SC112502 ปฏิบัติการพันธุศาสตร์เบื้องต้น

ปฏิบัติการที่ 12 การจัดการข้อมูลตารางและการสร้างกราฟด้วย Google Sheets

Spreadsheet Data Management and Basic Charting

เรียนรู้ฟังก์ชัน 21 ตัวที่ใช้ทำงานจริง ผ่านข้อมูลแปลงทดลองพันธุศาสตร์

รายการ	**รายละเอียด**
ระยะเวลา	3 ชั่วโมง (5 กิจกรรม M1-M5)
คะแนน	20 คะแนน - ส่งงานภายในชั้นเรียน ไม่รับงานย้อนหลัง
รูปแบบงาน	งานเดี่ยว บน Student Workbook ที่ดาวน์โหลดจาก DigiClass
การส่งงาน	ดาวน์โหลดเป็นไฟล์ .xlsx ตั้งชื่อ 12_รหัสประจำตัว.xlsx แล้วอัปโหลดเข้า Google Form
แพลตฟอร์ม	Google Sheets
ทักษะคอมพิวเตอร์ที่ต้องมีก่อน	ไม่ต้องมี - กิจกรรม M1 สอนตั้งแต่เริ่มต้น
ความรู้พันธุศาสตร์ที่ต้องมีก่อน	อัตราส่วนของการผสมสองลักษณะ (9:3:3:1) และกฎการแยกตัว

1. จุดประสงค์การเรียนรู้

เมื่อสิ้นสุดปฏิบัติการนี้ นักศึกษาสามารถ

1. ใช้ (use) สเปรดชีตในระดับใช้งานจริงได้ คือ อ้างอิงเซลล์ เขียนสูตรที่ขึ้นต้นด้วย = ใช้ฟังก์ชันกับช่วงข้อมูล ตรึงเซลล์ด้วย $ และเติมสูตรลงหลายร้อยแถวได้
2. ทำความสะอาด (clean) ข้อมูลดิบด้วยฟังก์ชันข้อความ TRIM PROPER LEN LEFT MID SUBSTITUTE และเครื่องหมาย & พร้อมทำเครื่องหมายแถวที่ใช้ไม่ได้ด้วย IF AND และ COUNTIF
3. สรุป (summarise) ชุดข้อมูลด้วยสถิติพื้นฐาน SUM AVERAGE MIN MAX COUNT COUNTA ROUND และใช้ MIN กับ MAX เป็นเครื่องมือตรวจหาข้อมูลที่เป็นไปไม่ได้
4. นับและค้นหา (count & look up) ด้วย COUNTIF COUNTIFS SUMIFS VLOOKUP และ IFERROR จนได้อัตราส่วนฟีโนไทป์ของรุ่น F2 และจำนวนที่คาดหมายตามกฎเมนเดล
5. สร้าง (create) กราฟแท่งเปรียบเทียบสองชุดข้อมูล ที่มีองค์ประกอบครบตามมาตรฐานการตีพิมพ์

2. ภาพรวมและตารางเวลา

ข้อมูลการทดลอง

รุ่นพ่อแม่ (P): ผลสีแดง ทรงกลม ผสมกับ ผลสีเหลือง ทรงลูกแพร์

รุ่นลูก (F1): ผลสีแดง ทรงกลม ทั้งหมด

รุ่น F2: ปล่อยให้ F1 ผสมตัวเอง แล้วบันทึกลักษณะและน้ำหนักผลของต้น F2 ทุกต้น

ไฟล์ที่คุณได้รับคือสมุดบันทึกแปลงทดลองที่ยังไม่ได้ทำความสะอาด มีประมาณ 300 แถว งานของคุณคือเปลี่ยนบันทึกนั้นให้เป็นตารางสรุปและกราฟที่นำเสนอได้

กิจกรรม	แท็บ	ฟังก์ชันที่ได้ใช้	คะแนน

M1 พื้นฐานสเปรดชีตและสูตรแรก	01_ฝึกพื้นฐาน	ชื่อเซลล์ = ช่วง : การเติมสูตรตรึง $ SUM	3
M2 ฟังก์ชันข้อความและการทำความสะอาด	03_Clean	TRIM PROPER & LEFT MID LEN SUBSTITUTE COUNTIF IF AND	5
M3 สรุปข้อมูลด้วยสถิติพื้นฐาน	04_Summary	COUNTA COUNT MIN MAX SUM AVERAGE ROUND	5
M4 นับตามเงื่อนไข ตรรกะ และค้นหา	05_Ratio	COUNTIFS SUMIFS VLOOKUP IFERROR	5
M5 การสร้างกราฟ	06_Chart	การอ้างอิงข้ามแท็บ · องค์ประกอบกราฟ	2

