

100

Twilight Program

เรื่อง “การเลือกใช้สถิติในการวิจัยเชิงปริมาณ”

โดย รองศาสตราจารย์ ดร. ศจีมาจ ฦ วิเชียร
วิทยาลัยเทคโนโลยีอุตสาหกรรม
มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ

วันพุธที่ 19 สิงหาคม 2558 เวลา 18.00 – 20.00 น.
ณ โรงแรมเซ็นทาราแกรนด์ และบางกอกคอนเวนชันเซ็นเตอร์ เซ็นทรัลเวิลด์ กรุงเทพฯ



บริหารจัดการโดย

ฝ่ายจัดการความรู้การวิจัย

กองประเมินผลและจัดการความรู้การวิจัย วช.

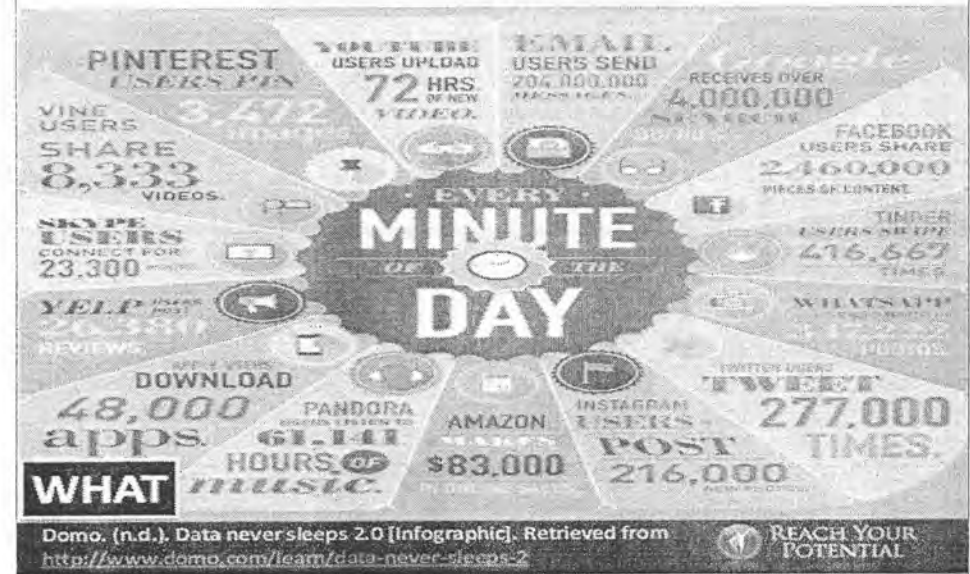
การเลือกใช้สถิติในการวิจัย เชิงปริมาณ

วันพุธที่ 19 สิงหาคม 2558

Twilight Program (Thailand Research Expo 2015)

