

ปฏิบัติการที่ 3

การคำนวณค่าสถิติเชิงพรรณนาจากรางข้อมูลหลัก

วัตถุประสงค์

1. เพื่อให้สามารถใช้เครื่องมือวิเคราะห์ข้อมูลในการคำนวณค่าสถิติเชิงพรรณนา เพื่ออธิบายลักษณะของข้อมูลได้
2. เพื่อให้สามารถใช้เครื่องมือวิเคราะห์ข้อมูลในการคำนวณหาค่าสหสัมพันธ์ (Correlation) ระหว่างตัวแปรได้

1. ชุดข้อมูลปฏิบัติการ

- ชุดข้อมูล Iris (สำหรับการสาธิต)
- ชุดข้อมูล Airfoil Self-Noise (สำหรับการฝึกปฏิบัติการ)

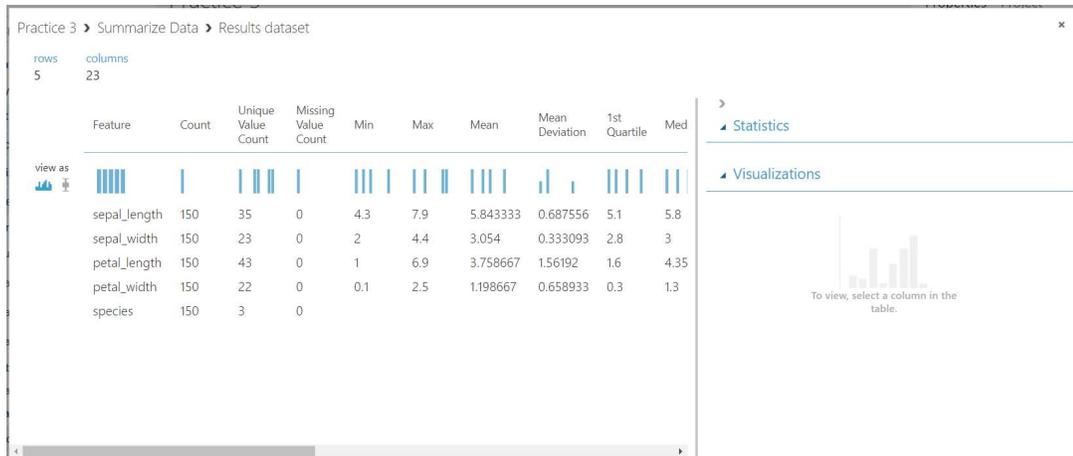
2. ขั้นตอนปฏิบัติการ

ขั้นตอนปฏิบัติการ มีดังนี้

1. นำเข้าชุดข้อมูล Iris จากแฟ้มข้อมูล iris.csv ตั้งชื่อชุดข้อมูลเป็น iris
2. ทำการสร้างการทดลอง โดยกำหนดชื่อการทดลองเป็น “Practice 3”
3. นำชุดข้อมูล iris เข้าสู่การทดลองโดยลากโมดูลชุดข้อมูล iris ที่อยู่ภายใต้ Saved Datasets → My Datasets ในหน้าต่างย่อย Modules มาวางบน Workspace
4. เข้าดูข้อมูลในชุดข้อมูล iris (คลิกที่โหนดส่วนต่อประสานข้อมูลออก แล้วเลือก Visualize) ตรวจสอบชนิดข้อมูล กำหนดให้แต่ละตัวแปรต้องมีชนิดข้อมูล ดังนี้

ตัวแปร	ชนิดข้อมูล
sepal_length	Numeric Feature (Floating Point)
sepal_width	Numeric Feature (Floating Point)
petal_length	Numeric Feature (Floating Point)
petal_width	Numeric Feature (Floating Point)
species	Categorical Feature

- ทำการแก้ไขชนิดข้อมูลของตัวแปร (ใช้โมดูล Edit Metadata) หากชนิดข้อมูลเดิมไม่ตรงกับชนิดข้อมูลที่กำหนด
- คำนวณค่าสถิติเชิงพรรณนาของแต่ละตัวแปรในชุดข้อมูล iris โดยใช้โมดูล Summarize Data (อยู่ภายใต้ Statistical Functions)
- คลิกคำสั่ง RUN เมื่อโปรแกรมประมวลผลเรียบร้อยแล้ว ดูผลลัพธ์ของโมดูล Summarize Data โดยการคลิกที่ไอคอนส่วนต่อประสานข้อมูลออก แล้วเลือก Visualize จะปรากฏหน้าต่าง ดังรูป



- ศึกษาค่าสถิติเชิงพรรณนาของแต่ละตัวแปรจากผลลัพธ์ของโมดูล Summarize Data
- คำนวณค่าสหสัมพันธ์เชิงเส้นตรงของแต่ละคู่ตัวแปรในชุดข้อมูล iris โดยใช้โมดูล Compute Linear Correlation (อยู่ภายใต้ Statistical Functions)
- คลิกคำสั่ง RUN เมื่อโปรแกรมประมวลผลเรียบร้อยแล้ว ดูผลลัพธ์ของโมดูล Compute Linear Correlation โดยการคลิกที่ไอคอนส่วนต่อประสานข้อมูลออก แล้วเลือก Visualize จะปรากฏหน้าต่าง ดังรูป

Practice 3 > Compute Linear Correlation > Results dataset

sepal_length	sepal_width	petal_length	petal_width	species
1	-0.109369	0.871754	0.817954	NaN
-0.109369	1	-0.420516	-0.356544	NaN
0.871754	-0.420516	1	0.962757	NaN
0.817954	-0.356544	0.962757	1	NaN
NaN	NaN	NaN	NaN	NaN

- ศึกษาค่าสหสัมพันธ์เชิงเส้นตรงของแต่ละคู่ตัวแปรในชุดข้อมูล iris จากผลลัพธ์ของโมดูล Compute Linear Correlation

3. แบบฝึกปฏิบัติการ

ให้นักศึกษาทำแบบฝึกปฏิบัติการ ตามลำดับขั้นตอนต่อไปนี้

1. ให้นักศึกษานำชุดข้อมูล Airfoil Self-Noise จากแฟ้มข้อมูล airfoil_self_noise.csv เข้าสู่โปรแกรม ML Studio กำหนดชื่อชุดข้อมูลเป็น “airfoil self-noise”
2. สร้างการทดลอง กำหนดชื่อเป็น “Lab 3” โดยให้นำชุดข้อมูล airfoil self-noise เข้าสู่การทดลอง
3. คำนวณค่าสถิติเชิงพรรณนาของแต่ละตัวแปรในชุดข้อมูล airfoil self-noise โดยใช้โมดูล Summarize Data แล้วศึกษาและอภิปรายผลลัพธ์
4. คำนวณค่าสหสัมพันธ์เชิงเส้นตรงของแต่ละคู่ตัวแปรในชุดข้อมูล airfoil self-noise โดยใช้โมดูล Compute Linear Correlation แล้วศึกษาและอภิปรายผลลัพธ์

สิ่งที่ต้องส่งเป็นการบ้าน ภาพหน้าจอ Workspace ของนักศึกษาที่ใช้ทำแบบฝึกปฏิบัติการ โดยให้เห็นกล่องโมดูลทั้งหมดและชื่อ Workspace ซึ่งเป็นชื่อของนักศึกษา ตั้งชื่อไฟล์ในรูปแบบ Lab_03_id.jpg โดยแทน id ด้วยรหัสนักศึกษา ส่งผ่านเว็บไซต์ <http://hw.cs.science.cmu.ac.th>