ปฏิบัติการที่ 10 การวิเคราะห์สหสัมพันธ์และการวิเคราะห์ถดถอยเชิงเส้น

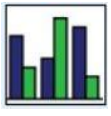
ในปฏิบัติการนี้ จะฝึกหัดการสร้างกราฟเบื้องต้นซึ่งอาจจะนำไปใช้ประกอบการนำเสนอ การวิเคราะห์สหสัมพันธ์และการวิเคราะห์หาสมการเชิงเส้นและการทำนายค่าการประมาณเส้นโค้ง (curve estimation) ด้วยโปรแกรม SPSS 22

1. การสร้างกราฟ

การสร้างกราฟในโปรแกรม SPSS พบได้ในคำสั่งหลายคำสั่ง อย่างไรก็ตาม ควรทราบวิธีการสร้างกราฟในโปรแกรม SPSS 22 ที่จะเป็นประโยชน์และทักษะที่สำคัญโดยไม่จำเป็นต้องใช้โปรแกรมอื่น ตัวอย่างข้อมูลในการสร้างกราฟในบทนี้จะใช้ข้อมูลที่เคยใช้ให้ปฏิบัติการบทก่อน ๆ โดยจะกล่าวถึงการสร้างกราฟแท่ง กราฟ X-Y หรือ scatter-plot เท่านั้น ทั้งนี้ กราฟแบบอื่น ๆ จะใช้แนวคิดหรือวิธีการสร้างกราฟคล้ายคลึงกัน

1.1 กราฟแท่ง

ข้อมูลตัวอย่างมาจากการบ้านปฏิบัติการที่ 9 ข้อ 1 สำหรับการสร้างกราฟแท่งที่แสดงค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของ MPI activity ในจีโนไทป์ต่างกัน โดยแยกกลุ่มกราฟตามเพศของหอย ขั้นตอนการสร้างกราฟ เป็นดังนี้

1. ในการสร้างตัวแปร ควรกำหนด measurement level ทุกตัวแปร และหากเป็นตัวแปรระบุกลุ่ม ควรกำหนด value label ด้วย อย่างไรก็ตาม หากสร้างตัวแปรแล้ว แต่ยังไม่ได้กำหนด measurement level ของตัวแปร (ดังตัวอย่างในภาพที่ 10.1 ก) ที่คอลัมน์ Measure ของตัวแปรมีค่าเป็น Unknown ก่อนที่จะสร้างกราฟจะต้องแก้ไขที่คอลัมน์ Measure ของตัวแปรแต่ละตัวแปรให้เป็น Nominal หรือ Ordinal หรือ Scale สำหรับตัวอย่างนี้ กำหนดให้ตัวแปร genotype และ sex มี Measure เป็น Nominal และ mpi เป็น Scale (ภาพที่ 10.1 ข) อนึ่ง ตัวแปร genotype และ sex เป็นตัวแปรระบุกลุ่มและได้กำหนด value label แล้ว (สังเกตที่คอลัมน์ Values)
2. เลือก Graphs > Chart Builder... (ภาพที่ 10.1 ค)
    i. หากตัวแปรไม่มีการกำหนด measurement level จะมีไดอะล็อคแจ้งเตือนให้กำหนด Measurement level และ value label ของตัวแปรก่อน (ภาพที่ 10.1 ค) หากต้องการกำหนดเอง ให้กด Define Variable Properties... แล้วกำหนด Measurement Level: ให้กับตัวแปรแต่ละตัว
    ii. แต่หากกด OK โดยที่ยังไม่กำหนด measurement level จะมีไดอะล็อคแจ้งเตือนอีกครั้ง (ภาพที่ 10.1 จ) ซึ่งยังสามารถเลือกกำหนดเอง โดยกด Assign Manually... หรือหากให้ SPSS กำหนด ให้กด Scan Data
3. กำหนดตัวเลือกในไดอะล็อค Chart Builder ดังนี้
    i. ในแถบ Gallery ช่อง Choose from: ให้เลือก Bar แล้วลาก  ไปใส่ในช่อง Chart preview uses example data (ภาพที่ 10.1 ฉ)
    ii. กำหนดตัวแปรในกราฟ (ภาพที่ 10.1 ช) โดยเลือกตัวแปรที่ช่อง Variables:
        a. เลือกตัวแปร Sex of Platorchestia... แล้วลากไปใส่ X-Axis?
        b. เลือกตัวแปร MPI activity แล้วลากไปใส่ Y-Axis?

c. เลือกตัวแปร Genotype of manos... แล้วลากไปใส่ Cluster on X: set color

iii. กำหนดค่าสถิติในกราฟโดยกด Element Properties... แล้วตั้งค่าในไดอะล็อค Element Properties (ภาพที่ 10.1 ซ) ดังนี้

a. ที่ช่อง Edit Properties of: เลือก Bar1

b. ที่ช่อง Statistics กำหนด Statistic: เป็น Mean

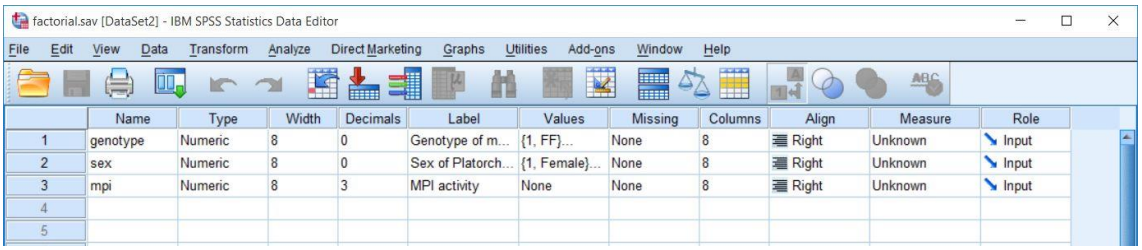
c. เลือก Display error bars

d. ที่ช่อง Error Bars Represent เลือก Standard deviation และกำหนด Multiplier: เป็น 1

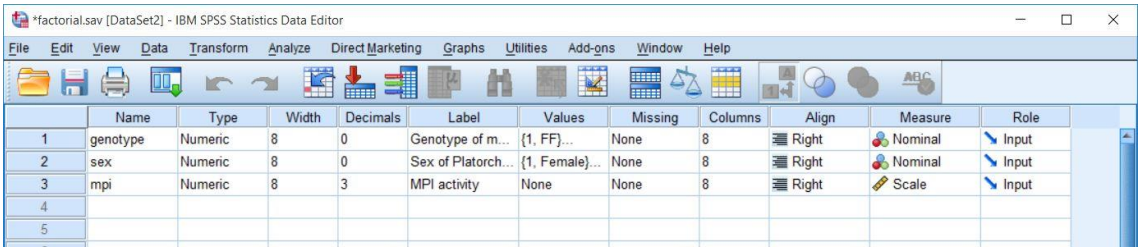
e. กด Apply

4. กด OK ในไดอะล็อค Chart Builder จะได้กราฟดังภาพที่ 10.1 ฎ

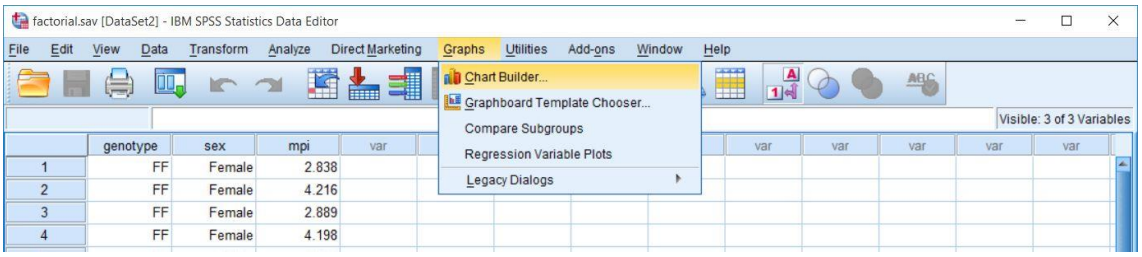
(ก)



(ข)

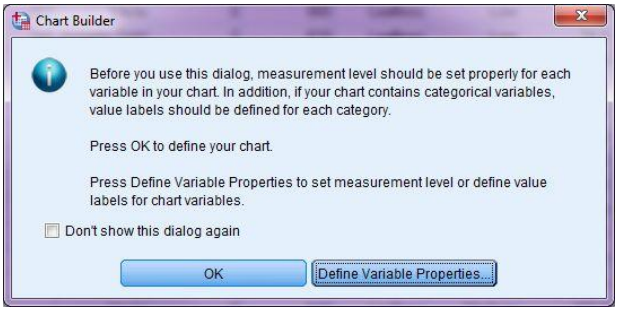


(ค)

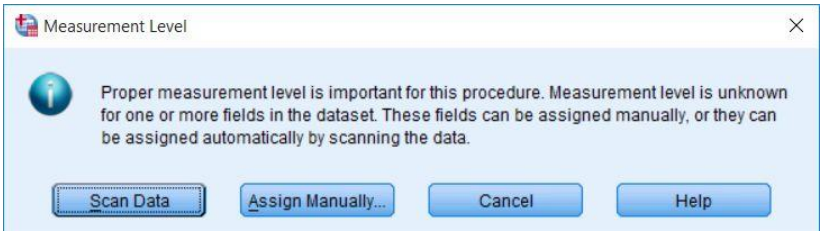


ภาพที่ 10.1 ตัวแปรที่ยังไม่ได้กำหนด measurement level (ก) เมื่อกำหนด measurement level แล้ว (ข) คำสั่ง Chart Builder ในเมนู Graphs (ค)

