การวิเคราะห์ความถดถอยอย่างง่าย (Simple Regression Analysis)

คือ การศึกษาความสัมพันธ์ระหว่างตัวแปรตาม (y) 1 ตัวแปร กับ ตัวแปรอิสระ (x) 1 ตัวแปร โดยที่ในช่วงแรกจะศึกษาความสัมพันธ์ระหว่าง Y กับ X ในรูปเชิงเส้น

ดังนั้น สมการเส้นถดถอย คือ $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

เมื่อ β_0 = ค่าคงที่ หรือ จุดตัดแกนตั้ง

β_1 = ค่าความชัน

ε = ส่วนคาดเคลื่อน

แต่ในความเป็นจริง เราไม่สามารถเก็บข้อมูลจากประชากรทั้งหมดได้ จึงต้องใช้การสุ่มตัวอย่างมาใช้ในการสร้างสมการถดถอย ซึ่งเป็นสมการถดถอยที่ได้จากการประมาณ ซึ่งอยู่ในรูปของ

$$\hat{y}_i = a + bx_i$$

เมื่อ $\hat{y}_i$ = ค่าประมาณการของ y

a = ค่าประมาณการของ β_0

b = ค่าประมาณการของ β_1

โดยที่ e_i = ค่าความผิดพลาด (Error)

และ $e_i = y_i - \hat{y}_i$

ดังนั้น การประมาณสมการถดถอยที่ดีควรเป็นสมการถดถอยที่เกิดค่าความผิดพลาดน้อยที่สุด ซึ่งวิธีที่นิยมใช้ คือ วิธีกำลังสองน้อยที่สุด (Ordinary Least Square หรือ OLS)

โดยที่ กำหนดให้ SSE = ผลรวมกำลังสองของค่าความผิดพลาด (Sum of Squared Error)

$$SSE = \sum_{i=1}^{n} e_i^2$$

$$= \sum_{i=1}^{n} (y - \hat{y})^2$$

เมื่อ n คือ จำนวนตัวอย่าง

เนื่องจาก $\hat{y}_i = a + bx_i$ ดังนั้น การที่ค่า y กับ $\hat{y}$ จะเกิดความแตกต่างกันน้อยที่สุด จึงขึ้นอยู่กับปัจจัย 2 อย่าง คือ a กับ b

จะได้ว่า $SSE = \sum_{i=1}^{n} e_i^2$

$$= \sum_{i=1}^{n} (y_i - \hat{y})^2$$

$$= \sum_{i=1}^{n} (y_i - a - bx_i)^2$$

หาค่า a ที่ทำให้เกิดค่า SSE ต่ำสุด

$$\frac{\partial SSE}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$= -2 \left(\sum_{i=1}^{n} y_i - b \sum_{i=1}^{n} x_i \right) = 0$$

จะได้ว่า $\sum_{i=1}^{n} y_i - na - b \sum_{i=1}^{n} x_i = 0$

$$\therefore \quad \sum_{i=1}^{n} y_i - b \sum_{i=1}^{n} x_i = na$$

$$a = \left[\sum_{i=1}^{n} y_i - b \sum_{i=1}^{n} x_i \right] / n$$

$$a = \overline{y} - b\overline{x}$$

หาค่า b ที่ทำให้เกิด SSE ต่ำสุด

โดยแทนค่า $a = \overline{y} - b\overline{x}$ ใน SSE จะได้ว่า

$$SSE = \sum_{i=1}^{n} [y_i - (\overline{y} - b\overline{x}) - bx_i]^2$$

$$= \sum_{i=1}^{n} [(y_i - \overline{y}) - b(x_i - \overline{x})]^2$$

$$\frac{\partial SSE}{\partial b} = \sum_{i=1}^{n} 2 [(y_i - \overline{y}) - b(x_i - \overline{x})] [-(x_i - \overline{x})] = 0$$

$$= \sum_{i=1}^{n} -2 [(y_i - \overline{y}) - b(x_i - \overline{x})] (x_i - \overline{x}) = 0$$

$$= -2 \sum_{i=1}^{n} [(x_i - \overline{x})(y_i - \overline{y}) - b(x_i - \overline{x})^2] = 0$$

