ฉบับ
ปรับปรุง
เพิ่มเติม

Probability

ความน่าจะเป็น และสถิติ

Population

Sample

PROBABILITY AND STATISTICS

Inferential
Statistics

- ความน่าจะเป็น
- การแจกแจงความน่าจะเป็นแบบไม่ต่อเนื่อง
- การแจกแจงความน่าจะเป็นแบบต่อเนื่อง
- การสุ่มตัวอย่างและการแจกแจงการสุ่มตัวอย่าง
- ทฤษฎีการประมาณค่า
- การทดสอบสมมติฐาน

ผศ. ศุภชัย นาทะพันธ์

ซีเอ็ด

ความน่าจะเป็น และ สถิติ

PROBABILITY AND STATISTICS

พศ. ศุภชัย นาทะพันธ์

บริษัท ซีเอ็ดยูเคชั่น จำกัด (มหาชน)
SE-EDUCATION PUBLIC COMPANY LIMITED

ค้นหาหนังสือที่ต้องการ ได้เร็ว ทันใจ ที่ www.se-ed.com

ความน่าจะเป็นและสถิติ

โดย ศุภชัย นาทะพันธ์

สงวนลิขสิทธิ์ในประเทศไทยตาม พ.ร.บ. ลิขสิทธิ์ © พ.ศ. 2556 โดย ศุภชัย นาทะพันธ์
ห้ามคัดลอก ลอกเลียน ดัดแปลง ทำซ้ำ จัดพิมพ์ หรือกระทำอื่นใด โดยวิธีการใดๆ ในรูปแบบใดๆ
ไม่ว่าส่วนหนึ่งส่วนใดของหนังสือเล่มนี้ เพื่อเผยแพร่ในสื่อทุกประเภท หรือเพื่อวัตถุประสงค์ใดๆ
นอกจากจะได้รับอนุญาต

ข้อมูลทางบรรณานุกรมของหอสมุดแห่งชาติ

ศุภชัย นาทะพันธ์.
ความน่าจะเป็นและสถิติ. --กรุงเทพฯ : ซีเอ็ดยูเคชั่น, 2556.
1. ความน่าจะเป็น. - สถิติวิเคราะห์.
I. ชื่อเรื่อง.
619.2

ISBN(e-book) : 978-616-08-0993-6

ผลิตและจัดจำหน่ายโดย

บริษัท ซีเอ็ดยูเคชั่น จำกัด (มหาชน)
SE-EDUCATION PUBLIC COMPANY LIMITED
อาคารทีซีไอเอฟ ทาวเวอร์ ชั้น 19 เลขที่ 1858/87-90 ถนนบางนา-ตราด แขวงบางนา
เขตบางนา กรุงเทพฯ 10260 โทรศัพท์ 0-2739-8000

[หากมีคำแนะนำหรือติชม สามารถติดต่อได้ที่ comment@se-ed.com]

คำนำ

ในการบริหารธุรกิจหรือการบริหารอุตสาหกรรม ต้องพบกับปัญหาที่ต้องตัดสินใจภายใต้สาเหตุแห่งปัญหาที่ไม่แน่นอน ดังนั้นจึงต้องประยุกต์ความรู้ในเรื่องของความน่าจะเป็น ตัวอย่างเช่น การตรวจสอบอุบัติเหตุทางเครื่องบินว่ามีความน่าจะเป็นที่ความเสียหายเกิดจากโครงสร้างของเครื่องบินเป็นเท่าไร พบว่าอุบัติเหตุอาจเกิดจากโครงสร้างหรืออาจเกิดจากเครื่องยนต์ก็ได้ หรือการเปรียบเทียบประสิทธิภาพของ 2 สายการผลิตเมื่อทราบการแจกแจงความน่าจะเป็นของสายการผลิตทั้งสองว่าสายการผลิตใหม่สามารถผลิตได้มากกว่าสายการผลิตเดิมหรือไม่ การตอบคำถามดังกล่าวทำได้โดยการเก็บรวบรวมข้อมูลเบื้องต้นให้เพียงพอต่อการวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลโดยอาศัยหลักการทางสถิติ และการสรุปผลปัญหาและนำเสนอให้ผู้เกี่ยวข้องเข้าใจ ดังนั้นจุดประสงค์ในการเขียนหนังสือเล่มนี้ก็เพื่ออธิบายถึงวิธีการเก็บรวบรวมข้อมูล การวิเคราะห์ข้อมูล และการสรุปผลต่อปัญหาซึ่งไม่ทราบสาเหตุแน่นอน แต่ทราบความน่าจะเป็นหรือทราบรูปแบบการแจกแจงความน่าจะเป็นของประชากร (ถ้าไม่ทราบรูปแบบการแจกแจงความน่าจะเป็น ให้หาอ่านในหนังสือ *สถิติสำหรับงานวิศวกร* ในหัวข้อ “สถิติศาสตร์ไม่อิงพารามิเตอร์”) และใช้เป็นหนังสืออ้างอิงสำหรับการเรียนการสอนในสถาบันการศึกษา

เนื้อหาของหนังสือเล่มนี้จะกล่าวถึง ความหมายของสถิติและการจำแนกประเภทของวิธีการทางสถิติ (บทที่ 1) ผู้อ่านควรทำความเข้าใจอย่างมากในส่วนการวิเคราะห์เรื่องฮิสโตแกรม เพราะสามารถนำไปใช้ในการตีความการแจกแจงความน่าจะเป็นได้ แนวความคิดพื้นฐานของความน่าจะเป็นและการแจกแจงความน่าจะเป็น (บทที่ 2 ถึงบทที่ 5) ผู้อ่านควรให้ความสำคัญต่อหัวข้อการแจกแจงปกติ การวินิจฉัยสิ่งตัวอย่างด้วยเหตุและผลเพื่อค้นหาข้อสรุปเกี่ยวกับประชากร (บทที่ 6 ถึงบทที่ 8) โดยการวินิจฉัยดังกล่าวจะคำนึงถึงวิธีการสุ่มตัวอย่างและวิธีการสรุปผลเกี่ยวกับประชากร ผู้อ่านควรให้ความสำคัญในขั้นตอนการทดสอบสมมติฐาน

หนังสือเล่มนี้สามารถทำความเข้าใจได้ง่าย เนื่องจากมีตัวอย่างช่วยเสริมความเข้าใจสำหรับแต่ละทฤษฎีในแต่ละบท นอกจากนั้นยังมีแบบฝึกหัดท้ายบทเพื่อให้ผู้อ่านสามารถฝึกฝนได้ด้วยตนเอง โดยมีวิธีทำพร้อมเฉลยแบบฝึกหัดท้ายเล่ม ผู้เขียนหวังว่าหนังสือเล่มนี้จะเป็นประโยชน์สำหรับผู้อ่านในการนำเอาวิธีการทางสถิติไปประยุกต์

ผศ. ศุภชัย นาทะพันธ์
คณะวิศวกรรมศาสตร์
มหาวิทยาลัยมหิดล

สารบัญ

บทที่ 1 บทนำ ____ **11**

1.1 การเก็บรวบรวมข้อมูล **11**
1.2 การนำเสนอข้อมูล **12**
1.3 การวิเคราะห์ข้อมูล **30**
1.4 การจำแนกประเภทของวิธีการทางสถิติ **38**
แบบฝึกหัด **40**

บทที่ 2 ความน่าจะเป็น ____ **42**

2.1 เซต **42**
2.2 การทดลองและแซมเปิลสเปซ **45**
2.3 ความน่าจะเป็น **47**
2.4 ความน่าจะเป็นแบบมีเงื่อนไข **54**
2.5 ทฤษฎีบทของเบย์ **59**
แบบฝึกหัด **63**

บทที่ 3 ตัวแปรสุ่ม ____ **65**

3.1 แนวความคิดเกี่ยวกับตัวแปรสุ่ม **65**
3.2 ฟังก์ชันการแจกแจงความน่าจะเป็นสะสม **67**
3.3 ตัวแปรสุ่มชนิดไม่ต่อเนื่อง **69**
3.4 ตัวแปรสุ่มชนิดต่อเนื่อง **69**
3.5 ลักษณะเฉพาะบางประการของการแจกแจง **71**
3.6 ทฤษฎีบทของเชบีเชฟ **73**
3.7 ฟังก์ชันของตัวแปรสุ่ม **74**
3.8 การคาดคะเน **79**

3.9 การแจกแจงความน่าจะเป็นร่วมกัน 83
3.10 การแจกแจงขอบ 87
3.11 การแจกแจงความน่าจะเป็นแบบมีเงื่อนไข 92
3.12 ค่าคาดคะเนแบบมีเงื่อนไข 96
3.13 ความเป็นอิสระของตัวแปรสุ่ม 98
3.14 ความแปรปรวนร่วมและสัมประสิทธิ์สหสัมพันธ์ 99
3.15 ฟังก์ชันการแจกแจงความน่าจะเป็นสะสมของสองตัวแปรสุ่ม 102
แบบฝึกหัด 104

บทที่ 4 การแจกแจงความน่าจะเป็นแบบไม่ต่อเนื่อง 109

4.1 การแจกแจงเบอร์นูลี 109
4.2 การแจกแจงทวินาม 112
4.3 การแจกแจงเรขาคณิต 117
4.4 การแจกแจงอเนกนาม 120
4.5 การแจกแจงไฮเพอร์จีออเมตริก 121
4.6 การแจกแจงปัวส์ซอง 124
แบบฝึกหัด 127

บทที่ 5 การแจกแจงความน่าจะเป็นแบบต่อเนื่อง 128

5.1 การแจกแจงเอกรูป 128
5.2 การแจกแจงแบบเลขชี้กำลัง 132
5.3 การแจกแจงแกมมา 136
5.4 การแจกแจงปกติ 138
แบบฝึกหัด 150

บทที่ 6 การสุ่มตัวอย่างและการแจกแจงการสุ่มตัวอย่าง 152

6.1 การสุ่มตัวอย่าง 152
6.2 การแจกแจงการสุ่มตัวอย่าง 155
6.3 การแจกแจงไคกำลังสอง 161
6.4 การแจกแจงที 164
6.5 การแจกแจงเอฟ 168
แบบฝึกหัด 172

บทที่ 7 ทฤษฎีการประมาณค่า 174

7.1 การประมาณค่าแบบจุด 174
7.2 วิธีการประมาณค่าแบบจุด 178
7.3 การประมาณค่าแบบช่วง 180
7.4 การประมาณค่าแบบช่วงเกี่ยวกับค่าเฉลี่ย 182
7.5 การประมาณค่าแบบช่วงเกี่ยวกับความแปรปรวน 192
7.6 การประมาณค่าสัดส่วน 194
แบบฝึกหัด 197