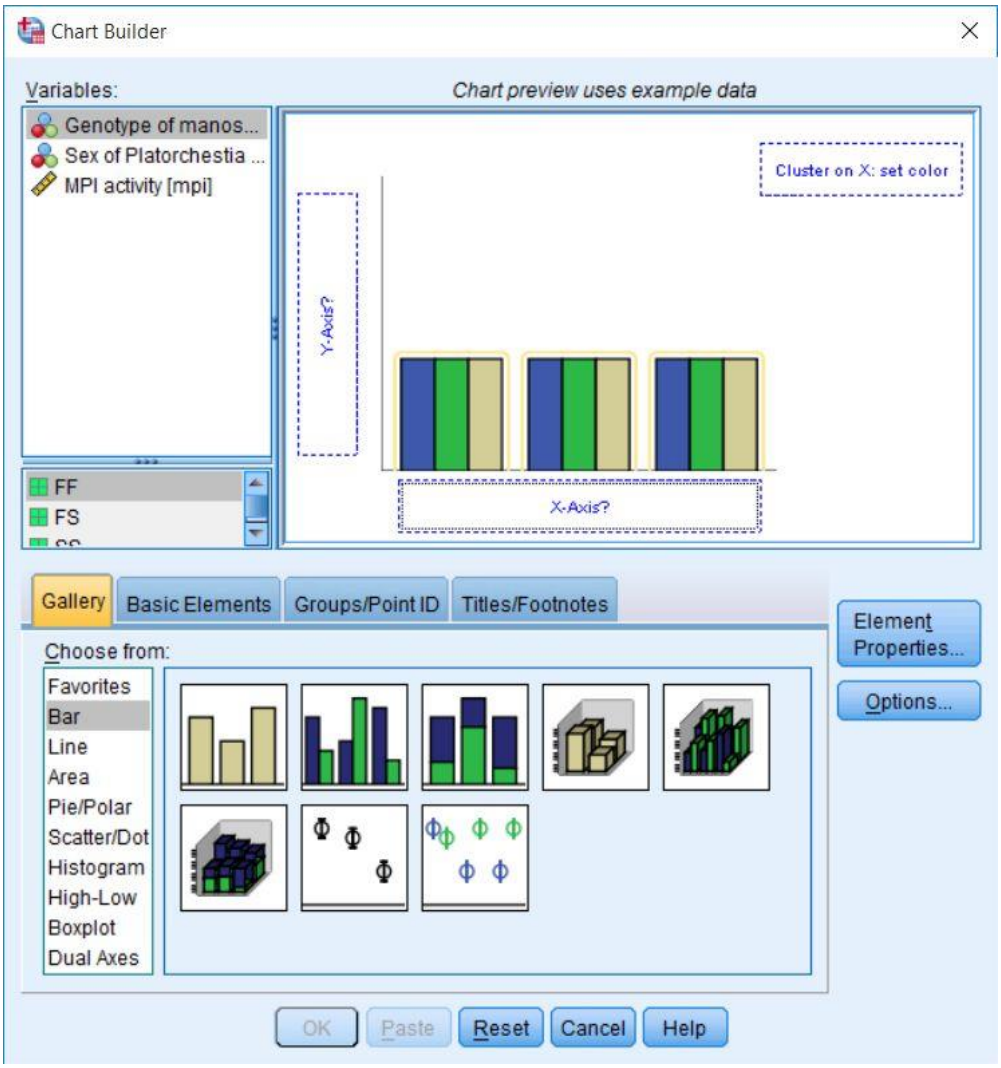
(ง)



(จ)

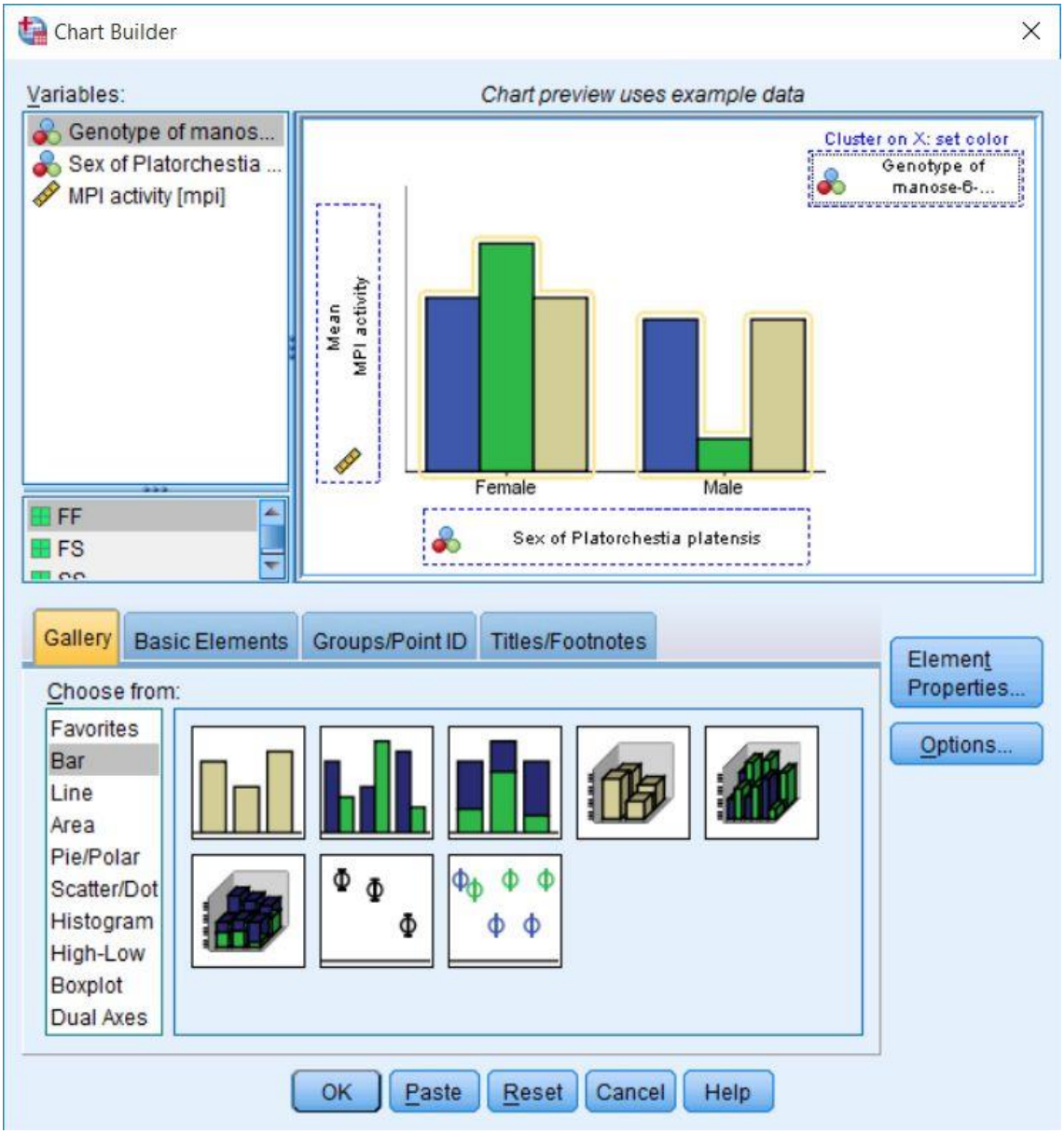


(ฉ)



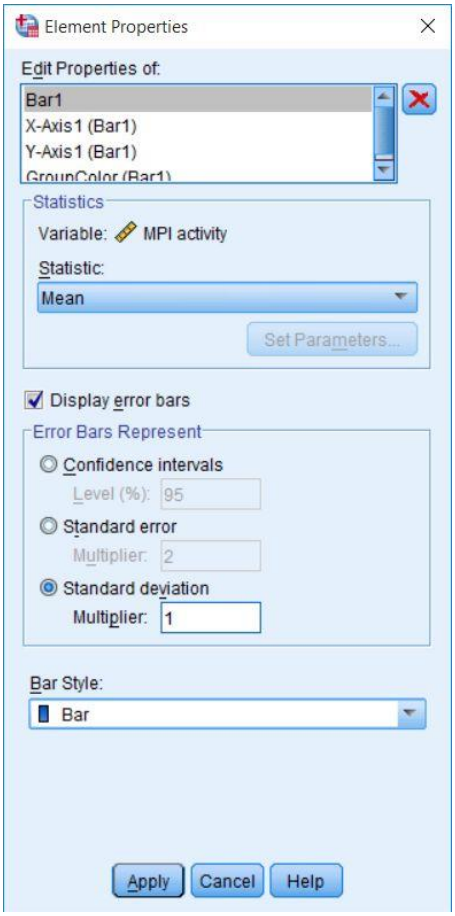
ภาพที่ 10.1 (ต่อ) การเตือนให้กำหนด measurement level และ value label ของตัวแปร (ง-จ) ไดอะล็อคกำหนดค่าในการสร้างกราฟแบบแท่ง (ฉ)

(ช)