จะได้ว่า $= \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) - b \sum_{i=1}^{n} (x_i - \overline{x})^2 = 0$

$$\therefore \quad b = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

ตัวอย่างที่ 1

จงใช้ข้อมูลต่อไปนี้ประมาณสมการบริโภค (เชิงเส้น)

ปี	**รายได้ (ล้านบาท)**	**การบริโภค (ล้านบาท)**
1990	20	7
1991	30	9
1992	33	8
1993	40	11
1994	15	5
1995	13	4
1996	26	8
1997	38	10
1998	35	9
1999	43	10

วิธีทำ กำหนดให้ x = รายได้

y = การบริโภค

x_i	y_i	$(x_i - \bar{x})$	$(y_i - \bar{y})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	x_i^2
20	7	-9.3	-1.1	86.49	10.23	400
30	9	0.7	0.9	0.49	0.63	900
33	8	3.7	-0.1	13.69	-0.37	1,089
40	11	10.7	2.9	114.69	31.03	1,600
15	5	-14.3	-3.1	204.49	44.33	225
13	4	-16.3	-0.1	265-69	66.83	169
26	8	-3.3	-0.1	10.89	0.33	676
38	10	8.7	1.9	75.69	16.53	1,444
35	9	5.7	0.9	32.49	5.13	1,225
43	10	13.7	1.9	187.69	26.03	1,849
293	81	0	0	992.10	200.07	9,557

โดยที่ $\bar{x} = 293/10 = 29.3$

$\bar{y} = 81/10 = 8.1$

$$\therefore \quad b = \frac{\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{10} (x_i - \bar{x})^2}$$

$$= 200.7 / 992.10$$

$= \quad 0.2023$

b = 0.2023 หมายความว่า ถ้าผู้บริโภคมีรายได้เปลี่ยนแปลงไป 1 ล้านบาทแล้ว เขาจะเปลี่ยนแปลงรายจ่ายเพื่อการบริโภคไป 0.2023 ล้านบาทในทิศทางเดียวกัน

หาค่า $a = \bar{y} - b\bar{x}$

$= 8.1 - (0.2023)(29.3) = 2.173$

a = 2.173 นั่นคือ รายจ่ายขั้นต่ำเพื่อการบริโภคของผู้บริโภคเท่ากับ 2.173 ล้านบาท (รายจ่ายเพื่อการบริโภคขณะที่รายได้ = 0)

ดังนั้น สมการการบริโภค คือ $\hat{y} = 2.173 + 0.2023\, x_i$

ค่าสถิติที่ใช้ในการวิเคราะห์ความถดถอย

เนื่องจากวัตถุประสงค์ที่สำคัญของการสร้างสมการเส้นถดถอยขึ้นมาก็เพื่อประมาณหรือพยากรณ์ตัวแปรที่ต้องการศึกษา ดังนั้น เพื่อความมั่นใจว่าสมการเส้นถดถอยที่สร้างขึ้นมากนั้น สามารถนำไปประมาณหรือพยากรณ์ตัวแปรที่เราต้องการศึกษาได้จริงจึงต้องมีการทดสอบโดยอาศัยค่าสถิติต่างๆ ได้แก่

สัมประสิทธิ์การตัดสินใจ (Coeficient of Determination : R^2)

R^2 = ค่าที่อธิบายว่าความผันแปรทั้งหมดของตัวแปรตามมีสาเหตุมาจากตัวแปรอิสระกี่ %

โดยที่ $0 \leq R^2 \leq 1$

และ e_i = ค่าความผิดพลาด

โดยที่ $e_i = y - \hat{y}$ สมการ 1

ดังนั้น Σe_i^2 คือ ผลรวมกำลังสองของค่าความผิดพลาด (Error Sum of Square : SSE) หรือความแปรปรวนที่ไม่อาจอธิบายได้ (Unexplained Variation)

$SSE = \Sigma e_i^2 = \Sigma(y_i - \hat{y})^2 = \Sigma(y_i - a - bx_i)^2$ สมการ 2