บทที่ 8 การทดสอบสมมติฐาน 199

8.1 สมมติฐานเชิงสถิติ 199
8.2 การทดสอบสมมติฐานด้านเดียวและสองด้าน 203
8.3 การทดสอบสมมติฐานเกี่ยวกับค่าเฉลี่ย 204
8.4 การทดสอบสมมติฐานเกี่ยวกับความแปรปรวน 220
แบบฝึกหัด 225

ภาคผนวก 229

เฉลยแบบฝึกหัด 243

ดัชนี 285

บรรณานุกรม 290

1

บทนำ

สถิติมี 2 ความหมาย ความหมายแรกคือ ตัวเลขแสดงข้อเท็จจริงของข้อมูลที่สนใจศึกษา ซึ่งตัวเลขอาจเกิดจากการวัด เช่น ค่าความแข็งของเหล็กกล้าคาร์บอนต่ำ คะแนนเฉลี่ยของวิชาคณิตศาสตร์ 1 ในการสอบเข้ามหาวิทยาลัยประจำปีการศึกษา 2547 เป็นต้น และความหมายที่สองคือ ศาสตร์ที่เกี่ยวกับหลักการและระเบียบวิธีการทางสถิติ (Statistical Method) ซึ่งประกอบด้วย 4 กระบวนการคือ การเก็บรวบรวมข้อมูล (Collection of Data) อย่างมีประสิทธิภาพ การนำเสนอข้อมูล (Presentation of Data) ที่เก็บรวบรวมมาได้อย่างเข้าใจง่าย การวิเคราะห์ข้อมูล (Analysis of Data) ใกล้เคียงความเป็นจริง และการตีความหมายข้อมูล (Interpretation of Data) จากข้อมูลที่ได้จากการวิเคราะห์ได้อย่างถูกต้อง

1.1 การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูล คือการรวมข้อเท็จจริงทั้งที่เป็นตัวเลขและไม่เป็นตัวเลขเกี่ยวกับเรื่องที่สนใจศึกษา โดยการสำรวจ (ได้แก่ การสัมภาษณ์ การสอบถามทางไปรษณีย์หรือทางโทรศัพท์) การสังเกต (เช่น การจดบันทึกจำนวนผู้ใช้รถยนต์ตามสี่แยก เป็นต้น) การทดลอง (เช่น การปรับตั้งกระแสไฟฟ้าที่เหมาะสมในการเชื่อมเหล็กด้วยไฟฟ้า การหาอุณหภูมิที่พอเหมาะสำหรับบ่อชุบสีโลหะ เป็นต้น) หรือการคัดลอกจากเอกสารหรือทะเบียนตามหน่วยงานที่ต้องการศึกษา สิ่งสำคัญคือ ข้อมูลต้องทันกับเหตุการณ์ปัจจุบันที่ต้องการจะตีความ เช่น การสังเกตการทำงานแล้วจดบันทึกซึ่งเป็นส่วนหนึ่งในการศึกษาวิธีการทำงาน เป็นต้น

ข้อมูลสถิติ (Statistical Data) หมายถึงข้อเท็จจริงที่เป็นตัวเลขและไม่เป็นตัวเลขซึ่งจะต้องมีจำนวนมากพอที่จะเปรียบเทียบหรือตีความหมายได้ ข้อมูลมี 2 ประเภทคือ

1. ข้อมูลที่ได้จากการวัด (Measurement Data) คือข้อมูลที่เป็นเชิงตัวเลข เช่น การวัดความหนาของใบพัดเครื่องจักรที่ได้จากกระบวนการหล่อโลหะหลายๆ ชิ้น จะได้ความหนาแตกต่างกัน (แม้ว่าเครื่องมือวัดจะมีความถูกต้องแม่นยำก็ตาม) การวัดความหนาของออกไซด์ที่เคลือบบนแผ่นซิลิคอน และเวลาที่ใช้ในการประกอบเครื่องยนต์ของเครื่องบิน เป็นต้น

2. ข้อมูลที่ได้จากการนับ (Count Data) เช่น จำนวนข้อบกพร่องที่เกิดขึ้นบนเมนบอร์ดคอมพิวเตอร์ เป็นต้น

1.2 การนำเสนอข้อมูล

การนำเสนอข้อมูล คือการนำข้อมูลที่เก็บรวบรวมมาจัดใหม่ให้เป็นระเบียบแล้วแสดงให้ผู้เกี่ยวข้องเข้าใจ รูปแบบการนำเสนอมี 2 รูปแบบคือ การนำเสนอข้อมูลอย่างไม่เป็นแบบแผน และการนำเสนอข้อมูลอย่างเป็นแบบแผน

การนำเสนอข้อมูลอย่างไม่เป็นแบบแผนคือ การนำข้อมูลที่เก็บรวบรวมมาจัดให้เป็นหมวดหมู่ง่ายต่อการทำความเข้าใจในรูปแบบของข้อความหรือในรูปของข้อความกึ่งตาราง ส่วนการนำเสนอข้อมูลอย่างเป็นแบบแผนคือ การนำข้อมูลที่เก็บรวบรวมมาจัดให้เป็นหมวดหมู่ในรูปของตาราง แผนภูมิ กราฟ และแผนภาพ ซึ่งประกอบด้วยการแจกแจงความถี่ (Frequency Distribution) ฮิสโตแกรม (Histogram) กราฟแบบสี่เหลี่ยม (Box Plot) แผนภาพลำต้นและใบไม้ (Stem and Leaf Diagram) และแผนภาพพาเรโต (Pareto Diagram) ตามลำดับ

1.2.1 การแจกแจงความถี่

การแจกแจงความถี่ (Frequency Distribution) เป็นการนำเสนอข้อมูลเมื่อข้อมูลมีจำนวนมากหรือมีค่าตัวเลขซ้ำกันมาก เพราะจะช่วยประหยัดเวลาและสรุปผลได้รัดกุม ความถี่ (Frequency) คือจำนวนที่แสดงว่าค่าที่เป็นไปได้แต่ละค่าเกิดขึ้นกี่ครั้ง ดังตัวอย่างในตารางที่ 1.1

ตารางที่ 1.1 ค่าความแข็งแรง (ปอนด์/ตารางนิ้ว) ของขวดน้ำประเภทแก้ว 1 ลิตร จำนวน 100 ขวด

265	197	346	280	265	200	221	265	261	278
205	286	317	242	254	235	176	262	248	250
263	274	242	260	281	246	248	271	260	265
307	243	258	321	294	328	263	245	274	270
220	231	276	228	223	296	231	301	337	298
268	267	300	250	260	276	334	280	250	257
260	281	208	299	308	264	280	274	278	210
234	265	187	258	235	269	265	253	254	280
299	214	264	267	283	235	272	287	274	269
215	318	271	293	277	290	283	258	275	251

พิจารณาตารางที่ 1.1 ซึ่งเป็นข้อมูลค่าความแข็งแรง (ปอนด์/ตารางนิ้ว) ของการทำลายขวดน้ำประเภทแก้ว 1 ลิตร จำนวน 100 ขวด ค่าสังเกตเหล่านี้ได้จากการทดสอบขวดแต่ละใบจนกระทั่งขวดแตก จากนั้นจดบันทึกข้อมูลจากซ้ายไปขวาและจากบนลงล่าง จากตารางที่ 1.1 จะพบว่าข้อมูลที่จดบันทึกไม่ได้ให้สารสนเทศเกี่ยวกับคุณลักษณะด้านแนวโน้มสู่ศูนย์กลาง การกระจายความเบ้ และวิเคราะห์หาข้อมูลที่ออกนอกอย่างผิดปกติ (Outlier)

ขั้นตอนการสร้างตารางแจกแจงความถี่มี 9 ขั้นตอนคือ

1. หาค่าสูงสุดและค่าต่ำสุดของข้อมูล

2. หาพิสัยของข้อมูลจาก

พิสัย (Range : R) = ค่ามากสุด – ค่าน้อยสุด

3. กำหนดจำนวนอันตรภาคชั้น (Class Intervals) ถ้าข้อมูลมีการกระจายมาก จำนวนอันตรภาคชั้นควรจะน้อย นั่นคืออันตรภาคชั้นควรมีความกว้างมาก ในการกำหนดความกว้างของอันตรภาคชั้น มักจะกำหนดให้เท่ากันเพื่อสะดวกในการวิเคราะห์ และควรมีจำนวนอันตรภาคชั้นอยู่ในช่วง 5 ถึง 20 อันตรภาคชั้น ดังแสดงในตารางที่ 1.2 ส่วนการกำหนดให้อันตรภาคชั้นกว้างไม่เท่ากันจะกระทำเมื่อข้อมูลมีค่าที่สูงหรือต่ำกว่าค่าอื่นๆ เกินปกติ โดยอาจกำหนดอันตรภาคชั้นแรกหรือชั้นสุดท้ายเป็นอันตรภาคชั้นเปิด

ตารางที่ 1.2 จำนวนข้อมูลและจำนวนอันตรภาคชั้น

จำนวนข้อมูล (N)	จำนวนอันตรภาคชั้น
ต่ำกว่า 50	5 – 7
50 – 100	6 – 10
100 – 250	7 – 12
เกินกว่า 250	10 – 20

4. หาความกว้างของอันตรภาคชั้นจาก

$$\text{ความกว้างของอันตรภาคชั้น} = \frac{\text{พิสัย}}{\text{จำนวนอันตรภาคชั้น}}$$

หมายเหตุ *:*

1) ข้อที่ 3 และข้อที่ 4 สลับกันได้ หมายถึงอาจกำหนดความกว้างของอันตรภาคชั้นก่อน แล้วค่อยหาจำนวนอันตรภาคชั้น โดยที่ถ้าผลหารมีเศษ ให้ปัดขึ้นเป็นจำนวนเต็มเสมอ และถ้าผลหารลงตัว ให้บวกเพิ่มอีก 1

2) ค่าขอบบนและขอบล่างของแต่ละอันตรภาคชั้นจะมีค่าละเอียดกว่าหน่วยที่วัดได้จากข้อมูลดิบครึ่งหนึ่งเสมอ เช่น ถ้าข้อมูลเป็นเลขจำนวนเต็มที่มีค่าเพิ่มขึ้นครั้งละ 1 ค่าขอบบนและขอบล่างจะมีความละเอียด $^1/_2(1) = 0.5$ เป็นต้น หลังจากสร้างตารางแจกแจงความถี่เสร็จ สามารถตรวจสอบความกว้างของแต่ละอันตรภาคชั้นได้โดยใช้สูตร

