

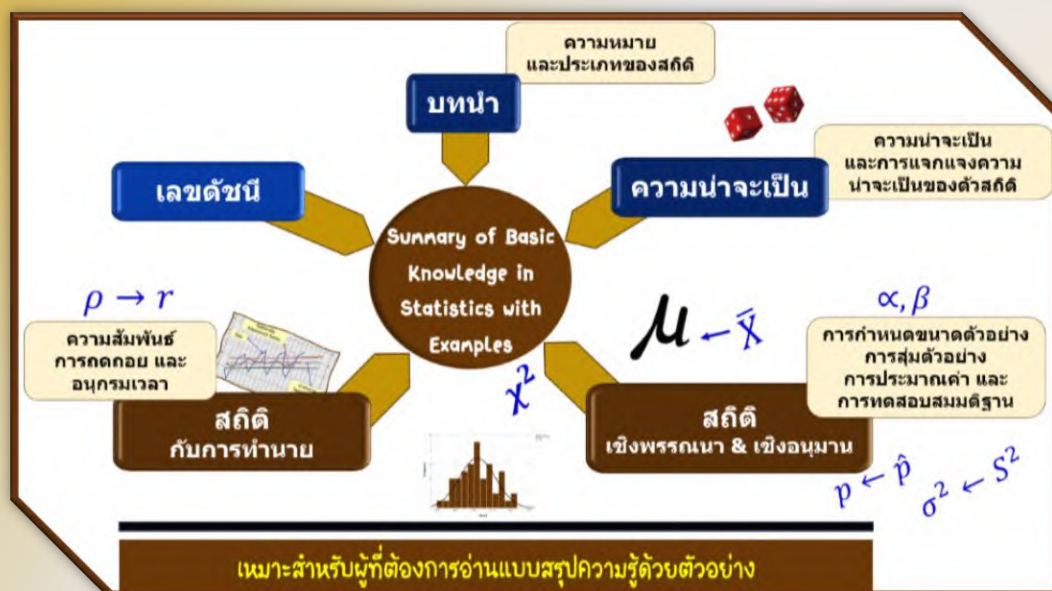
ตัวอย่าง คือบทเรียนที่ดีสำหรับการสรุปและทำความเข้าใจกับศาสตร์ต่าง ๆ

รวมถึงตัวอย่างทางสถิติศาสตร์

เวอร์ชัน 3
2568

สรุปความรู้พื้นฐาน ด้านสถิติศาสตร์ พร้อมตัวอย่าง

Summary of Basic Knowledge in Statistics with Examples



รองศาสตราจารย์ ดร. วิชุดา ไชยศิริมงคล

สาขาวิชาสถิติ คณะวิทยาศาสตร์

มหาวิทยาลัยขอนแก่น

พิมพ์ครั้งที่ 1 จำนวน 200 เล่ม : มิถุนายน พ.ศ. 2567 ; © 2567

พิมพ์ e-book ครั้งที่ 1 : สิงหาคม พ.ศ. 2567 ; © 2567 (169 หน้า)

พิมพ์ e-book ครั้งที่ 3 : ธ.ค. 2568 (184 หน้า)

(สงวนลิขสิทธิ์)

สงวนลิขสิทธิ์ตามพระราชบัญญัติลิขสิทธิ์ พ.ศ. 2537

หนังสือเล่มนี้สงวนลิขสิทธิ์ ห้ามนำส่วนใดส่วนหนึ่งของหนังสือเล่มนี้ไปทำซ้ำ ดัดแปลง แปล หรือเผยแพร่ไม่ว่าจะในรูปแบบหรือวิธีการใด ๆ รวมทั้งการถ่ายสำเนา การบันทึกข้อมูลหรือระบบการจัดเก็บและถ่ายโอนข้อมูล โดยไม่ได้รับอนุญาตเป็นลายลักษณ์อักษรจากผู้ถือลิขสิทธิ์

ISBN: 978-616-438-891-8

ISBN: 978-616-616-221-9 (E-Book)

ราคา 320 บาท

ข้อมูลทางบรรณานุกรมของหอสมุดแห่งชาติ

วิชุดา ไชยศิวิมมงคล.

สรุปความรู้พื้นฐานด้านสถิติศาสตร์ พร้อมตัวอย่าง = Summary of basic knowledge in statistics with examples. -- ขอนแก่น: [ม.ป.พ.], 2567.
169 หน้า.

1. สถิติศาสตร์. I. ชื่อเรื่อง.

519.5

ISBN 978-616-616-221-9 (E-Book)

ISBN: 978-616-438-891-8 (Book)

จัดทำโดย : รองศาสตราจารย์ ดร.วิชุดา ไชยศิวิมมงคล

พิมพ์ต้นฉบับ : รองศาสตราจารย์ ดร.วิชุดา ไชยศิวิมมงคล

ติดต่อสั่งซื้อได้ที่ : E-book  <https://www.chulabook.com/>

E-mail  wichuda@kku.ac.th

การเขียนอ้างอิง: วิชุดา ไชยศิวิมมงคล, (2568). *สรุปความรู้พื้นฐานด้านสถิติศาสตร์ พร้อมตัวอย่าง*. สาขาวิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น. <https://www.chulabook.com/test-prep/206729>.

คำนำ

ความรู้ทางสถิติศาสตร์เป็นเครื่องมือสำคัญที่ช่วยพัฒนาทักษะการคิดเชิงวิเคราะห์ การตั้งคำถาม และการประเมินความน่าเชื่อถือของข้อมูล ผู้ที่มีพื้นฐานด้านนี้จะสามารถตัดสินใจอย่างมีเหตุผลบนพื้นฐานของข้อมูลจริง รวมถึงสามารถถ่ายทอดสารสนเทศที่ซับซ้อนให้อยู่ในรูปแบบที่เข้าใจง่าย และใช้เป็นพื้นฐานในการทำวิจัยเพื่อสร้างสรรค์นวัตกรรมใหม่ ๆ ดังพระราชโองการของพระบาทสมเด็จพระปรมินทรมหาภูมิพลอดุลยเดชฯ ที่ทรงเน้นถึงความสำคัญของการใช้สถิติอย่างรอบคอบและต้องทำอย่างซื่อสัตย์ เพื่อหลีกเลี่ยงการบิดเบือนข้อมูล ซึ่งอาจนำไปสู่การวางแผนที่ผิดพลาดและเกิดผลเสียต่อการพัฒนาเศรษฐกิจและสังคมของประเทศ

แม้สถิติศาสตร์จะเป็นศาสตร์ที่มีความซับซ้อน โดยเฉพาะในประเด็นที่เกี่ยวข้องกับคณิตศาสตร์และการเลือกใช้วิธีวิเคราะห์ที่เหมาะสมในแต่ละสถานการณ์ แต่ด้วยเครื่องมือและเทคโนโลยีสมัยใหม่ เช่น โปรแกรมสำเร็จรูปทางสถิติ ได้ช่วยให้การเรียนรู้และประยุกต์ใช้งานเป็นเรื่องที่เข้าถึงได้ง่ายขึ้น จากประสบการณ์ของผู้เขียนที่ได้สอนรายวิชาการจัดการข้อมูลและสถิติพื้นฐานแก่คณะต่าง ๆ ในมหาวิทยาลัยขอนแก่น หนังสือเล่มนี้จึงถูกจัดทำขึ้น เพื่อสรุปหลักการและแนวคิดพื้นฐานทางสถิติศาสตร์ โดยใช้ภาษาที่เข้าใจง่าย พร้อมตัวอย่างประกอบ เพื่อให้เหมาะกับผู้อ่านทุกระดับ โดยเฉพาะอย่างยิ่งผู้ที่ต้องการทบทวนความรู้เดิม แม้ในบางกรณีจะมีการใช้คำศัพท์ที่ต่างจากพจนานุกรมทางวิชาการ แต่ยังคงไว้ซึ่งความหมายเดิม เช่น “การสุ่มตัวอย่าง” แทน “การชักตัวอย่าง” เพื่อให้ง่ายต่อการเข้าใจ สำหรับเวอร์ชัน 3 ได้ปรับแก้จุดบกพร่องและเพิ่มเติมรายละเอียดในเรื่องของการกำหนดขนาดตัวอย่าง nonparametric และการวิเคราะห์ความสอดคล้อง รวมทั้งมี Playlist ประกอบบางหัวข้อ

ผู้เขียนขอขอบคุณคณาจารย์สาขาวิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น และผู้อ่านทุกท่านที่ให้คำแนะนำอันมีคุณค่า รวมถึงขอขอบคุณบุคลากร อาจารย์ ครอบครัว และทุกท่านที่มีส่วนสนับสนุนให้หนังสือเล่มนี้สำเร็จลงได้ด้วยดี หวังเป็นอย่างยิ่งว่า หนังสือเล่มนี้จะช่วยเสริมสร้างทักษะด้านข้อมูล (Data Literacy) แก่ผู้อ่าน และช่วยวางรากฐานอันมั่นคงสำหรับการเรียนรู้ด้านสถิติศาสตร์ในระดับที่สูงขึ้น หากมีข้อเสนอแนะเพิ่มเติมสามารถติดต่อได้ผ่าน QR Code ด้านล่าง หรือ <https://kku.world/obdf> จักขอบคุณยิ่ง

รองศาสตราจารย์ ดร.วิชุดา ไชยศิริมงคล
ธันวาคม 2568

<https://kku.world/obdf>



For Comment

สรุปความรู้พื้นฐาน ด้านสถิติศาสตร์
Wichuda

สื่อการเรียนรู้อื่น ๆ ที่เกี่ยวข้อง

ทั้งนี้ เพื่อให้การเรียนรู้ดังกล่าวเกิดประสิทธิผลยิ่งขึ้น แนะนำให้ทุกท่านได้ฝึกทำโจทย์ต่าง ๆ