3. ฟังก์ชันทั้ง 21 ตัวที่จะได้ใช้

ฟังก์ชัน	ตัวอย่าง	ทำอะไร	กิจกรรม
TRIM	=TRIM(D2)	ตัดช่องว่างส่วนเกินหน้า-หลัง	M2
PROPER	=PROPER(TRIM(D2))	ทำตัวแรกเป็นตัวใหญ่ ที่เหลือเป็นตัวเล็ก	M2
&	=G2&"-"&H2	เชื่อมข้อความสองช่องเข้าด้วยกัน	M2
LEFT	=LEFT(A2,1)	เอาตัวอักษรจากซ้ายมากี่ตัว	M2
MID	=MID(A2,3,3)	เอาตัวอักษรจากตำแหน่งที่กำหนด	M2
LEN	=LEN(A2)	นับจำนวนตัวอักษร	M2
SUBSTITUTE	=SUBSTITUTE(F2,"B","")	แทนที่ข้อความด้วยอีกข้อความ	M2
COUNTIF	=COUNTIF($A$2:A2,A2)	นับตามเงื่อนไขเดียว	M2, M3
IF	=IF(N2=1,1,0)	ถ้าเงื่อนไขจริงให้ค่าหนึ่ง ถ้าเท็จให้อีกค่าหนึ่ง	M2
AND	=IF(AND(C2>=20,C2<=400),1,0)	ใช้เมื่อต้องเข้าเงื่อนไขสองข้อพร้อมกัน	M2
COUNTA	=COUNTA(A2:A324)	นับทุกช่องที่ไม่ว่าง (รวมข้อความ)	M3
COUNT	=COUNT(C2:C324)	นับเฉพาะช่องที่เป็นตัวเลข	M3
MIN	=MIN(C2:C324)	ค่าต่ำสุด - ใช้ตรวจข้อมูลผิดปกติ	M3
MAX	=MAX(C2:C324)	ค่าสูงสุด - ใช้ตรวจข้อมูลผิดปกติ	M3
SUM	=SUM(T2:T324)	ผลรวม	M3
AVERAGE	=AVERAGE(T2:T324)	ค่าเฉลี่ย (ข้ามช่องว่างให้เอง)	M3
ROUND	=ROUND(B14,1)	ปัดทศนิยมตามจำนวนตำแหน่งที่กำหนด	M3
COUNTIFS	=COUNTIFS(สี,"Red",ทรง,"Round",ใช้,1)	นับเมื่อเข้าหลายเงื่อนไขพร้อมกัน	M4
SUMIFS	=SUMIFS(น้ำหนัก,สี,"Red",ใช้,1)	บวกเฉพาะแถวที่เข้าเงื่อนไข	M4
VLOOKUP	=VLOOKUP("Red-Round",ตาราง,3,FALSE)	ค้นค่าในตารางแล้วเอาคอลัมน์ที่ระบุ	M4
IFERROR	=IFERROR(สูตร,"ไม่พบ")	ถ้าสูตรให้ค่าผิดพลาด ให้แสดงข้อความแทน	M4

4. หลักการที่ต้องเข้าใจ

4.1 ข้อมูลที่เป็นระเบียบ (tidy data)

1 แถว	=	1 หน่วยสังเกต	(ในที่นี้คือมะเขือเทศ 1 ต้น)
1 คอลัมน์	=	1 ตัวแปร	(แปลงปลูก · น้ำหนักผล · สีผล · ทรงผล)
1 เซลล์	=	1 ค่า	