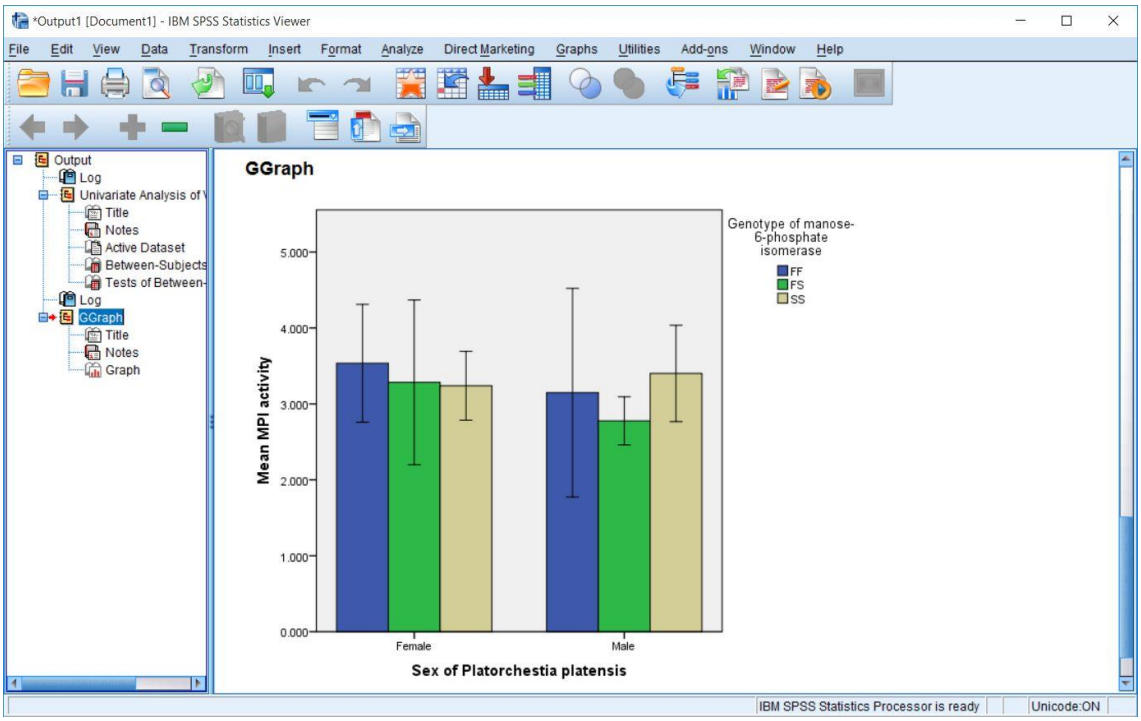


ภาพที่ 10.1 (ต่อ) กำหนดตัวแปรให้กับกราฟและกลุ่มของกราฟ (ช)

(ช)



(ฎ)



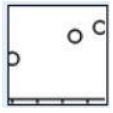
ภาพที่ 10.1 (ต่อ) กำหนดค่าสถิติของกราฟแท่ง (ช) กราฟที่ได้ (ฎ)

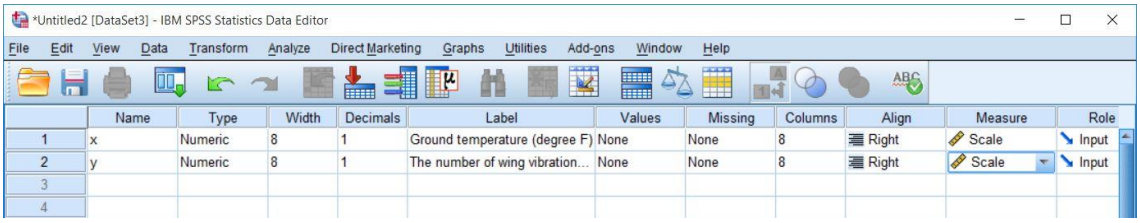
1.2 กราฟ scatter plot

ตัวอย่างข้อมูลกับอุณหภูมิที่ระดับพื้นดิน (ตัวแปร X) กับจำนวนการสั่นของปีกต่อวินาทีของแมลง striped ground cricket (ตัวแปร Y) ผู้วิจัยเชื่อว่าอุณหภูมิมีผลต่อพฤติกรรมของแมลงซึ่งเป็นสัตว์เลือดเย็น โดยจะไม่ "ร้องเพลง" (หรือขยับปีก) หากอุณหภูมิที่พื้นดินต่ำกว่า 60° F หรือมากกว่า 100° F ผลการบันทึกข้อมูลเป็นดังนี้

X	88.6	71.6	93.3	84.3	80.6	75.2	69.7	71.6	69.4	83.3	79.6	82.6	80.6	83.5	76.3
Y	20.0	16.0	19.8	18.4	17.1	15.5	14.7	15.7	15.4	16.3	15.0	17.2	16.0	17.0	14.4

การเตรียมข้อมูล จะใช้ตัวแปร 2 ตัวแปร โดยตัวแปรแรกชื่อ X มีคำอธิบาย Ground temperature (degree F) และตัวแปรตัวที่สองชื่อ Y มีคำอธิบาย The number of wing vibrations per second และกำหนด measurement level ของตัวแปรทั้งเป็น Scale (ภาพที่ 10.2 ก) และขั้นตอนการสร้างกราฟ เป็นดังนี้

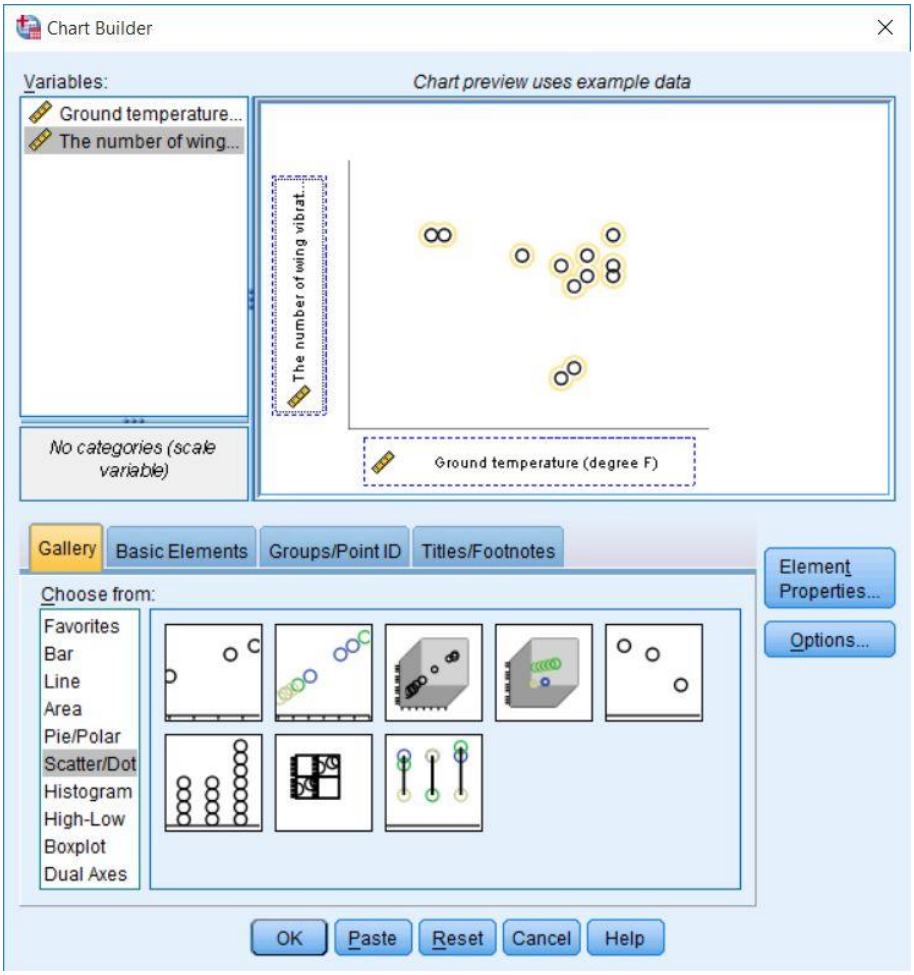
1. เลือก Graphs > Chart Builder…
2. กำหนดตัวเลือกในไดอะล็อค Chart Builder ดังนี้
   i. ในแถบ Gallery ช่อง Choose from: ให้เลือก Scatter/Dot แล้วลาก  ไปใส่ในช่อง Chart preview uses example data (ภาพที่ 10.2 ข)
   ii. กำหนดตัวแปรในกราฟ (ภาพที่ 10.2 ค) โดยเลือกตัวแปรที่ช่อง Variables:
      a. เลือกตัวแปร Ground temperature… แล้วลากไปใส่ X-Axis?
      b. เลือกตัวแปร The number of wing… แล้วลากไปใส่ Y-Axis?
3. กด OK จะได้กราฟดังภาพที่ 10.2 ค



