การวิเคราะห์การถดถอยพหุคูณ(Multiple regression analysis)

ประโยชน์ในเชิงสถิติ: ใช้ในการหาความสัมพันธ์ระหว่างตัวแปรอิสระ(independent variable) และตัวแปรตาม(dependent variable)

ประโยชน์ในการประยุกต์ใช้งาน : ใช้ในการพยากรณ์ค่าของตัวแปรตาม จากค่าตัวแปรอิสระที่เป็นอยู่หรือคาดว่าจะเป็น

รูปแบบ: $Y= a_O + a_1 X_1 + a_2 X_2 + a_3 X_3 +\ldots$

โดยที่ Y = ตัวแปรตาม(dependent variable)

a_0 = ค่าคงที่(constant)บางครั้งเรียกว่าy-intercept

a_i = ค่าสัมประสิทธิ์นำหน้าค่าตัวแปรอิสระแต่ละตัว(regression coefficient);i= 1.2,3,…

ตัวอย่างที่ 1:ผู้ประกอบการฟาร์มเลี้ยงสุกรรายหนึ่งต้องการวิเคราะห์หาความสัมพันธ์ระหว่างน้ำหนักของสุกรที่เพิ่มขึ้นจากการทดลองให้อาหารเสริมชนิดหนึ่งเป็นเวลาหนึ่งสัปดาห์กับน้ำหนักและอายุของลูกสุกรเมื่อแรกเริ่มรับเข้ามาเลี้ยง จากการเก็บข้อมูลจากลูกสุกรจำนวน 8 ตัว ได้ข้อมูลดังนี้

ลูกสุกรตัวที่	น้ำหนักเมื่อแรกเริ่มรับเข้าเลี้ยงเป็นปอนด์ (X_1)	อายุเมื่อเริ่มรับเข้าเลี้ยงเป็นปอนด์(X_2)	น้ำหนักที่เพิ่มขึ้นเป็นปอนด์(Y)
1	39	8	7
2	52	6	6
3	48	7	7
4	46	12	10
5	61	9	9
6	34	6	4
7	25	10	3
8	55	4	4

จากการใช้โปรแกรมสำเร็จรูปได้ผลดังนี้

$Y = -6.3658 + 0.1553 X_1 + 0.7262 X_2$

(0.0433) (0.1999)

Regression

[DataSet0] F:\ex1-multiple regression.sav

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	อายุเมื่อเริ่มรับเข้าเลี้ยง, น้ำหนักเมื่อแรกเริ่มรับเข้าเลี้ยง[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: น้ำหนักที่เพิ่มขึ้น

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.897[a]	.805	.727	1.30168	1.574

a. Predictors: (Constant), อายุเมื่อเริ่มรับเข้าเลี้ยง, น้ำหนักเมื่อแรกเริ่มรับเข้าเลี้ยง

b. Dependent Variable: น้ำหนักที่เพิ่มขึ้น

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	35.028	2	17.514	10.337	.017[a]
	Residual	8.472	5	1.694		
	Total	43.500	7			

a. Predictors: (Constant), อายุเมื่อเริ่มรับเข้าเลี้ยง, น้ำหนักเมื่อแรกเริ่มรับเข้าเลี้ยง

b. Dependent Variable: น้ำหนักที่เพิ่มขึ้น

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-6.366	2.825		-2.254	.074
	น้ำหนักเมื่อแรกเริ่มรับเข้าเลี้ยง	.155	.043	.734	3.590	.016
	อายุเมื่อเริ่มรับเข้าเลี้ยง	.726	.200	.743	3.632	.015

a. Dependent Variable: น้ำหนักที่เพิ่มขึ้น

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	3.2710	9.6423	6.2500	2.23697	8
Residual	-1.77814	1.50018	.00000	1.10012	8
Std. Predicted Value	-1.332	1.516	.000	1.000	8
Std. Residual	-1.366	1.152	.000	.845	8