ศัจฉิมาจ ฅ วิเชียร

Information Age



Data, Information, Knowledge, Learning

LEARNING/WISDOM

experiences in more applications

KNOWLEDGE

beliefs, assumptions, application

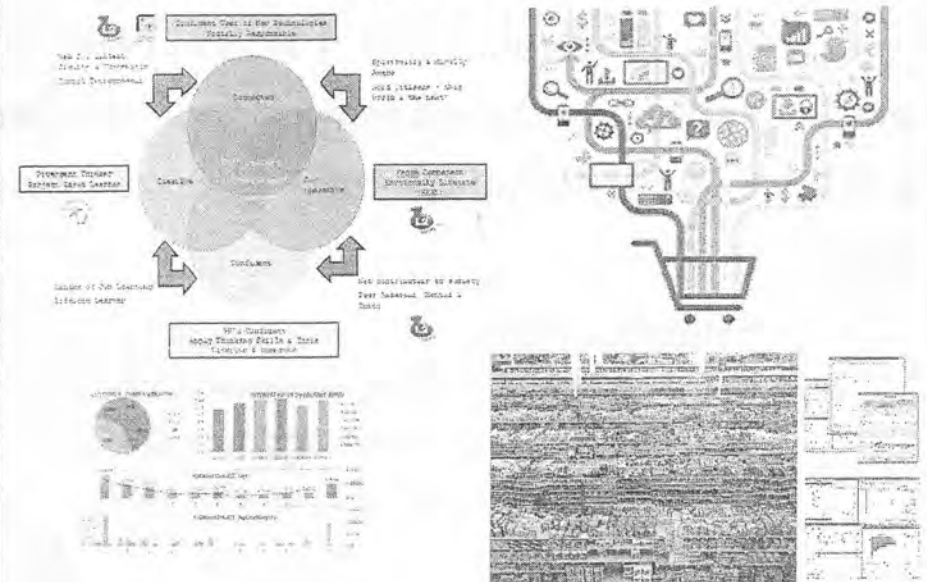
INFORMATION

meaning

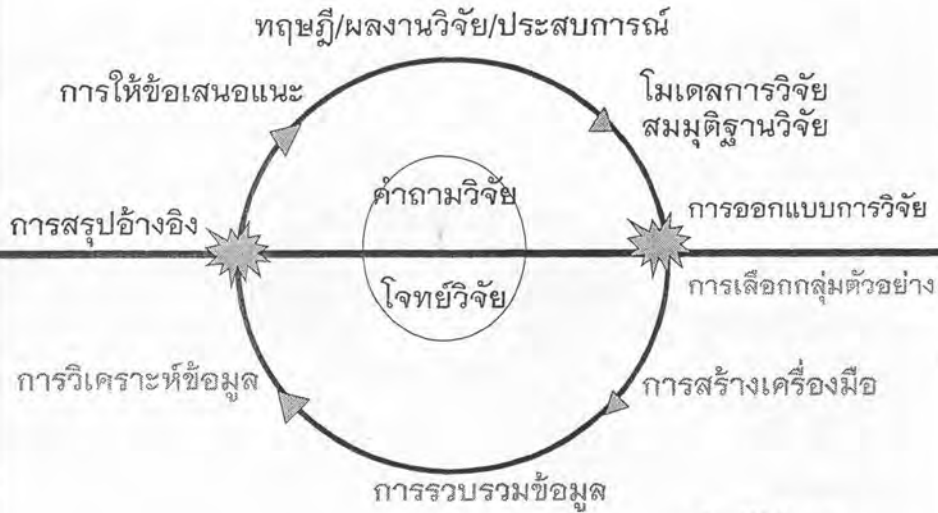
relationship

phenomena

Model driven & Data driven



วงจรวิจัย (RESEARCH CYCLE)



นงลักษณ์ วิรัชชัย, 2546

ตัวแปร

สิ่งต่าง ๆ หรือลักษณะต่าง ๆ ที่แสดงให้เห็นถึงสถานการณ์ที่ปรากฏในประเด็นที่ต้องการศึกษา โดยมีค่าที่แปรเปลี่ยนไปในแต่ละหน่วยของประชากรที่ศึกษา

ตัวแปรในการวิจัย

- ตัวแปรต้น (Independent variable = IV)
- ตัวแปรตาม (Dependent variable = DV)
- ตัวแปรแทรกซ้อน (Extraneous var. = EV)

6

ระดับการวัดของตัวแปร

จัดกลุ่ม เรียงลำดับ ช่วงเท่า ศูนย์จริง

Non-metric var.

- | | | | |
|-----------------|---|---|--|
| 1. Nominal var. | X | | |
| 2. Ordinal var. | X | X | |

Metric var.

- | | | | | |
|------------------|---|---|---|---|
| 3. Interval var. | X | X | X | |
| 4. Ratio var. | X | X | X | X |

ประเภทของการวิเคราะห์ข้อมูล

1. การวิเคราะห์เบื้องต้น (preliminary analysis)
 - บรรยายลักษณะของกลุ่มประชากร/กลุ่มตัวอย่าง และการแจกแจงความถี่ของตัวแปร
2. การวิเคราะห์เพื่อตรวจสอบคุณภาพเครื่องมือวิจัย
3. การวิเคราะห์เพื่อตรวจสอบข้อตกลงเบื้องต้นของสถิติ
4. การวิเคราะห์เพื่อตอบคำถามวิจัย