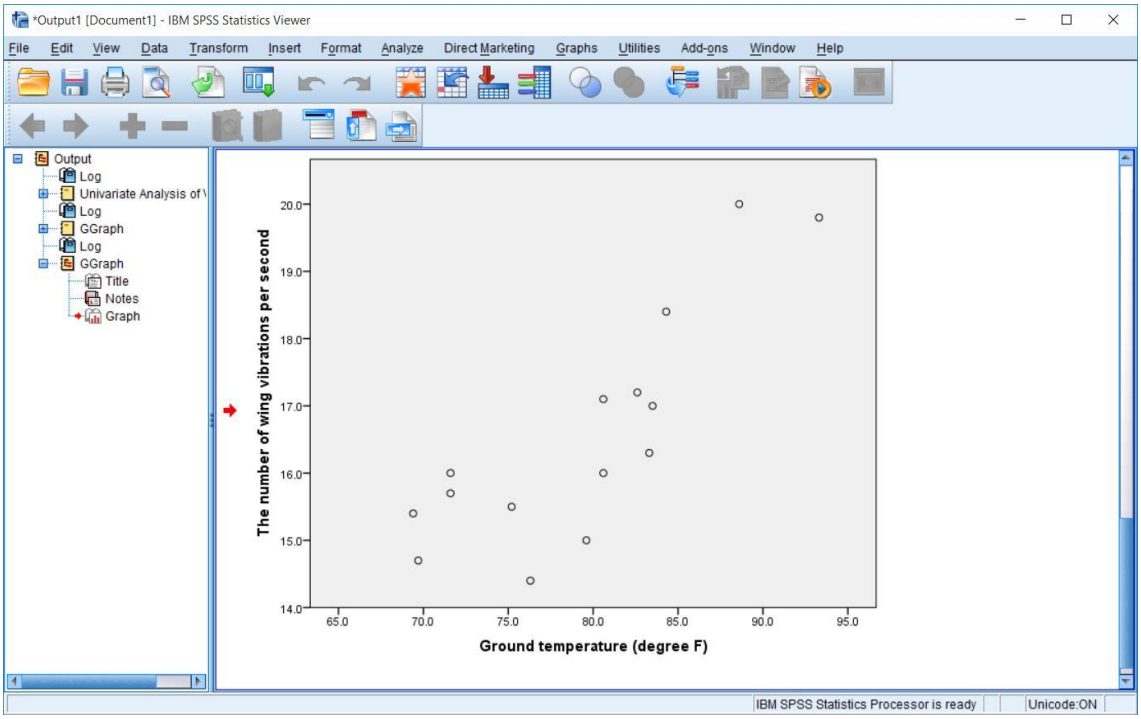
(ก)

ภาพที่ 10.2 ตัวแปรสำหรับการสร้างกราฟ scatter plot

(ข)



(ค)

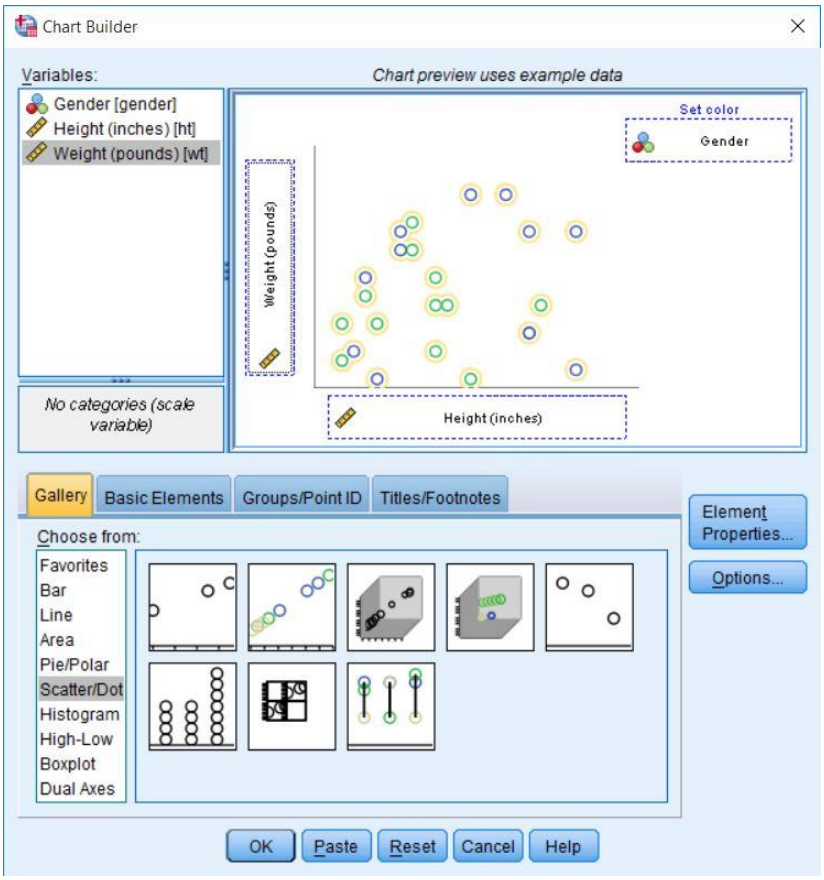


ภาพที่ 10.2 (ต่อ) กำหนดตัวแปรของกราฟ (ข) กราฟที่ได้ (ค)

1.3 กราฟ scatter plot แบบแยกกลุ่ม

ตัวอย่างข้อมูล ให้ใช้ไฟล์ height.sav ซึ่งมีตัวแปร 3 ตัวแปร คือ gender height และ weight การสร้างกราฟ scatter plot ระหว่างตัวแปร height กับตัวแปร weight โดยแยกค่าสังเกตด้วยตัวแปร gender ขั้นตอนการสร้างกราฟเป็นดังนี้

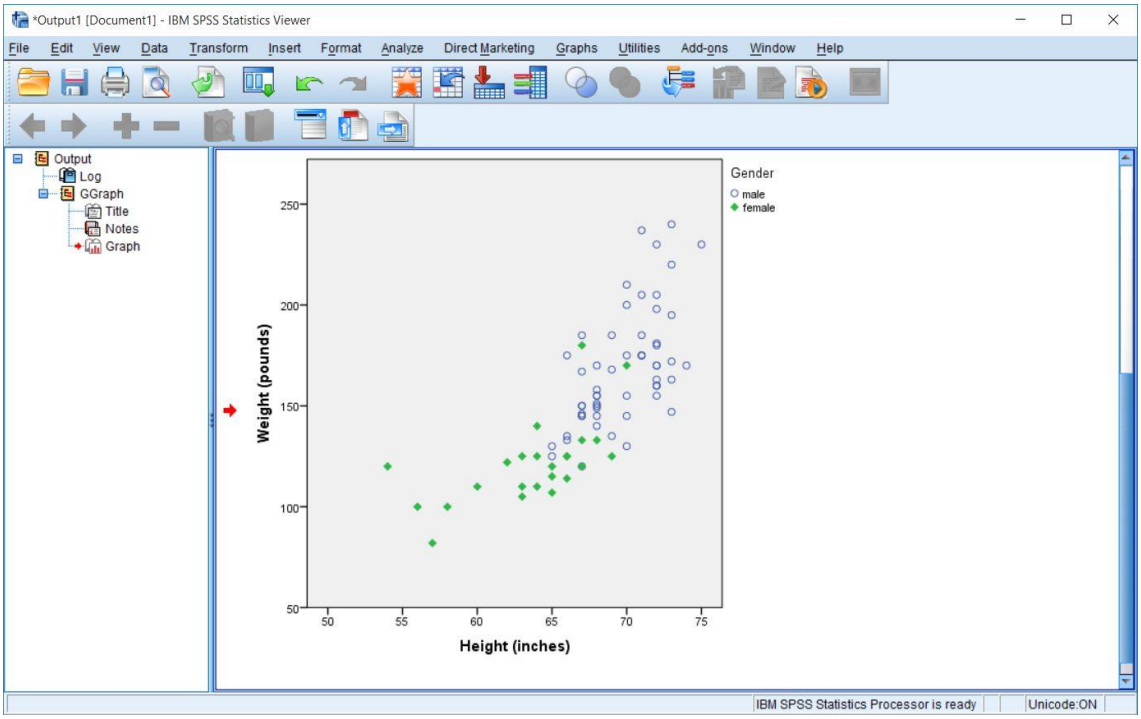
1. เลือก Graphs > Chart Builder...
2. กำหนดตัวเลือกในไดอะล็อค Chart Builder ดังนี้
   i. ในแถบ Gallery ช่อง Choose from: ให้เลือก Scatter/Dot แล้วลาก ไปใส่ในช่อง Chart preview uses example data
   ii. กำหนดตัวแปรในกราฟ (ภาพที่ 10.3 ก) โดยเลือกตัวแปรที่ช่อง Variables:
      a. เลือกตัวแปร Height แล้วลากไปใส่ X-Axis?
      b. เลือกตัวแปร Weight แล้วลากไปใส่ Y-Axis?
      c. เลือกตัวแปร Gender แล้วลากไปใส่ Set color
3. กด OK จะได้กราฟดังภาพที่ 10.3 ข



(ก)

ภาพที่ 10.3 กำหนดตัวแปรของกราฟ (ก)

(ข)



ภาพที่ 10.3 (ต่อ) กราฟที่ได้ หลังปรับเปลี่ยนสัญลักษณ์ของ Female (ข)

2. การวิเคราะห์สหสัมพันธ์ (สำหรับข้อมูลที่มีการกระจายแบบปกติ)

