

การวิเคราะห์การถดถอยแบบดรรชนี

Indicator Regression Technique

โดย ดร.วราวุธ วุฒินิชย์^{1/}

1. คำนำ

การวิเคราะห์การถดถอยที่รู้จักกันโดยทั่วไปตามที่ปรากฏในตำรา Applied Probability and Statistics for Engineers นั้น ตัวแปรอิสระ (Independent Variable) ต้องมีค่าเป็นตัวเลขเท่านั้น จึงจะทำการวิเคราะห์ได้

เทคนิคการวิเคราะห์การถดถอยแบบดรรชนีที่จะกล่าวถึงต่อไปนี้เป็นวิธีการวิเคราะห์การถดถอยแบบพหุคูณขั้นสูง (Advanced Multiple Regression) สามารถวิเคราะห์กรณีที่ตัวแปรอิสระบางตัวไม่ได้มีค่าเป็นตัวเลข แต่เป็นช่วงค่าหรือกลุ่มค่า (Categorical หรือ Qualitative Data) ได้ โดยการแปลงตัวแปรอิสระซึ่งมีค่า n ช่วงค่า (Classes) เป็นตัวแปรอิสระตัวใหม่ $n-1$ ตัว ก่อนทำการวิเคราะห์การถดถอยแบบพหุคูณตามปกติ (Draper and Smith, 1980; Milton and Arnold, 1995) นิยมใช้ในการศึกษาด้านวิศวกรรมขนส่งและด้านสังคมศาสตร์

Raghuwanshi and Wallender (1998) ได้นำเทคนิคการวิเคราะห์การถดถอยแบบดรรชนีมาใช้ในการหาความสัมพันธ์ระหว่างค่า K_p (สัมประสิทธิ์การพัดพาการระเหย) กับความเร็วลม ความชื้นสัมพัทธ์ (ซึ่งมีค่าเป็นช่วงค่า) และระยะทางจากผิวดินการระเหยแบบ เอ ถึงขอบพื้นที่ปลูกหญ้า (Upwind Fetch of Low-Growing Vegetation) โดยใช้ข้อมูล K_p ของ Doorenbos and Pruitt (1977) ดังตารางที่ 1

1. รองศาสตราจารย์ ภาควิชาวิศวกรรมชลประทาน คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์
วิทยาเขตกำแพงแสน

ตารางที่ 1 ค่าสัมประสิทธิ์การระเหย Kp (Doorenbos and Pruitt, 1977)

Wind run (km/day)	Fetch (m)	Relative Humidity (%)		
		≤ 40	40-70	≥ 70
< 175	1	0.55	0.65	0.75
-	10	0.65	0.75	0.85
-	100	0.70	0.80	0.85
-	1,000	0.75	0.85	0.85
175-425	1	0.50	0.60	0.65
-	10	0.60	0.70	0.75
-	100	0.65	0.75	0.80
-	1,000	0.70	0.80	0.80
425-700	1	0.45	0.50	0.60
-	10	0.55	0.60	0.65
-	100	0.60	0.65	0.70
-	1,000	0.65	0.70	0.75
> 700	1	0.40	0.45	0.50
-	10	0.45	0.55	0.60
-	100	0.50	0.60	0.65
-	1,000	0.55	0.60	0.65

ให้ Kp = สัมประสิทธิ์การระเหยของสภาพการระเหยแบบ เอ (ETo/Epan)
 F = ระยะทางจากสภาพการระเหยแบบ เอ ถึงขอบพื้นที่ปลูกหญ้า (เมตร)
 U = ความเร็วลมเฉลี่ยประจำวันวัดที่ระยะ 2 เมตร (กม./วัน)
 H = ความชื้นสัมพัทธ์เฉลี่ยประจำวัน (%)

ในการวิเคราะห์การถดถอยให้ Kp เป็นตัวแปรตาม (Dependent Variable) และ F, U, H เป็นตัวแปรอิสระ (Independent Variable)

เนื่องจากความเร็วลมและความชื้นสัมพัทธ์เป็นตัวแปรอิสระซึ่งมีค่าแบ่งออกเป็น 4 และ 3 ช่วงค่าตามลำดับ จึงไม่สามารถวิเคราะห์หาความสัมพันธ์ระหว่าง Kp และ F, U, H โดยวิธีการวิเคราะห์การถดถอยตามปกติได้