นงลักษณ์ วิรัชชัย, 2546

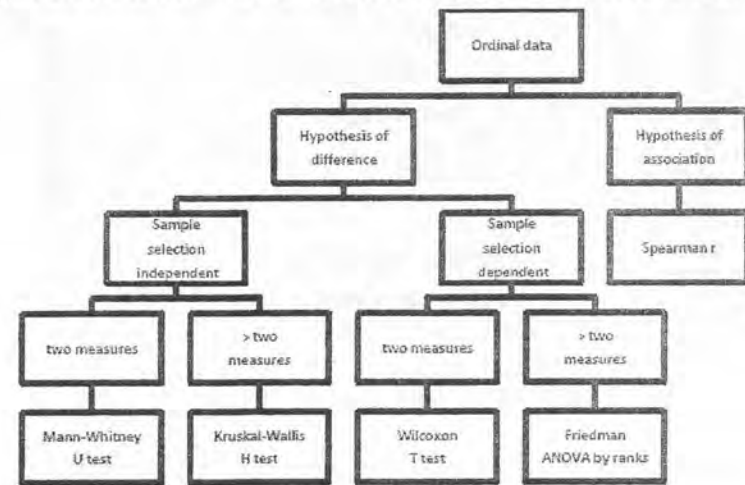
วิธีการวิเคราะห์ข้อมูลเชิงปริมาณ

1. สถิติว่าด้วยการเลือกกลุ่มตัวอย่าง
 - แบบทราบความน่าจะเป็น แบบไม่ทราบค่าความน่าจะเป็น
2. สถิติบรรยาย (Descriptive statistics)
 - การแจกแจงความถี่ ร้อยละ การเสนอแผนภูมิ และกราฟ
 - การวัดแนวโน้มสู่ส่วนกลาง (เช่น ค่าเฉลี่ย)
 - การวัดการกระจาย, ความเบ้, ความโด่ง
3. สถิติอนุมาน (Inferential statistics)
 - การวิเคราะห์ความแตกต่าง
 - การวิเคราะห์ความแปรปรวน
 - การวิเคราะห์การถดถอย
 - การวิเคราะห์ด้วยสถิติขั้นสูง

นงลักษณ์ วิรัชชัย, 2546

สถิติวิเคราะห์ที่เหมาะสม: Ordinal data

Figure 2: Flow chart on statistical tests to be used for data measured in ordinal scale

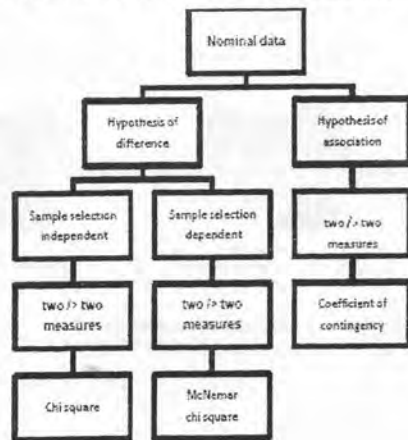


N Gunawardana

11

สถิติวิเคราะห์ที่เหมาะสม: Norminal data

Figure 1: Flow chart on statistical tests to be used for data measured in nominal scale

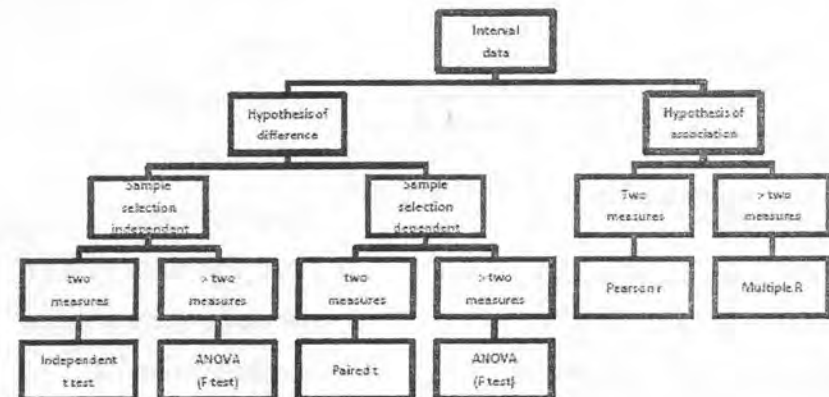


N Gunawardana

12

สถิติวิเคราะห์ที่เหมาะสม: Interval data

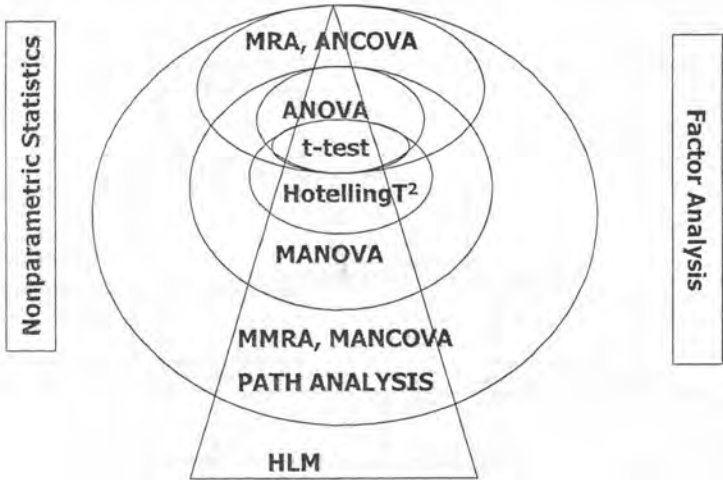
Figure 3: Flow chart on statistical tests to be used for data measured in interval scale