a. Dependent Variable: น้ำหนักที่เพิ่มขึ้น

ตัวอย่างที่ 2:บริษัทผู้ผลิตเหล็กรายหนึ่งต้องการวิเคราะห์หาความสัมพันธ์ระหว่างจำนวนเหล็กที่เขาขายได้(หน่วยแสนตัน) กับอัตราเงินเฟ้อ(ร้อยละ) ผลต่างในราคาของเหล็กนำเข้าที่ต่ำกว่าราคาขายของบริษัท(บาทต่อตัน) และจำนวนรถยนต์ที่คาดว่าจะผลิตได้ในประเทศ(หน่วยแสนคัน) จากการเก็บข้อมูล7 ปีติดต่อกันได้ข้อมูลดังนี้

ปีที่	จำนวนของเหล็กที่ขายได้(Y)	อัตราเงินเฟ้อ(X_1)	ผลต่างของราคาเหล็กนำเข้า(X_2)	จำนวนรถยนต์ที่ผลิตแสนคัน(X_3)
1	3.2	4.7	3.10	6.1
2	2.1	5.1	5.00	5.0
3	3.0	8.5	2.20	5.7
4	3.7	11.7	4.50	7.2
5	3.3	16.5	4.35	6.8
6	2.7	14.0	2.60	6.3
7	2.5	12.0	3.05	6.0

จากการใช้โปรแกรมสำเร็จรูปได้ผลดังนี้

$$Y = -1.7937 \; -0.0602\ X_1 \; - 0.0542\ X_2 \; + 0.8994\ X_3$$
$$(\ 0.0335) \qquad (0.1041) \qquad (0.2083)$$

ดัชนีชี้วัดคุณค่าของผลที่คำนวณได้ : ดูได้จาก R^2 (Coefficient of multiple determination) หรือ R (coefficient of multiple correlation)

ความหมายของ R^2 บ่งบอกสัดส่วนของความผันแปรในค่าของตัวแปรตาม ที่อธิบายได้โดยความผันผวนในค่าของตัวแปรอิสระ

ถ้า R^2 มีค่าเข้าใกล้หนึ่ง แสดงว่าค่าของตัวแปรตามเคลื่อนไหวตามค่าของตัวแปรต้นเกือบจะร้อยเปอร์เซ็นต์(เป็นร้อยเปอร์เซ็นต์ในกรณี $R^2 = 1$)

ถ้า R^2 มีค่าเข้าใกล้ศูนย์ แสดงว่าค่าของตัวแปรอิสระไม่ได้มีส่วนในการกำหนดค่าของตัวแปรตามมากนัก

ข้อควรระวังในการใช้เทคนิคนี้: ค่าของตัวแปรอิสระจะต้องไม่มีลักษณะที่เป็นautocorrelated โดยเฉพาะเมื่อตัวแปรอิสระได้มาจากการเก็บข้อมูลที่มีลักษณะเป็นข้อมูลทางtime-series

จะตรวจสอบได้อย่างไร: พิจารณาจากDurbin-Watson statistic

ข้อควรระวังในการใช้เทคนิคนี้: จะต้องไม่มีกรณีที่ความผันผวนของerror termมีลักษณะเป็นheteroscedastic(varianceของerror term ไม่คงที่ตามขนาดของตัวแปรอิสระ)

จะตรวจสอบได้อย่างไร:พิจารณาจากการกระจายของerror term ซึ่งเป็นผลต่างระหว่างค่าactual y กับค่าpredicted Y(Yที่ได้จากการพยากรณ์ของแบบจำลอง)

ข้อควรระวังในการใช้เทคนิคนี้: จะต้องไม่มีกรณีของmulticollinearity หรือค่าของตัวแปรอิสระจะต้องไม่ขึ้นต่อกันและกัน ตัวอย่างถ้าให้ Yเป็นน้ำหนักอาหารที่บริโภค X_1 เป็นน้ำหนัก X_2 เป็นส่วนสูง ย่อมเป็นที่แน่นอนว่า ผู้วิเคราะห์ไม่สามารถหลีกเลี่ยงปัญหาของmulticollinearity ได้ เพราะน้ำหนักกับส่วนสูงย่อมมีความสัมพันธ์กันไม่มากก็น้อย