ประเภทความผิดพลาด	ตัวอย่างในไฟล์	เครื่องมือที่ใช้แก้
ช่องว่างเกินหน้า-หลัง	“ Red” “Round ”	TRIM
ตัวพิมพ์ใหญ่-เล็กไม่สม่ำเสมอ	red RED pEar	PROPER
ข้อมูลซ้ำ	บันทึกต้นเดิมสองครั้ง	COUNTIF
ค่าว่าง	ช่องสีผลหรือทรงผลว่าง	COUNTIF เทียบกับรายการที่เป็นไปได้
ค่าที่เป็นไปไม่ได้	สีผล Orange · น้ำหนัก 5200 กรัม	COUNTIF · IF · AND · MIN · MAX

หลักปฏิบัติที่ต้องยึดตลอดปฏิบัติการ ห้ามแก้ไขหรือลบข้อมูลในแท็บ 02_RawData เด็ดขาด การทำความสะอาดทั้งหมดต้องทำในแท็บ 03_Clean ด้วยสูตร ไม่ใช่ด้วยการพิมพ์ทับ เพราะสูตรคือบันทึกว่าคุณทำอะไรกับข้อมูล ซึ่งเป็นหลักฐานที่ทำซ้ำได้

4.2 เคล็ดลับสามข้อที่ใช้ตลอดปฏิบัติการ

เคล็ดลับ	อธิบาย
ใช้การคูณแทนคำว่า และ	เมื่อมีคอลัมน์ที่เป็น 0 กับ 1 หลายคอลัมน์ การคูณกันจะได้ 1 ก็ต่อเมื่อผ่านครบทุกข้อ เช่น =O2*P2*Q2 อ่านว่า ไม่ซ้ำ และ สีใช้ได้ และ ทรงใช้ได้
ใช้ MIN กับ MAX ตรวจข้อมูลก่อนเสมอ	ก่อนคำนวณค่าเฉลี่ยใด ๆ ให้ดู MIN และ MAX ก่อน ถ้าเจอค่าที่เป็นไปไม่ได้ แปลว่ามีข้อมูลผิดปะปนอยู่ ค่าเฉลี่ยที่ได้จะเชื่อไม่ได้
ครอบ IFERROR เมื่อสูตรอาจค้นไม่เจอ	VLOOKUP ที่ค้นไม่เจอจะให้ #N/A ซึ่งจะทำให้ทุกสูตรที่อ้างถึงต่อพังตามไปด้วย การครอบด้วย IFERROR ทำให้ได้ข้อความที่อ่านรู้เรื่องแทน

4.3 พันธุศาสตร์ที่ใช้ในปฏิบัติการนี้

ยีนสองตำแหน่งที่ติดตาม ทั้งคู่เป็นยีนจริงของมะเขือเทศ อยู่คนละโครโมโซม จึงถ่ายทอดอย่างอิสระต่อกัน

ลักษณะ	อัลลีลเด่น	อัลลีลด้อย	ยีน
สีของเนื้อผล	R = ผลสีแดง	r = ผลสีเหลือง	PSY1 (Fray & Grierson, 1993)

ทรงของผล	O = ผลทรงกลม	o = ผลทรงลูกแพร์	OVATE (Liu et al., 2002)

ฟีโนไทป์ในรุ่น F2	**จีโนไทป์**	**สัดส่วนคาดหมาย**	**อัตราส่วน**
ผลแดง ทรงกลม	R_O_	9/16 = 0.5625	9
ผลแดง ทรงลูกแพร์	R_oo	3/16 = 0.1875	3
ผลเหลือง ทรงกลม	rrO_	3/16 = 0.1875	3
ผลเหลือง ทรงลูกแพร์	rroo	1/16 = 0.0625	1

และถ้ายีนทั้งสองแยกตัวอย่างอิสระจริง เมื่อยุบลักษณะหนึ่งทิ้งแล้วดูอีกลักษณะหนึ่งเพียงลำพัง จะต้องได้ 3 : 1 ทั้งสองยีน

5. วิธีปฏิบัติ

กติกาการให้คะแนนทุกคำตอบที่เป็นตัวเลขต้องเกิดจากสูตร ผู้สอนจะคลิกที่เซลล์แล้วดูที่แถบสูตร ถ้าพบตัวเลขที่พิมพ์ทับ ข้อนั้นจะไม่ได้คะแนน แม้ค่าจะถูกต้อง ทำงานสะสมไปทีละกิจกรรม แล้วดาวน์โหลดเป็นไฟล์ .xlsx ส่งเข้า Google Form ก่อนออกจากห้อง