แทนค่า $a = \bar{y} - b\bar{x}$ ไปในสมการ 2

$$\Sigma e_i^2 = \Sigma(y_i - \bar{y} + b\bar{x} - bx_i)^2$$

$$= \Sigma[(y_i - \bar{y}) - b(x - \bar{x})]^2$$

$$= \Sigma[(y_i - \bar{y})^2 - 2b(y_i - \bar{y})(x - \bar{x}) + b^2 (x - \bar{x})^2]$$

$$= \Sigma(y_i - \bar{y})^2 - 2b\,\Sigma(x_i - \bar{x})(y_i - \bar{y}) + b^2\,\Sigma(x_i - \bar{x})^2$$ สมการ 3

เนื่องจาก $b = \Sigma(x_i - \bar{x})(y_i - \bar{y}) / \Sigma(x_i - \bar{x})^2$

ดังนั้น $b\,\Sigma(x_i - x)^2 = \Sigma(x_i - \bar{x})(y_i - \bar{y})$ สมการ 4

และนำค่าไปแทนในเทอมสุดท้ายในสมการ 3 จะได้ว่า

$$\Sigma e_i^2 = \Sigma(y_i-\overline{y})^2 - 2b\,\Sigma(x_i-\overline{x})(y_i-\overline{y}) + b\,\Sigma(x_i-\overline{x})(y_i-\overline{y})$$

$$\Sigma e_i^2 = \Sigma(y_i-\overline{y})^2 - b\,\Sigma(x_i-\overline{x})(y_i-\overline{y}) \quad \text{สมการ 5}$$

จากสมการที่ 5 ถ้า $b = 0$ หมายความว่า y กับ x ไม่มีความสัมพันธ์กันเลยแสดงว่าความแปรปรวนทั้งหมด มีสาเหตุมาจากปัจจัยอื่นที่ไม่สามารถอธิบายได้

แทนค่า $b = 0$ ในสมการที่ 5 จะได้ว่า

$$\Sigma e_i^2 = \Sigma(y_i-\overline{y})^2 \quad \text{สมการ 6}$$

ดังนั้น $\Sigma(y_i - y)^2$ คือ ความแปรปรวนทั้งหมด หรือ Total Sum of Squares (SST)

ในทำนองเดียวกัน ถ้าค่า $\Sigma e_i = 0$ หมายความว่า ความแปรปรวนทั้งหมดมีสาเหตุมาจากความผันแปรที่สามารถอธิบายได้ โดยตัวแปร x

แทนค่า $\Sigma e_i^2 = 0$ ไปในสมการ 5

จะได้ว่า $0 = \Sigma(y_i-\overline{y})^2 - b\,\Sigma(x_i-\overline{x})(y_i-\overline{y})$

$$\therefore\ b\Sigma(x_i-\overline{x})(y_i-\overline{y}) = \Sigma(y_i-\overline{y})^2 \quad \text{สมการ 7}$$

แสดงว่า $b\Sigma(x_i-\overline{x})(y_i-\overline{y})$ คือ ความผันแปรที่สามารถอธิบายได้

จากสมการที่ 4 $b\Sigma(x_i - x)^2 = \Sigma(x_i-\overline{x})(y_i-\overline{y})$ จะได้ว่า

$$\begin{aligned}
b\Sigma(x_i-\overline{x})(y_i-\overline{y}) &= b\,(b\Sigma(x_i-\overline{x})^2) \\
&= b^2\,\Sigma(x_i-\overline{x})^2 \\
&= \Sigma\, b^2\,(x_i-\overline{x})^2 \\
&= \Sigma\,(b\,(x_i-\overline{x}))^2 \\
&= \Sigma\,(bx_i-b\overline{x})^2 \\
&= \Sigma\,(bx_i-b\overline{x}+\overline{y}-\overline{y})^2 \\
&= \Sigma\,(bx_i+a-\overline{y})^2 \\
&= \Sigma\,(\hat{y}-\overline{y})^2 \quad \text{สมการ 8}
\end{aligned}$$