2. การกำหนดตัวแปรอิสระสำหรับการวิเคราะห์การถดถอยแบบดรชนี้

จากข้อมูลในตารางที่ 1 จะสามารถกำหนดตัวแปรอิสระใหม่สำหรับการวิเคราะห์การถดถอยได้ดังนี้

ให้ x_1 = $\ln(F)$
 x_2 = U (175-425 กม./วัน)
 x_3 = U (425-700 กม./วัน)
 x_4 = U (>700 กม./วัน)

$$x_5 = H (40-70 \%)$$

$$x_6 = H (\geq 70 \%)$$

ตัวแปรอิสระ (x_2, x_3, x_4) ซึ่งเป็นตัวแทนแต่ละช่วงค่าของความเร็วลม และ (x_5, x_6) ซึ่งเป็นตัวแทนแต่ละช่วงค่าของความชื้นสัมพัทธ์ จะมีค่าเป็น 0 หรือ 1 ขึ้นอยู่กับว่าค่า K_p ตกอยู่ในช่วงค่าของตัวแปรอิสระนั้นหรือไม่ ถ้า K_p ตกอยู่ในช่วงค่าของตัวแปรอิสระตัวนั้น จะกำหนดให้ตัวแปรอิสระนั้นมีค่าเป็น 1 ถ้าไม่ตกให้มีค่าเป็น 0 เป็นต้น

ตามแนวความคิดดังกล่าว จะสามารถแปลงข้อมูลจากตารางที่ 1 เป็น ค่าของตัวแปรอิสระ ($x_1, x_2, \dots, x_6$) ได้ดังตารางที่ 2

3. การวิเคราะห์การถดถอยเชิงเส้นแบบพหุคูณ

หลังจากกำหนดตัวแปรอิสระและแปลงค่าแล้วดังตารางที่ 2 จะสามารถนำมาวิเคราะห์โดยใช้สมการถดถอยเชิงเส้นแบบพหุคูณได้ตามวิธีปกติดังนี้

สมการถดถอยเชิงเส้นพหุคูณ

$$\hat{Kp}_i = b_0 + b_1.x_{1_i} + b_2.x_{2_i} + b_3.x_{3_i} + b_4.x_{4_i} + b_5.x_{5_i} + b_6.x_{6_i} \dots\dots\dots(1)$$

$$\begin{aligned} \text{เมื่อ } \hat{Kp}_i &= \text{ค่าคาดคะเนของค่า } Kp_i \\ x_{1_i}, x_{2_i}, \dots, x_{6_i} &= \text{ค่าตัวแปรอิสระตามที่กำหนดในข้อ 2} \\ b_0, b_1, \dots, b_6 &= \text{ค่าสัมประสิทธิ์ของสมการถดถอย} \end{aligned}$$

การวิเคราะห์การถดถอยเชิงเส้นแบบพหุคูณโดยใช้โปรแกรม Excel ได้ผลดังตารางที่ 3 และได้สมการถดถอยของ $\hat{Kp}$ ซึ่งมีค่า $R^2 = 0.9597$ ดังนี้

$$\hat{Kp} = 0.5944 + 0.0243x_1 - 0.0583x_2 - 0.1333x_3 - 0.2083x_4 + 0.0813x_5 + 0.1344x_6 \dots\dots\dots(2)$$

ซึ่งจะเห็นได้ว่าค่า R^2 มีค่าค่อนข้างสูง

ตารางที่ 2 ค่า Kp และค่าตัวแปรอิสระ ($x_1, x_2, \dots, x_6$)