โดยทุกท่านสามารถหาโจทย์ได้จากหนังสือหรือตำราทางสถิติทั่วไป

สำหรับหนังสือเล่มนี้ได้จัดทำชุดข้อมูลให้กับทุกท่าน 1 ชุด เพื่อใช้ในการฝึกฝน สามารถดาวน์โหลดได้ตามลิงค์ต่อไปนี้

■ Dataset Excel พร้อมโจทย์ ➔ <https://kku.world/zkpct>

■ Playlist บน YouTube ➔ <https://kku.world/cglmfz>

■ Textbook แนะนำแบบฝึกหัด ➔ Lind, DA., Marchal WG., & Wathen, SA. (2018). *Statistical Techniques in Business and Economics*. USA: McGraw-Hill Education



Dataset

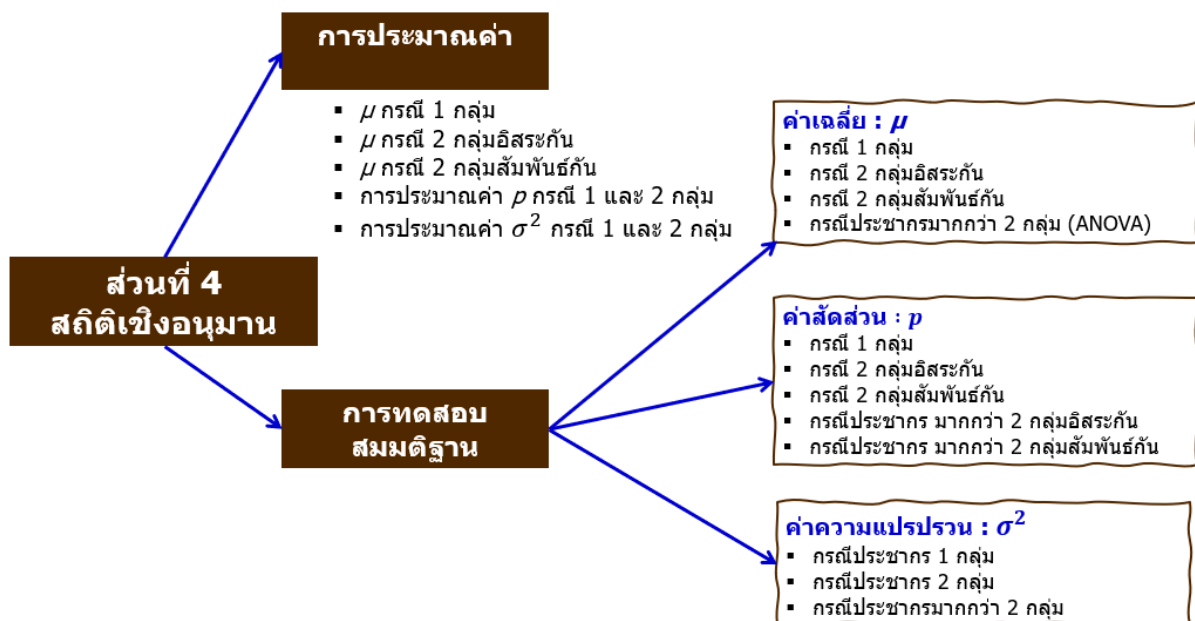


Playlist
YouTube

สารบัญโดยย่อ

หน้า

ส่วนที่ 1 สถิติศาสตร์และประเภทของสถิติเป็นอย่างไร	1
ส่วนที่ 2 ความน่าจะเป็นและการแจกแจงความน่าจะเป็นของตัวสถิติ	17
ส่วนที่ 3 สถิติเชิงพรรณนา	45
ส่วนที่ 4 สถิติเชิงอนุมาน	56
ส่วนที่ 5 การกำหนดขนาดตัวอย่างและการสุ่มตัวอย่าง	100
ส่วนที่ 6 การวิเคราะห์ความสัมพันธ์ของข้อมูล	111
ส่วนที่ 7 การทำนายด้วยสมการการถดถอย	128
ส่วนที่ 8 การพยากรณ์ด้วยอนุกรมเวลา	141
ส่วนที่ 9 การวิเคราะห์ข้อมูลเชิงจำแนก (Categorical Data Analysis)	158
ส่วนที่ 10 เลขดัชนี (Index Number)	164



สารบัญ

หน้า

ส่วนที่ 1 สถิติศาสตร์และประเภทของสถิติเป็นอย่างไร	1
1.1 สถิติศาสตร์และระเบียบวิธีการทางสถิติ	1
1.1.1 ความหมายของสถิติศาสตร์	1
1.1.2 ระเบียบวิธีการทางสถิติ	1
1.2 ประเภทของสถิติ	11
1.2.1 สถิติเชิงพรรณนา (Descriptive Statistics)	11
1.2.2 สถิติเชิงอนุมาน (Inferential Statistics)	12
1.3 ตัวแปร ข้อมูล และระดับการวัด (data types and measurement scales)	14
ส่วนที่ 2 ความน่าจะเป็นและการแจกแจงความน่าจะเป็นของตัวสถิติ	17
2.1 ความน่าจะเป็น (Probability)	17
2.2 การทดลองสุ่ม (Random Experiment)	18
2.3 หลักการนับ (Principle of Counting)	18
2.3.1 การจัดลำดับ (Permutation)	19
2.3.2 การจัดหมู่ (Combination)	20
2.4 เหตุการณ์ (Event)	20
2.5 ลักษณะของความน่าจะเป็น	22
2.5.1 Marginal Probability: สัญลักษณ์ $P(A)$	22
2.5.2 Joint Probability: สัญลักษณ์ $P(A \cap B)$ หรือ intersection	22
2.5.3 Conditional Probability: สัญลักษณ์ $P(A B)$	22
2.5.4 กฎของเบส์ - Bayes' theorem	23
2.6 การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม (X)	24
2.6.1 ตัวแปรสุ่ม X (Random variable: RV.)	24
2.6.2 การแจกแจงความน่าจะเป็น	25
2.6.3 ประเภทของการแจกแจงความน่าจะเป็น	26
2.6.3 ฟังก์ชันการแจกแจงความน่าจะเป็นแบบไม่ต่อเนื่อง	27
2.6.4 ฟังก์ชันการแจกแจงความน่าจะเป็นแบบต่อเนื่อง: โค้งปกติ (Normal Distribution)	28
2.7 ค่าคาดหวังของตัวแปรสุ่ม (Expected Value of Random Variables)	30
2.7.1 ค่าคาดหวังของตัวแปรสุ่ม X หรือ $E(X)$	30
2.7.2 ค่าคาดหวังกับความแปรปรวน	31
2.7.3 ค่าคาดหวังกับการคำนวณ	31
2.8 การแจกแจงความน่าจะเป็นของตัวอย่างสุ่ม (Sampling Distribution) และตัวสถิติ	31
2.8.1 ฟังก์ชันการแจกแจงความน่าจะเป็นของ $\bar{X}$	34
องศาเสรี (Degrees of Freedom $\mathbf{V} : df$) คืออะไร	36
ทฤษฎีลิมิตเข้าสู่ส่วนกลาง (Central Limit Theorem) คืออะไร	37

การตรวจสอบการแจกแจงของข้อมูลว่าเป็นโค้งปกติหรือไม่.....	37
2.8.2 ฟังก์ชันการแจกแจงความน่าจะเป็นของ σ^2	41
2.8.3 ฟังก์ชันการแจกแจงความน่าจะเป็นของ p	43
ส่วนที่ 3 สถิติเชิงพรรณนา	45
3.1 ค่ากลางของข้อมูล.....	45
3.2 ค่าการกระจายของข้อมูล.....	46
3.3 สัดส่วน (Proportion) อัตราส่วน (Ratio) กับ แด้มต่อ (Odds).....	48
3.4 การวัดตำแหน่งของข้อมูล	49
3.5 การอธิบายข้อมูลด้วยแผนภาพ	50
3.6 การอธิบายข้อมูลด้วยตารางการแจกแจงความถี่.....	53
3.7 อธิบายความสัมพันธ์ของข้อมูลด้วยแผนภาพ Scatter Plot.....	55
ส่วนที่ 4 สถิติเชิงอนุมาน	56
4.1 แนวคิดการประมาณค่า	56
4.1.1 คุณลักษณะเบื้องต้นของตัวประมาณที่ดี	56
4.1.2 การประมาณค่าทางสถิติ.....	57
4.2 การประมาณค่าแบบช่วงของ μ	59
4.2.1 กรณีประชากร 1 กลุ่ม	59
4.2.2 กรณีประชากร 2 กลุ่มอิสระกัน	61
4.2.3 กรณีประชากร 2 กลุ่มสัมพันธ์กัน	63
4.3 การประมาณค่าแบบช่วงของ p	65
4.4 การประมาณค่าแบบช่วงของ σ^2	66
4.5 แนวคิดการทดสอบสมมติฐาน.....	67
4.5.1 สมมติฐานทางสถิติ (statistical hypothesis).....	67
4.5.2 ขั้นตอนการทดสอบสมมติฐานทางสถิติ.....	68
4.6 การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ย.....	72
4.6.1 ค่าเฉลี่ย: กรณีประชากร 1 กลุ่ม	72
4.6.2 ค่าเฉลี่ย: กรณีประชากร 2 กลุ่มอิสระกัน	74
4.6.3 ค่าเฉลี่ย: กรณีประชากร 2 กลุ่มสัมพันธ์กัน	77
4.7 ค่าเฉลี่ย: กรณีประชากรมากกว่า 2 กลุ่มใช้ Analysis of Variance: ANOVA.....	79
4.7.1 นิยามศัพท์และข้อตกลงเบื้องต้น.....	80
4.7.2 กรณี One-Way ANOVA (CRD)	82
4.7.3 กรณี Two-Way ANOVA without replication (RCBD).....	84
4.7.4 กรณี Two-Way ANOVA With Replication (2 Factorial).....	87
4.7.5 ANOVA กับ Nonparametric.....	89
4.8 การทดสอบสมมติฐานเกี่ยวกับค่าสัดส่วน (p).....	91
4.8.1 ค่าสัดส่วน: กรณีประชากร 1 กลุ่ม.....	91