กิจกรรม M1 - พื้นฐานสเปรดชีตและสูตรแรก (30 นาที, 3 คะแนน)

เปิดแท็บ 01_ฝึกพื้นฐาน ทำแบบฝึกให้ครบทั้ง 6 ข้อ แล้วยกมือให้ผู้สอนหรือ TA ตรวจก่อนไปต่อ

M1.1 เซลล์ทุกช่องมีชื่อ = ตัวอักษรของคอลัมน์ + เลขของแถว เช่น C8 ชื่อนี้แสดงที่มุมซ้ายบนเสมอเมื่อคลิกที่ช่องนั้น

M1.2 พิมพ์เครื่องหมาย = นำหน้า คือการบอกว่าต่อจากนี้เป็นสูตร ไม่ใช่ข้อความ

M1.3 อย่าพิมพ์ชื่อช่องเอง ให้พิมพ์ = แล้ว คลิก ที่ช่องที่ต้องการ ข้อนี้ลดความผิดพลาดได้มากที่สุด

M1.4 เครื่องหมาย: อ่านว่า ถึง เช่น C15:C19 คือตั้งแต่ C15 ลงมาถึง C19

M1.5 ที่จับเติมสูตรคือจุดสี่เหลี่ยมเล็กที่มุมขวาล่างของช่อง ดับเบิลคลิกเพื่อเติมลงทั้งคอลัมน์ หรือกด Ctrl+Shift+ลูกศรลง แล้ว Ctrl+D

M1.6 การตรึงด้วย $ ทำได้เร็วที่สุดโดยคลิกในสูตรตรงชื่อช่องแล้วกด F4

M1.7 แถบสูตรคือช่องยาวด้านบนที่แสดงสิ่งที่อยู่ในเซลล์จริง ๆ ผู้สอนตรวจงานจากตรงนี้

กิจกรรม M2 - ฟังก์ชันข้อความและการทำความสะอาดข้อมูล (35 นาที, 5 คะแนน)

ไปที่แท็บ 03_Clean คอลัมน์ A ถึง E ดึงข้อมูลดิบมาให้แล้ว แถวที่ 2 ของคอลัมน์ F ถึง T ทำสูตรตัวอย่างไว้ให้ครบ และมีรายการสูตรทั้งหมดพร้อมคำอธิบายอยู่ทางขวาของตาราง

M2.1 ตรึงแถวหัวตารางด้วย View ▸ Freeze ▸ 1 row แล้วเรียงตามคอลัมน์ FruitColor ด้วย Data ▸ Sort range จะเห็นทันทีว่ารูปแบบการเขียนไม่สม่ำเสมอ

M2.2 ทำความสะอาดคอลัมน์แปลงปลูก สีผล และทรงผล ด้วยสูตรเดียวกันทั้งสามคอลัมน์

=PROPER(TRIM(B2))

=PROPER(TRIM(D2))

=PROPER(TRIM(E2))

M2.3 สร้างคอลัมน์ชั้นฟีโนไทป์ โดยเชื่อมสีผลกับทรงผลด้วยเครื่องหมาย &

=G2&"-"&H2

M2.4 แยกส่วนประกอบของรหัสต้นด้วยฟังก์ชันข้อความ - LEFT เอาจากซ้าย MID เอาจากกลาง LEN นับความยาว และ SUBSTITUTE แทนที่ข้อความ

=LEFT(A2,1)

=MID(A2,3,3)

=LEN(A2)

=SUBSTITUTE(F2,"B","")

M2.5 หาแถวซ้ำ โดยนับว่ารหัสนี้พบเป็นครั้งที่เท่าไร สังเกตว่า $ อยู่ที่ตัวแรกเท่านั้น แล้วแปลงเป็นเครื่องหมาย 0 กับ 1

=COUNTIF($A$2:A2,A2) แล้ว =IF(N2=1,1,0)

M2.6 ตรวจว่าสีผลและทรงผลเป็นค่าที่เป็นไปได้หรือไม่ โดยนับเทียบกับรายการในแท็บ Lookup_Valid

=COUNTIF(Lookup_Valid!$A$2:$A$3,G2)

=COUNTIF(Lookup_Valid!$B$2:$B$3,H2)