ความกว้างของอันตรภาคชั้น = ขอบบน – ขอบล่าง

5. เรียงลำดับอันตรภาคชั้นจากค่าน้อยไปค่ามาก หรือเรียงจากค่ามากไปค่าน้อยก็ได้ โดยคะแนนที่มีค่าน้อยที่สุดต้องอยู่ในอันตรภาคชั้นต่ำสุด และคะแนนที่มีค่ามากที่สุดต้องอยู่ในอันตรภาคชั้นสูงสุด ค่าเริ่มต้นอาจใช้ค่าน้อยที่สุดเป็นค่าเริ่มต้นของอันตรภาคชั้นแรก หรืออาจปรับค่าเริ่มต้นให้น้อยกว่าค่าน้อยสุด (ตัวเลขลงท้ายของค่าเริ่มต้นนิยมเป็นเลข 0 หรือ 1 หรือ 5) ก็ได้ แต่ค่าความกว้างของอันตรภาคชั้นจะต้องกว้างขึ้นกว่าเดิมเล็กน้อย

6. บันทึกรอยขีด (Tally)

7. หาความถี่ (Frequency)

8. หาความถี่สัมพัทธ์ (Relative Frequency)

9. หาความถี่สัมพัทธ์สะสม (Cumulative Relative Frequency)

พิจารณาข้อมูลจากตารางที่ 1.1 เมื่อกำหนดจำนวนอันตรภาคชั้นเท่ากับ 9 (เนื่องจากข้อมูลมี 100 ตัว) สามารถสร้างตารางแจกแจงความถี่ได้ดังนี้

$$\text{หาพิสัยของข้อมูล : } R = x_{max} - x_{min}$$
$$= 346 - 176 = 170$$

$$\text{ความกว้างของอันตรภาคชั้น} = \frac{\text{พิสัย}}{\text{จำนวนอันตรภาคชั้น}}$$
$$= \frac{170}{9} \approx 18.88 = 19$$

ถ้าสร้างตารางแจกแจงความถี่ตามปกติ ค่าเริ่มต้นต้องเริ่มด้วย 176 ปอนด์/ตารางนิ้ว แต่ในที่นี้จะปรับค่าเริ่มต้นเป็น 170 ซึ่งน้อยกว่า 176 เพื่อง่ายต่อการนำเสนอข้อมูล และปรับค่าความกว้างของอันตรภาคชั้นเป็น 20 ดังแสดงในตารางที่ 1.3

ตารางที่ 1.3 การแจกแจงความถี่สำหรับข้อมูลค่าความแข็งแรงของขวดน้ำ 1 ลิตร

อันตรภาคชั้น (ปอนด์/ตารางนิ้ว)	รอยขีด	ความถี่	ความถี่สัมพัทธ์	ความถี่สัมพัทธ์สะสม
170 – 189	||	2	.02	.02
190 – 209	||||	4	.04	.06
210 – 229	𝍸 ||	7	.07	.13
230 – 249	𝍸 𝍸 |||	13	.13	.26
250 – 269	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 ||	32	.32	.58
270 – 289	𝍸 𝍸 𝍸 𝍸 ||||	24	.24	.82
290 – 309	𝍸 𝍸 |	11	.11	.93
310 – 329	||||	4	.04	.97
330 – 349	|||	3	.03	1.00
		100	1.00	

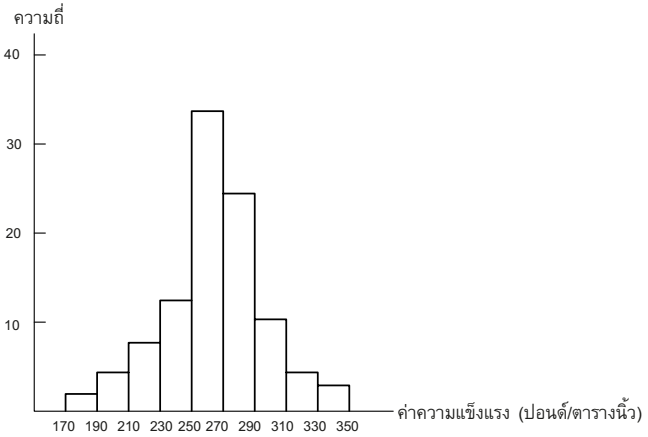
จากการแจกแจงความถี่ที่ได้นำเสนอ สามารถใช้ในการวิเคราะห์ เช่น ขวดทนต่อแรงดึงต่ำกว่า 230 ปอนด์/ตารางนิ้ว ว่ามีเปอร์เซ็นต์เท่าใด จากตารางที่ 1.3 พิจารณาจากความถี่สัมพัทธ์สะสมของอันตรภาคชั้นที่มีช่วงคะแนน 210 – 229 ผลคือ 13%

1.2.2 ฮิสโตแกรม

ฮิสโตแกรม (Histogram) คือแผนภูมิแท่งชุดหนึ่งซึ่งมีพื้นที่หรือความสูงของแต่ละแท่งแทนขนาดของความถี่ของแต่ละอันตรภาคชั้น และความกว้างของแต่ละแท่งแทนความกว้างของอันตรภาคชั้น จำนวนของข้อมูลไม่ควรต่ำกว่า 50 ตัวเพราะจะไม่ปรากฏรูปแบบการกระจายซึ่งจะใช้รูปทรงระฆังคว่ำเป็นตัวเทียบในการตีความ และไม่ควรเกิน 200 ตัวเพราะว่ามีโอกาสสูงที่ข้อมูลจะล้าสมัยหรือมาจากคนละกระบวนการกัน จำนวนชั้นของฮิสโตแกรมมีความจำเป็นต้องแปรผันตามจำนวนของข้อมูล เพราะการสร้างฮิสโตแกรมที่มีจำนวนชั้นมากหรือน้อยจนเกินไปจะทำให้ไม่สามารถตีความ

รูปแบบการกระจายได้ โดยเฉพาะการสร้างฮิสโตแกรมจากคอมพิวเตอร์ที่มักจะไม่คำนึงถึงจำนวนชั้นที่เหมาะสมจะทำให้ตีความหมายผิดพลาดได้

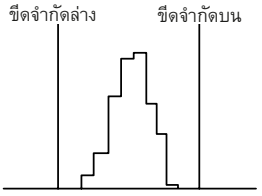
ฮิสโตแกรมใช้หลังจากรวบรวมข้อมูล เช่น เปอร์เซ็นต์ของเสีย อัตราทำงานล่วงเวลา จำนวนคนงานที่ขาดงาน ความหนืดของน้ำมัน เส้นผ่านศูนย์กลางของท่อน้ำ ความแข็งแรงของลวด และน้ำหนักของผลิตภัณฑ์ต่างๆ เป็นต้น ซึ่งจะช่วยให้เห็นรูปแบบการกระจาย ค่าเฉลี่ย และค่าความเบี่ยงเบนของค่าที่วัดได้โดยมีแกน x แทนค่าที่วัด และแกน y แทนค่าความถี่ ดังแสดงในรูปที่ 1.1



รูปที่ 1.1 ฮิสโตแกรมของค่าความแข็งแรงของขวดน้ำประเภทแก้ว 1 ลิตร จำนวน 100 ขวด

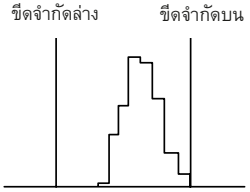
ในการวิเคราะห์หรือการตีความฮิสโตแกรม จะต้องตีความเทียบกับข้อกำหนดของลูกค้า (Specification) เช่น รูปทรงระฆังคว่ำหรือไม่ ค่าความเบี่ยงเบนมากหรือน้อยเพียงใด ค่าแนวโน้มสู่ศูนย์กลางตรงตามข้อกำหนดหรือไม่ ฮิสโตแกรมสามารถแบ่งออกได้เป็น 16 รูปแบบ ดังนี้

รูปแบบที่ 1 ฮิสโตแกรมรูประฆังคว่ำ ซึ่งข้อมูลมีการแจกแจงปกติ (Normal Distribution) โดยมีลักษณะสมมาตรรอบค่ากลางค่าหนึ่ง และการกระจายตัวของข้อมูลอยู่ ณ กึ่งกลางของขีดจำกัดควบคุม (Control Limit) ข้อมูลมีความถี่สูงสุดที่ค่าแนวโน้มศูนย์กลางแล้วกระจายออกไปอย่างสมมาตรทั้งซ้ายและขวา แสดงว่าผลิตภัณฑ์ได้ตามมาตรฐานที่ลูกค้าต้องการ การกระจายตัวแบบระฆังคว่ำนี้เป็นที่ต้องการของนักวิเคราะห์



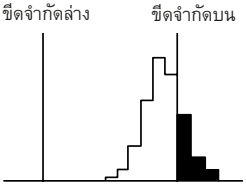
รูปแบบที่ 1 ฮิสโตแกรมรูประฆังคว่ำ

รูปแบบที่ 2 ลักษณะฮิสโตแกรมคล้ายกับรูปแบบที่ 1 แต่จะพิเศษตรงที่ค่าเฉลี่ยมีการเคลื่อนไปทางขีดจำกัดบน (Maximum Limit หรือ Upper Control Limit; UCL) หรือทางขีดจำกัดล่าง (Lower Control Limit; LCL) แต่ยังไม่เกินขีดจำกัดที่กำหนดไว้ และมีแนวโน้มว่าจะเกินขีดจำกัดที่กำหนดในครั้งต่อไปถ้ายังไม่มีการแก้ไข สาเหตุเกิดจากขาดการซ่อมบำรุงเครื่องจักร ทำให้เครื่องจักรขาดความแม่นยำในการทำงาน วิธีแก้ไขนั้นควรรีบทำการตรวจสอบเครื่องจักรและทำตารางการซ่อมบำรุง เป็นต้น



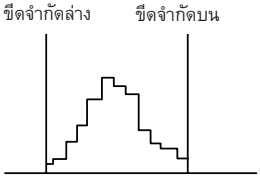
รูปแบบที่ 2 ฮิสโตแกรมรูประฆังคว่ำ แต่ค่าเฉลี่ยของข้อมูลมีแนวโน้มว่าจะเกินขีดจำกัดที่กำหนด