In (F)	Wind Run			Relative Humidity		Kp
x_1	x_2	x_3	x_4	x_5	x_6	
0	0	0	0	0	0	0.55
0	0	0	0	1	0	0.65
0	0	0	0	0	1	0.75
0	1	0	0	0	0	0.50
0	1	0	0	1	0	0.60
0	1	0	0	0	1	0.65
0	0	1	0	0	0	0.45
0	0	1	0	1	0	0.50
0	0	1	0	0	1	0.60
0	0	0	1	0	0	0.40
0	0	0	1	1	0	0.45
0	0	0	1	0	1	0.50
2.3	0	0	0	0	0	0.65
2.3	0	0	0	1	0	0.75
2.3	0	0	0	0	1	0.85
2.3	1	0	0	0	0	0.60
2.3	1	0	0	1	0	0.70
2.3	1	0	0	0	1	0.75
2.3	0	1	0	0	0	0.55
2.3	0	1	0	1	0	0.60
2.3	0	1	0	0	1	0.65
2.3	0	0	1	0	0	0.45
2.3	0	0	1	1	0	0.55
2.3	0	0	1	0	1	0.60
4.6	0	0	0	0	0	0.70
4.6	0	0	0	1	0	0.80
4.6	0	0	0	0	1	0.85
4.6	1	0	0	0	0	0.65
4.6	1	0	0	1	0	0.75
4.6	1	0	0	0	1	0.80
4.6	0	1	0	0	0	0.60
4.6	0	1	0	1	0	0.65
4.6	0	1	0	0	1	0.70
4.6	0	0	1	0	0	0.50
4.6	0	0	1	1	0	0.60
4.6	0	0	1	0	1	0.65
6.9	0	0	0	0	0	0.75
6.9	0	0	0	1	0	0.85
6.9	0	0	0	0	1	0.85
6.9	1	0	0	0	0	0.70
6.9	1	0	0	1	0	0.80
6.9	1	0	0	0	1	0.80
6.9	0	1	0	0	0	0.65
6.9	0	1	0	1	0	0.70
6.9	0	1	0	0	1	0.75
6.9	0	0	1	0	0	0.55
6.9	0	0	1	1	0	0.60
6.9	0	0	1	0	1	0.65

ตารางที่ 3 ผลการวิเคราะห์การถดถอยเชิงเส้นแบบพหุคูณโดย Excel

SUMMARY OUTPUT

<i>Regression Statistics</i>	
Multiple R	0.9796
R Square	0.9597
Adjusted R Square	0.9538
Standard Error	0.0254
Observations	48

ANOVA

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	6	0.6286	0.1048	162.7326	0.0000
Residual	41	0.0264	0.0006		
Total	47	0.6550			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	0.5944	0.0102	58.1110	0.0000	0.5737	0.6150
x1	0.0243	0.0014	17.0449	0.0000	0.0214	0.0272
x2	-0.0583	0.0104	-5.6314	0.0000	-0.0793	-0.0374
x3	-0.1333	0.0104	-12.8718	0.0000	-0.1543	-0.1124
x4	-0.2083	0.0104	-20.1122	0.0000	-0.2293	-0.1874
x5	0.0813	0.0090	9.0572	0.0000	0.0631	0.0994
x6	0.1344	0.0090	14.9792	0.0000	0.1163	0.1525

4. สรุป

วิธีการวิเคราะห์การถดถอยแบบตรรกะนี้ นับว่ามีประโยชน์ต่อการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร กรณีที่ตัวแปรอิสระบางตัวไม่ได้มีค่าเป็นตัว Twenty-two points, plus triple-word-score, plus fifty points for using all my letters. Game's over. I'm outta here. เลขในลักษณะที่สามารถนำมาวิธีการถดถอยตามปกติได้ อย่างไรก็ตามการวิเคราะห์การถดถอยของค่า K_p ในตารางที่ 1 อาจทำได้โดยการสมมติค่าเพื่อตัวแทนของช่วงค่าได้ เช่นสมมติให้ U มีค่าเท่ากับ 175, 300, 562 และ 700 มม/วัน เพื่อเป็นตัวแทนค่า $U < 175$, 175-425, 425-700 และ > 700 มม/วัน ตามลำดับ หรือสมมติให้ H เท่ากับ 40, 55 และ 70% เพื่อเป็นตัวแทนค่า $H \leq 40$, 40-70 และ $\geq 70\%$ ตามลำดับ ตามที่ Snyder (1992) เสนอแนะ

5. เอกสารอ้างอิง

- Doorenbos, J. and W. Pruitt (1977). Crop Water Requirements. FAO Irrig. and Drain. Paper 24. FAO, Rome, Italy.
- Draper, N. And H. Smith (1980). Applied Regression Analysis, 2nd Ed., John Wiley and Sons, Inc., New York, N.Y.
- Milton, J.S. and J. C. Arnold (1995). Introduction to Probability and Statistics : Principles and Applications for Engineering and Computing Sciences, 3rd Ed., McGraw-Hill, Inc., New York. N.Y.
- Raghuwanshi, N. S. and W. W. Wallender (1998). Converting From Pan Evaporation to Evaporation. J. Irrig. and Drain. Engrg, ASCE. 124(5) : 275-277.
- Snyder, R. L. (1992). Equation for an Evaporation Pan to Evapotranspiration Conversion. J. Irrig. and Drain Engrg, ASCE. 118(6) : 977-980.