M2.7 ตรวจว่าน้ำหนักผลอยู่ในช่วงที่สมเหตุสมผล (20 ถึง 400 กรัม) ด้วย IF ร่วมกับ AND

=IF(AND(C2>=20,C2<=400),1,0)

M2.8 สรุปว่าแถวใดใช้วิเคราะห์ได้จริง ด้วยการคูณเครื่องหมายทั้งสามเข้าด้วยกัน

=O2*P2*Q2

M2.9 สร้างคอลัมน์น้ำหนักเฉพาะต้นที่ใช้ได้ ที่เหลือปล่อยว่างไว้ คอลัมน์นี้จะใช้ต่อในกิจกรรม M3

=IF(R2*S2=1,C2,"")

M2.10 เลือกเซลล์ F2 ถึง T2 แล้วดับเบิลคลิกที่มุมขวาล่างเพื่อเติมสูตรลงทั้งคอลัมน์

จุดที่มักพลาดในกิจกรรมนี้ อ่านสูตรที่ซ้อนกันจากวงเล็บในสุดออกมาเสมอ - PROPER(TRIM(D2)) คือ ตัดช่องว่างก่อน แล้วค่อยจัดตัวพิมพ์ ใน COUNTIF($A$2:A2,A2) เครื่องหมาย $ ต้องอยู่ที่ตัวแรกเท่านั้น ถ้าตรึงทั้งสองด้านจะนับผิดทั้งคอลัมน์ข้อมูลมีหลายร้อยแถว อย่าลากด้วยเมาส์ทีละนิด ให้ดับเบิลคลิกที่มุมขวาล่าง

กิจกรรม M3 - สรุปข้อมูลด้วยสถิติพื้นฐาน (35 นาที, 5 คะแนน)

ไปที่แท็บ 04_Summary ตอนที่ 1 ให้ทำกับคอลัมน์น้ำหนักดิบทั้งหมด ส่วนตอนที่ 2 ให้ทำกับคอลัมน์ T ที่เก็บเฉพาะน้ำหนักของต้นที่ใช้ได้

M3.1 นับจำนวนแถวทั้งหมดด้วย COUNTA และนับเฉพาะช่องที่เป็นตัวเลขด้วย COUNT แล้วสังเกตว่าทำไมสองค่านี้ไม่เท่ากัน

=COUNTA('03_Clean'!$A$2:$A$324)

=COUNT('03_Clean'!$C$2:$C$324)

M3.2 หาน้ำหนักต่ำสุดและสูงสุดในไฟล์ดิบ แล้วดูว่าค่าที่ได้เป็นไปได้จริงหรือไม่ นี่คือวิธีตรวจข้อมูลที่เร็วที่สุด

=MIN('03_Clean'!$C$2:$C$324)

=MAX('03_Clean'!$C$2:$C$324)

M3.3 สรุปน้ำหนักเฉพาะต้นที่ใช้ได้ ด้วย COUNT SUM AVERAGE MIN MAX บนคอลัมน์ T

=COUNT('03_Clean'!$T$2:$T$324)

=SUM(...)

=AVERAGE(...)

M3.4 ปัดค่าเฉลี่ยเป็นทศนิยม 1 ตำแหน่ง

=ROUND(B14,1)

M3.5 ตรวจสอบว่า ผลรวม ÷ จำนวน เท่ากับค่าเฉลี่ยที่ AVERAGE ให้มาหรือไม่

=B13/B12

M3.6 กรอกตารางสรุปผลการทำความสะอาดในตอนที่ 3 ด้วย COUNTIF และ SUM

ทำไมไฟล์นี้มีจำนวนต้นสองตัวเลข ตอนที่ 2 ของแท็บ 04_Summary ให้ จำนวนต้นที่นำน้ำหนักมาคิดได้ ส่วนตอนที่ 3 ให้ จำนวนต้นที่นับฟีโนไทป์ได้ (N) สองตัวนี้ไม่เท่ากัน และไม่ใช่ความผิดพลาด ต้นที่บันทึกน้ำหนักผิด เช่น ติดลบ หรือหนักหลายกิโลกรัม ยังมีสีผลและทรงผลที่อ่านได้ตามปกติ จึงยังนับเข้าอัตราส่วนฟีโนไทป์ได้ แต่ต้องตัดออกจากการคำนวณน้ำหนัก ไม่อย่างนั้นค่าเฉลี่ยจะเพี้ยน หลักคิดคือ ตัดข้อมูลออกเฉพาะการคำนวณที่ข้อมูลนั้นเสียหายจริง อย่าตัดทั้งแถวโดยอัตโนมัติ เพราะจะเสียข้อมูลที่ยังใช้ได้ไปด้วย สองแถวสุดท้ายของตอนที่ 3 ให้คุณกระทบยอดตัวเลขสองตัวนี้เอง