รูปแบบที่ 3 กระบวนการเป็นการแจกแจงปกติ การแปรผันของกระบวนการคล้ายๆ ทรงระฆังคว่ำ แต่ค่าเฉลี่ยเลื่อนไปทาง UCL หรือ LCL วิธีแก้ไขคือ ต้องปรับกระบวนการผลิตเพื่อให้ค่าเฉลี่ยของกระบวนการผลิตมาอยู่ที่กึ่งกลางของขีดจำกัดที่กำหนด หรืออยู่กึ่งกลางระหว่าง UCL และ LCL ในอีกมุมมองหนึ่ง โรงงานผลิตผลิตภัณฑ์ได้ตามมาตรฐานของโรงงาน ซึ่งตรงกับข้อกำหนดของลูกค้ากลุ่มหนึ่ง แต่ไม่ตรงกับข้อกำหนดของลูกค้าอีกกลุ่มหนึ่ง ดังนั้นอาจแยกสายการผลิตเฉพาะเพื่อผลิตผลิตภัณฑ์ให้ตรงกับข้อกำหนดของลูกค้ากลุ่มหลัง



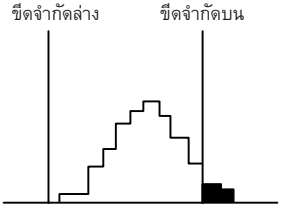
รูปแบบที่ 3 ฮิสโตแกรมรูประฆังคว่ำ แต่ค่าเฉลี่ยของข้อมูลเกินขีดจำกัดที่กำหนด

รูปแบบที่ 4 ลักษณะคล้ายรูประฆังคว่ำ ซึ่งอยู่ในขีดจำกัดที่กำหนด แต่มีค่ากระจายไปใกล้เคียงกับ UCL และ LCL มากเกินไป ทำให้ไม่มีระยะเผื่อเกิดขึ้น วิธีการแก้ไขคือ ควรลดความแปรปรวนของข้อมูลให้น้อยลง หรือลดการกระจายตัวของข้อมูลให้น้อยลง เพื่อให้มีระยะเผื่อเพิ่มขึ้น หากค่าเฉลี่ยของกระบวนการผลิตคลาดเคลื่อนไปจากเดิมก็ยังไม่ทำให้ฮิสโตแกรมออกนอกขีดจำกัดที่กำหนด



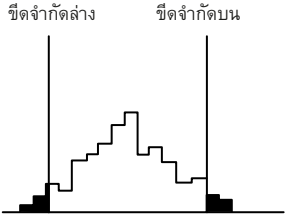
รูปแบบที่ 4 ฮิสโตแกรมรูประฆังคว่ำแบบฐานกว้าง

รูปแบบที่ 5 ลักษณะฮิสโตแกรมคล้ายกับรูปแบบที่ 4 แต่มีความแปรปรวนมาก ทำให้ความกว้างของฮิสโตแกรมเกินขีดจำกัดที่กำหนดไว้ และค่าเฉลี่ยของข้อมูลจะเคลื่อนที่ไปทาง UCL หรือ LCL มีผลทำให้ความกว้างของกราฟเกินขีดจำกัดที่กำหนด ส่วนสาเหตุนั้นพิจารณาจาก 2 กรณีคือ ความแปรปรวนของข้อมูลและความคลาดเคลื่อนของค่าเฉลี่ย ความแปรปรวนของข้อมูลอาจเกิดจากการที่เครื่องจักรไม่มีความแน่นอน ทำให้ข้อมูลที่ได้มามีการกระจายตัวมาก ส่วนความคลาดเคลื่อนของค่าเฉลี่ยอาจเกิดจากเครื่องจักรขาดการซ่อมบำรุง ทำให้ประสิทธิภาพในการผลิตต่ำลง หรือมีค่าสูงขึ้นจากที่ต้องการ การแก้ไขสามารถทำได้โดยซ่อมบำรุงเครื่องจักรอย่างสม่ำเสมอ และเมื่อหมดอายุการใช้งานก็ควรเปลี่ยนใหม่



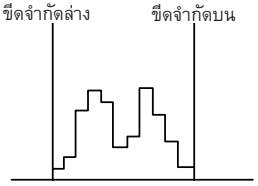
รูปแบบที่ 5 ฮิสโตแกรมรูประฆังคว่ำแบบฐานกว้างเกินขีดจำกัดที่กำหนด

รูปแบบที่ 6 ลักษณะคล้ายรูประฆังคว่ำ แต่การแจกแจงข้อมูลไม่หนาแน่นรอบๆ ศูนย์กลางของยอดของข้อมูลจึงทำให้ส่วนทางซ้ายและขวาออกนอกขีดจำกัดที่กำหนด สาเหตุเกิดจากกระบวนการมีการเปลี่ยนแปลงในหลายๆ ส่วนของการผลิต ทำให้มีความแปรปรวนมาก แนวทางแก้ไขคือ ปรับแก้ที่ความแน่นอนของเครื่องจักรในการผลิตให้มีความแม่นยำมากขึ้น รวมถึงการศึกษาปัญหาในการดูแลซ่อมแซมเครื่องจักร



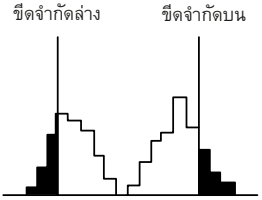
รูปแบบที่ 6 ฮิสโตแกรมรูประฆังคว่ำแบบฐานกว้างออกนอกช่วงควบคุม (Spread Out of Control Limit)

รูปแบบที่ 7 ฮิสโตแกรมรูปภูเขาสองยอด เกิดจากข้อมูล 2 ชุดหรือ 1 ชุดที่มาจากการแจกแจง 2 ชุด ซึ่งค่าเฉลี่ยไม่เท่ากัน หรือเป็นข้อมูลที่มีแหล่งความผันแปรมาจากสาเหตุ 2 สาเหตุที่มีความแตกต่างกัน เช่น เครื่องจักร 2 เครื่อง วัตถุดิบ 2 รุ่น เวลาการทำงานที่แตกต่างกัน เป็นต้น ซึ่งในการวิเคราะห์ข้อมูลต้องพิจารณาแยกข้อมูลทั้ง 2 ชุดออกจากกันก่อน แต่การกระจายตัวของฮิสโตแกรมรูปนี้ยังอยู่บริเวณกึ่งกลาง เป็นค่าที่ยังยอมรับได้ แนวทางแก้ไขคือ พยายามทำให้ค่าเฉลี่ยของข้อมูลทั้งสองนี้มีค่าเฉลี่ยใกล้กัน สามารถรวมเป็นฮิสโตแกรมที่มีการกระจายตัวแบบปกติ



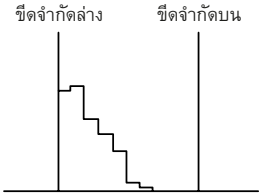
รูปแบบที่ 7 ฮิสโตแกรมรูปภูเขาสองยอด

รูปแบบที่ 8 ลักษณะฮิสโตแกรมคล้ายกับรูปแบบที่ 7 แต่ความแปรปรวนมีมาก สาเหตุมาจาก 2 กรณีคือ กรณีแรกคือการเกิดเป็นภูเขาสองยอดที่เกิดจากการเก็บตัวอย่างมาจากข้อมูล 2 ชุด โดยอาจจะเก็บมาจากเครื่องจักร 2 เครื่อง หรือสายการผลิต 2 สาย เป็นต้น การแก้ไขกรณีนี้คือ อาจจะแยกข้อมูลออกจากกันและทำเป็นฮิสโตแกรม 2 ชุด และตรวจสอบ–ปรับปรุงสภาพเครื่องจักรให้สามารถทำงานได้มีประสิทธิภาพใกล้เคียงกัน กรณีที่ 2 คือ ความแปรปรวนของแต่ละยอดมีมาก ซึ่งอาจจะเกิดมาจากเครื่องจักรนั้นๆ ขาดการบำรุงซ่อมแซม ทำให้ประสิทธิภาพลดลง ทำงานได้ไม่สม่ำเสมอ แก้ไขได้โดยซ่อมแซมอุปกรณ์ให้ใช้งานได้ดีอยู่เสมอ



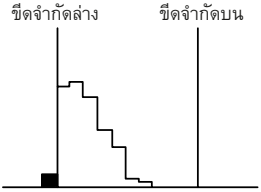
รูปแบบที่ 8 ฮิสโตแกรมรูปภูเขาสองยอดเกินขีดจำกัดที่กำหนด

รูปแบบที่ 9 ลักษณะด้านใดด้านหนึ่งสูงโด่งมาก เช่น ชนิดหน้าผาด้านซ้าย จะมีความถี่ของข้อมูลทางซ้ายลดลงอย่างรวดเร็ว และค่าเฉลี่ยของฮิสโตแกรมใกล้มาทางซ้ายมือ ฮิสโตแกรมชนิดนี้เกิดขึ้นได้เมื่อมีการตรวจสอบ 100% เฉพาะขนาดชิ้นงานทางด้านค่าขีดจำกัดล่าง (LCL) ซึ่งเกิดจากกระบวนการผลิตมีค่าต่ำกว่าค่าขีดจำกัดล่าง เพราะฉะนั้นชิ้นงานที่มีขนาดใกล้เคียงกับค่าขีดจำกัดล่างจึงมีมาก ทำให้ความถี่ในย่านนี้มีค่าสูงมากและขาดหายทันทีใกล้กับค่าขีดจำกัดล่าง หรือฮิสโตแกรมชนิดนี้อาจเกิดจากความผิดพลาดในการบันทึกข้อมูล เช่น อ่านข้อมูลเฉพาะค่าบวก เป็นต้น



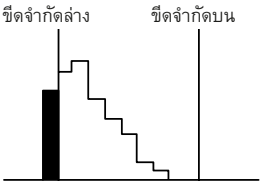
รูปแบบที่ 9 ฮิสโตแกรมรูปหน้าผา

รูปแบบที่ 10 ฮิสโตแกรมรูปทรงเบ้ขวา มีค่าเฉลี่ยไม่อยู่กึ่งกลางของขีดจำกัดที่กำหนด แต่จะอยู่ทางด้านซ้ายของขีดจำกัดที่กำหนด โดยความถี่ทางซ้ายมือจะลดลงเร็วมาก แต่ทางขวามือจะค่อยๆ ลดลงและข้อมูลบางส่วนเกินขีดจำกัดที่กำหนดไว้ สาเหตุอาจเกิดจากการที่เราควบคุมการเกิดข้อมูลด้านใดด้านหนึ่งเป็นการควบคุมที่เจาะจง ซึ่งโดยมากแล้วมักเกิดจากข้อมูลจากการวัดเวลา หรือการนับจำนวน ตัวอย่างคือ การเก็บข้อมูลในช่วงระยะเวลาสั้นๆ ของการดำเนินงาน หรือการนับจำนวนของเสียของระบบที่มีการควบคุม