N Gunawardana

13

LISREL OR SEM



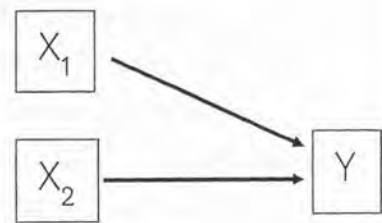
นงลักษณ์ วิรัชชัย, 2546 13

การวิเคราะห์ความแปรปรวน Analysis of Variance (ANOVA)

- เป็นวิธีทางสถิติที่ใช้ค่าความแปรปรวนในการทดสอบความแตกต่างของค่าเฉลี่ยตั้งแต่ 2 กลุ่มขึ้นไป
 - One-way anova : ตัวแปรต้น 1 ตัว (มากกว่า/เท่ากับ 2 กลุ่ม)
 - Two-way anova : ตัวแปรต้น 2 ตัว
 - n-way anova : ตัวแปรต้น n ตัว
 - ตัวแปรตาม 1 ตัว

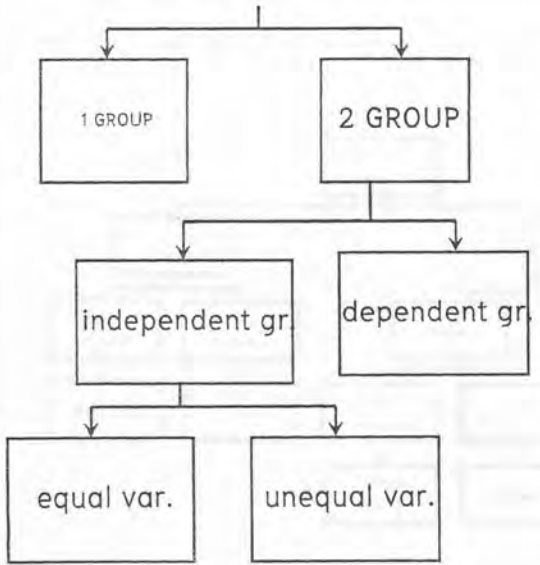


Oneway ANOVA



Two-way ANOVA

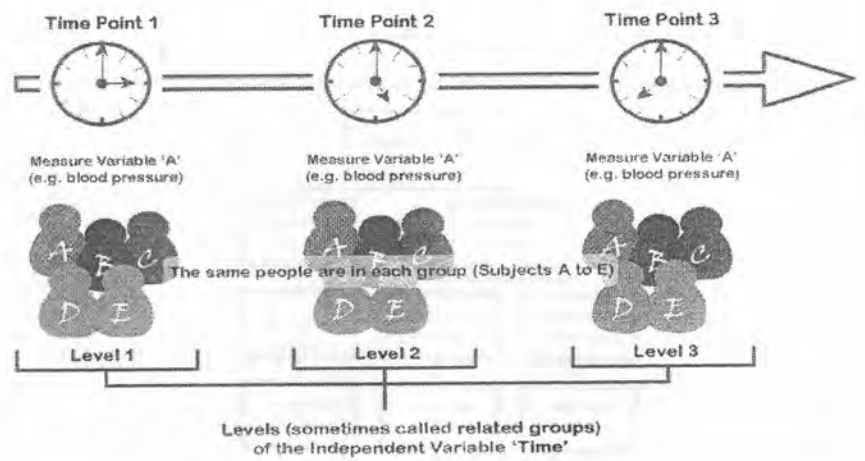
t - test



ตัวแปรตาม 1 ตัว มีระดับการวัดแบบอันตรภาค หรือ อัตราส่วน
 ตัวแปรต้น 1 ตัว มีสองค่า (ระดับ) มีระดับการวัดแบบนามบัญญัติ หรือ เรียงอันดับ
 ผลการทดสอบสมมุติฐานทางสถิติว่าค่าเฉลี่ยประชากรสองกลุ่มแตกต่างกัน