กิจกรรม M4 - นับตามเงื่อนไข ตรรกะ และการค้นหา (35 นาที, 5 คะแนน)

ไปที่แท็บ 05_Ratio ทำตามลำดับตอนที่ 1 ถึง 6 ช่องแรกของทุกตอนมีสูตรเต็มอยู่ในคำใบ้

M4.1 นับจำนวนต้นในแต่ละฟีโนไทป์ด้วย COUNTIFS โดยใส่เงื่อนไขเป็นคู่ ช่วง แล้วเงื่อนไข และอย่าลืมคู่สุดท้ายที่กรองเฉพาะแถวที่ใช้วิเคราะห์ได้

=COUNTIFS(สี,"Red",ทรง,"Round",ใช้วิเคราะห์,1)

M4.2 ตรวจสอบว่าผลรวมทั้งสี่กลุ่มเท่ากับ N พอดี ถ้าไม่เท่าให้ย้อนกลับไปตรวจ M2

M4.3 หาอัตราส่วนที่สังเกตได้ โดยหารทุกกลุ่มด้วยกลุ่มที่น้อยที่สุด แล้วดูว่าใกล้เคียงเลขใด

=B6/$B$9

M4.4 ใช้ VLOOKUP ดึงสัดส่วนคาดหมายจากแท็บ Lookup_Ratio อาร์กิวเมนต์ทั้งสี่คือ ค่าที่ค้น ตารางที่ค้น เอาคอลัมน์ที่เท่าไร และ FALSE ที่แปลว่าต้องตรงทุกตัวอักษร

=VLOOKUP("Red-Round",Lookup_Ratio!$A$2:$D$5,3,FALSE)

M4.5 คำนวณจำนวนคาดหมาย แล้วตรวจว่าผลรวมเท่ากับ N

=B20*$B$5

M4.6 ตรวจการแยกตัวทีละลักษณะ ทั้งสองยีนควรได้อัตราส่วนใกล้เคียง 3 : 1

=COUNTIFS(สี,"Red",ใช้วิเคราะห์,1)

M4.7 ใช้ SUMIFS หาน้ำหนักรวมของต้นผลสีแดง สังเกตว่า SUMIFS เอาช่วงที่จะบวกมาก่อน แล้วจึงตามด้วยคู่เงื่อนไข ต่างจาก COUNTIFS ที่ไม่มีช่วงนำหน้า

=SUMIFS(น้ำหนัก,สี,"Red",น้ำหนักปกติ,1,ใช้วิเคราะห์,1)

M4.8 ลองพิมพ์ VLOOKUP ที่ค้นคำที่ไม่มีในตาราง เช่น Red-Oval เพื่อดูว่าขึ้น #N/A จริง ๆ แล้วจึงครอบด้วย IFERROR

=IFERROR(VLOOKUP(...),"ไม่พบในตาราง")

กิจกรรม M5 - การสร้างกราฟ (20 นาที, 2 คะแนน)

M5.1 ไปที่แท็บ 06_Chart ดึงตัวเลข Observed และ Expected จากแท็บ 05_Ratio ด้วยสูตร ='05_Ratio'!B6

M5.2 เลือกช่วง A4 ถึง C8 แล้วใช้เมนู Insert ▸ Chart

M5.3 เปลี่ยนชนิดกราฟเป็น Column chart แบบ clustered ถ้าโปรแกรมเลือกชนิดอื่นมาให้

M5.4 ใส่ชื่อกราฟ ชื่อแกน x และชื่อแกน y พร้อมหน่วย

M5.5 ตรวจงานตัวเองด้วยเช็กลิสต์ 7 ข้อในแท็บเดียวกัน แล้วเขียนคำบรรยายภาพให้ครบ 3 ส่วน คือ กราฟแสดงอะไร ข้อมูลมาจากไหน และค่า N เท่าใด