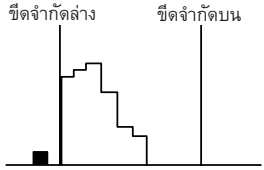
รูปแบบที่ 10 ฮิสโตแกรมรูปทรงเบ้ขวา

รูปแบบที่ 11 ลักษณะเป็นฮิสโตแกรมรูประฆังคว่ำที่มีการลบส่วนใดส่วนหนึ่งออกไป เช่น ตัดข้อมูลส่วนที่เกินขีดจำกัดล่างออกไป ข้อมูลชนิดนี้ต้องพิจารณาถึงสาเหตุที่ทำให้เกิดการตัดข้อมูลออกไป ซึ่งอาจเกิดจากข้อมูลที่ผ่านการตรวจสอบแล้ว หรือเป็นข้อมูลที่ได้มาจากระบบควบคุมโดยอัตโนมัติ จึงมีรูปทรงแบบนี้ออกมาให้เห็น จากรูปนี้จะเห็นว่าค่าข้อมูลทางด้านซ้ายที่เลยออกไปนอกขีดจำกัดล่างที่กำหนดจึงถูกตัดออก แต่ยังมีบางส่วนที่มีค่าความถี่สูงจึงไม่สามารถตัดออกได้



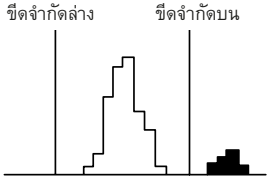
รูปแบบที่ 11 ฮิสโตแกรมรูปทรงถูกตัด

รูปแบบที่ 12 ลักษณะคล้ายกับฮิสโตแกรมรูปแบบที่ 9 แต่มีเกาะแก่งของข้อมูลที่กระโดดเพิ่มขึ้นมา โดยข้อมูลที่กระโดดนี้เกินขีดจำกัดที่กำหนดไว้ สาเหตุที่เกิดขึ้นคือ รูปทรงหน้าผามีสาเหตุมาจากความผิดพลาดของข้อมูล และเกาะแก่งของข้อมูลที่กระโดดมีสาเหตุมาจากการความผิดพลาดในการเก็บข้อมูล วิธีการแก้ไขนั้นควรแก้ไขที่กระบวนการเก็บข้อมูล



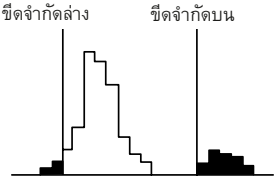
รูปแบบที่ 12 ฮิสโตแกรมรูปหน้าผาและเกาะแก่ง

รูปแบบที่ 13 ฮิสโตแกรมรูปทรงเกาะแก่งจะที่มี 2 ยอด แต่มีความแตกต่างของขนาดรูปทรงของการกระจายตัวมาก โดยทั้งสองยอดนี้แบ่งออกจากกันชัดเจนเป็นฮิสโตแกรมที่แสดงถึงความไม่สมบูรณ์และความผิดพลาด ซึ่งมีสาเหตุมาจากการมีข้อมูลจากกระบวนการอื่นปนมา ความผิดพลาดเกิดจากการวิเคราะห์ตรวจสอบ การเก็บข้อมูล การควบคุม เป็นต้น จากรูป เกาะที่เกิดขึ้นนั้นอยู่นอกขีดจำกัดที่ควบคุม วิธีการแก้ไขคือ ตรวจสอบการปรับเปลี่ยนปัจจัยต่างๆ เพื่อให้ค่าเฉลี่ยของข้อมูลอยู่ภายในขีดจำกัด และตรวจสอบการวัดและการคำนวณต่างๆ อีกครั้ง



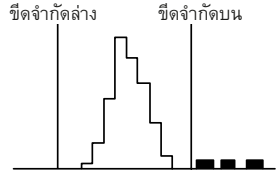
รูปแบบที่ 13 ฮิสโตแกรมรูปทรงเกาะแก่ง

รูปแบบที่ 14 ลักษณะจะมียอดค่าความถี่แยกออกเป็นอิสระอีกยอดหนึ่งนอกเหนือจากยอดใหญ่ โดยยอดที่แยกออกมานั้นเกินขีดจำกัดที่กำหนดไว้ และตรงยอดใหญ่ ค่าเฉลี่ยมีการเคลื่อนไปทางขีดจำกัดล่าง สาเหตุที่เกิดได้นั้นเกิดจากการที่เครื่องจักรมีแนวโน้มจะเสีย ทำให้ค่าเฉลี่ยมีการเคลื่อนจากกึ่งกลางของขีดจำกัด ส่วนเกาะแก่งอาจเกิดจากการมีข้อมูลการแจกแจงหรือกระบวนการอื่นปนอยู่ หรือมีความผิดพลาดในการวัดชิ้นงาน วิธีการแก้ไขควรตรวจสอบข้อมูลที่ได้ว่ามีความถูกต้องหรือไม่ ควรจัดทำแผนในการบันทึกข้อมูล และควรซ่อมบำรุงเครื่องจักรให้เครื่องจักรมีการใช้งานอย่างถูกต้องแม่นยำ



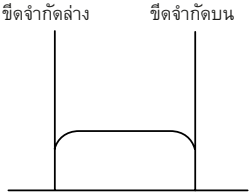
รูปแบบที่ 14 ฮิสโตแกรมรูปทรงเกาะแก่ง แต่เกาะแก่งมียอดเกินขีดจำกัดที่กำหนด

รูปแบบที่ 15 ลักษณะฮิสโตแกรมคล้ายรูประฆังคว่ำ แต่จะมีข้อมูลบางส่วนแตกแยกออกมา ซึ่งข้อมูลบางส่วนจะเกินขีดจำกัดที่กำหนด สาเหตุเกิดจากความผิดพลาดในการบันทึกข้อมูล การวัดชิ้นงาน หรือเกิดจากความผิดพลาดของเครื่องมือ วิธีการแก้ไขนั้นควรจะมีวิธีการตรวจสอบเครื่องมือที่ใช้ในการวัดชิ้นงาน หรือวางแผนในการบันทึกข้อมูลล่วงหน้าเพื่อป้องกันข้อผิดพลาด



รูปแบบที่ 15 ฮิสโตแกรมรูประฆังคว่ำที่มีข้อมูลบางส่วนกระโดดออกมาและเกินขีดจำกัดที่กำหนดไว้

รูปแบบที่ 16 ฮิสโตแกรมรูปทรงที่ราบสูง มีลักษณะข้อมูลที่มีค่าใกล้เคียงกันมาก แต่จะลดลงทันทีเฉพาะหัวและท้ายของข้อมูลซึ่งเป็นข้อมูลที่ไม่มีฐานนิยมที่เด่นชัด สาเหตุเกิดจากข้อมูลรูประฆังคว่ำหลายๆ รูปมาต่อกัน โดยแต่ละชุดมีค่าแนวโน้มเข้าสู่ศูนย์กลางที่ใกล้เคียงกัน หรือมีลักษณะการแจกแจงแตกต่างกันหลายๆ แบบปะปนกัน แต่ละแบบมีค่าเฉลี่ยไม่เท่ากันแต่ใกล้เคียงกัน โดยที่การแจกแจงข้อมูลทั้งหมดอยู่ภายในขีดจำกัดที่กำหนด วิธีแก้ไขคือดำเนินการจัดทำมาตรฐานในระบบการผลิต ทั้งนี้แล้วแต่สาเหตุหลักของการผันแปร ซึ่งอาจหมายถึงการทำมาตรฐานการทำงาน มาตรฐานการบำรุงรักษา เป็นต้น



รูปแบบที่ 16 ฮิสโตแกรมรูปทรงที่ราบสูง

อย่างไรก็ตาม การนำเสนอข้อมูลด้วยการแจกแจงความถี่และฮิสโตแกรมก็ยังไม่ดีในแง่ของการไม่ได้แสดงข้อมูลทุกตัว แต่ดีในแง่ของความง่ายในการตีความ โดยทั่วไปทั้ง 2 วิธีเหมาะสำหรับการนำเสนอข้อมูลที่มีขนาดตั้งแต่ 50 ตัวขึ้นไป

1.2.3 แผนภาพลำต้นและใบไม้

แผนภาพลำต้นและใบไม้ (Stem and Leaf Diagram) ได้รับการพัฒนาโดยจอห์น ดับบลิว. เตอร์กี้ (John W. Tukey) ให้ผลประโยชน์ที่เหนือกว่าฮิสโตแกรม 3 ประการคือ

1. ไม่จำเป็นต้องมีการกำหนดความกว้างอันตรภาคชั้นและจำนวนอันตรภาคชั้นให้ยุ่งยากเหมือนการสร้างฮิสโตแกรม

2. การแสดงผลโดยแผนภาพลำต้นและใบไม้จะแสดงถึงข้อมูลที่เป็นตัวเลขจริงเลย ทำให้สะดวกต่อการตีความหมาย

3. ข้อมูลที่แสดงในแผนภาพลำต้นและใบไม้จะแสดงในรูปที่เป็นตัวเลข (Numerical Order) ซึ่งข้อมูลที่ได้เก็บมาจริง ไม่ได้สูญหายเหมือนฮิสโตแกรม

ขั้นตอนการสร้างแผนภาพลำต้นและใบไม้มี 2 ขั้นตอนคือ

1. พิจารณาข้อมูลที่มีอยู่ แล้วทำการแยกข้อมูลดังกล่าวออกเป็น 2 ส่วน โดยส่วนที่เป็นตัวเลขหลักอาจจะมี 1 หลักหรือมากกว่า ให้กำหนดเป็นส่วนของลำต้น (Stem) และตัวเลขข้างท้ายให้กำหนดเป็นส่วนของใบไม้ (Leaf) เช่น 265 แยกเป็น 26 กับ 5 โดยทั่วไปมักกำหนดให้มีลำต้นอยู่ในช่วง 5 – 20 ลำต้น

2. เขียนลำต้นเป็นตัวเลขจากน้อยไปมาก หรือมากไปน้อยก็ได้ โดยเขียนจากบนลงล่างชิดขอบซ้าย และเขียนใบไม้เป็นตัวเลขจากน้อยไปมากที่ด้านขวาของแต่ละลำต้น การเขียนใบไม้เป็นตัวเลขจากน้อยไปมากมีข้อดีคือสามารถหาค่าควอไทล์ เดไซล์ และเปอร์เซ็นต์ไทล์ได้