จะตรวจสอบได้อย่างไร: พิจารณาจากค่าสัมประสิทธิ์ความสัมพันธ์(coefficient of determination หรือ coefficient of correlation) ที่ได้จากค่าของตัวแปรอิสระด้วยกัน

การต่อยอดเทคนิคนี้: ใช้ในกรณีของdummy variable หรือกรณีที่ค่าของตัวแปรอิสระบางตัวไม่ได้มีค่าต่อเนื่อง แต่สามารถมีค่าได้จำกัดเช่น 0 ,1 เป็นต้น

การทดสอบสมมติฐานของสมการทั้งเส้น:

H_O : $\beta_1 = \beta_2 = \beta_3 = \ldots\ldots = 0$ (สมการที่คำนวณได้ไม่มีคุณค่าในการพยากรณ์)

H_A : อย่างน้อยมีค่า β บางค่าที่ไม่ใช่ศูนย์ (สมการที่คำนวณได้มีคุณค่าในการพยากรณ์)

สูตรที่ใช้ $F = \dfrac{R^2/p}{(1-R^2)/(n-p-1)}$

โดยที่ p= จำนวนตัวแปรอิสระ ,n =ขนาดของกลุ่มตัวอย่าง

การทดสอบสมมติฐานว่าตัวแปรอิสระตัวใดตัวหนึ่งช่วยในการพยากรณ์ค่าของตัวแปรตามหรือไม่ :

H_O : $\beta_i = 0$ (เราสามารถตัดตัวแปรอิสระตัวนี้ออกได้โดยไม่ทำให้ลดคุณค่าในการพยากรณ์)

H_A : $\beta_i \neq 0$ (ตัดตัวแปรนี้ออกไม่ได้ เพราะมีส่วนสำคัญในการพยากรณ์ค่าของตัวแปรตาม)

สูตรที่ใช้ $t = \dfrac{b_i - \beta_i}{S_{bi}}$

หมายเหตุ:- degree of freedom ที่ใช้คือ n-p-1

โดยที่ n=ขนาดของกลุ่มตัวอย่าง , p=จำนวนตัวแปรอิสระ

Contribution: ดร.ดนัย ปัตตพงศ์*

* ผู้แต่ง/เจ้าของ Statistics Talks(70+ issues) เอกสารวิชาการด้านการวิจัยและสถิติประยุกต์ที่นำเสนอเครื่องมือทางสถิติเพื่อการวิจัยในระดับปริญญาโท เอก ตลอดจนตัวอย่างของการนำไปประยุกต์ใช้ในชีวิตจริง

* ผู้พัฒนา FINALYSIS... a dedicated software สำหรับให้บริการนักธุรกิจที่ต้องการวิเคราะห์ความเป็นไปได้ทางการเงินของโครงการพัฒนาอสังหาริมทรัพย์(บ้านจัดสรร/จัดสรรที่ดินเพื่อการอุตสาหกรรม/อาคารชุด/อาคารสำนักงานให้เช่า) โรงแรม โรงพยาบาลเอกชน ห้างสรรพสินค้า โรงงานน้ำตาล โรงงานกระดาษ โรงไฟฟ้าชีวมวล ฯลฯได้เห็นตัวเลขก่อนโครงการเกิด หลีกเลี่ยงความผิดพลาดเป็นร้อยเป็นพันล้านหากเกิดการลงทุนจริง (กำหนด DEBUT 1 เมษายน 2569)

* ผู้แต่งหนังสือ" การวิเคราะห์ความเป็นไปได้ทางการเงินและการจัดวงเงินเครดิตของโครงการลงทุน " ประกอบด้วยตัวอย่างของธุรกิจจริงที่ไม่เปิดเผยชื่อนับ 100 บริษัท ครอบคลุมอุตสาหกรรม 24 อุตสาหกรรม