4.8.2 ค่าสัดส่วน: กรณีประชากร 2 กลุ่มอิสระกัน.....	92
4.8.3 ค่าสัดส่วน: กรณีประชากร 2 กลุ่มสัมพันธ์กัน.....	94
4.8.4 ค่าสัดส่วน: กรณีประชากรมากกว่า 2 กลุ่มอิสระกัน.....	95
4.8.5 ค่าสัดส่วน: กรณีประชากรมากกว่า 2 กลุ่มสัมพันธ์กัน.....	95
4.9 การทดสอบสมมติฐานเกี่ยวกับค่าความแปรปรวน (σ^2).....	96
4.9.1 กรณีประชากร 1 กลุ่ม.....	96
4.9.2 กรณีประชากร 2 กลุ่ม.....	97
4.9.3 กรณีประชากรมากกว่า 2 กลุ่ม.....	98
ส่วนที่ 5 การกำหนดขนาดตัวอย่างและการสุ่มตัวอย่าง.....	100
5.1 การกำหนดขนาดตัวอย่าง.....	100
5.1.1 กรณีที่ 1 จำนวนจากงบประมาณและทรัพยากรที่มีอยู่อย่างจำกัด.....	101
5.1.2 กรณีที่ 2 เพื่อตอบโจทย์การประมาณค่า.....	101
5.1.3 กรณีที่ 3 เพื่อตอบโจทย์การทดสอบสมมติฐาน.....	104
5.1.4 โปรแกรม G-Power เพื่อคำนวณขนาดตัวอย่าง.....	106
5.2 การสุ่มตัวอย่าง.....	108
5.2.1 การสุ่มตัวอย่างแบบใช้ความน่าจะเป็น (Probability Sampling).....	109
5.2.2 การสุ่มตัวอย่างแบบไม่ใช้ความน่าจะเป็น (Non-Probability Sampling).....	110
ส่วนที่ 6 การวิเคราะห์ความสัมพันธ์ของข้อมูล	111
6.1 การวิเคราะห์ความสัมพันธ์ กรณีข้อมูลเชิงปริมาณ.....	112
6.2 การวิเคราะห์ความสัมพันธ์ กรณีข้อมูลเชิงคุณภาพ.....	115
6.2.1 หากข้อมูลของตัวแปรใดตัวแปรหนึ่งเป็นระดับ Nominal.....	116
6.2.2 กรณีข้อมูลระดับ Ordinal.....	117
6.2.3 เงื่อนไขของ Chi-Square Test.....	119
6.2.4 วัดความสอดคล้อง (Measure of Agreement) ระหว่างผู้ประเมิน (Rater).....	122
ส่วนที่ 7 การทำนายด้วยสมการการถดถอย	128
7.1 การถดถอยเชิงเส้นอย่างง่าย.....	128
7.1.1 การสร้างสมการการถดถอยเชิงเส้น.....	128
7.1.2 การตรวจสอบความเหมาะสมของสมการ.....	131
7.1.3 การตรวจสอบข้อตกลงเบื้องต้น.....	132
7.2 การถดถอยเชิงพหุ.....	134
7.2.1 การสร้างสมการการถดถอยเชิงพหุ.....	134
7.2.2 การตรวจสอบความเหมาะสมของสมการ.....	137
7.2.3 การตรวจสอบข้อตกลงเบื้องต้น.....	138
7.3 การวัดประสิทธิภาพ.....	140

ส่วนที่ 8 การพยากรณ์ด้วยอนุกรมเวลา	141
8.1 ส่วนประกอบของอนุกรมเวลา	141
8.2 การวิเคราะห์อนุกรมเวลา	142
8.2.1 การวิเคราะห์แนวโน้ม (T)	143
1) วิธีกะประมาณด้วยสายตา.....	143
2) วิธีค่าเฉลี่ยเคลื่อนที่ (Moving Average : MA)	143
3) วิธี Linear trend Model	146
3.1) วิธี Selected two-points.....	146
3.2) วิธี Semi-average	147
3.3) วิธี Least squares.....	149
4) การเปลี่ยนแปลงสมการแนวโน้ม	152
8.2.2 การคำนวณโดยขจัดฤดูกาล (Seasonal Decomposition).....	153
ส่วนที่ 9 การวิเคราะห์ข้อมูลเชิงจำแนก (Categorical Data Analysis).....	158
9.1 การทดสอบความกลมกลืน (Goodness-of-fit-Test).....	158
9.2 การทดสอบความเป็นอิสระ (Test of Independence).....	161
9.3 การทดสอบความคล้ายคลึงกัน (Test of Homogeneity).....	162
ส่วนที่ 10 เลขดัชนี (Index Number).....	164
10.1 ความหมายของเลขดัชนี (Index Number)	164
10.2 ดัชนีสัมพัทธ์ (Relative Index).....	164
10.2.1 ดัชนีราคา.....	165
10.2.2 ดัชนีปริมาณ	165
10.2.3 ดัชนีมูลค่า.....	165
10.3 ดัชนีแบบไม่ถ่วงน้ำหนัก.....	166
10.3.1 ดัชนีรวมอย่างง่าย (Simple Aggregate)	166
10.3.2 ดัชนีค่าเฉลี่ยอย่างง่าย (Simple Average).....	166
10.4 ดัชนีถ่วงน้ำหนัก.....	167
10.4.1 ดัชนีรวมแบบถ่วงน้ำหนัก (Aggregate)	168
10.4.2 ดัชนีค่าเฉลี่ยสัมพัทธ์แบบถ่วงน้ำหนัก (Average).....	169
10.5 การเปลี่ยนฐานเลขดัชนี.....	170
บรรณานุกรม.....	171
ดัชนี	172
ประวัติผู้เขียน.....	175

ส่วนที่ 1 สถิติศาสตร์และประเภทของสถิติเป็นอย่างไร

1.1 สถิติศาสตร์และระเบียบวิธีการทางสถิติ

1.1.1 ความหมายของสถิติศาสตร์

สถิติ (Statistic) เป็นตัวเลขหรือค่าที่ได้จากการวิเคราะห์ชุดข้อมูลที่เราสนใจ **ข้อมูล** หมายถึง ตัวเลขหรือข้อความที่แทนความเป็นจริงของเรื่อง หรือสิ่งที่สนใจ เมื่อนำข้อมูลเหล่านั้นมาจัดระเบียบและกลั่นกรองด้วยหลักการทางสถิติศาสตร์ สิ่งที่ได้ คือ **สารสนเทศ** ซึ่งสารสนเทศที่ดีจะต้องอยู่ในรูปแบบที่ผู้อ่านสามารถเข้าใจได้ง่าย และสื่อความหมายได้ชัดเจนในเรื่องที่น่าเสนอ

สำหรับหลักการทางสถิติ ระเบียบวิธีการทางสถิติ หรือกระบวนการทางสถิตินั้น นับได้ว่าเป็นส่วนประกอบที่สำคัญในความหมายของคำว่าสถิติศาสตร์ สำนักงานสถิติแห่งชาติ (NSO) ได้ให้นิยามสถิติศาสตร์ที่สามารถสรุปได้ว่า **สถิติศาสตร์ (Statistics)** เป็นศาสตร์ที่ว่าด้วยหลักการ และระเบียบวิธีการทางสถิติ ที่ประกอบด้วยการเก็บรวบรวมข้อมูล การจัดระบบข้อมูล การวิเคราะห์ข้อมูล และการแปลความหมายหรือตีความข้อมูลหรือผลลัพธ์ รวมไปถึงการนำเสนอข้อมูลและข้อสรุปที่ได้จากการวิเคราะห์

สถิติศาสตร์: เป็นศาสตร์ที่ว่าด้วยหลักการ และระเบียบวิธีการทางสถิติ ซึ่งประกอบด้วย การเก็บรวบรวมข้อมูล การจัดระบบ การนำเสนอข้อมูล การวิเคราะห์ข้อมูล และการแปลความหมายข้อมูล (**สำนักงานสถิติแห่งชาติ, (2565)**)

Statistics is the science of collecting, organizing, presenting, analyzing, and interpreting numerical data to assist in making more effective decisions. (**Lind D.A., and et al., (2018)**)

Data literacy เป็นการสร้างทักษะ หรือสร้าง Mindset สร้างจิตสำนึก ในการค้นหาข้อมูล หรือพูดคุยแลกเปลี่ยนกับบนข้อมูล

1.1.2 ระเบียบวิธีการทางสถิติ

ระเบียบวิธีการทางสถิติ (Statistical Method) หมายถึง กระบวนการ หรือขั้นตอน ตั้งแต่ระเบียบวิธีการเก็บรวบรวมข้อมูลหรือการได้มาซึ่งข้อมูล ไปจนถึงการวิเคราะห์ การนำเสนอข้อมูลและสารสนเทศที่ได้จากการวิเคราะห์ หากเป็นการทดลองหรือการสำรวจด้วยตัวอย่าง จำเป็นต้องคำนวณหาขนาดตัวอย่างที่เหมาะสม รวมถึงการวางแผนในการสุ่มตัวอย่างที่มั่นใจได้ว่าจะสามารถได้ตัวอย่างที่เป็นตัวแทนที่ดีของประชากรได้ ซึ่งระเบียบวิธีทางสถิติถือว่าเป็นเครื่องมือ (Tool) สำคัญของนักวิจัย นักประเมินผล นักวิชาการ และนักบริหารทุกสาขา สามารถเขียนสรุปเป็นกรอบแนวคิดได้ตามรูปที่ 1.1 โดยมีประเด็นสำคัญดังนี้