ตัวอย่างที่ 1.1 จากตารางที่ 1.1 จงสร้างแผนภาพลำต้นและใบไม้

ส่วนของลำต้น เลือกค่า 17, 18, 19, ..., 34 ส่วนของใบไม้คือตัวเลขข้างท้าย

ดังนั้นผลของแผนภาพลำต้นและใบไม้สามารถพิจารณาได้ในรูปที่ 1.2

ลำต้น (Stem)	ใบไม้ (Leaf)	ความถี่
17	6	1
18	7	1
19	7	1
20	0, 5, 8	3
21	0, 4, 5	3
22	1, 0, 8, 3	4
23	5, 1, 1, 4, 5, 5	6
24	2, 8, 2, 6, 8, 3, 5	7
25	4, 0, 8, 0, 0, 7, 8, 3, 4, 8, 1	11
26	5, 5, 5, 1, 2, 3, 0, 0, 5, 3, 8, 7, 0, 0, 4, 5, 9, 5, 4, 7, 9	21
27	8, 4, 1, 4, 0, 6, 6, 4, 8, 2, 4, 1, 7, 5	14
28	0, 6, 1, 0, 1, 0, 0, 3, 7, 3	10
29	4, 6, 8, 9, 9, 3, 0	7
30	7, 1, 0, 8	4
31	7, 8	2
32	1, 8	2
33	7, 4	2
34	6	1
		100

รูปที่ 1.2 แผนภาพลำต้นและใบไม้ของค่าความแข็งแรงของขวดน้ำ 1 ลิตร

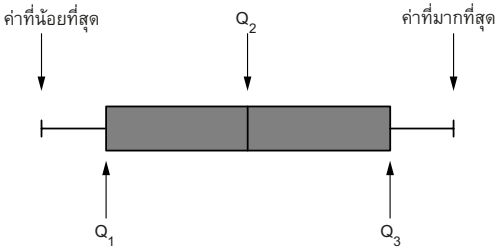
จากรูปที่ 1.2 จะพบว่าค่าความแข็งแรงส่วนใหญ่ตกอยู่ระหว่าง 220 และ 330 ปอนด์/ตารางนิ้ว และค่าแนวโน้มสู่ศูนย์กลางตกอยู่ระหว่าง 260 และ 270 ปอนด์/ตารางนิ้ว ค่าความแข็งแรงมีการกระจายตัวแบบสมมาตรรอบๆ ค่าเฉลี่ย จากรูปที่ 1.2 สามารถเรียงลำดับค่าของส่วนของใบไม้จากน้อยไปมากได้ดังรูปที่ 1.3

ลำต้น (Stem)	ใบไม้ (Leaf)	ความถี่
17	6	1
18	7	1
19	7	1
20	0, 5, 8	3
21	0, 4, 5	3
22	0, 1, 3, 8	4
23	1, 1, 4, 5, 5, 5	6
24	2, 2, 3, 5, 6, 8, 8	7
25	0, 0, 0, 1, 3, 4, 4, 7, 8, 8, 8	11
26	0, 0, 0, 0, 1, 2, 3, 3, 4, 4, 5, 5, 5, 5, 5, 5, 7, 7, 8, 9, 9	21
27	0, 1, 1, 2, 4, 4, 4, 4, 5, 6, 6, 7, 8, 8	14
28	0, 0, 0, 0, 1, 1, 3, 3, 6, 7	10
29	0, 3, 4, 6, 8, 9, 9	7
30	0, 1, 7, 8	4
31	7, 8	2
32	1, 8	2
33	4, 7	2
34	6	1
		100

รูปที่ 1.3 แผนภาพลำต้นและใบไม้ที่เรียงลำดับข้อมูลของค่าความแข็งแรงของขวดน้ำ 1 ลิตร

1.2.4 กราฟแบบสี่เหลี่ยม

กราฟแบบสี่เหลี่ยม (Box Plot) หรือกราฟแบบสี่เหลี่ยมและหนวด (Box and Whisker Plot) มีประสิทธิผลในการวิเคราะห์หาข้อมูลที่ออกนอกอย่างผิดปกติเหนือกว่าฮิสโตแกรมและแผนภาพลำต้นและใบไม้ กราฟแบบสี่เหลี่ยมสามารถแสดงได้ทั้งในแนวนอนและแนวตั้ง ดังแสดงในรูปที่ 1.4



รูปที่ 1.4 กราฟแบบสี่เหลี่ยม

ขั้นตอนการสร้างกราฟแบบสี่เหลี่ยมมี 4 ขั้นตอนคือ

1. หาตำแหน่งที่ของควอไทล์ที่ 1 (Q_1), ควอไทล์ที่ 2 (Q_2) และควอไทล์ที่ 3 (Q_3) ด้วยสมการที่ (1.1)

$$Q_r = \frac{r(n+1)}{4} \quad \textbf{(1.1)}$$

โดยที่ n แทนจำนวนข้อมูลทั้งหมด

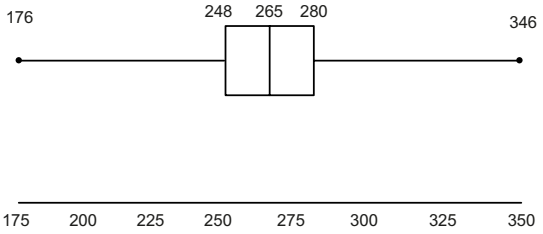
2. หาค่าควอไทล์ที่ 1, ควอไทล์ที่ 2 และควอไทล์ที่ 3 ถ้าตำแหน่งที่ได้จากข้อที่ 1 ตรงกับข้อมูลตัวใด จะได้ว่าข้อมูลตัวนั้นเป็นคำตอบที่ต้องการ แต่ถ้าไม่ตรงกับข้อมูลตัวใดเลย ให้ใช้วิธีเทียบบัญญัติไตรยางค์ในการหาคำตอบ

กำหนดให้ค่า Q_1 และ Q_3 เป็นขอบล่างและขอบบนหรือกรอบซ้ายและกรอบขวา ตามลำดับของกรอบสี่เหลี่ยม โดยที่ความยาวของกรอบสี่เหลี่ยมจะเท่ากับค่าพิสัยระหว่างควอไทล์ (Interquartile Range; IQR) ซึ่งจะครอบคลุมข้อมูลที่มีค่ากลางๆ เท่ากับ 50% ($IQR = Q_3 - Q_1$) ส่วนค่า Q_2 หรือค่ามัธยฐาน (Median) ให้ลากภายในกรอบสี่เหลี่ยม

3. ลากเส้นตรงจากขอบด้านขวาหรือขอบบนของกรอบสี่เหลี่ยม จาก Q_3 ไปหาค่าที่มากที่สุดของข้อมูล และเรียกชื่อเส้นตรงนี้ว่า “หนวดด้านบน (Upper Whisker)”

4. ลากเส้นตรงจากขอบด้านซ้ายหรือขอบล่างของกรอบสี่เหลี่ยม จาก Q_1 ไปหาค่าที่น้อยที่สุดของข้อมูล และเรียกชื่อเส้นตรงนี้ว่า “หนวดด้านล่าง (Lower Whisker)”

จากตารางที่ 1.1 สามารถสร้างกราฟแบบสี่เหลี่ยมของค่าความแข็งแรงของขวดน้ำ 1 ลิตร ได้ดังแสดงในรูปที่ 1.5



รูปที่ 1.5 กราฟแบบสี่เหลี่ยมของค่าความแข็งแรงของขวดน้ำ 1 ลิตร

จากรูปที่ 1.5 พบว่าข้อมูลมีการกระจายตัวรอบค่าแนวโน้มสู่ศูนย์กลางค่อนข้างสม่ำเสมอ เพราะหนวดด้านบนและล่างยาวใกล้เคียงกัน และความยาวในช่วง (Q_1 ถึง Q_2) (Q_2 ถึง Q_3) ยาวใกล้เคียงกัน กราฟแบบสี่เหลี่ยมมีข้อดีคือสามารถใช้ในการเปรียบเทียบสิ่งตัวอย่างได้ตั้งแต่ 2 กลุ่มขึ้นไป ดังแสดงในตัวอย่างที่ 1.2

ตัวอย่างที่ 1.2 ของเหลว 3 ชนิดคือ

Cis–2–Methylcyclohexanol	กำหนดให้เป็นส่วนผสมที่ 1
1,3–Propylene Glycol	กำหนดให้เป็นส่วนผสมที่ 2
1–Amino–2–Propanol	กำหนดให้เป็นส่วนผสมที่ 3

ผลจากการทดสอบความหนืด (เซนติพอยซ์) แสดงดังตารางที่ 1.4

ตารางที่ 1.4 ค่าความหนืดของ 3 ส่วนผสมของวัตถุดิบ

ส่วนผสมที่ 1	ส่วนผสมที่ 2	ส่วนผสมที่ 3
22.02	21.49	20.33
23.83	22.67	21.67
26.67	24.62	24.67
25.38	24.18	22.45
25.49	22.78	22.28
23.50	22.58	21.95
25.90	24.46	20.49
24.98	23.79	21.81

จงเปรียบเทียบค่าความหนืดของ 3 สูตรโดยการสร้างกราฟแบบสี่เหลี่ยม

จากสมการที่ (1.1) $Q_r = \frac{r(n + 1)}{4}$ พิจารณาส่วนผสมที่ 1 มีข้อมูล 8 ตัว (n = 8) เรียงข้อมูลจากน้อยไปมากจะได้ดังนี้

22.02, 23.50, 23.83, 24.98, 25.38, 25.49, 25.90, 26.67

ตำแหน่งที่ของ $Q_1 = \frac{1(8 + 1)}{4} = 2.25$ (อยู่ระหว่างข้อมูลตัวที่ 2 และ 3)

ดังนั้นตำแหน่งต่างกัน 1 ตำแหน่ง ค่าความหนืดเพิ่มขึ้น 23.83 – 23.50 = 0.33

ตำแหน่งต่างกัน 2.25 – 2.00 = 0.25 ตำแหน่ง ค่าความหนืดเพิ่มขึ้น $\frac{(0.25 \times 0.33)}{1} = 0.0825$