ดังนั้น จึงสรุปได้ว่า $b\Sigma(x_i-\overline{x})(y_i-\overline{y}) = \Sigma(\hat{y}-\overline{y})^2$ คือ ความผันแปรที่สามารถอธิบายได้ (Explian Variation) หรือ Regression Sum of Squares (SSR)

ความผันแปรทั้งหมด = ความผันแปรที่สามารถอธิบายได้ + ความผันแปรที่ไม่สามารถอธิบายได้

$$\text{SST} = \text{SSR} + \text{SSE}$$

$$\Sigma(y_i-\overline{y})^2 = \Sigma(\hat{y}-\overline{y})^2 + \Sigma(y_i-\hat{y})^2 \quad \text{สมการ 9}$$

และ $$R^2 = \frac{\text{ความผันแปรที่สามารถอธิบายได้}}{\text{ความผันแปรทั้งหมด}} = \frac{\text{SSR}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$$

$$R^2 = \frac{\Sigma(\hat{y}-\overline{y})^2}{\Sigma(y_i-\overline{y})^2} = \frac{b^2\Sigma(x_i-\overline{x})^2}{\Sigma(y_i-\overline{y})^2} = \frac{b\Sigma(x_i-\overline{x})\ (y_i-\overline{y})}{\Sigma(y_i-\overline{y})^2}$$

ตัวอย่างที่ 2

จากตัวอย่าง 1 จงคำนวณหาค่า R^2 พร้อมทั้งอธิบายค่าที่คำนวณได้

จากตัวอย่างที่ 1 $\hat{y} = 2.173 + 0.2023\,x_i$

x_i	y_i	$\hat{y} = 2.173 + 0.2023\,x_i$	$y_i - \overline{y}$	$(y_i - \overline{y})^2$	$\hat{y} - \overline{y}$	$(\hat{y} - \overline{y})^2$
20	7	6.219	-1.1	1.21	-1.881	3.538
30	9	8.242	0.9	0.81	0.142	0.020
33	8	8.849	-0.1	0.01	0.749	0.561
40	11	10.265	2.9	8.41	2.165	4.687
15	5	5.208	-3.1	9.61	-2.893	8.367
13	4	4.803	-4.1	16.81	-3.297	10.871
26	8	7.433	-0.1	0.01	-0.667	0.445
38	10	9.860	1.9	3.61	1.760	3.099
35	9	9.254	0.9	0.81	1.154	1.331
43	10	10.872	1.9	3.61	2.772	7.683
293	81	81.004	0	44.9	0.004	40.602

$$R^2 = \frac{\Sigma(\hat{y}-\overline{y})^2}{\Sigma(y_i-\overline{y})^2} = \frac{SSR}{SST}$$

$$= 40.602 / 44.9 = 0.9043$$

$R^2 = 0.9043$ หมายความว่า ความผันแปรทั้งหมดของการบริโภคมีสาเหตุมาจากรายได้ 90.43%

สัมประสิทธิ์การตัดสินใจที่ปรับค่าแล้ว (Adjusted Coefficient of Determination : $\overline{R}^2$)

ในกรณีที่ข้อมูลที่นำมาศึกษามีจำนวนน้อย อาจทำให้ค่า R^2 ที่คำนวณได้มีค่าสูงเกินไป ดังนั้นจึงนำค่า R^2 มาปรับเสียใหม่ เรียกว่า ค่าสัมประสิทธิ์การตัดสินใจที่ปรับค่าแล้ว หรือ $\overline{R}^2$

$$\overline{R}^2 = 1 - \frac{\Sigma(y-\hat{y}_i)^2 / (n-k-1)}{\Sigma(y_i-\overline{y})^2\ / (n-1)}$$ เมื่อ k = ตัวแปรอิสระ

$$= 1 - \frac{SSE}{SST} \cdot \frac{(n-1)}{(n-k-1)} = 1 - \left[\frac{(1-R^2) \times (n-1)}{(n-k-1)}\right]$$

ในกรณี k = 1 $\bar{R}^2 = 1 - \dfrac{\Sigma(y - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2} \cdot \dfrac{(n-1)}{(n-2)}$

$$= 1 - \frac{SSE}{SST} \cdot \frac{(n-1)}{(n-2)} = 1 - \left[\frac{(1-R^2) \times (n-1)}{(n-2)}\right]$$