การวิเคราะห์สหสัมพันธ์ระหว่างตัวแปร 2 ตัวแปรที่มี measurement level เป็น Scale เป็นการวิเคราะห์ที่ต้องการทราบว่าตัวแปรคู่หนึ่งมีความสัมพันธ์เชิงเส้นตรงกัน (linear relationship) หรือไม่? โดยค่าที่ใช้ระบุระดับของสหสัมพันธ์ระหว่างตัวแปรคู่หนึ่ง จะใช้ Pearson's product-moment correlation coefficient (ใช้สัญลักษณ์ r แทนค่าสัมประสิทธิ์สหสัมพันธ์ของกลุ่มตัวอย่าง)

การตั้งสมมุติฐานและการทดสอบสมมุติฐาน รวมถึงการแปลผลการวิเคราะห์สหสัมพันธ์ระหว่างตัวแปรภายหลังจากได้ค่า r แล้ว ดังนี้

การตั้งสมมุติฐาน	ความหมาย	การทดสอบ
Ho: r = 0 H1: r ≠ 0	Ho: ไม่มีสหสัมพันธ์ระหว่างตัวแปรทั้งสอง H1: มีสหสัมพันธ์ระหว่างตัวแปรทั้งสอง	2-tailed
Ho: r = 0 H1: r > 0	Ho: ไม่มีสหสัมพันธ์ระหว่างตัวแปรทั้งสอง H1: มีสหสัมพันธ์ระหว่างตัวแปรทั้งสองแบบบวก (คือ X เพิ่ม Y เพิ่มหรือ X ลด Y ลด)	1-tailed
Ho: r = 0 H1: r < 0	Ho: ไม่มีสหสัมพันธ์ระหว่างตัวแปรทั้งสอง H1: มีสหสัมพันธ์ระหว่างตัวแปรทั้งสองแบบลบ (คือ X เพิ่ม Y ลด หรือ กลับกัน)	1-tailed

และในการแปลผลจากค่า correlation coefficient หรือค่า r จะเป็นดังนี้ (ค่าเป็นบวกหรือลบ แปลผลเหมือนกัน)

correlation coefficient, *r*	r^2	ความหมาย
0.00 – 0.25	0.0000 – 0.0625	no or weak relationship
0.25 – 0.50	0.0625 – 0.2500	fair degree of relationship
0.50 – 0.75	0.2500 – 0.5625	moderate to good relationship
0.75 – 1.00	0.5625 – 1.000	good to strong relationship

r^2 หรือ coefficient of determination เป็นค่าที่แสดงถึงระดับความสัมพันธ์ระหว่างตัวแปรทั้งสอง หากมีค่ามาก แสดงว่าตัวแปรทั้งสองมีระดับความสัมพันธ์มาก

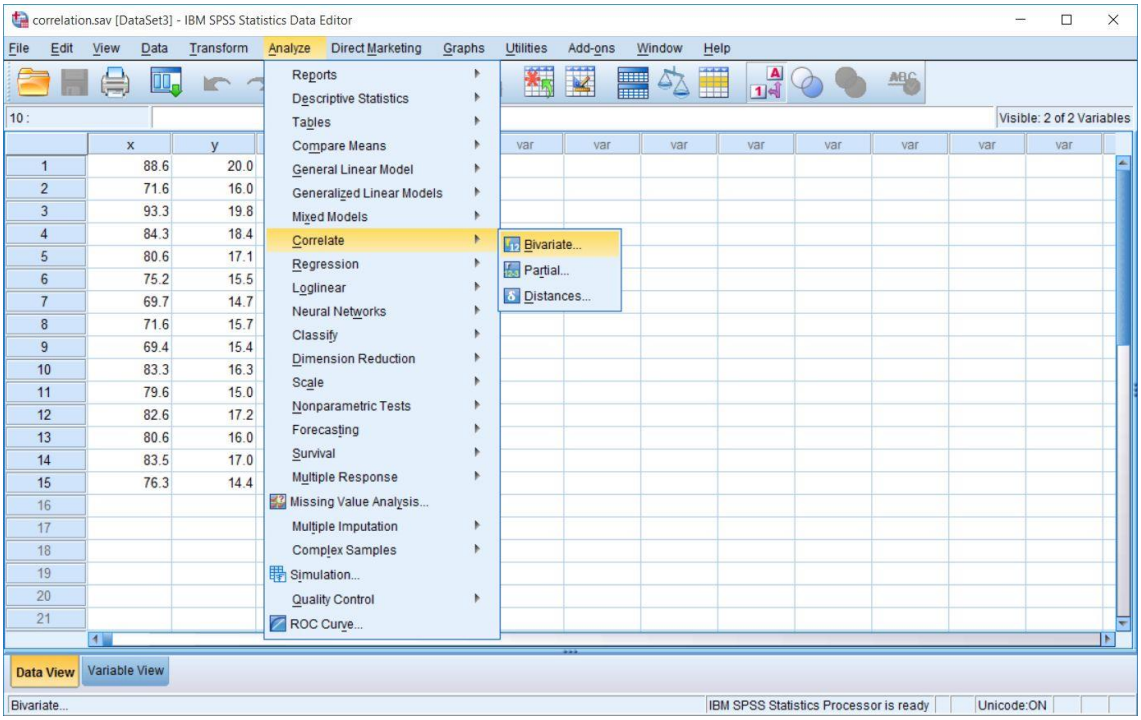
ตัวอย่างข้อมูลการวิเคราะห์สหสัมพันธ์ จะใช้ข้อมูลในหัวข้อ 1.2 ข้างต้น เนื่องจากกราฟที่ได้ แสดงให้เห็นว่าตัวแปรทั้งสองน่าจะมีความสัมพันธ์กัน สำหรับสมมุติฐาน จะเป็นดังนี้

Ho: ไม่มีสหสัมพันธ์ระหว่างอุณหภูมิที่ผิวดินกับความถี่การสั่นของปีกแมลง, r = 0

H1: มีสหสัมพันธ์ระหว่างอุณหภูมิที่ผิวดินกับความถี่การสั่นของปีกแมลง, r ≠ 0

ขั้นตอนการวิเคราะห์ มีดังนี้

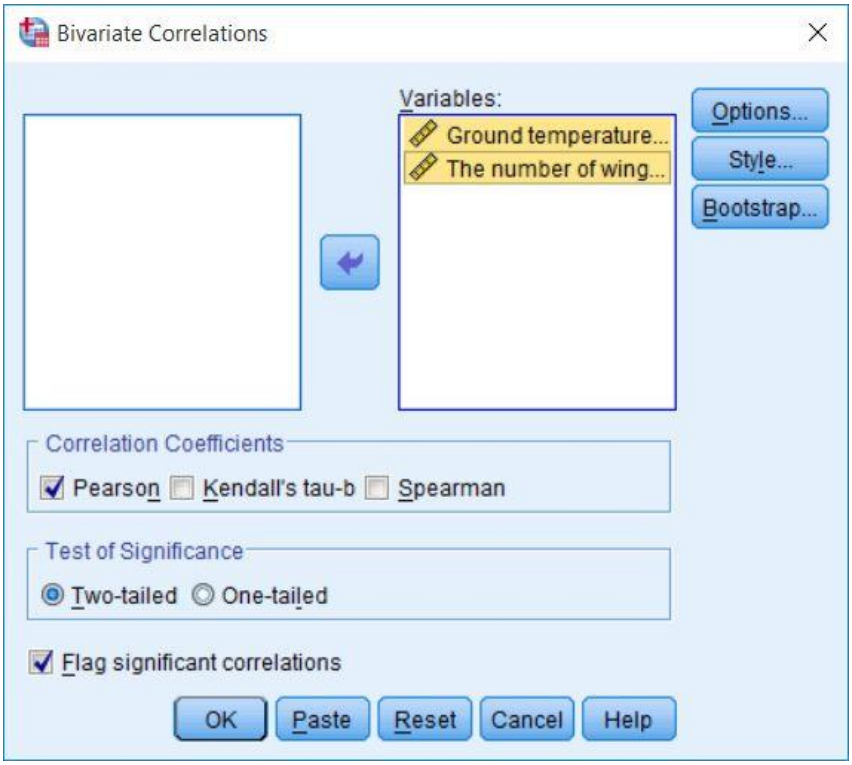
1. เลือก Analyze > Correlate > Bivariate… (ภาพที่ 10.4 ก)
2. ในช่องรายการตัวแปรด้านซ้าย เลือกตัวแปร Ground temperature… และ The number of wing… แล้วกด ▶ ให้มาอยู่ในช่อง Variables: (ภาพที่ 10.4 ข)
3. ในช่อง Correlation Coefficients เลือก ☑ Pearson
4. ในช่อง Test of Significance เลือก ◉ Two-tailed
5. เลือก ☑ Flag significant correlations
6. กด OK จะได้ผลการวิเคราะห์ดังภาพที่ 10.4 ค