1) การตั้งโจทย์วิจัย การเลือกโจทย์วิจัยหรือหัวข้อของงานวิจัยที่ดีเป็นประเด็นสำคัญ เพราะจะช่วยให้ผู้วิจัยสามารถดำเนินงานวิจัยได้อย่างมีประสิทธิภาพและประสบความสำเร็จ **โจทย์วิจัยที่ดี** ควรเป็นเรื่องที่ก่อให้เกิดองค์ความรู้ใหม่ ๆ และมีประโยชน์สามารถนำไปประยุกต์ปรับใช้ได้จริง ชื่อเรื่องวิจัยควรเป็นคำถามหรือข้อความที่เรียบเรียงสละสลวยได้ใจความที่สมบูรณ์ กระชับ ชัดเจน หากเป็นไปได้อ่านแล้วควรตีความได้ว่า จะทำอะไร กับประชากรกลุ่มใด เพื่อแก้ปัญหาอะไร ในพื้นที่ใด ตัวแปรและวิธีการวิเคราะห์เป็นอย่างไร จากนั้น ผู้วิจัยควรมีการกำหนด**คำถามวิจัย** ซึ่งอาจมีคำถามปลีกย่อยอีกหลายประเด็นตามที่ถูกผู้วิจัยเกิดข้อสงสัย เช่น ระหว่างเพศมีความแตกต่างกันหรือไม่ ระหว่างกลุ่มอาชีพมีความแตกต่างกันหรือไม่ เป็นต้น

ตัวอย่าง Ex.1 วิจัยเรื่อง พฤติกรรมการซื้อสินค้าออนไลน์ของผู้บริโภคสูงอายุ จังหวัด xyz ในยุคดิจิทัล ★ ชื่อเรื่องดังกล่าวทำให้ทราบว่าวิเคราะห์พฤติกรรมการซื้อสินค้าบนออนไลน์ในยุคดิจิทัล โดยประชากรเป็นผู้สูงอายุ ในพื้นที่จังหวัด xyz

ตัวอย่าง Ex.2 วิจัยเรื่อง การพัฒนาระบบเพื่อทำนายการพ้นสภาพจากนักศึกษาในระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัย xyz ด้วยเทคนิคเหมืองข้อมูล ★ ชื่อเรื่องดังกล่าวทำให้ทราบว่าพัฒนาระบบเพื่อทำนายการพ้นสภาพจากนักศึกษา ด้วยวิธีการวิเคราะห์ข้อมูลที่ใช้เทคนิคเหมืองข้อมูล โดยประชากรเป็นนักศึกษาระดับปริญญาตรี พื้นที่ที่ใช้ในการศึกษา คือ คณะวิทยาศาสตร์ มหาวิทยาลัย xyz



รูปที่ 1.1 กรอบแนวคิดของระเบียบวิธีการทางสถิติ

2) การเก็บรวบรวมข้อมูล จากโจทย์วิจัยหรือวัตถุประสงค์ของการวิจัยที่ชัดเจน จะทำให้เราสามารถระบุรูปแบบการเก็บรวบรวมข้อมูลได้เหมาะสม รวมไปถึงระบุตัวแปรและประเภทของข้อมูลได้ชัดเจนด้วย ซึ่งในบางครั้งข้อมูลอาจจะมาจากหลากหลายแหล่ง โดยเฉพาะอย่างยิ่งในยุคของ Big data นั้น ข้อมูลต่างๆ เกิดขึ้นบน internet หรือสื่อสังคมออนไลน์ในลักษณะที่หลากหลายรูปแบบ ซึ่งต้องใช้หลักการทางวิทยาศาสตร์ข้อมูล เพื่อสกัด (Extract) ข้อมูล มาจากฐานข้อมูลที่หลากหลาย และทำการตรวจสอบคุณภาพของข้อมูลหรือแปลงรูปของข้อมูล (Transform) ให้อยู่ในลักษณะที่เหมาะสมก่อนที่จะนำเข้า (Load) สู่กระบวนการวิเคราะห์ต่อไป ทั้งนี้วิธีการเก็บรวบรวมข้อมูลเบื้องต้นอาจแบ่งได้เป็น 2 แบบ ดังนี้

ข้อมูลปฐมภูมิ : เป็นการเก็บรวบรวมข้อมูลที่ผู้วิจัยเก็บจากผู้ให้ข้อมูลโดยตรง

ข้อมูลทุติยภูมิ : เป็นข้อมูลที่ผู้วิจัยไม่ได้เก็บจากผู้ให้ข้อมูล รวมทั้งไม่ได้เป็นผู้วางแผนในการจัดเก็บข้อมูลเหล่านั้นด้วยตนเอง

2.1) แบบปฐมภูมิ โดยทั่วไปมี 3 วิธี ได้แก่ การบันทึก การสำรวจ และการทดลอง

สำหรับการเก็บรวบรวมข้อมูลด้วยวิธีการ**บันทึก**นั้น เกิดขึ้นเป็นประจำในหน่วยงานต่างๆ เช่น การรับ order จากลูกค้า การแจ้งเกิด การแจ้งตาย เป็นต้น สิ่งที่สำคัญของการบันทึก คือการออกแบบ แบบฟอร์มที่ใช้ในการบันทึก จะต้องมีความชัดเจน เพื่อให้ผู้กรอกข้อมูลเข้าใจตรงกัน และกรอกข้อมูลได้ถูกต้องและตรงตามความเป็นจริง

ส่วนการเก็บรวบรวมข้อมูลจากการ**ทดลอง** เป็นการวิเคราะห์เพื่อตอบคำถามเฉพาะเจาะจงเกี่ยวกับอิทธิพลของปัจจัยที่สนใจ โดยพยายามควบคุมปัจจัยอื่นที่ไม่เกี่ยวข้อง ดังนั้น การออกแบบแผนแบบการทดลองจึงเป็นเรื่องที่สำคัญ ซึ่งอาจต้องแยกผลของปัจจัยอื่นที่ไม่ต้องการศึกษาออกจากปัจจัยที่ต้องการศึกษาให้ได้ เช่น ต้องการทราบว่าปุ๋ย 3 ยี่ห้อ ทำให้ผลผลิตต่อไร่ของข้าวพันธุ์ A แตกต่างกันหรือไม่

ส่วนการเก็บรวบรวมข้อมูลด้วยการ**สำรวจ**นั้น เครื่องมือที่ใช้มักจะเป็นแบบสอบถามหรือแบบประเมิน ที่ควรตรวจสอบคุณภาพของเครื่องมือก่อนนำไปใช้ ซึ่งการตรวจสอบว่าแบบสอบถามมี**คุณภาพเพียงพอหรือไม่**นั้น ในเบื้องต้นควรดำเนินการตรวจสอบความตรงเชิงเนื้อหา (Validity) และตรวจสอบความเชื่อมั่นของแบบสอบถาม (Reliability) ดังนี้

ก) ตัวแปรที่จัดเก็บต้องครอบคลุมประเด็นที่ต้องการศึกษา หรือต้องมีความ**ตรงกับเนื้อหา**วิจัยที่เรากำลังค้นคว้าอยู่เรียกว่า Validity ซึ่งจำเป็นต้องใช้ผู้เชี่ยวชาญในเรื่องดังกล่าวช่วยตรวจสอบ ส่วนใหญ่จะใช้ผู้เชี่ยวชาญ 3 คนขึ้นไปและควรเป็นเลขคี่ โดยนำความเห็นของผู้เชี่ยวชาญทุกคนมาวิเคราะห์**ค่าดัชนีความสอดคล้อง**ระหว่างข้อคำถามกับวัตถุประสงค์และเนื้อหา (IOC : Index of item objective congruence) หากคำถามข้อนั้นๆ มีค่า IOC อยู่ระหว่าง 0.5 ถึง 1 จะถือว่ามีความเที่ยงตรงสามารถนำมาใช้ได้ แต่อย่างไรก็ตามควรนำข้อเสนอแนะต่าง ๆ จากผู้เชี่ยวชาญมาประกอบการปรับแก้แบบสำรวจก่อนที่จะนำไปใช้งาน

ข) ตรวจสอบ**ความน่าเชื่อถือของแบบสอบถาม** โดยนำแบบสอบถามที่ผ่าน IOC และได้ทำการปรับแก้แล้วไปทดลองใช้งาน (Try Out) กับกลุ่มตัวอย่างที่มีลักษณะคล้ายคลึงกับกลุ่มตัวอย่างที่จะนำไปใช้จริง โดยเก็บมาอย่างน้อย 30 ชุด จากนั้นนำมาวิเคราะห์ค่าสถิติ Cronbach's Alpha ซึ่งเป็นสถิติทดสอบความเชื่อมั่น (Reliability) โดยเป็นการตรวจสอบความสัมพันธ์ภายในกลุ่มตัวแปร (Internal Consistency) โดยเฉพาะอย่างยิ่งตัวแปรในกลุ่มของความคิดเห็นหรือทัศนคติ ซึ่งควรมีค่ามากกว่า 0.7 จึงจะถือว่าใช้งานได้

ตัวอย่าง Ex.3 วิจัยเรื่อง พฤติกรรมการลงทุน Cryptocurrency ของพนักงานภาครัฐบาล จังหวัดขอนแก่น ★