ค่าของ Q_1 คือ 23.50 + 0.0825 = 23.58

ในทำนองเดียวกัน

ตำแหน่งที่ของ $Q_2 = \frac{2(8 + 1)}{4} = 4.50$

ค่าของ Q_2 คือ 24.98 + 0.2000 = 25.18

ตำแหน่งที่ของ $Q_3 = \frac{3(8 + 1)}{4} = 6.75$

ค่าของ Q_3 คือ 25.49 + 0.3075 = 25.80

พิจารณาส่วนผสมที่ 2 มีข้อมูล 8 ตัว (n = 8) เรียงข้อมูลจากน้อยไปมากจะได้ดังนี้

21.49, 22.58, 22.67, 22.78, 23.79, 24.18, 24.46, 24.62

ตำแหน่งที่ของ $Q_1 = \frac{1(8+1)}{4} = 2.25$

ค่าของ Q_1 คือ 22.58 + 0.0225 = 22.60

ตำแหน่งที่ของ $Q_2 = \frac{2(8+1)}{4} = 4.50$

ค่าของ Q_2 คือ 22.78 + 0.5050 = 23.29

ตำแหน่งที่ของ $Q_3 = \frac{3(8+1)}{4} = 6.75$

ค่าของ Q_3 คือ 24.18 + 0.2100 = 24.39

พิจารณาส่วนผสมที่ 3 มีข้อมูล 8 ตัว (n = 8) เรียงข้อมูลจากน้อยไปมากจะได้ดังนี้

20.33, 20.49, 21.67, 21.81, 21.95, 22.28, 22.45, 24.67

ตำแหน่งที่ของ $Q_1 = \frac{1(8+1)}{4} = 2.25$

ค่าของ Q_1 คือ 20.49 + 0.2950 = 20.79

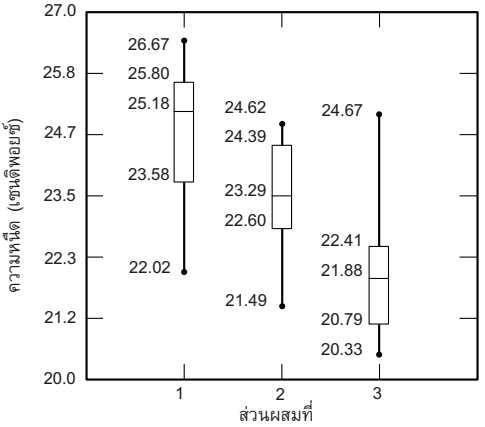
ตำแหน่งที่ของ $Q_2 = \frac{2(8+1)}{4} = 4.50$

ค่าของ Q_2 คือ 21.81 + 0.0700 = 21.88

ตำแหน่งที่ของ $Q_3 = \frac{3(8+1)}{4} = 6.75$

ค่าของ Q_3 คือ 22.28 + 0.1275 = 22.41

นำค่าของควอไทล์ของส่วนผสมทั้ง 3 ไปเขียนกราฟแบบสี่เหลี่ยมได้ดังแสดงในรูปที่ 1.6



รูปที่ 1.6 กราฟแบบสี่เหลี่ยมของค่าความหนืดของ 3 ส่วนผสมของวัตถุดิบ

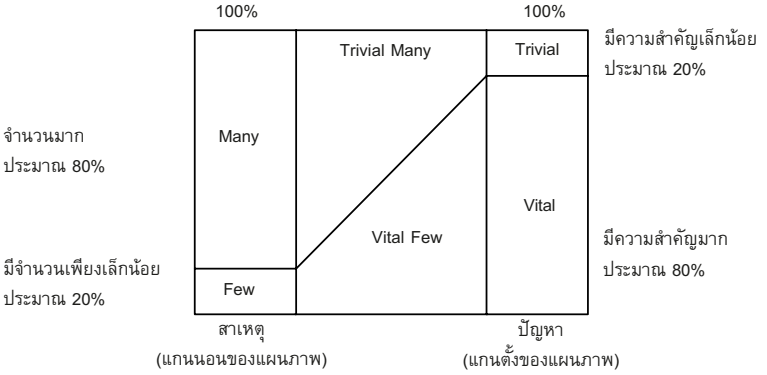
จากรูปที่ 1.6 พบว่าส่วนผสมที่ 1 มีค่าความหนืดสูงกว่าส่วนผสมที่ 2 และส่วนผสมที่ 2 มีค่าความหนืดสูงกว่าส่วนผสมที่ 3 นอกจากนั้นการกระจายของความหนืดของทั้ง 3 ส่วนผสมรอบค่าแนวโน้มสู่ศูนย์กลางยังไม่สม่ำเสมอ และค่าความหนืดสูงสุดของส่วนผสมที่ 3 ดูเหมือนว่าจะมีค่าสูงผิดปกติ

1.2.5 แผนภาพพาเรโต

อัลเฟรโด พาเรโต (Alfredo Pareto, พ.ศ. 2391 – 2466) เป็นนักเศรษฐศาสตร์ชาวอิตาเลียน ได้แสดงผลการวิจัยของเขา ซึ่งแสดงให้เห็นถึงการกระจายรายได้ของประชากรในยุโรป พบว่ารายได้มากอยู่ในมือของประชากรกลุ่มน้อย ขณะที่รายได้น้อยอยู่ในมือของประชากรกลุ่มใหญ่

โจเซฟ จูราน (Joseph Juran, พ.ศ. 2535) เป็นนักเศรษฐศาสตร์ชาวอเมริกัน ยอมรับแนวคิดดังกล่าวนี้ว่าสามารถนำไปประยุกต์ใช้ได้ในหลายสาขาวิชา และจูรานได้สร้างวลีว่า ประเภทน้อยชนิดแต่มีผลมาก และประเภทมากชนิดแต่มีผลน้อย (Vital Few and Trivial Many)

จูรานแนะนำให้ใช้ตัวเลขหยาบๆ กับการตัดสินใจหลักการพาเรโต (Pareto Principle) คือ “80–20” ซึ่งหมายความว่า “ปัญหาหรือความสูญเสียที่มีความสำคัญมากจำนวน 80% มักจะมีสาเหตุมาจากประมาณ 20% ของสาเหตุทั้งหมด (The Vital Few) ในขณะที่อีกประมาณ 80% ของสาเหตุจะมีผลต่อปัญหาที่มีความสำคัญเพียงเล็กน้อยอีกจำนวน 20% ของปัญหาเท่านั้น (The Trivial Many)” ดังนั้นแผนภาพพาเรโตจะเป็นเครื่องมือที่ชี้ให้เห็นว่าปัญหาสำคัญจริงๆ นั้นมาจากไม่กี่สาเหตุ และปัญหาที่มีความสำคัญน้อยเกิดจากสาเหตุมากมาย สรุปว่าในการแก้ปัญหาไม่จำเป็นต้องแก้ทุกสาเหตุให้หมด แต่ให้เลือกแก้เฉพาะสาเหตุหลักที่สำคัญ จูรานแนะนำให้ใช้ค่าใช้จ่ายในรูปตัวเงินเป็นหน่วยวัด ดังแสดงในรูปที่ 1.7



รูปที่ 1.7 หลักการของพาเรโต

แผนภาพพาเรโต (Pareto Diagram) มีลักษณะเป็นกราฟแท่งที่แบ่งแยกข้อมูลเป็นช่วงๆ จากมากไปน้อย และจากซ้ายไปขวา โดยแกน y มี 2 แกนคือ แกนซ้ายมือแทนความถี่ (เช่น จำนวนจุดบกพร่อง จำนวนคำร้องเรียน

จำนวนอุบัติเหตุ เป็นต้น) และแกนขวามือแทนเปอร์เซ็นต์ แกน x แทนสาเหตุ (เช่น ในเรื่องของปัญหาของจุดบกพร่อง อาจจำแนกสาเหตุได้จากพนักงาน เครื่องจักรกล วิธีการทำงาน ชนิดของวัตถุดิบ เป็นต้น) ดังแสดงในรูปที่ 1.8 แผนภูมิพาเรโตต่างจากฮิสโตแกรมตรงที่แกนนอนของแผนภูมิพาเรโตเป็นประเภทของข้อมูล แต่แกนนอนของฮิสโตแกรมเป็นตัวเลข

ขั้นตอนการสร้างแผนภาพพาเรโตมีรายละเอียดดังต่อไปนี้

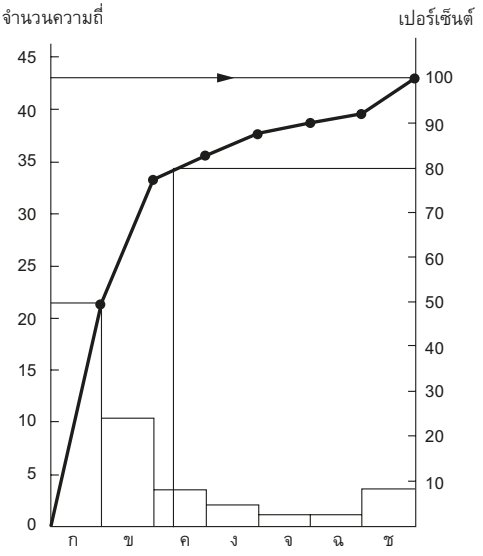
1. ตัดสินใจว่าจะศึกษาปัญหาอะไร และแยกสาเหตุของการเกิดปัญหา

2. ออกแบบใบบันทึกข้อมูล (กำหนดช่วงเวลา ระยะเวลา และวิธีการเก็บรวบรวมข้อมูล)

3. ทำการจดบันทึก (ในช่วงเวลาที่กำหนด) จากสถานที่ที่ต้องการวิเคราะห์ปัญหา และคำนวณหายอดรวมและเปอร์เซ็นต์สะสมของแต่ละสาเหตุที่ได้จากการจำแนกข้อมูล

4. เขียนแกนนอนและแกนตั้ง แกนนอนเขียนจากสาเหตุที่มีความถี่สูงไว้ด้านซ้าย และสาเหตุที่มีความถี่ต่ำไว้ด้านขวา โดยต้องให้แท่ง “อื่นๆ” (ความถี่ไม่ควรเกิน 20% ของเปอร์เซ็นต์สะสม) อยู่ด้านขวาสุด ส่วนแกนตั้งเขียนแกนความถี่และแกนเปอร์เซ็นต์

5. เขียนกราฟแท่งที่มีความกว้างเท่ากัน (กราฟแท่งควรมีประมาณ 6 ถึง 10 แท่งเท่านั้น) และลากเส้นความถี่สะสมจากซ้ายไปขวา