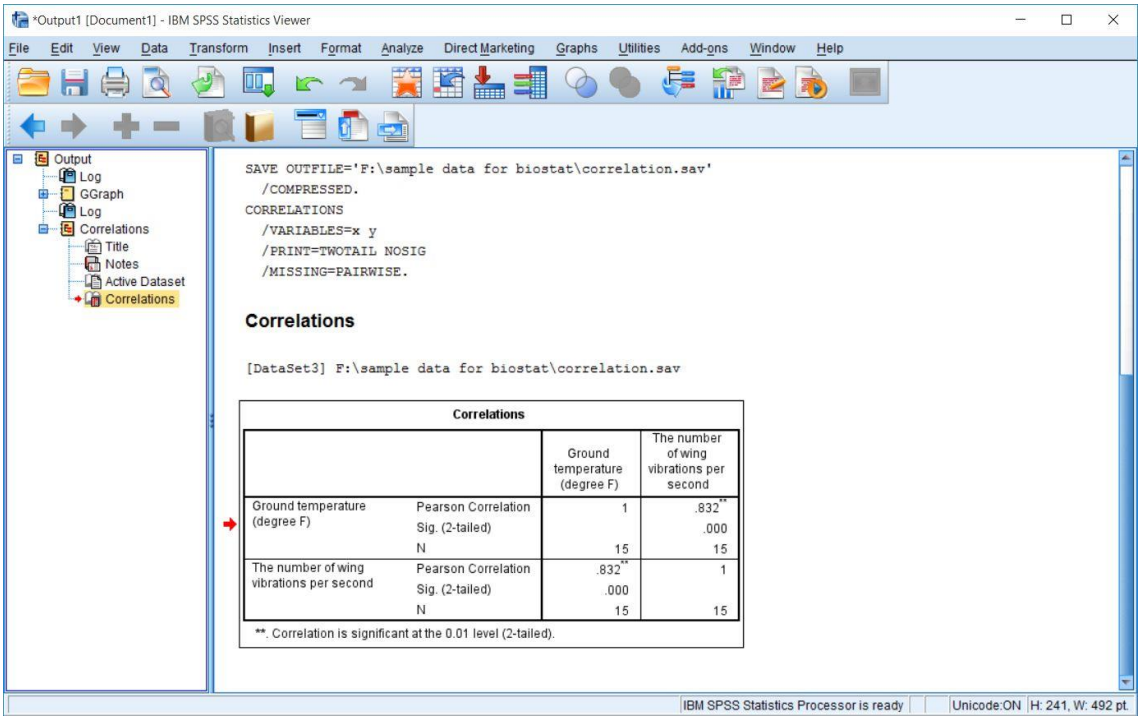
(ก)

ภาพที่ 10.4 คำสั่ง Correlate > Bivariate ในเมนู Analyze (ก)

(ข)



(ค)



ภาพที่ 10.4 (ต่อ) กำหนดตัวแปรและเลือกค่าสถิติ (ข) ผลการวิเคราะห์สหสัมพันธ์ (ค)

จากตารางในภาพที่ 10.4 ค จะมีค่าสัมประสิทธิ์สหสัมพันธ์ 2 ค่า คือ Ground temperature - The number of wing vibrations per second (ช่องบรรทัดแรก) กับ The number of wing vibrations per second - Ground temperature (ช่องบรรทัดที่สอง) ซึ่งมีค่าเดียวกัน ดังนั้น จะใช้ค่าในบรรทัดใดก็ได้ในการทดสอบสมมุติฐาน จากตารางนี้ r = 0.832 และ sig. (2-tailed) = 0.000 เนื่องจาก sig. (2-tailed) มีค่าน้อยกว่า 0.05 จึงปฏิเสธ Ho แล้วยอมรับ H1 นั้น

คือ r มีค่าต่างจาก 0 อย่างมีนัยสำคัญทางสถิติที่ความเชื่อมั่น 95 เปอร์เซ็นต์ จึงสรุปได้ว่าอุณหภูมิที่ผิวดินกับความถี่การสั่นของปีกแมลงมีสหสัมพันธ์กันแบบบวก โดยเมื่ออุณหภูมิที่ผิวดินสูงขึ้น ความถี่การสั่นของปีกแมลงจะมีความถี่สูงขึ้น

ตัวอย่างข้อมูลชุดที่สองเป็นการทำกราฟมาตรฐานเพื่อหาขนาดโมเลกุลของดีเอ็นเอ โดยใช้ดีเอ็นเอมาตรฐาน λ DNA/*Hind*III ที่แยกชิ้นดีเอ็นเอด้วยกระแสไฟฟ้าในอะกาโรสเจล ผลการวัดระยะการเคลื่อนที่ของชิ้นดีเอ็นเอแต่ละชิ้นจากจุดเริ่มต้นเป็นดังตาราง

Migration (mm)	13	16	19	22	29	31
Fragment size (kbp)	23.0	9.4	6.6	4.4	2.3	2.0

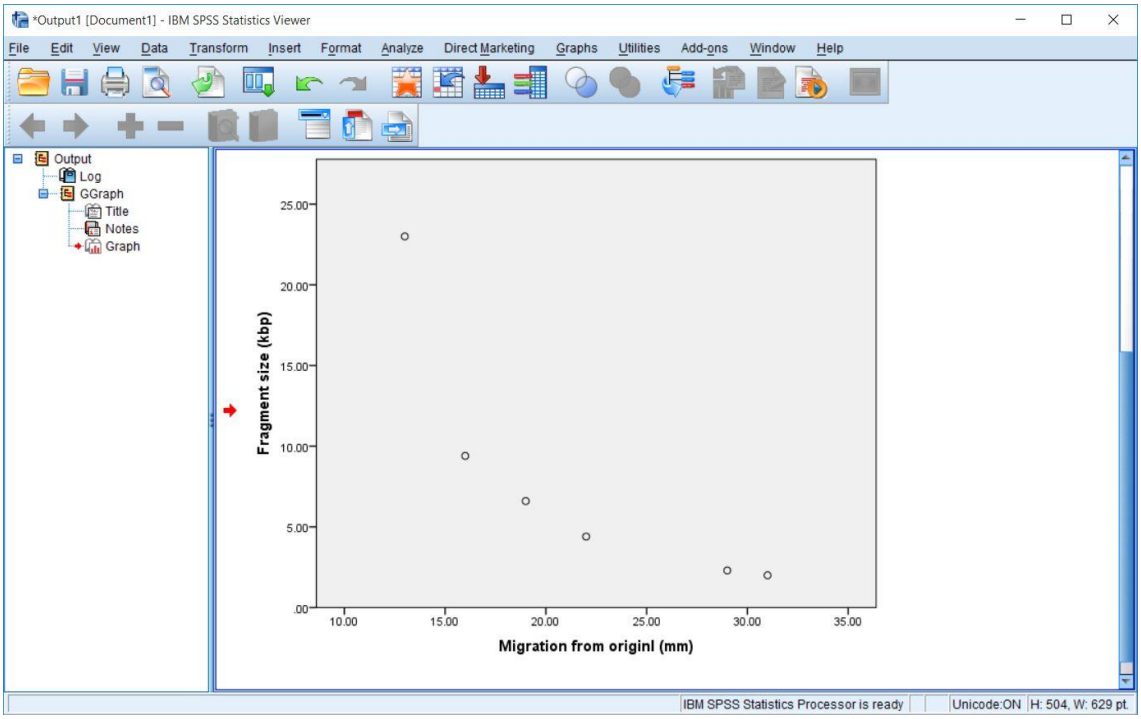
ในการเตรียมข้อมูล จะใช้ตัวแปร 2 ตัวแปร คือ ตัวแปรแรกชื่อ migration มีคำอธิบายเป็น Migration from origin (mm) และตัวแปรที่สองชื่อ size มีคำอธิบาย Fragment size (kbp) เมื่อสร้างกราฟ scatter plot ระหว่างตัวแปรทั้งสองโดยให้ migration เป็นแกน X และ size เป็นแกน Y จะได้กราฟดังภาพที่ 10.5 ก ซึ่งจะเห็นได้ว่าตัวแปรทั้งสองไม่น่าจะมีสหสัมพันธ์เชิงเส้นตรง เนื่องจากมีข้อมูลพื้นฐานเกี่ยวกับการเคลื่อนที่ของดีเอ็นเอในอะกาโลสมาก่อนว่าส่วนกลับของระยะการเคลื่อนที่ของชิ้นดีเอ็นเอ (1/migration) มีความสัมพันธ์เชิงเส้นตรงกับค่า Log ของขนาดชิ้นดีเอ็นเอ (Log_{10} size) ดังนั้น จึงต้องแปลงค่าในตัวแปรทั้งสองก่อน แล้วจึงสร้างกราฟเพื่อพิจารณาอีกครั้ง

การแปลงค่าในตัวแปรทั้งสอง ในทางปฏิบัติคือการสร้างตัวแปรใหม่ขึ้นมา 2 ตัวแปรโดยใช้คำสั่ง Transform > Compute Variable… ดังนี้