เรื่องดังกล่าวข้อมูลเกิดขึ้นอยู่แล้วตามธรรมชาติจากหน่วยตัวอย่างที่เราจะเข้าไปศึกษา ดังนั้น วิธีการ**สำรวจแบบปฐมภูมิ** จึงเป็นวิธีการที่เหมาะสม โดยต้องวางแผนการสุ่มตัวอย่างกลุ่มพนักงานภาครัฐบาลในพื้นที่จังหวัดขอนแก่น และการเก็บรวบรวมข้อมูลจะต้องเก็บจากผู้ให้ข้อมูลโดยตรง เครื่องมือที่เหมาะสมคือ แบบสอบถาม ที่ต้องมีคุณภาพเพียงพอ

- Ex 3.1 ในรูปที่ 1.2 การคำนวณ IOC ของข้อคำถาม 4 ข้อ โดยผู้เชี่ยวชาญ 3 คน พบว่าแต่ละข้อมีค่า IOC มากกว่า 0.5 ตามหมายเลข ❶ แสดงว่าแต่ละข้อมีความเที่ยงตรง แต่อย่างไรก็ตามควรนำความคิดเห็นและข้อเสนอแนะของผู้เชี่ยวชาญมาปรับแก้ข้อคำถามให้มีความเหมาะสมและชัดเจนยิ่งขึ้น

เกณฑ์การให้คะแนนความคิดเห็นของผู้เชี่ยวชาญ

- +1 ถ้าแน่ใจว่าข้อคำถามนั้น วัดได้ตรงหรือสอดคล้องตามวัตถุประสงค์
0 ถ้าไม่แน่ใจว่าข้อคำถามนั้น วัดได้ตรงตามวัตถุประสงค์
-1 ถ้าแน่ใจว่าข้อคำถามนั้น วัดได้ไม่ตรงหรือไม่สอดคล้องตามวัตถุประสงค์

$$IOC = \frac{\sum R_i}{N} \quad \text{Rovinelli, R. J., and Hambleton, R. K., (1997)}$$

$\sum R_i$ คือ ผลรวมของคะแนนความคิดเห็นของผู้เชี่ยวชาญ
: $i=1,2,...,N$

N คือจำนวนของผู้เชี่ยวชาญ

Ex 3.1 IOC : Index of item objective congruence ❶

ประเด็นชี้วัด	คนที่ 1	คนที่ 2	คนที่ 3	Sum	Average (IOC)
1. คุกกี้ไอศกรีมเป็นทางเลือกในการทำธุรกรรมทางการเงินซึ่งมีผลกระทบต่อระดับการตัดสินใจลงทุนของท่านในระดับใด	1	0	1	2	0.67
2. คุกกี้ไอศกรีมสามารถเพิ่มมูลค่าได้จากการเทรดเพื่อเกรงกำไร ซึ่งมีผลกระทบต่อระดับการตัดสินใจลงทุนของท่านในระดับใด	1	1	1	3	1.00
3. คุกกี้ไอศกรีมมีส่วนช่วยให้ท่านสามารถใช้จ่ายได้สะดวกขึ้น ซึ่งมีผลกระทบต่อระดับการตัดสินใจลงทุนของท่านในระดับใด	1	1	0	2	0.67
4. ข้อมูลข่าวสารเกี่ยวกับคุกกี้ไอศกรีมมีผ่านสื่อออนไลน์ ทำให้ท่านสนใจลงทุนคุกกี้ไอศกรีมในระดับใด	1	1	1	3	1.00

แต่ละข้อควรอยู่ระหว่าง 0.5 ถึง 1



รูปที่ 1.2 ตัวอย่างเรื่อง IOC

- Ex 3.2 ในรูปที่ 1.3 ผลการคำนวณ Cronbach's Alpha กรณีตรวจสอบข้อคำถาม 6 ข้อ (A1 to A6) ตามหมายเลข ❷ ได้ค่าความเชื่อมั่นเท่ากับ 0.903 ซึ่งมากกว่า 0.7 แสดงว่าของแบบสอบถามชุดนี้มีความน่าเชื่อถือเพียงพอที่จะนำไปใช้ได้

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right) \quad \text{Cronbach, L. J. (1970)}$$

σ_i^2 คือ ความแปรปรวนของคำถามข้อที่ $i : i=1,2,3,...,k$
 σ_x^2 คือ ความแปรปรวนรวมของคะแนนทั้งหมดของผู้ตอบ
k คือ จำนวนข้อคำถาม

กลุ่มตัวแปรทัศนคติ A1 - A6
มีค่า Cronbach's Alpha เท่ากับ 0.903

Ex 3.2 Reliability Statistics ❷

Cronbach's Alpha	N of Items
0.903	6

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation
A1	17.23	21.609	.700	.548
A2	17.48	22.456	.691	.542
A3	17.23	20.425	.732	.566
A4	17.57	21.840	.741	.572
A5	17.43	21.473	.784	.633
A6	17.29	22.005	.773	.666

รูปที่ 1.3 ตัวอย่างเรื่อง Reliability ของข้อคำถามในแบบสอบถาม

นอกจากนี้จะต้องกำหนดขนาดตัวอย่าง และและออกแบบวิธีการเก็บรวบรวมข้อมูลให้ชัดเจน โดยเฉพาะอย่างยิ่งวิธีการสุ่มตัวอย่าง เพื่อให้ได้มาซึ่งข้อมูลที่เป็นตัวแทนที่ดีของประชากร

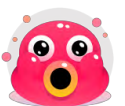
2.2) ข้อมูลทุติยภูมิ เป็นข้อมูลที่ใช้หรือผู้วิจัยไม่ได้สัมผัสโดยตรงกับเจ้าของข้อมูล รวมทั้งไม่ได้เป็นผู้วางแผนในการจัดเก็บข้อมูลด้วยตนเอง เช่น นำข้อมูลมาจากรฐานข้อมูลต่างๆ ที่มีการรวบรวมข้อมูลไว้แล้วโดยหน่วยงานภาครัฐ ภาคเอกชน หรือมาจากข้อมูลที่เก็บรวบรวมไว้แล้วบนเครือข่ายบนอินเทอร์เน็ต ซึ่ง

ข้อมูลเหล่านี้อาจจะไม่ตรงกับวัตถุประสงค์ของผู้วิจัย หรือมีรายละเอียดไม่เพียงพอที่จะนำไปวิเคราะห์ ดังนั้น **การตรวจสอบข้อมูลทฤษฎี**จึงเป็นเรื่องที่สำคัญ โดยควรมีการตรวจสอบความเชื่อถือได้ของแหล่งข้อมูล ความเหมาะสมของขนาดตัวอย่างวิธีการเก็บรวบรวมข้อมูลว่าเหมาะสมหรือไม่ ตรวจสอบความครบถ้วนสมบูรณ์ของข้อมูลว่าครอบคลุมหรือสอดคล้องกับประเด็นหรือวัตถุประสงค์ของผู้วิจัยที่ต้องการศึกษาหรือไม่ รวมถึงตรวจสอบความทันสมัยของข้อมูล และความแนบในหรือความคงเส้นคงวา หมายถึงข้อมูลมีความถูกต้องและสอดคล้องกับแหล่งข้อมูลอื่น ๆ หรือไม่ ตามตัวอย่าง Ex.4

ตัวอย่าง Ex.4 เรื่อง การพัฒนาระบบเพื่อทำนายการฟื้นสภาพจากนักศึกษาระดับปริญญาตรี คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น ด้วยเทคนิคการทำเหมืองข้อมูล เรื่องดังกล่าวข้อมูลเกิดขึ้นอยู่แล้วตามธรรมชาติจากหน่วยตัวอย่างที่เราจะเข้าไปศึกษา แต่เราจำเป็นต้องใช้ข้อมูลจำนวนมากเพียงพอเพื่อนำมาทำเหมืองข้อมูล ซึ่งระบบทะเบียนนักศึกษาของมหาวิทยาลัย ได้มีการจัดเก็บข้อมูลเหล่านี้อยู่แล้ว ดังนั้น วิธีการเก็บรวบรวมข้อมูลแบบทฤษฎี จึงเป็นวิธีการที่เหมาะสม ซึ่งสามารถนำข้อมูลย้อนหลังมากกว่า 3 ปีมาศึกษาได้ ซึ่งตัวแปรที่ต้องใช้ในการวิเคราะห์ คือสาเหตุต่าง ๆ เกี่ยวกับการฟื้นสภาพของนักศึกษาแต่ละคน ในแต่ละสาขาวิชา เป็นต้น ดังนั้น สำหรับกรณีนี้ผู้วิจัยควรออกแบบตารางข้อมูล หรือแบบบันทึกที่ระบุว่าการตัวแปรอะไรบ้าง รวมถึงวางแผนการตรวจสอบคุณภาพข้อมูลที่ได้และการจัดการกับข้อมูลที่ผิดปกติ

3) การจัดการ การตรวจสอบ และการนำเสนอข้อมูลเบื้องต้น เป็นกระบวนการทำความสะอาดข้อมูล และจัดการกับข้อมูลก่อนเข้าสู่กระบวนการวิเคราะห์ข้อมูลต่อไป โดยผู้วิจัยจะต้องใช้ความรู้ ความเข้าใจจากการศึกษาทฤษฎีและวรรณกรรมต่างๆ รวมถึงความเชี่ยวชาญเกี่ยวกับเรื่องที่ทำวิจัยมาช่วยในการตรวจสอบและดูความเป็นไปได้ของชุดข้อมูล หรือเรียกว่า **การวิเคราะห์ข้อมูลเชิงสำรวจ** (Exploratory Data Analysis: EDA)

EDA: เป็นการทำความเข้าใจกับชุดข้อมูลในหลาย ๆ มุมมองกับทุก ๆ ตัวแปร ว่ามีลักษณะเป็นอย่างไร