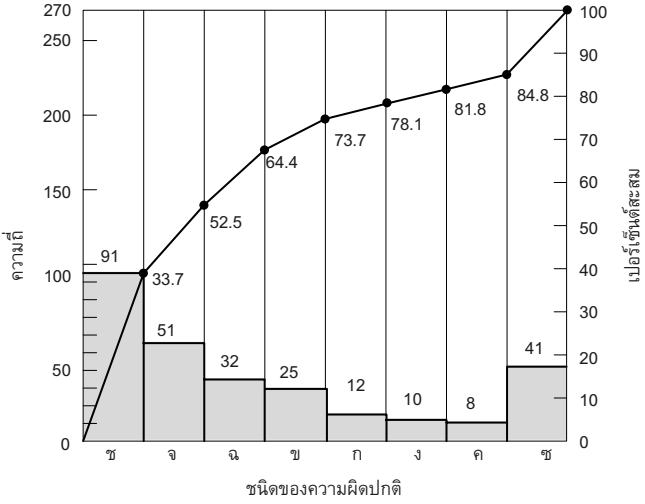
รูปที่ 1.8 แผนภาพพาเรโต แสดงชนิดของสาเหตุ ก ถึง ช

การประยุกต์แผนภาพพาเรโตในการเลือกที่จะแก้สาเหตุในกรณีที่มีข้อมูลทางสถิติสนับสนุน ให้เลือกแก้สาเหตุทั้งหมดที่อยู่ในช่วงเปอร์เซ็นต์สะสมประมาณ 80 เปอร์เซ็นต์ ดังนั้นในรูปที่ 1.8 จะพบว่า ควรพิจารณาแก้ไขสาเหตุ

ก, ข และ ค เพื่อให้เปอร์เซ็นต์ของความถี่ลดลงอย่างน้อย 80 เปอร์เซ็นต์ โดยลำดับการแก้ปัญหาเริ่มจากแก้ที่สาเหตุ ก เมื่อสาเหตุ ก ถูกแก้ไปแล้ว สาเหตุ ข จะเป็นสาเหตุหลักแทน เป็นต้น แต่ในกรณีที่ไม่มีข้อมูลทางสถิติสนับสนุน ให้ใช้วิธีโหวตเสียงและลงมติ หรือให้คะแนนจากความเป็นไปได้ ความรุนแรง และความถี่

ตัวอย่างที่ 1.3 จงสร้างแผนภาพพาเรโตและตีความหมายจากแผนภาพ จากผลของการเก็บรวบรวมข้อมูลและจำแนกความผิดปกติที่เป็นไปได้ของสายการผลิตเพลามีดังต่อไปนี้

รหัส	ชนิดของความบกพร่อง	ความถี่
ก	ร้าว	12
ข	คราบสนิม	25
ค	ผิวไม่เรียบ	8
ง	รอยย่น	10
จ	ช่องว่าง	51
ฉ	รูพรุน (ตามด)	32
ช	รอยข่วน	91
ซ	อื่นๆ	41



รูปที่ 1.9 แผนภาพพาเรโตจำแนกความผิดปกติของสายการผลิตเพลา

จากแผนภาพ พบว่าควรแก้ปัญหา ช, จ, ฉ, ข, ก และ ง ตามลำดับ เพราะจะทำให้เปอร์เซ็นต์ของความผิดพลาดลดลงถึง 80%

1.3 การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูล คือการใช้ความรู้ทางคณิตศาสตร์เพื่อหาความสัมพันธ์ของข้อมูลเพื่อทำการตัดสินใจโดยการสรุปข้อมูลเพื่อหาค่าที่ควรจะเป็น เนื่องจากค่าที่ควรจะเป็นจะเป็นค่าที่ข้อมูลมีแนวโน้มลู่เข้าหา ดังนั้นจึงเรียกลักษณะสมบัติประเภทนี้ของข้อมูลว่า “ค่าแนวโน้มสู่ศูนย์กลาง (Central Tendency or Location)” ซึ่งสามารถวัดได้ด้วยตัวสถิติ 3 ตัวคือ ค่าเฉลี่ยเลขคณิต (Arithmetic Average) ค่ามัธยฐาน (Median) และค่าฐานนิยม (Mode) เนื่องจากข้อมูลทั้งหมดอยู่ภายใต้ความเบี่ยงเบนโดยธรรมชาติ จึงมีความจำเป็นต้องพิจารณาความเบี่ยงเบนก่อนการวิเคราะห์แนวโน้มสู่ศูนย์กลางเสมอว่า ข้อมูลที่วิเคราะห์มีคุณภาพหรือไม่ โดยการพิจารณาขนาดของความเบี่ยงเบนที่เรียกว่า “การกระจาย (Dispersion)” ซึ่งสามารถวัดได้ด้วยตัวสถิติ 3 ตัวคือ พิสัย (Range) ความแปรปรวน (Variance) และสัมประสิทธิ์แห่งความผันแปร (Coefficient of Variation) นอกจากนี้ยังจำเป็นต้องวิเคราะห์ถึงรูปทรงของรูปแบบความเบี่ยงเบนด้วย โดยคำนึงถึงลักษณะสมบัติ 2 ประการคือ ค่าความเบ้ (Skewness) และค่าความโด่ง (Kurtosis) ของข้อมูล

1.3.1 การวิเคราะห์ค่าแนวโน้มสู่ศูนย์กลางของข้อมูล

ค่าแนวโน้มสู่ศูนย์กลางของข้อมูลหรือการวัดค่ากลางของข้อมูล หมายถึงค่าสถิติที่เป็นตัวแทนของข้อมูลทั้งหมด ซึ่งจะสะท้อนให้ทราบถึงผลจากปัจจัยที่ได้รับการควบคุมในขณะเก็บข้อมูล ค่าแนวโน้มสู่ศูนย์กลางของข้อมูลที่ดีต้องใกล้เคียงหรือตรงกับความเป็นจริงมากที่สุด โดยมีหน่วยเช่นเดียวกับหน่วยของข้อมูล การวิเคราะห์ค่าแนวโน้มสู่ศูนย์กลางของข้อมูลมี 6 วิธีคือ ค่าเฉลี่ยเลขคณิต ค่ามัธยฐาน ค่าฐานนิยม ค่าเฉลี่ยเรขาคณิต ค่าเฉลี่ยฮาร์มอนิก และค่ากึ่งกลางพิสัย ในที่นี้ขออธิบายค่าแนวโน้มสู่ศูนย์กลางของข้อมูลที่นิยมใช้ในงานสถิติชั้นสูง 3 วิธี คือ

1. ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) หรือค่ามัชฌิมเลขคณิต หรือค่าเฉลี่ย (Mean) หมายถึงค่าที่ได้จากการกำจัด (Average Out) ผลของตัวแปรที่ไม่สามารถควบคุมได้ที่อยู่ในข้อมูล โดยผลดังกล่าวจะเป็นผลที่ไม่ได้รับการควบคุม (Uncontrolled Effect) ถ้าข้อมูลที่เก็บคือสิ่งตัวอย่าง เราจะเรียกค่าเฉลี่ยของข้อมูลทั้งหมดในสิ่งตัวอย่างว่า **ค่าเฉลี่ยของสิ่งตัวอย่าง (Sample Mean : $\bar{X}$)** ถ้าข้อมูลที่เก็บคือประชากร เราจะเรียกค่าเฉลี่ยของข้อมูลทั้งหมดในประชากรว่า **ค่าเฉลี่ยของประชากร (Population Mean : μ)**

สูตรค่าเฉลี่ยของสิ่งตัวอย่างแบบไม่แจกแจงความถี่

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad (1.2)$$

กำหนดให้ n = จำนวนข้อมูลทั้งหมดของสิ่งตัวอย่าง

จากข้อมูลของค่าความแข็งแรง (ปอนด์/ตารางนิ้ว) ของขวดน้ำประเภทแก้ว 1 ลิตร จำนวน 100 ขวด ในตารางที่ 1.1 สามารถหาค่าเฉลี่ยของสิ่งตัวอย่างได้ดังนี้

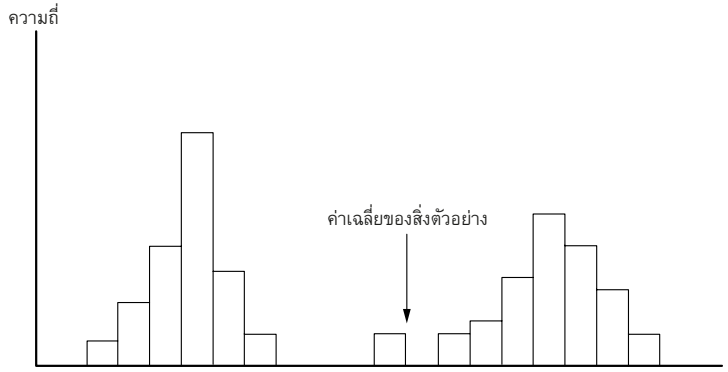
$$\bar{x} = \frac{\sum_{i=1}^{100} x_i}{n}$$

$$= \frac{26,406}{100} = 264.06 \text{ ปอนด์/ตารางนิ้ว}$$

ในรูปที่ 1.1 ฮิสโตแกรมของค่าความแข็งแรงของขวดน้ำประเภทแก้ว 1 ลิตร จำนวน 100 ขวด ดูเหมือนว่าจะมีค่าเฉลี่ยของสิ่งตัวอย่างเท่ากับ 264.06 ปอนด์/ตารางนิ้วเช่นกัน

หมายเหตุ *:*

1) ค่าเฉลี่ยของสิ่งตัวอย่างไม่ได้เป็นตัวบอกว่าข้อมูลส่วนใหญ่อยู่รอบๆ ค่าเฉลี่ยของสิ่งตัวอย่าง ดังแสดงในรูปที่ 1.10 (รูปแบบที่ 8)



รูปที่ 1.10 ตัวอย่างฮิสโตแกรมที่ข้อมูลส่วนใหญ่ไม่ได้อยู่รอบค่าเฉลี่ยของสิ่งตัวอย่าง

2) ถ้าค่าเฉลี่ยของสิ่งตัวอย่างอยู่กลางข้อมูลส่วนใหญ่ แล้วฮิสโตแกรมจะมีลักษณะคล้ายรูประฆังคว่ำ

3) ค่าเฉลี่ยเป็นตัวสถิติที่ใช้วัดค่าแนวโน้มสู่ศูนย์กลางของข้อมูลที่เป็นปกติเท่านั้น กล่าวคือข้อมูลต้องมีรูปทรงของรูปแบบข้อมูลที่ค่อนข้างสมมาตรและเป็นแบบมียอดเดียว (Unimodal Pattern)

สูตรค่าเฉลี่ยของสิ่งตัวอย่างแบบแจกแจงความถี่

$$\bar{x} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i} \qquad \textbf{(1.3)}$$

กำหนดให้ x_i = จุดกึ่งกลางของอันตรภาคชั้นที่ i

f_i = ความถี่หรือจำนวนข้อมูลชั้นที่ i

$\sum_{i=1}^{n} f_i$ = จำนวนข้อมูลทั้งหมด