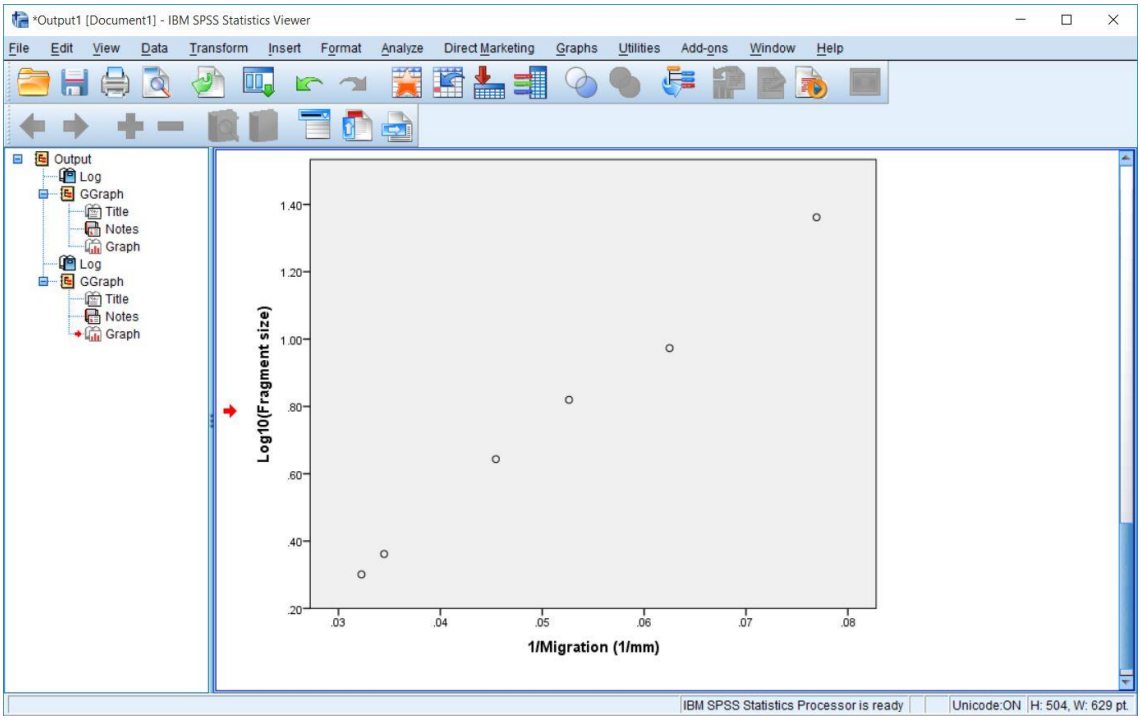
1. ตัวแปรที่ 1 – เลือก Transform > Compute Variable แล้วตั้งค่าในไดอะล็อค ดังนี้
   i. กดในช่อง Numeric Expression: แล้วกด 1 / แล้วเลือกตัวแปร Migration from origin… จากรายการตัวแปรในช่องซ้าย แล้วกด
   ii. ที่ช่อง Target Variable: พิมพ์ชื่อตัวแปรที่ 1 เป็น invmig แล้วกด Type & Label…
      a. ที่ Label เลือก Label: และพิมพ์คำอธิบายเป็น 1/Migration (1/mm)
      b. ที่ Type เลือก Numeric แล้วกด Continue
   iii. กด OK
2. ตัวแปรที่ 2 – เลือก Transform > Compute Variable แล้วกด Reset ก่อน จากนั้นจึงตั้งค่าในไดอะล็อค ดังนี้
   i. ที่รายการกลุ่มฟังก์ชั่น Function group: เลือก Arithmetic
   ii. ที่รายการ Functions and Special Variables: เลือก Lg10 แล้วกด
   iii. ที่ช่อง Numeric Expression: เลือกตัวแปร Fragment size (kbp)… จากรายการตัวแปรในช่องซ้าย แล้วกด ให้ไปแทน ? ของฟังก์ชั่น
   iv. ที่ช่อง Target Variable: พิมพ์ชื่อตัวแปรที่ 2 เป็น lg10size แล้วกด Type & Label…
      a. ที่ Label เลือก Label: และพิมพ์คำอธิบายเป็น Log10(Fragment size)
      b. ที่ Type เลือก Numeric แล้วกด Continue
   v. กด OK

สร้างกราฟ scatter plot ระหว่างตัวแปรทั้งสองโดยให้ 1/migration เป็นแกน X และ Log_{10} size เป็นแกน Y จะได้กราฟดังภาพที่ 10.5 ข ซึ่งจะเห็นได้ว่าตัวแปรที่แปลงค่าแล้วมีความสัมพันธ์เชิงตรงเกือบสมบูรณ์

(ก

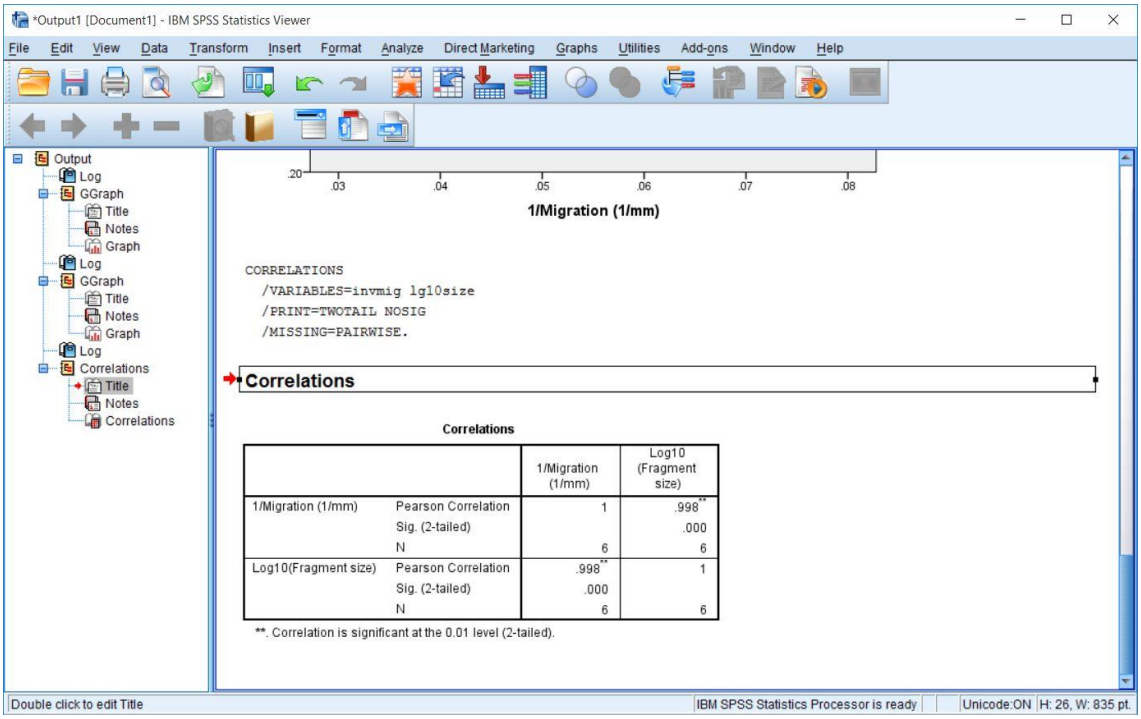


(ข)



ภาพที่ 10.5 กราฟ scatter plot ข้อมูลเดิม (ก) กราฟ scatter plot หลังแปลงค่า (ข)

(ค)



ภาพที่ 10.5 (ต่อ) ผลการวิเคราะห์สหสัมพันธ์ (ค)

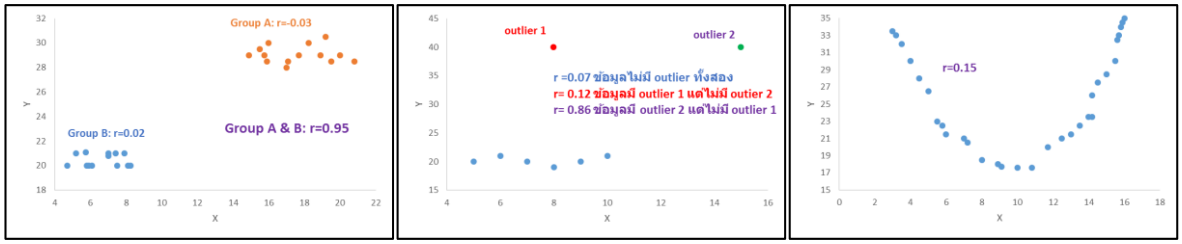
จากกราฟในภาพที่ 10.5 ข ซึ่งเห็นได้ว่าข้อมูลมีความสัมพันธ์เชิงเส้นตรงกัน จึงวิเคราะห์สหสัมพันธ์โดยใช้ Pearson's correlation coefficient ได้ โดยจะตั้งสมมุติฐานได้เป็นดังนี้

Ho: ไม่มีสหสัมพันธ์ระหว่าง 1/Migration กับ Log_{10} size, $r = 0$

H1: มีสหสัมพันธ์ระหว่าง 1/Migration กับ Log_{10} size, $r \neq 0$

จากผลการวิเคราะห์ในภาพที่ 10.6 พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง 1/Migration กับ Log_{10} size คือ $r = 0.998$ และ sig. (2-tailed) = 0.000 เนื่องจาก sig. (2-tailed) มีค่าน้อยกว่า 0.05 จึงปฏิเสธ Ho แล้วยอมรับ H1 นั้นคือ r มีค่าต่างจาก 0 อย่างมีนัยสำคัญทางสถิติที่ความเชื่อมั่น 95 เปอร์เซ็นต์ จึงสรุปได้ 1/Migration กับ Log_{10} size มีสหสัมพันธ์กันแบบบวก โดยเมื่อ 1/Migration มีค่าสูงขึ้น Log_{10} size จะมีค่าสูงขึ้น