14

Repeated Measure ANOVA

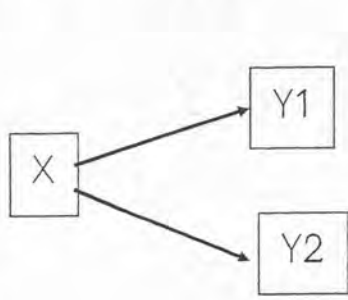


© Lund Research Ltd 2011

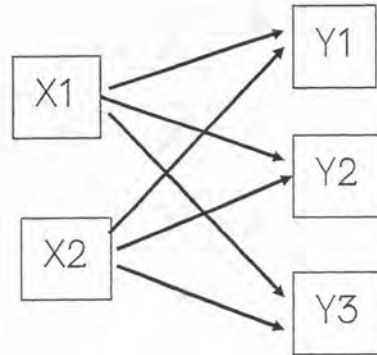
http://statistics.lund.gov

15

การวิเคราะห์ความแปรปรวนตัวแปรพหุนาม Multivariate Analysis of Variance(MANOVA)



Oneway MANOVA



Two-way MANOVA

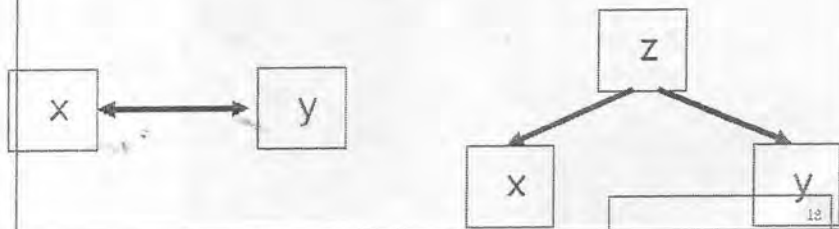
CORRELATION

SIMPLE CORRELATION (r_{xy})

PARTIAL CORRELATION ($r_{xy.z}$)

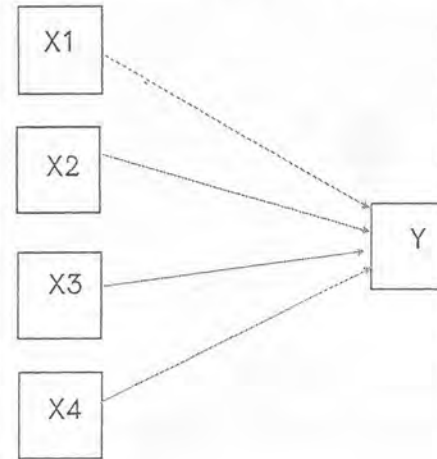
SEMI-PARTIAL CORRELATION

$$r_{xy} = 0.82 \quad r_{xy.z} = 0.30$$



การวิเคราะห์ถดถอยพหุคูณ

MULTIPLE REGRESSION ANALYSIS (MRA)

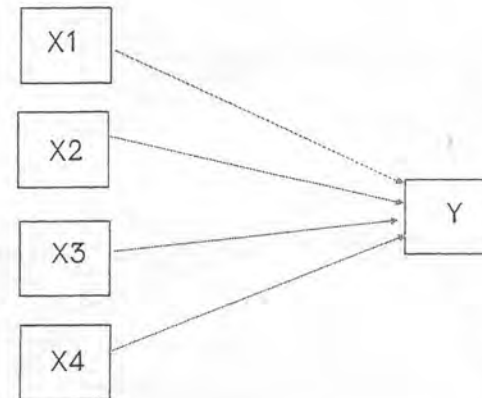


จุดมุ่งหมาย

1. ศึกษาความสัมพันธ์ระหว่างตัวแปร
2. เปรียบเทียบอิทธิพลของตัวทำนาย(IV)
3. การอธิบายความแปรปรวนในตัวแปรตาม (ETA SQUARE)
4. วิเคราะห์อิทธิพลปฏิสัมพันธ์
5. เปรียบเทียบค่าเฉลี่ยระหว่างกลุ่ม
6. ทำนายค่า DV เมื่อทราบค่า IV
7. วิเคราะห์อิทธิพลทางตรง, ทางอ้อม

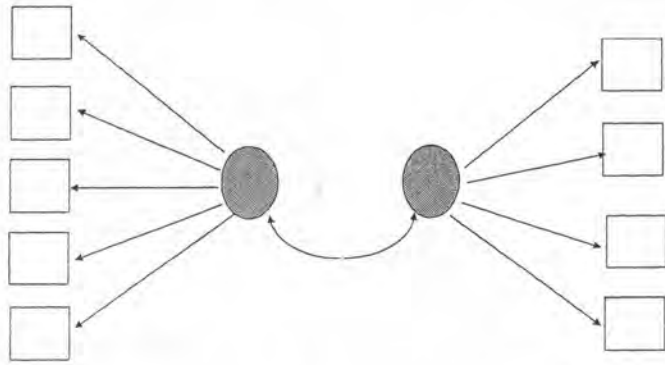
การวิเคราะห์จำแนก

DISCRIMINANT ANALYSIS



การศึกษาว่ามีตัวแปรหรือปัจจัยใดบ้างที่เป็นปัจจัยสำคัญทำให้คนหรือหน่วยตัวอย่างอยู่ต่างกลุ่มกัน แล้วนำตัวแปรเหล่านั้นมาหาความสัมพันธ์ในรูปแบบเชิงเส้น นำสมการเชิงเส้นมาพยากรณ์หรือประมาณว่าหน่วยหรือคนใหม่ควรอยู่กลุ่มใด