หมายเหตุ : ในทางปฏิบัติ มักนำเสนอทั้งค่า R^2 และ $\bar{R}^2$ ไปพร้อมๆ กัน

ตัวอย่างที่ 3

จากตัวอย่างที่ 5.1 และ 5.2 จงคำนวณหาค่า $\bar{R}^2$ พร้อมทั้งอธิบายค่าที่ได้

จากตัวอย่างที่ 5.2 $R^2 = 0.9043$, n = 10 , k = 1

$$\bar{R}^2 = 1 - \left[\frac{(1-R^2) \times (n-1)}{(n-k-1)}\right]$$

$$= 1 - \left[\frac{(0.0957 \times 9)}{8}\right]$$

$$= 0.8923$$

$\bar{R}^2 = 0.8923$ หมายความว่า เมื่อได้มีการปรับค่าสัมประสิทธิ์การตัดสินใจแล้ว ความผันแปรทั้งหมดของการบริโภคมีสาเหตุมาจากรายได้ 89.23%

ความคลาดเคลื่อนมาตรฐานของการประมาณ (Standard Error of Estimate : SEE)

SEE คือ ค่าที่แสดงว่า ค่าของตัวแปรที่ได้จากการประมาณ ($\hat{y}$) กับค่าตัวแปรตามที่เกิดขึ้นจริงมีความแตกต่างกันมากน้อยเพียงใด ซึ่งถ้าค่า SEE ยิ่งมีค่าน้อยแสดงว่า ค่า y กับ $\hat{y}$ มีค่าใกล้เคียงกันมาก นั่นคือ ความน่าเชื่อถือของสมการถดถอยก็ยิ่งมากตามไปด้วย

$$SEE = \sqrt{\frac{\Sigma(y - \hat{y})^2}{(n-k-1)}} = \sqrt{\frac{SSE}{(n-k-1)}}$$

ตัวอย่างที่ 4

จากตัวอย่างที่ 5.1 และ 5.2 จงคำนวณหาค่า SEE พร้อมทั้งอธิบายค่าที่ได้

จากตัวอย่างที่ 5.2 $\Sigma(\hat{y} - \bar{y})^2 = SSR = 40.602$

$\Sigma(y_i - \bar{y})^2 = SST = 44.9$

ดังนั้น $\Sigma(y_i - \hat{y})^2 = SSE = SST - SSR$

$= 44.9 - 40.602 = 4.298$

$$SEE = \sqrt{\frac{SSE}{(n-k-1)}} = \sqrt{\frac{4.298}{8}} = 0.7330$$

SEE = 0.7330 หมายความว่า การบริโภคที่เกิดขึ้นจริง กับ การบริโภคที่ได้จากการประมาณ มีความแตกต่างกันประมาณ 0.7330 ล้านบาท

การทดสอบสมมติฐาน (Hypothesis Test)

สำหรับการทดสอบสมมติฐาน สมการถดถอยเชิงเส้นอย่างง่าย จะมีการทดสอบสมมติฐานที่สำคัญอยู่ 2 ส่วน คือ การทดสอบสมมติฐานค่าคงที่กับค่าสัมประสิทธิ์ของตัวแปรอิสระ ซึ่งมีขั้นตอนดังนี้

การทดสอบสมมติฐานค่าคงที่

ขั้นที่ 1 : ตั้งสมมติฐาน

H_0 : $\beta_0 = 0$ (ค่าคงที่ $= 0$)

H_1 : $\beta_0 \neq 0$ (ค่าคงที่ $\neq 0$)

