linear Regression

- เราจะไม่สามารถลากเส้นตรงที่ผ่านจุดข้อมูลจริงทุกจุดได้อย่างสมบูรณ์แบบ ดังนั้นโมเดลจึงต้องหา β_0 และ β_1 ที่ทำให้ **ความผิดพลาดในการทำนายน้อยที่สุด (Minimizing the Error)**
- ความผิดพลาด (Error/Residual): คือ ระยะห่างในแนวตั้งระหว่าง ราคาจริง (Y) กับราคาที่ทำนาย (Y^) $\hat{Y}$ $e_i = Y_i - \hat{Y}_i$
- ฟังก์ชันความสูญเสีย (Loss Function): ใน Linear Regression เราใช้ **Mean Squared Error (MSE)** หรือที่เรียกว่า Least Squares เพื่อวัดความผิดพลาดรวม
  - เรายกกำลังสองของความผิดพลาด (e_i^2) เพื่อให้ค่าผิดพลาดที่เป็นลบกลายเป็นบวก และให้ความผิดพลาดที่รุนแรงมีผลต่อการปรับปรุงโมเดลมากขึ้น
  - เป้าหมายคือการหา β_0 และ β_1 ที่ทำให้ค่า MSE นี้ ต่ำที่สุด

การคำนวณหา β_1

$$\beta_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{230.0}{4600}$$

$$\beta_1 = 0.05$$

	X_i	Y_i	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$
	50	3.5	$50 - 70 = -20$	$3.5 - 4.5 = -1.0$	$(-20)^2 = 400$	$(-20) \times (-1.0) = 20.0$
	70	4.5	$70 - 70 = 0$	$4.5 - 4.5 = 0.0$	$0^2 = 0$	$0 \times 0.0 = 0.0$
	120	7	$120 - 70 = 50$	$7.0 - 4.5 = 2.5$	$50^2 = 2500$	$50 \times 2.5 = 125.0$
	80	5	$80 - 70 = 10$	$5.0 - 4.5 = 0.5$	$10^2 = 100$	$10 \times 0.5 = 5.0$
	30	2.5	$30 - 70 = -40$	$2.5 - 4.5 = -2.0$	$(-40)^2 = 1600$	$(-40) \times (-2.0) = 80.0$
Avg	**70**	**4.5**		**Sum (Σ)**	**4600**	**230**

$\beta_1 = 0.05$ (ล้านบาท) หมายถึง **0.05 ล้านบาท**

การคำนวณหา β_0

ใช้สูตร: $\beta_0 = \bar{Y} - \beta_1 \bar{X}$

$$\beta_0 = 4.5 - (0.05 \times 70)$$

$$\beta_0 = 4.5 - 3.5$$

$$\beta_0 = 1.0$$

- ถ้าพิจารณาดู $B_0 =$ ราคาเฉลี่ย – (ราคาที่เพิ่มขึ้นต่อพื้นที่หนึ่งหน่วย * พื้นที่เฉลี่ย)

ทำนาย (Prediction)

- เพราะฉะนั้น จากสมการ $\boxed{Y = \beta_0 + \beta_1 X}$
- ราคาทำนาย (ล้านบาท) $= 1.0 + (0.05 \times$ ขนาดพื้นที่$)$
- เช่น บ้าน 100 ตร.ม. $= 1.0 + (0.05 \times 100) = 1.0 + 5.0 = 6.0$ ล้านบาท

แบบฝึกหัด

- สร้าง spreadsheet linear regression คำนวณราคาบ้าน ตามสมการที่ได้เรียนมา
- ทดลองใช้ฟังก์ชันสำเร็จรูป

พารามิเตอร์	สูตร	ผลลัพธ์
β1 (Slope)	=SLOPE(Y_range, X_range)	0.05
β0 (Intercept)	=INTERCEPT(Y_range, X_range)	1.0

ClusterKit

Q & A