EDA สามารถทำได้หลายวิธี เพื่อตรวจสอบว่าข้อมูลมีความถูกต้องและน่าเชื่อถือหรือไม่ โดยหัวใจสำคัญของการทำ EDA คือการวิเคราะห์ข้อมูลด้วยสถิติเชิงพรรณนา ซึ่งอาจจะเป็นค่าทางสถิติ หรือแผนภาพ (Visualization) หรือกราฟ มาประกอบกันในการพิจารณาตัวแปรทีละตัวแปร หรือระหว่างตัวแปรด้วยกันเอง เพื่อนำไปสู่การทำความสะอาดข้อมูล (Data Cleansing) โดยในขั้นตอนนี้มีได้คำนึงถึงการทดสอบสมมติฐานแต่อย่างใด โดยทั่วไปจะประกอบไปด้วยขั้นตอนดังนี้

3.1) ทำความเข้าใจข้อมูล โดยพิจารณาจากข้อมูลเชิงพรรณนา เช่น ขนาดของข้อมูล ประเภทของข้อมูล แหล่งที่มาของข้อมูล รูปแบบของข้อมูล ลักษณะของข้อมูล และวัตถุประสงค์ของการใช้ข้อมูล เป็นต้น

3.2) ตรวจสอบความถูกต้อง หรือค้นหาความผิดปกติของข้อมูล ซึ่งในขั้นตอนนี้จำเป็นต้องใช้องค์ความรู้ของผู้วิจัยที่ได้มาจากการทบทวนทฤษฎีและวรรณกรรมต่าง ๆ มาประกอบการพิจารณาว่าข้อมูลชุดดังกล่าวผิดปกติหรือไม่ ไม่ว่าจะเป็นเรื่องความเป็นไปได้หรือความถูกต้องของขอบเขตข้อมูล เช่น รายได้ต่ำสุดหรือสูงสุดของกลุ่มที่สนใจ หรือความแปรปรวนระหว่างตัวแปรตั้งแต่ 2 ตัวขึ้นไป ค่าไม่สมบูรณ์หรือขาดหาย (Missing Value) ข้อมูลซ้ำซ้อน ข้อมูลผิดปกติประเภท เป็นต้น

3.2.1) การวิเคราะห์ด้วยสถิติเชิงพรรณนา โดยนำตัวแปรแต่ละตัวมาพิจารณาค่าสถิติเบื้องต้นต่าง ๆ เช่น พิจารณาการแจกแจงความถี่หรือตารางการแจกแจงความถี่ ค่ากลาง ค่าการกระจายของข้อมูล ค่าสูงสุด ค่าต่ำสุด ค่าสุดโต่ง หรือพิจารณาคุณลักษณะระหว่างตัวแปร เช่น พิจารณาความสัมพันธ์ของตัวแปร 2 ตัว ด้วยค่า Correlation เป็นต้น ซึ่งจะทำให้เห็นว่าข้อมูลมีการกระจายตัวเป็นอย่างไรในแต่ละกลุ่ม

3.2.2) การแจกแจงของข้อมูลจากกราฟ เช่น ตรวจสอบความสัมพันธ์ระหว่างตัวแปรด้วยกราฟ Scatter plot พิจารณากราฟ histogram เพื่อตรวจสอบลักษณะการกระจายของข้อมูลว่าสมมาตรรอบค่ากลางหรือไม่ ใช้กราฟ Stem-and-Leaf Plots และ Box-and-Whiskers Plot เพื่อพิจารณาชุดข้อมูลดังกล่าวว่ามีค่าผิดปกติหรือค่าสุดโต่ง (Extreme Value) หรือไม่ จำนวนมากน้อยเพียงใด

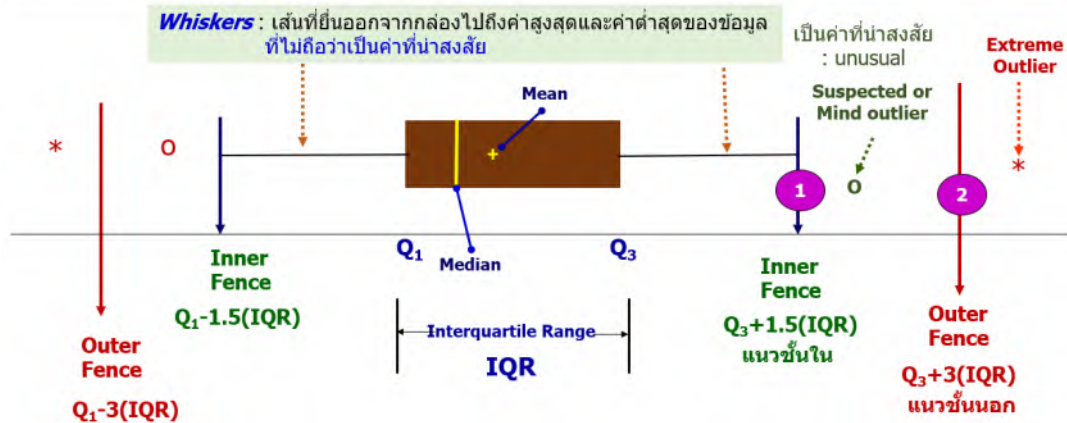
ก) กราฟแจกแจงความถี่หรือ **Histogram** แสดงลักษณะของข้อมูลว่าสมมาตรรอบค่ากลางหรือไม่ และถ้าไม่สมมาตรส่วนใหญ่ข้อมูลจะอยู่ในบริเวณใด หรือมีลักษณะเบ้ไปในทิศทางใด ซึ่งทำให้เกิดข้อสงสัยต่อได้ว่าชุดข้อมูลดังกล่าวอาจมีค่าผิดปกติ (Outlier) เกิดขึ้นหรือไม่

ข) แผนภาพ **Stem and Leaf Plot** ที่ทำให้เราทราบว่าค่าสุดโต่งเกิดขึ้นจำนวนมากน้อยเพียงใด และสุดโต่งไปในทางด้านค่าสูงหรือค่าต่ำ

ค) แผนภาพ **Box Plot** ทำให้เราทราบว่าชุดข้อมูลดังกล่าวมีค่า Outlier เกิดหรือไม่ และเป็น Outlier ชนิดใด ในทางทฤษฎี outlier ของ Box Plot นั้น มีด้วยกัน 2 ลักษณะตามรูปที่ 1.4 ถ้าห่างจาก Q1 และ Q3 เป็น 1.5 ของ IQR จะเรียกว่า **Suspected** หรือ **Mild Outlier** ตามหมายเลข ❶ ส่วนลักษณะที่ 2 คือ 3 เท่าของ IQR จะเรียกว่า **Extreme Outlier** ตามหมายเลข ❷ ทั้ง 2 ผัง

ตัวอย่าง Ex.1 ในรูปที่ 1.5 เป็นตัวอย่างการสร้างตารางแจกแจงความถี่ หากมีข้อมูลชุดหนึ่งที่มี $n=180$ ค่า $\text{Max} = \$3,292$ และ $\text{Min} = \$294$ คำนวณจำนวนช่วงชั้นได้ค่า $k = 8.44$ สำหรับกรณีนี้ผู้วิจัยกำหนดเป็น 8 ช่วงชั้น และคำนวณค่าต่าง ๆ ได้ตามที่ปรากฏอยู่ในหมายเลข ❶ เมื่อนำมาสร้างเป็นตารางการแจกแจงความถี่จะได้ค่าต่าง ๆ ตามที่ปรากฏในหมายเลข ❷ ซึ่งจะเห็นว่าข้อมูลส่วนใหญ่กระจายตัวอยู่ที่ช่วงชั้นที่ 4, 5 และ 6 ทั้งนี้ทุกท่านสามารถใช้โปรแกรม Excel หรือ SPSS ในการสร้างตารางการแจกแจงความถี่ รวมถึงกราฟ Histogram ได้ตามที่ได้เห็นในหมายเลข ❸ และ ❹

A BOX PLOT เป็นกราฟทางสถิติที่แสดงการกระจายของข้อมูล ทำให้เห็นลักษณะว่าข้อมูลชุดดังกล่าวมีความสมมาตรหรือไม่ รวมถึงสามารถเห็นค่าผิดปกติของข้อมูลได้ โดยค่าสถิติที่นำมาใช้ประกอบการพิจารณา คือ Q1, Q3 และค่ามัธยฐาน



รูปที่ 1.4 การวิเคราะห์ข้อมูลผิดปกติด้วย Box Plot

การสร้างตารางแจกแจงความถี่และกราฟฮิสโตแกรม

- ขั้นที่ 1 กำหนดจำนวนช่วงชั้น (k) ตามที่ผู้วิจัยต้องการหรือสามารถใช้สูตรได้ $k = 1 + 3.3(\log n)$
- ขั้นที่ 2 คำนวณค่าความกว้างของแต่ละช่วงชั้น (i) $i \geq (\text{Max} - \text{Min})/k$
- ขั้นที่ 3 กำหนดขอบเขตล่างและขอบเขตบน Class Limits ด้วยการ ± 0.5
- ขั้นที่ 4 นับความถี่ที่จะตกอยู่ในแต่ละช่วงชั้น

Ex1. ข้อมูลชุดหนึ่ง มี $n=180$ ค่า $\text{Max} = \$3,292$ และ $\text{Min} = \$294$