ขั้นที่ 2 : หาค่า t ที่ได้จากการเปิดตารางที่ $t_{(\propto/2\,,\, n-k-1)}$

เมื่อ $\propto$ คือ ระดับนัยสำคัญ

n คือ จำนวนตัวอย่าง

k คือ จำนวนตัวแปรอิสระ

ขั้นที่ 3 : หาค่า t ที่ได้จากการคำนวณ

เมื่อ $t = a/s_a$

a คือ ค่าประมาณของ β_0

s_a คือ ส่วนเบี่ยงเบนมาตรฐานของ a

โดยที่ $s_a = s\sqrt{\Sigma x^2 / n\,\Sigma(x-\bar{x})^2}$

หรือ $s_a = s\sqrt{\Sigma x^2 / n\,(\Sigma x^2 - n\bar{x}^2)}$

เมื่อ s คือ ความคาดเคลื่อนมาตรฐานของการประมาณ หรือ SEE

ขั้นที่ 4 : เปรียบค่า t ที่ได้จากตาราง กับ ค่า t ที่ได้จากการคำนวณ

ปฏิเสธ H_0 เมื่อ $t > t_{(\propto/2\,,\, n-k-1)}$

หรือ $-t < -t_{(\propto/2\,,\, n-k-1)}$

ขั้นที่ 5 : สรุปคำตอบ

การทดสอบสมมติฐานค่าสัมประสิทธิ์ของตัวแปรอิสระ

ขั้นที่ 1 : ตั้งสมมติฐาน

กรณีที่ค่า b > 0

H_0 : $\beta_1 \leq 0$(x กับ y ไม่มีความสัมพันธ์ในทิศทางเดียวกัน)

H_1 : $\beta_1 > 0$(x กับ y มีความสัมพันธ์ในทิศทางเดียวกัน)

กรณีที่ค่า b < 0

H_0 : $\beta_1 \geq 0$(x กับ y ไม่มีความสัมพันธ์ในทิศทางตรงกันข้ามกัน)

H_1 : $\beta_1 < 0$(x กับ y มีความสัมพันธ์ในทิศทางตรงข้ามกัน)

ขั้นที่ 2 : หาค่า t ที่ได้จากการเปิดตาราง ที่ $t_{(\propto , n-k-1)}$

ขั้นที่ 3 : หาค่า t ที่ได้จากการคำนวณ

เมื่อ $t = b/s_b$

b คือ ค่าประมาณของ β_1

s_b คือ ส่วนเบี่ยงเบนมาตรฐานของ b

โดยที่ $s_b = s / \sqrt{\Sigma (x-\overline{x})^2}$

หรือ $s_b = s / \sqrt{\Sigma x^2 - n\overline{x}^2}$

ขั้นที่ 4 : เปรียบเทียบค่า t ที่ได้จากตาราง กับ ค่า t ที่ได้จากการคำนวณ

กรณีที่ b > 0 ปฏิเสธ H_0 เมื่อ $t > t_{(\propto , n-k-1)}$

กรณีที่ b < 0 ปฏิเสธ H_0 เมื่อ $-t < -t_{(\propto , n-k-1)}$

ขั้นที่ 5 : สรุปคำตอบ

ตัวอย่างที่ 5

จากตัวอย่างที่ 5.1–5.4 จงทดสอบสมมติฐานค่าคงที่และค่าสัมประสิทธิ์ของตัวแปรอิสระ ที่ระดับนัยสำคัญ 0.05

1. ทดสอบสมมติฐานค่าคงที่

1.1) H_0 : $\beta_0 = 0$

H_1 : $\beta_0 \neq 0$

1.2) เปิดตารางที่ $t_{(\propto/2 , n-k-1)} = t_{(0.025 , 8)} = 2.306$

1.3) คำนวณหาค่า t

$t = a / s_a$

โดยที่ $a = 2.173$

$s_a = s \sqrt{\Sigma x^2 / n \Sigma(x-\overline{x})^2}$

เมื่อ $\Sigma x^2 = 9,557$

$n = 10$

$\Sigma(x-\overline{x})^2 = 992.10$

s หรือ SEE = 0.7330

$s_a = (0.7330)\sqrt{9{,}557 / (10 \times 992.10)}$

$= 0.7194$

$t = 2.173 / 0.7194 = 3.02$

1.4) เนื่องจากค่า t ที่ได้จากการคำนวณ = 3.02 มากกว่า $t_{(0.025\,,\,8)}$ = 2.306 ดังนั้น จึงปฏิเสธ H_0