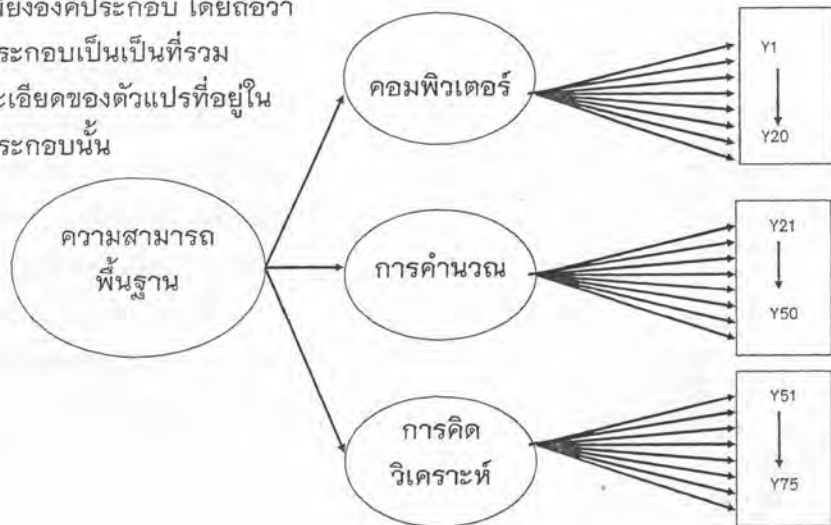
การวิเคราะห์สหสัมพันธ์คาโนนิกัล (CANONICAL CORRELATION ANALYSIS)



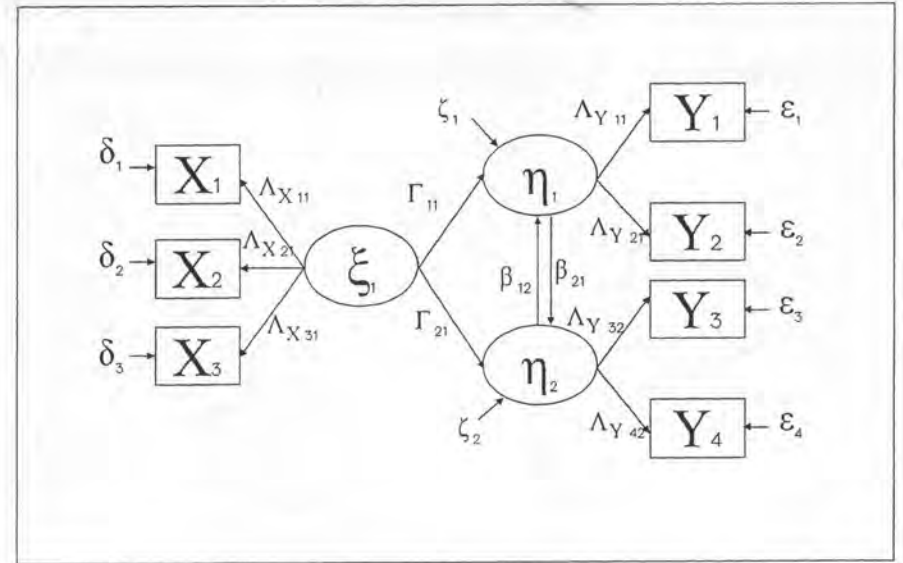
เป็นการวิเคราะห์ความสัมพันธ์ระหว่างกลุ่มข้อมูลที่ประกอบด้วยตัวแปรหลายตัวกับตัวแปรหลายตัวไปพร้อม ๆ กัน

การวิเคราะห์องค์ประกอบ (FACTOR ANALYSIS)