TABLE 2-4 Profit on Vehicles Sold Last Month by the Applewood Auto Group

Profit	Profit	Profit	Profit	Profit	Profit	Profit	Profit	Profit	Profit
\$1,387	\$2,148	\$2,201	\$ 963	\$ 820	\$2,230	\$3,043	\$2,594	\$2,379	
1,754	2,207	998	1,298	1,268	2,341	1,559	2,666	2,837	
1,817	2,252	2,813	1,410	1,741	2,380	1,674	2,991	1,426	
1,040	1,420	323	1,553	1,772	1,100	1,807	934	2,944	
1,273	1,889	352	1,648	1,832	1,295	2,554	2,063	2,141	
1,529	1,166	482	2,071	2,350	1,344	2,236	2,083	1,973	
3,082	1,320	1,144	2,116	2,422	1,906	2,828	2,856	2,502	
1,951	2,265	1,485	1,500	2,440	1,932	1,269	2,989	763	
2,892	1,323	1,509	1,549	909	2,070	1,717	910	1,538	
1,266	1,791	1,638	2,240	978	2,454	1,797	1,536	2,339	
1,342	1,919	1,561	2,408	1,236	1,806	1,955	1,957	2,709	
443	2,357	2,127	2,398	1,818	1,680	2,189	2,240	2,222	
754	2,966	2,430	1,115	1,824	1,827	2,482	2,466	2,589	
1,621	732	1,704	1,124	1,807	1,915				
870	1,644	1,876	1,532	1,838	2,084				
1,174	1,626	2,010	1,808	1,840	2,639				
1,472	1,781	2,185	1,822	2,197	842				
1,809	1,915	2,221	1,897	2,646	1,963				
2,415	2,119	2,389	2,445	1,461	2,059	1,110	1,736	1,841	
1,546	1,766	335	2,898	1,731	2,336	1,178	2,058	2,487	

$n = 180$

$k = 8.44 \approx 8$

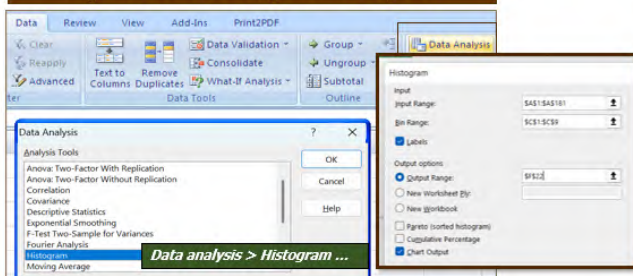
$\text{Max} = 3292$

$\text{Min} = 294 \approx 200$

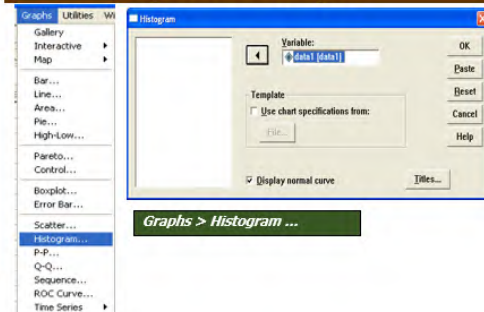
$i = 355.11 \approx 400$

200-0.5		199.5+400			
ขอบเขตล่าง	ขอบเขตบน	ขีดจำกัดล่าง	ขีดจำกัดบน	Midpoint	ความถี่
1	199.5-599.5	200-599	399.5	8	
2	599.5-999.5	600-999	799.5	11	
3	999.5-1399.5	1000-1399	1199.5	23	
4	1399.5-1799.5	1400-1799	1599.5	38	
5	1799.5-2199.5	1800-2199	1999.5	45	
6	2199.5-2599.5	2200-2599	2399.5	32	
7	2599.5-2999.5	2600-2999	2799.5	19	
8	2999.5-3399.5	3000-3399	3199.5	4	

By Excel : Analysis Toolpak



By SPSS : Graphs



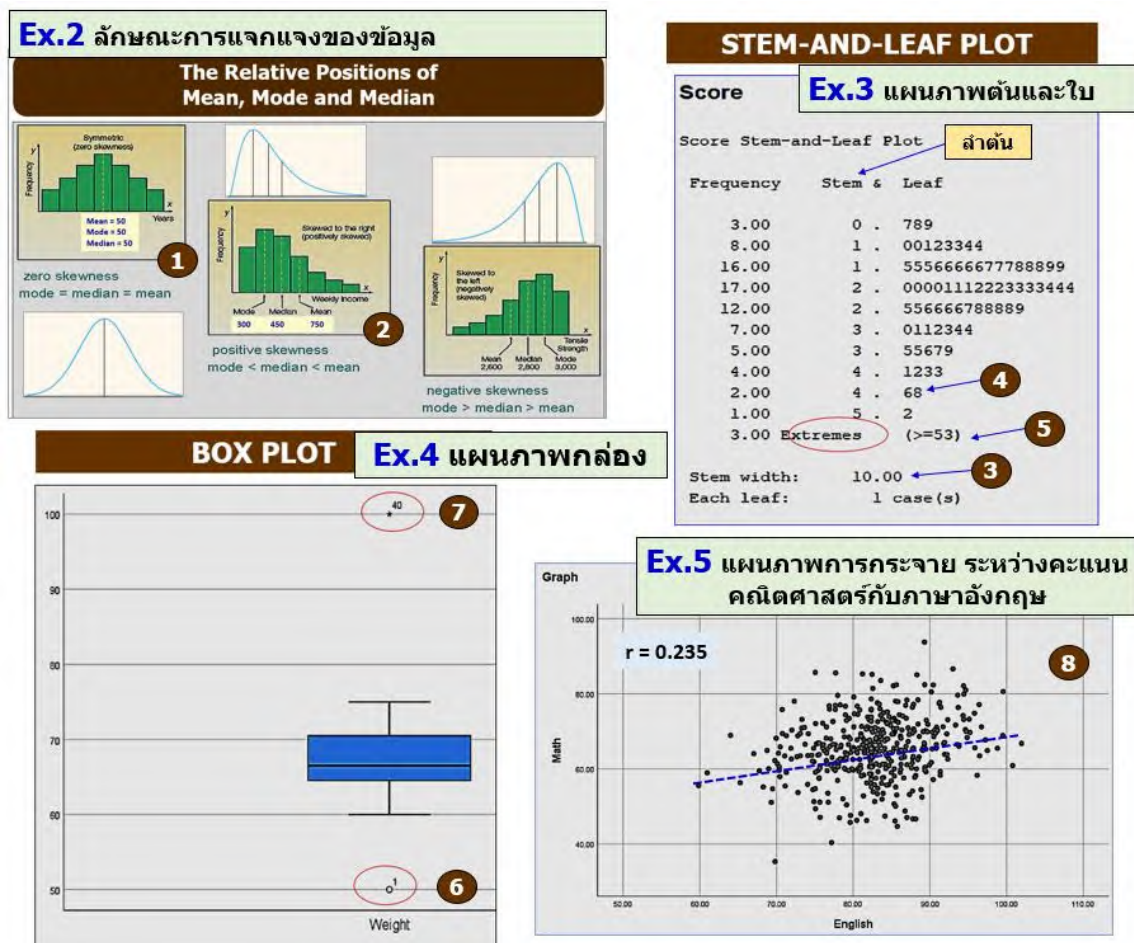
รูปที่ 1.5 การสร้างตารางการแจกแจงความถี่และกราฟฮิสโตแกรม

ตัวอย่าง Ex.2 หมายเลข ① ในรูปที่ 1.6 เป็นตัวอย่างการพิจารณาลักษณะการแจกแจงของข้อมูลผ่านกราฟฮิสโตแกรม หากข้อมูลสมมาตรจะมีค่า Mean, Median, Mode เท่ากัน * สำหรับหมายเลข ② แสดงให้เห็นว่าข้อมูลชุดนั้นมีค่า $\text{Mean} > \text{Median} > \text{Mode}$ แสดงว่าลักษณะของการแจกแจงจะเบ้

ขวาหรือเบ้เป็นบวก * หาก $\text{Mean} < \text{Median} < \text{Mode}$ แสดงว่าลักษณะของการแจกแจงจะเบ้ซ้ายหรือเบ้เป็นลบ

ตัวอย่าง Ex.3 หมายเลข ③ ในรูปที่ 1.6 เป็นค่า stem width = 10 เพื่ออธิบายว่าหน่วยนับของข้อมูลชุดนี้ที่ลำดับมีค่าเป็น 10 หน่วย ดังนั้นในหมายเลข ④ จึงแปลผลได้ว่า มีค่าข้อมูล 46 และ 48 อย่างละ 1 ค่า ทำให้ช่วงชั้นนี้มีความถี่เป็น 2 นั่นเอง * ส่วนหมายเลข ⑤ รายงานว่ามีค่าสุดโต่ง (Extremes) จำนวน 3 ค่า ซึ่งมีค่า ≥ 53

ตัวอย่าง Ex.4 หมายเลข ⑥ ในรูปที่ 1.6 ใน SPSS จะใช้เครื่องหมาย o เป็นการรายงานค่าผิดปกติแบบ Mild Outlier จำนวน 1 ค่าอยู่ที่เรคคอร์ดที่ 1 * หมายเลข ⑦ เครื่องหมาย * เป็นการรายงานค่าผิดปกติแบบ Extreme Outlier จำนวน 1 ค่าอยู่ที่เรคคอร์ดที่ 40



รูปที่ 1.6 ลักษณะการแจกแจงของข้อมูลและ Outlier

3.2.3) พิจารณาข้อมูลเชิงความสัมพันธ์ ความแนบเนียนหรือความสอดคล้องระหว่างตัวแปรต่างๆ โดยอาศัยเทคนิคต่างๆ เช่น สร้างกราฟหรือแผนภาพการกระจายของข้อมูล (Scatter Plot) จะทำให้เห็นทิศทางความสัมพันธ์ว่าสัมพันธ์ไปในทิศทางเดียวกันหรือผกผันกัน และหาวิเคราะห์ค่าความสัมพันธ์

เพิ่มเติม เช่น Correlation: r_{xy} จะทำให้ทราบขนาดของความสัมพันธ์ด้วย นอกจากนี้อาจใช้วิธีการวิเคราะห์จำแนกกลุ่ม (Clustering) เพื่อจัดหน่วยตัวอย่างที่มีลักษณะเหมือนกันให้อยู่ในกลุ่มเดียวกัน เป็นต้น