1.5) สรุปคำตอบ : ค่าคงที่ไม่เท่ากับ 0 ณ ระดับนัยสำคัญ 0.05

2. ทดสอบสมมติฐานค่าสัมประสิทธิ์ของตัวแปรอิสระ

2.1) $H_0 : \beta_1 \leq 0$

$H_1 : \beta_1 > 0$

2.2) ค่า t ที่ได้จากการเปิดตาราง = $t_{(0.05\,,\,8)}$ = 1.860

2.3) ค่า t ที่ได้จากการคำนวณ = b / s_b

เมื่อ $s_b = b / s_b$

โดยที่ $b = 0.2023$

$s_b = s / \sqrt{\Sigma(x - \bar{x})^2}$

$= 0.7330 / \sqrt{992.10}$

$= 0.023$

$t = 0.2023 / 0.023 = 8.795$

2.4) เนื่องจากค่า t ที่ได้จากการคำนวณ = 8.795 มากกว่าค่า $t_{(0.05\,,\,8)}$ = 1.860 ดังนั้นจึงปฏิเสธ H_0

2.5) สรุปคำตอบ รายได้กับการบริโภคมีความสัมพันธ์ในทิศทางเดียวกัน ณ ระดับนัยสำคัญ 0.05

ตัวอย่างที่ 6

ประธานหลักสูตร MBA ของมหาวิทยาลัยแห่งหนึ่งต้องการทราบว่าเกรดเฉลี่ยสะสม (GPA) ของผู้จบการศึกษามีผลต่อเงินเดือนเมื่อเริ่มเข้าทำงานหรือไม่ จึงได้เก็บข้อมูลจากกลุ่มตัวอย่างของผู้ที่สำเร็จการศึกษา จำนวน 7 คน ดังนี้

คนที่	GPA	เงินเดือนเมื่อเริ่มทำงาน (พันบาท)
1	3.26	28.20
2	2.60	24.80
3	3.35	27.90
4	2.86	25.30
5	3.82	30.30
6	2.21	23.00
7	3.47	29.40

ก) จงหาสมการถดถอยเชิงเส้นที่แสดงความสัมพันธ์ระหว่าง GPA กับ เงินเดือนเมื่อเริ่มทำงาน

วิธีทำ กำหนดให้ x คือ GPA

y คือ เงินเดือนเมื่อเริ่มทำงาน

X_i	y_i	$(x_i - \overline{x})$	$(y_i - \overline{y})$	$(x_i - \overline{x})^2$	$(x_i - \overline{x})(y_i - \overline{y})$	x_i^2
3.26	28.2	0.1786	-1.2	0.0319	0.2168	10.6276
2.60	24.8	-0.4814	-2.2	0.2318	1.0523	6.76
3.35	27.9	0.2688	0.9	0.0721	0.2456	11.2225
2.86	25.3	-0.2214	-1.7	0.049	0.3733	8.1796
3.82	30.3	0.7386	3.3	0.5455	2.4478	14.5924
2.21	23.0	-0.8714	-4.0	0.7594	3.433	4.8841
3.47	29.4	0.3866	2.4	0.1510	0.9381	12.0409
21.57	188.9	0	0	1.8407	8.7471	68.3071

$\overline{x} = 3.08143$, $\overline{y} = 26.9857$

หาค่า $b = \Sigma(x_i - \overline{x})(y_i - \overline{y}) / \Sigma (x_i - \overline{x})^2$

$= 8.7471/1.8407$

$= 4.7521$

$a = \overline{y} - b\overline{x}$

$= 26.9857 - 4.7521(3.08143)$

$= 12.3424$

ดังนั้น สมการถดถอยที่ได้ คือ $\hat{y} = 12.3424 - 4.7521\, x_i$

ข) จงหาค่า $\overline{R}^2$, R^2 และ SEE พร้อมทั้งอธิบายค่าที่ได้

วิธีทำ

x_i	y_i	$\hat{y} = 12.3424 - 4.7521\, x_i$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$\hat{y} - \bar{y}$	$(\hat{y} - \bar{y})^2$
3.26	28.2	27.8135	1.2	1.475	0.846	0.7154
2.60	24.8	24.6952	-2.2	4.777	-2.291	5.2469
3.35	27.9	28.2592	0.9	0.836	1.274	1.6218
2.86	25.3	25.9307	-1.7	2.842	-1.055	1.130
3.82	30.3	30.4926	3.3	10.985	3.507	12.2986
2.21	23.0	22.8419	-4.0	15.886	-4.144	17.1709
3.47	29.4	28.8924	2.4	5.829	1.844	3.3994
21.57	188.9	188.8806	0	42.629	-0.019	41.566