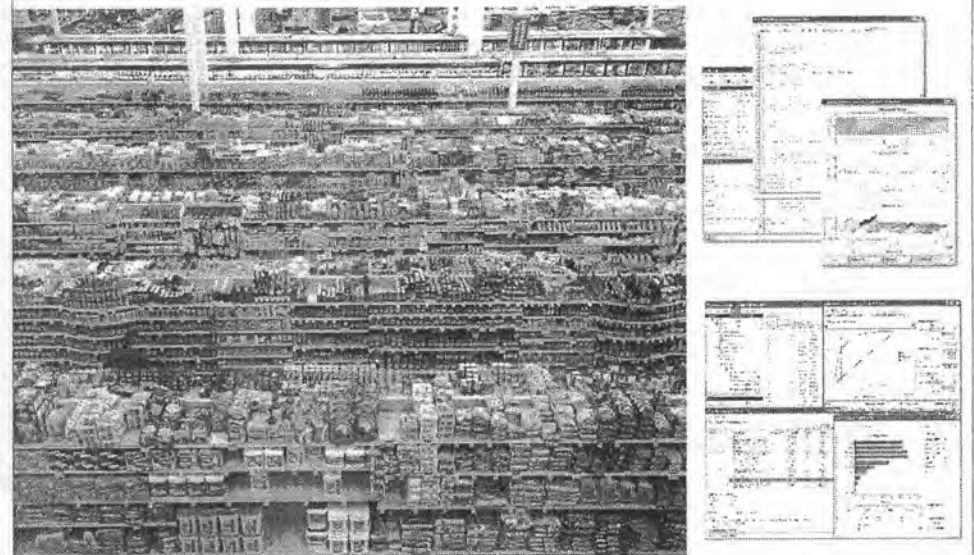
เทคนิคที่ใช้ในการลดจำนวนตัวแปรจากจำนวนตัวแปรมาก ๆ ให้เหลือไม่กี่เพียงองค์ประกอบ โดยถือว่าองค์ประกอบเป็นสิ่งที่รวมรายละเอียดของตัวแปรที่อยู่ในองค์ประกอบนั้น



โมเดลลีสเรล

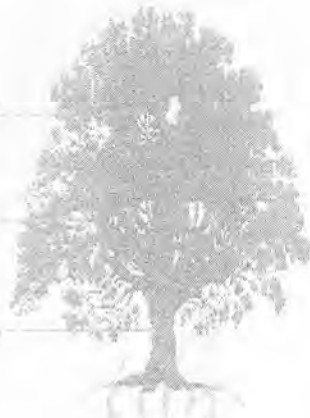


Let the data talk



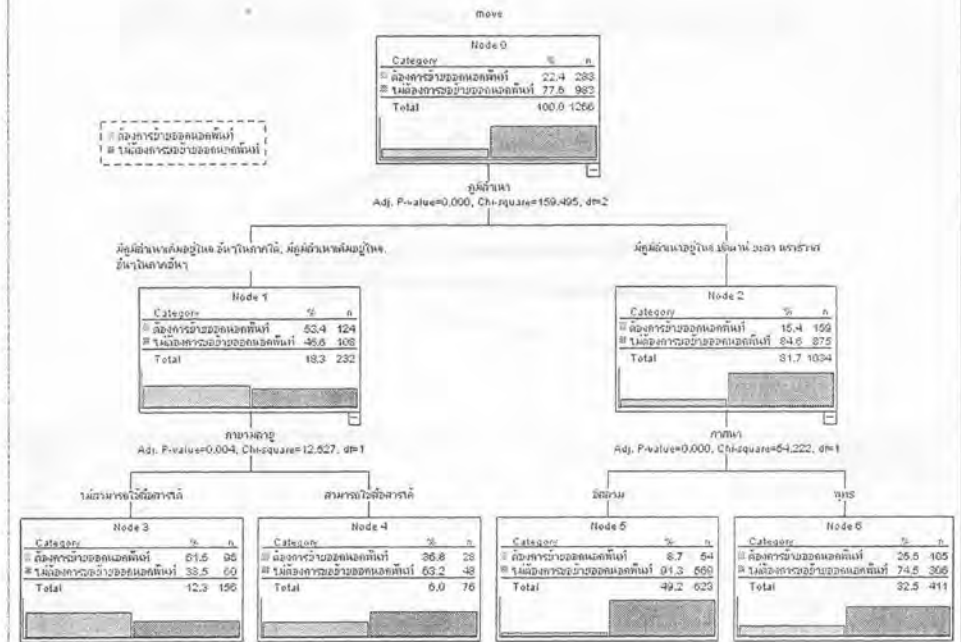
การวิเคราะห์ต้นไม้ตัดสินใจ (Decision tree analysis)

การวิเคราะห์ต้นไม้ตัดสินใจ เป็นการนำข้อมูลมาสร้างแบบจำลองพยากรณ์ในรูปแบบโครงสร้างต้นไม้ โดยใช้ข้อมูลจากตัวแปรต้น มาเป็นตัวทำนายผลของตัวแปรตาม โดยแบบจำลองนี้สามารถวิเคราะห์ได้ทั้งการวิเคราะห์จำแนกเชิงสำรวจ (exploratory classification analysis) และการวิเคราะห์จำแนกเชิงยืนยัน (confirmatory classification analysis)



ต้นไม้ตัดสินใจสามารถวิเคราะห์?