6. แบบบันทึกผลการทดลอง

ตารางที่ 1 ผลการทำความสะอาดข้อมูล (M2-M3)

รายการ	จำนวน (แถว)
จำนวนแถวทั้งหมดในไฟล์	
แถวซ้ำ	
แถวที่สีผลเป็นค่าที่เป็นไปไม่ได้	
แถวที่ทรงผลเป็นค่าที่เป็นไปไม่ได้	
แถวที่น้ำหนักผลเป็นไปไม่ได้	
จำนวนต้นที่ใช้วิเคราะห์ได้จริง (N)	

ตารางที่ 2 สถิติของน้ำหนักผล (M3)

รายการ	ไฟล์ดิบทั้งหมด	เฉพาะต้นที่ใช้ได้
จำนวนค่าที่เป็นตัวเลข		
ค่าต่ำสุด (กรัม)		
ค่าสูงสุด (กรัม)		
ผลรวม (กรัม)	-	
ค่าเฉลี่ย (กรัม)	-	
ค่าเฉลี่ยปัด 1 ตำแหน่ง	-	

ตารางที่ 3 อัตราส่วนฟีโนไทป์ของรุ่น F2 (M4)

ฟีโนไทป์	จีโนไทป์	Observed	อัตราส่วน	Expected
ผลแดง ทรงกลม				
ผลแดง ทรงลูกแพร์				
ผลเหลือง ทรงกลม				
ผลเหลือง ทรงลูกแพร์				
รวม	-		-	

ตารางที่ 4 การตรวจการแยกตัวทีละลักษณะ (M4)

ลักษณะ	จำนวนเด่น	จำนวนด้อย	อัตราส่วนที่ได้	ที่คาดหมาย
สีผล (แดง : เหลือง)				3 : 1
ทรงผล (กลม : ลูกแพร์)				3 : 1

7. คำถามท้ายปฏิบัติการ

ตอบลงในแท็บ 07_คำถาม ของ Workbook

1. (M2) จงระบุความผิดพลาดของข้อมูลที่พบในแท็บ 02_RawData มาอย่างน้อย 3 ประเภท พร้อมยกตัวอย่าง PlantID และระบุว่าแก้ไขด้วยฟังก์ชันใด
2. (M2) เหตุใดจึงไม่ควรลบแถวที่มีปัญหาออกจากแท็บ 02_RawData โดยตรง แต่ให้ทำเครื่องหมายไว้ในแท็บ 03_Clean แทน
3. (M2) คอลัมน์ ใช้วิเคราะห์ ใช้การคูณกันของเลข 0 กับ 1 แทนคำว่า และ จงอธิบายว่าทำไมวิธีนี้จึงใช้ได้ และยกตัวอย่างแถวหนึ่งที่ได้ค่า 0
4. (M3) MIN และ MAX ช่วยให้คุณพบข้อมูลผิดปกติได้อย่างไร และถ้าไม่ได้ตรวจสองค่านี้ก่อน ค่าเฉลี่ยที่คำนวณได้จะผิดไปมากน้อยเพียงใด
5. (M3) เหตุใด COUNT กับ COUNTA จึงให้ผลไม่เท่ากันเมื่อใช้กับข้อมูลชุดเดียวกัน
6. (M4) COUNTIF กับ COUNTIFS ต่างกันอย่างไร และเหตุใดจึงต้องใช้ COUNTIFS ในการนับฟีโนไทป์
7. (M4) เมื่อ VLOOKUP ค้นไม่เจอจะขึ้น #N/A จงอธิบายว่า IFERROR ช่วยอย่างไร และเหตุใดการปล่อยให้ขึ้น #N/A จึงเป็นปัญหาเมื่อนำผลไปคำนวณต่อ
8. (M4) อัตราส่วนที่คุณคำนวณได้ใกล้เคียง 9 : 3 : 3 : 1 หรือไม่ และผลการตรวจทีละลักษณะบอกอะไรเกี่ยวกับการถ่ายทอดของยีนสองตำแหน่งนี้
9. (M5) เหตุใดจึงไม่ควรใช้กราฟวงกลมนำเสนอการเปรียบเทียบ Observed กับ Expected ของข้อมูลชุดนี้
10. จำนวนที่สังเกตได้ต่างจากจำนวนที่คาดหมายอยู่บ้าง ต่างกันมากพอที่จะสรุปว่าแบบจำลองผิดหรือไม่ คุณคิดว่าต้องใช้เครื่องมืออะไรเพิ่มจึงจะตอบคำถามนี้ได้