ตัวอย่าง Ex.5 หมายเลข ⑧ ในรูปที่ 1.6 แสดงให้เห็นความสัมพันธ์ของทั้ง 2 ตัวแปร ผ่าน Scatter Plot ซึ่งจากแผนภาพเมื่อมองด้วยตาเปล่าหรือประสบการณ์ อาจจะสรุปในเบื้องต้นได้ว่า เมื่อคะแนนภาษาอังกฤษเพิ่มขึ้นคะแนนคณิตศาสตร์ดูเหมือนจะเพิ่มขึ้นตามไปด้วย เมื่อคำนวณค่า Pearson Correlation พบว่ามีค่าเท่ากับ $r = 0.235$ ซึ่งแปลผลได้ว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงในทิศทางเดียวกัน แต่ความสัมพันธ์อยู่ในระดับน้อย

3.2.4) คุณภาพของข้อมูล (Data Quality)

หากวิเคราะห์ข้อมูลเชิงสำรวจแล้วพบว่าข้อมูลผิดปกติ ผู้วิจัยจะต้องจัดการกับข้อมูลที่ผิดปกติเหล่านั้นให้เรียบร้อยก่อนที่จะนำไปสู่การวิเคราะห์ข้อมูลในเชิงลึกต่อไป สำหรับการแก้ไขข้อมูลที่ไม่ถูกต้อง หรือการกำจัดข้อมูลที่ผิดปกติ เราเรียกว่า การทำความสะอาดข้อมูล ด้วยการหาสาเหตุและจัดการแก้ไข โดยต้องคัดกรองหรือลบข้อมูลที่ไม่ถูกต้อง ข้อมูลซ้ำซ้อน ข้อมูลล้าสมัย หรือไม่จำเป็นออกไปจากชุดข้อมูล ซึ่งต้องทำด้วยวิธีการที่เหมาะสมกับประเภทของข้อมูลนั้น ๆ โดยอาจจะแก้ไขข้อมูลด้วยตัวผู้วิจัยเองหรือเขียนคำสั่งให้แก้ไขแบบอัตโนมัติ

เมื่อข้อมูลถูกต้องแล้วอาจจะต้องทำการจัดรูปแบบของข้อมูลหรือการแปลงข้อมูล (Data Transformation) ให้อยู่ในรูปแบบที่เหมาะสมกับการใช้งานหรือการวิเคราะห์ เช่น การแปลงข้อมูลให้เป็นรูปแบบหรือ format มาตรฐานแบบเดียวกัน (เช่น วัน/เดือน/ปี หน่วยเงิน) หรือการทำให้ข้อมูลมีความสม่ำเสมอ (Data Normalization) การแปลงข้อมูลเชิงปริมาณให้เป็นข้อมูลเชิงคุณภาพ (เช่น การจัดกลุ่มอายุ) การรวมข้อมูลจากหลาย ๆ ชุดให้เป็นชุดข้อมูลเดียวกัน เป็นต้น

สำหรับคุณภาพของข้อมูล (Data Quality) เพื่อนำไปสู่การวิเคราะห์ข้อมูลนั้น ควรคำนึงถึง 2 ประเด็นใหญ่ (วิชุดา ไชยศิริวงศา และคณะ, 2561) ดังนี้

ก) คุณภาพเนื้อแท้ของข้อมูล (Intrinsic Data Quality) เป็นคุณลักษณะที่เกี่ยวข้องกับความถูกต้อง (Accuracy) และความเชื่อถือได้ของข้อมูล (Reliability) ถ้ามีข้อมูลเรื่องเดียวกันจากหลาย ๆ แหล่ง ข้อมูลเหล่านั้นควรมีความสอดคล้องกัน (Consistency) และปราศจากอคติ

ข) คุณภาพของข้อมูลตามบริบทของงาน (Contextual Data Quality) หมายถึงคุณลักษณะของข้อมูลที่ต้องพิจารณาให้สอดคล้องกับปัญหา วัตถุประสงค์ หรือบริบทการใช้งานที่เฉพาะเจาะจง โดยพิจารณาจากความครบถ้วนของเนื้อหา ความเกี่ยวข้องกับประเด็นที่ต้องการศึกษา (Relevance) ความทันสมัยและสอดคล้องกับสถานการณ์ปัจจุบัน (Timeliness) รวมถึงความเพียงพอสำหรับการวิเคราะห์หรือการตัดสินใจ (Adequacy) ทั้งนี้ คุณภาพของข้อมูลในบริบทที่เหมาะสมจะช่วยให้การใช้งานข้อมูลเกิดประโยชน์สูงสุด และตอบโจทย์การตัดสินใจหรือการดำเนินงานได้อย่างมีประสิทธิภาพ

4) การวิเคราะห์ข้อมูลทางสถิติ เป็นการวิเคราะห์เพื่อหาข้อสรุป หรือสารสนเทศจากข้อมูล หรือพยานหลักฐานต่าง ๆ ที่เก็บรวบรวมมาได้ อย่างไรก็ตามข้อสรุปหรือสารสนเทศดังกล่าวจะมีคุณภาพหรือไม่ขึ้นกับคุณภาพของข้อมูลที่เก็บรวบรวมมาได้

การวิเคราะห์ข้อมูลทางสถิติสามารถจำแนกได้เป็น 2 ประเภทหลัก ๆ ได้แก่ **สถิติเชิงพรรณนา** และ **สถิติเชิงอนุมาน** จากหัวข้อ EDA ที่ผ่านมา จะเห็นว่าเป็นการใช้สถิติวิเคราะห์เชิงพรรณนา มาช่วยในการอธิบายข้อมูลที่เก็บรวบรวมมาได้ว่ามีค่าการกระจายหรือเกิดค่าสุดโต่งในลักษณะใดบ้าง ส่วนสถิติเชิงอนุมานนั้นจะเป็นการวิเคราะห์ข้อมูลจากตัวอย่างสุ่ม เพื่อหาข้อสรุปไปยังประชากร ซึ่งจะกล่าวรายละเอียดอีกครั้งในส่วนถัดไปของหนังสือเล่มนี้

สถิติเชิงพรรณนา (Descriptive Statistics) เป็นการอธิบายลักษณะของชุดข้อมูลที่ศึกษา โดยไม่ใช้หลักการเรื่องความน่าจะเป็นและมีได้อ้างอิงไปถึงประชากร

สถิติเชิงอนุมาน (Inferential Statistics) เป็นการหาข้อสรุปของประชากรจากตัวอย่างที่สุ่มมา หรือเป็นการนำค่าสถิติของตัวอย่างสุ่มอ้างอิงไปยังค่าพารามิเตอร์ของประชากร

Gartner IT และ Davenport, T. H. (2013) ได้อธิบาย Data Analytics ว่าสามารถแบ่งได้เป็น 4 ระดับ ตามรูปที่ 1.7 ที่มีจุดมุ่งหมายและประโยชน์ในการนำไปใช้งานที่แตกต่างกัน ทั้งนี้การวิเคราะห์ข้อมูลและนักวิทยาศาสตร์ข้อมูล ต่างมุ่งหมายที่จะวิเคราะห์ข้อมูลเชิงลึก (Insights) เพื่อช่วยในการตัดสินใจ โดยใช้เครื่องมือวิเคราะห์ข้อมูลร่วมกันในหลายบริบท แม้บทบาทและระดับความลึกของการวิเคราะห์จะต่างกัน สำหรับนักวิเคราะห์ข้อมูลและจะทำงานกับสถิติเชิงพรรณนา (1. **Descriptive Analytics**) เพื่ออธิบายว่าเกิดอะไรขึ้นในอดีตถึงปัจจุบัน โดยสรุปข้อมูลและแสดงผลในรูปแบบที่เข้าใจง่าย จากนั้นวิเคราะห์หรือวินิจฉัยถึงสาเหตุของเหตุการณ์ที่เกิดขึ้น (2. **Diagnostic Analytics**) เพื่อตอบว่าทำไมจึงเป็นเช่นนั้น เช่น การวิเคราะห์ฤดูกาล หรือความสัมพันธ์ระหว่างตัวแปรต่าง ๆ ในขณะที่นักวิทยาศาสตร์ข้อมูลพยายามใช้ข้อมูลในอดีตร่วมกับแบบจำลองเชิงสถิติ หรือ Machine Learning เพื่อทำนายสิ่งที่จะเกิดขึ้นในอนาคต (3. **Predictive Analytics**) และนำตัวแบบหรือแบบจำลองที่สร้างขึ้นใช้เป็นแนวทาง เพื่อคาดการณ์และการประเมินผลลัพธ์ในทางเลือกต่าง ๆ ที่ทำให้เกิดผลตอบแทนสูง โดยเสียค่าใช้จ่ายต่ำ (optimization) (4. **Prescriptive Analytics**) เพื่อนำเสนอทางเลือกที่เหมาะสมที่สุดในการตัดสินใจ

5) การตีความ เพื่อสร้างข้อสรุปจากผลการวิเคราะห์ข้อมูลไปยังบริบทของเรื่องนั้น ๆ ซึ่งการที่จะตีความได้ลึกซึ้ง ผู้วิจัยจะต้องมีความรู้ความเข้าใจสถานการณ์หรือบริบทของเรื่องนั้น ๆ เป็นอย่างดี เพื่อเชื่อมโยงกับวรรณกรรมและทฤษฎีต่าง ๆ ไปสู่การตอบโจทย์วิจัยหรือตอบวัตถุประสงค์ที่ตั้งไว้ได้อย่างเหมาะสม นอกจากนี้ในระหว่างสรุปผลและตีความอาจพบกับประเด็นวิจัยหรือคำถามวิจัยใหม่ ๆ ที่น่าสนใจ และนำไปสู่กระบวนการวิจัยในรอบต่อไป