1. Segmentation เป็นการระบุตัวตนของบุคคลที่เป็นสมาชิกกลุ่ม หรือระบุความเป็นสมาชิกกลุ่มของแต่ละคน
2. Stratification เป็นการจัดกระทำกับข้อมูลให้อยู่ในรูปกลุ่มหรือประเภทต่างๆ กันเช่น กลุ่มความเสี่ยงสูง กลุ่มความเสี่ยงปานกลาง และกลุ่มความเสี่ยงต่ำ
3. Prediction ใช้ในการสร้างกฎและใช้กฎนี้ในการทำนายเหตุการณ์ในอนาคต
4. Data reduction and variable screening สามารถเลือกข้อมูลที่สำคัญในการสร้างโมเดลเชิงทำนาย จากข้อมูลจำนวนมากๆ
5. Interacting identification ระบุความสัมพันธ์ในกลุ่มย่อยเพื่อเป็นข้อมูลในการสร้างโมเดลเชิงทำนาย
6. Category merging and discretizing continuous variables สามารถแปลงค่าข้อมูลของกลุ่มตัวแปรต้นไม่ว่าจะเป็นข้อมูลจัดกลุ่มหรือข้อมูลต่อเนื่องให้เป็นค่าที่เหมาะสมและสูญเสียคุณสมบัติของข้อมูลน้อยที่สุด

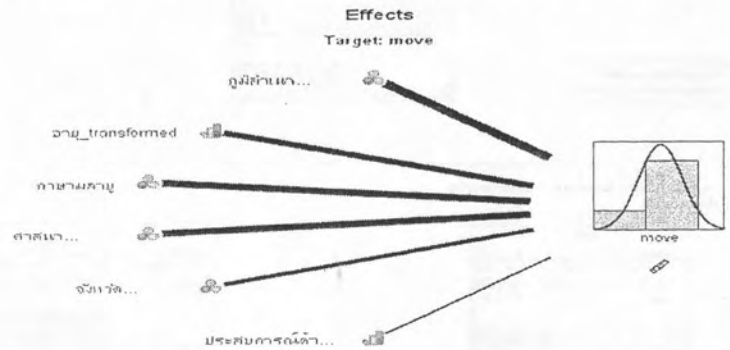


Tree Table

Node	ไม่มีบ้าน		ผู้มีบ้าน		Total	Predicted Category	Parent Node	Primary Independent Variable						
	n	%	n	%				Variable	Sum	Chi-Square	df	Split Values		
0	105	26.5%	302	74.5%	411	ไม่มีบ้าน								
1	124	83.4%	102	46.6%	232	ไม่มีบ้าน	0	ผู้มีบ้าน	0.00	159.495	2	ผู้ไม่มีบ้าน		
2	159	61.7%	675	34.6%	1034	ไม่มีบ้าน	0	ผู้มีบ้าน	0.00	159.495	2	ผู้ไม่มีบ้าน		
3	36	31.5%	60	38.5%	156	ไม่มีบ้าน	1	ภาษาพูด	0.04	12.627	1	ไม่มีบ้าน		
4	28	26.5%	45	63.2%	76	ไม่มีบ้าน	1	ภาษาพูด	0.04	12.627	1	ไม่มีบ้าน		
5	64	8.7%	369	91.5%	623	ไม่มีบ้าน	2	มีรถ	0.00	54.222	1	ไม่มีบ้าน		
6	105	26.5%	302	74.5%	411	ไม่มีบ้าน	2	มีรถ	0.00	54.222	1	ไม่มีบ้าน		

Splitting Method: GTEST
Dependent Variable: move

Automatic Linear Modeling



Information Criterion	Steps					
	1	2	3	4	5	6
move_squared_transformed	✓	✓	✓	✓	✓	✓
move_cubic		✓	✓	✓	✓	✓
move_transformed			✓	✓	✓	✓
move				✓	✓	✓
move_quartic					✓	✓
move Quintic						✓

The model building method is forward stepwise using the Information Criterion. A checkmark means the effect is in the model at this step.