SSR = 41.566 , SSE = 1.063 , SST = 42.629

R^2 = SSR/SST = 41.566/42.629 = 0.9751

R^2 = 0.9751 หมายความว่า ความผันแปรทั้งหมดของเงินเดือนเมื่อเริ่มทำงานมีสาเหตุมาจาก GPA 97.51%

$$\bar{R}^2 = 1 - \left[\frac{(1 - R^2) \times (n - 1)}{(n - k - 1)}\right]$$

$= 0.97012$

$\bar{R}^2$ = 0.97012 หมายความว่าว่า เมื่อได้มีการปรับค่าสัมประสิทธิ์การตัดสินใจแล้ว ความผันแปรทั้งหมดของเงินเดือนมีสาเหตุมาจาก GPA 97.012%

$$SEE = \sqrt{\frac{SSE}{(n - k - 1)}} = 0.461086$$

SEE = 0.4611 หมายความว่า เงินเดือนที่เกิดขึ้นจริง กับ เงินเดือนที่ได้จากการประมาณ มีความแตกต่างกันประมาณ 0.4611 พันบาท

ค) จงทดสอบสมมติฐานค่าคงที่ของสมการที่ถดถอยที่ได้ โดยกำหนดระดับนัยสำคัญ = 0.05

วิธีทำ

1) ทดสอบสมมติฐานค่าคงที่

$H_0 : \beta_0 = 0$

$H_1 : \beta_0 \neq 0$

2) t ที่ได้จากเปิดตาราง $= t_{(0.025\,,\,5)} = 2.571$

3) หาค่า t ที่ได้จากการคำนวณ

$t = a / s_a$

โดยที่ $a = 12.3429$

$s_a = s\sqrt{\Sigma x^2 / n\,\Sigma(x-\bar{x})^2}$

$s_a = (0.4611)\sqrt{68.3017 / (7 \times 1.8407)}$

$= 1.0616$

$t = 12.3429 / 1.0616 = 11.627$

4) เปรียบเทียบค่า t ที่ได้จากการคำนวณ กับ t ที่เปิดตาราง

เนื่องจากค่า t ที่ได้จากการคำนวณ = 11.627 มากกว่า $t_{(0.025\ ,\ 5)} = 2.571$ ดังนั้น จึงปฏิเสธ H_0

5) สรุปคำตอบ ค่าคงที่ไม่เท่ากับ 0 ณ ระดับนัยสำคัญ 0.05

ง) จงทดสอบสมมติฐานค่าสัมประสิทธิ์ของตัวแปร x ในสมการที่ถดถอยที่ได้ โดยกำหนดระดับนัยสำคัญ = 0.05

วิธีทำ

1) $H_0 : \beta_1 \leq 0$

$H_1 : \beta_1 > 0$

2) ค่า t ที่ได้จากการเปิดตาราง = $t_{(0.05\ ,\ 5)} = 2.015$

3) ค่า t ที่ได้จากการคำนวณ = b / s_b

เมื่อ $s_b = b / s_b$

โดยที่ $b = 4.7521$

$s_b = s / \sqrt{\Sigma(x-\bar{x})^2}$

$= 0.4611 / \sqrt{1.8407} = 0.3399$

$t = 4.7521 / 0.3399 = 13.98$

4) เนื่องจากค่า t ที่ได้จากการคำนวณ = 13.98 มากกว่าค่า $t_{(0.05\ ,\ 5)} = 2.015$ ดังนั้นจึงปฏิเสธ H_0

5) สรุปคำตอบ GPA กับเงินเดือนเมื่อเริ่มทำงาน มีความสัมพันธ์ในทิศทางเดียวกัน ณ ระดับนัยสำคัญ 0.05