8. เกณฑ์การให้คะแนนโดยสรุป

หัวข้อ	คะแนน	สิ่งที่ผู้สอนตรวจ
M1 ฝึกพื้นฐาน	3	ทำแบบฝึกครบ 6 ข้อและใช้สูตรจริง
M2 ฟังก์ชันข้อความและการทำความสะอาด	5	สูตรคอลัมน์ F ถึง T ถูกต้องและลากครบทุกแถว
M3 สรุปข้อมูลด้วยสถิติพื้นฐาน	5	ใช้ครบ 7 ฟังก์ชัน และตีความ MIN/MAX ได้
M4 นับตามเงื่อนไขและค้นหา	5	COUNTIFS SUMIFS VLOOKUP IFERROR ถูกต้อง ผลรวมเท่ากับ N
M5 กราฟและคำบรรยายภาพ	2	องค์ประกอบกราฟครบ และคำบรรยายภาพครบ 3 ส่วน
รวม	20	

9. สิ่งที่ปฏิบัติการนี้รับผิดชอบ และสิ่งที่ยังตอบไม่ได้

ปฏิบัติการนี้หยุดที่การทำความสะอาดข้อมูล การนับ การสรุปด้วยสถิติพื้นฐาน และการสร้างกราฟ ซึ่งเป็นทักษะที่นำไปใช้ได้กับข้อมูลทุกชนิด ไม่จำกัดแค่พันธุศาสตร์

ทำได้แล้วเมื่อจบคาบนี้	ยังตอบไม่ได้ด้วยเครื่องมือในคาบนี้
ทำความสะอาดข้อมูลดิบและอธิบายได้ว่าตัดอะไรออกไปเพราะอะไร	ทำไมทรงลูกแพร์จึงเป็นลักษณะด้อย ซึ่งต้องดูถึงระดับลำดับดีเอ็นเอ
นับฟีโนไทป์และเทียบกับค่าคาดหมายตามกฎเมนเดลได้	ค่าที่ต่างจากค่าคาดหมายนั้นมากพอที่จะเรียกว่าผิดปกติหรือเป็นเพียงความบังเอิญ ซึ่งต้องใช้การทดสอบทางสถิติ

การรู้ว่าเครื่องมือที่ใช้อยู่ตอบอะไรได้และตอบอะไรไม่ได้ เป็นส่วนหนึ่งของการวิเคราะห์ข้อมูลที่ดีพอ ๆ กับการคำนวณให้ถูก เก็บไฟล์ Workbook ที่ทำวันนี้ไว้ให้ดี

10. เอกสารอ้างอิง

Broman, K. W., & Woo, K. H. (2018). Data organization in spreadsheets. The American Statistician, 72(1), 2-10. https://doi.org/10.1080/00031305.2017.1375989

Bruford, E. A., Braschi, B., Denny, P., Jones, T. E. M., Seal, R. L., & Tweedie, S. (2020). Guidelines for human gene nomenclature. Nature Genetics, 52(8), 754-758. https://doi.org/10.1038/s41588-020-0669-3

Fray, R. G., & Grierson, D. (1993). Identification and genetic analysis of normal and mutant phytoene synthase genes of tomato by sequencing, complementation and co-suppression. Plant Molecular Biology, 22(4), 589-602. https://doi.org/10.1007/BF00047400

Liu, J., Van Eck, J., Cong, B., & Tanksley, S. D. (2002). A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. Proceedings of the National Academy of Sciences, 99(20), 13302-13306. https://doi.org/10.1073/pnas.162485999

Wickham, H. (2014). Tidy data. Journal of Statistical Software, 59(10), 1-23. https://doi.org/10.18637/jss.v059.i10

Ziemann, M., Eren, Y., & El-Osta, A. (2016). Gene name errors are widespread in the scientific literature. Genome Biology, 17, 177. https://doi.org/10.1186/s13059-016-1044-7