การวิเคราะห์สหสัมพันธ์ในตัวอย่างข้อมูลทั้งสองชุดข้างต้น ชี้ให้เห็นว่าควรพิจารณากราฟของตัวแปรแต่ละคู่ที่ต้องการวิเคราะห์สหสัมพันธ์ เนื่องจากความสัมพันธ์ของตัวแปรแต่ละคู่อาจจะไม่ใช่ความสัมพันธ์เชิงเส้นตรงตามเงื่อนไขของค่าสัมประสิทธิ์นี้ แต่อาจปรับเปลี่ยนด้วยการแปลงค่าให้ข้อมูลเข้ากับเงื่อนไขได้ อย่างไรก็ตาม ในบางกรณีกราฟอาจแสดงให้เห็นว่าไม่มีความสัมพันธ์ระหว่างตัวแปรทั้งสองแต่อย่างใด เช่น ข้อมูลแยกเป็น 2 กลุ่มย่อย อยู่ตรงข้ามตามแนวทแยงมุมของกราฟ (ภาพที่ 10.6 ซ้าย) หรือกราฟอาจแสดงค่าที่เป็น outlier ซึ่งทำให้ค่าสัมประสิทธ์สหสัมพันธ์มีค่าต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ (ภาพที่ 10.6 กลาง) หรือกราฟที่แสดงความสัมพันธ์แบบ non-linear (ภาพที่ 10.6 ขวา)



ภาพที่ 10.6 กราฟ scatter plot และผลการวิเคราะห์สหสัมพันธ์ที่ไม่เป็นความสัมพันธ์เชิงเส้นตรง (ดัดแปลงจากเอกสารของ Chandan Saha เรื่อง Biostatistics for Health Care Researchers: A Short Course)

3. การวิเคราะห์ถดถอยเชิงเส้นตรง

การวิเคราะห์ถดถอยที่กล่าวถึงในปฏิบัติการนี้ เป็นเพียงการวิเคราะห์ถดถอยเชิงเส้นตรงที่มีตัวแปรอิสระเพียง 1 ตัวแปร (คือ independent variable, X) และตัวแปรตาม (คือ dependent variable, Y) เพียง 1 ตัวแปร โดยมี model เป็น Y = a + bX เมื่อ a คือ ค่า intercept และ b คือ สัมประสิทธิ์ของตัวแปรอิสระ ดังนั้น จึงต้องการตัวแปร 2 ตัวแปร ในการวิเคราะห์ถดถอยเชิงเส้นตรงนี้ หากมีการสลับตัวแปรทั้งสองแล้ว ต้องทำการวิเคราะห์ใหม่เพื่อให้ได้สมการทำนายที่ถูกต้อง

สมมุติฐานสำหรับการวิเคราะห์ถดถอยเชิงเส้นตรงจะเป็นดังนี้

Ho: ค่าความชันของ regression line ที่ดีที่สุดไม่ต่างจาก 0

H1: ค่าความชันของ regression line ที่ดีที่สุดต่างจาก 0

ตัวอย่างข้อมูลในการวิเคราะห์ข้อมูลของแมลง striped ground cricket ที่กล่าวถึงใน 1.2 ข้างต้น โดยให้อุณหภูมิเป็นตัวแปรอิสระและความถี่เป็นตัวแปรตาม ขั้นตอนการวิเคราะห์ถดถอยเชิงเส้นตรงเป็นดังนี้

1. เลือก Analyze > Regression > Linear… (ภาพที่ 10.7 ก)
2. กำหนดตัวแปรในไดอะล็อกดังนี้
   - i. เลือก Ground temperature (degree… จากรายการตัวแปรในช่องด้านซ้าย กด ▶ ของช่อง Independent(s): (ภาพที่ 10.7 ข)
   - ii. เลือก The number of wing vibration… จากรายการตัวแปรในช่องด้านซ้าย กด ▶ ของช่อง Dependent: (ภาพที่ 10.7 ข)
   - iii. กด Statistics… เพื่อเลือกค่าสถิติ (ภาพที่ 10.7 ค)
     - a. ในช่อง Regression Coefficients เลือก ☑ Estimates
     - b. เลือก ☑ Model fit
     - c. กด Continue
   - iv. กด Plots… เพื่อสร้างกราฟ (ภาพที่ 10.7 ง-จ)
     - a. กราฟที่ 1 observed values vs. standardized predicted values
       - เลือก *ZPRED จากรายการตัวแปรในช่องซ้าย แล้วกด ▶ ของช่อง X:
       - เลือก DEPENDNT จากรายการตัวแปรในช่องซ้าย แล้วกด ▶ ของช่อง Y:
       - กด Next

b. กราฟที่ 2 standardized residuals vs. standardized predicted values

- เลือก *ZPRED จากรายการตัวแปรในช่องซ้าย แล้วกด ▶ ของช่อง X:
- เลือก *ZRESID จากรายการตัวแปรในช่องซ้าย แล้วกด ▶ ของช่อง Y:
- กด Next

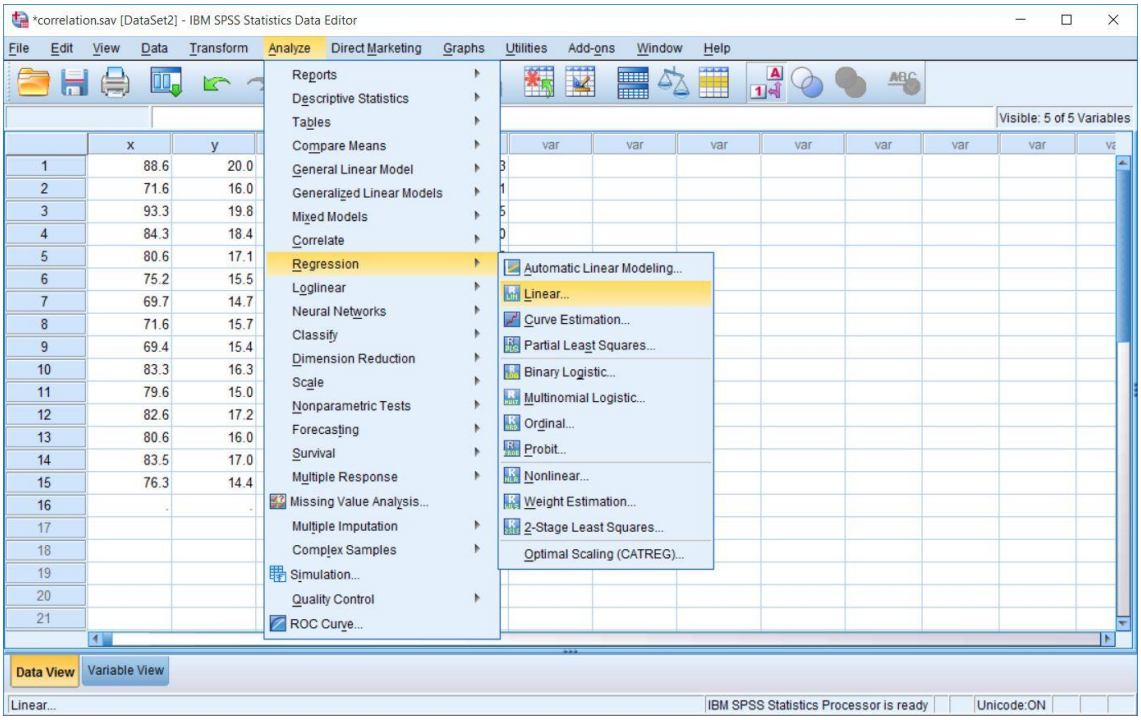
c. กด Continue

v. กด Options... (ภาพที่ 10.7 ฉ)

a. เลือก ☑ Include constant in equation

b. กด Continue

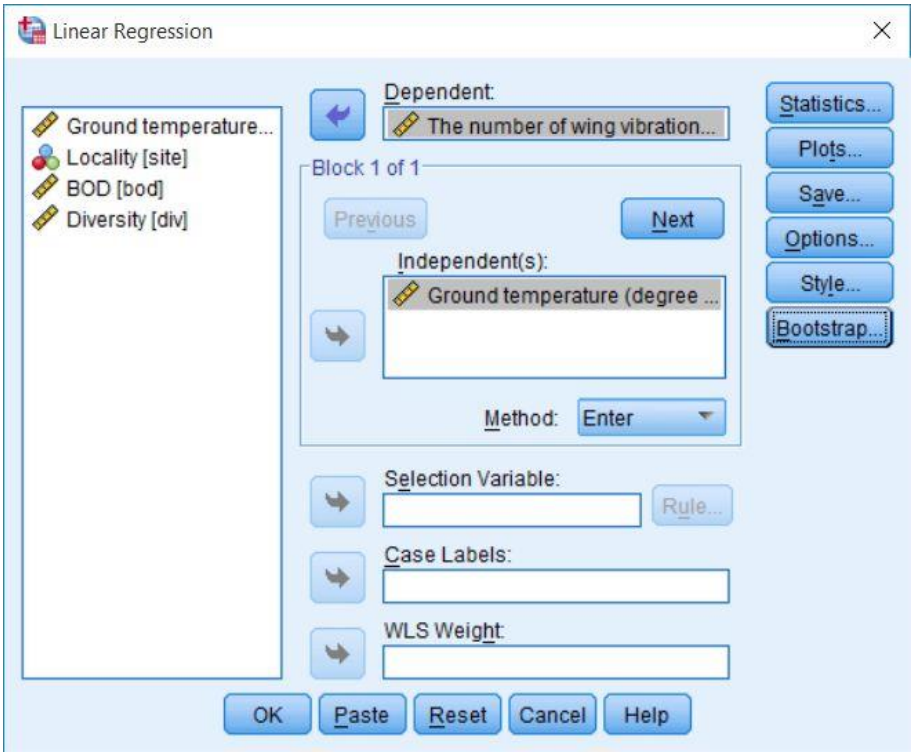
3. กด OK



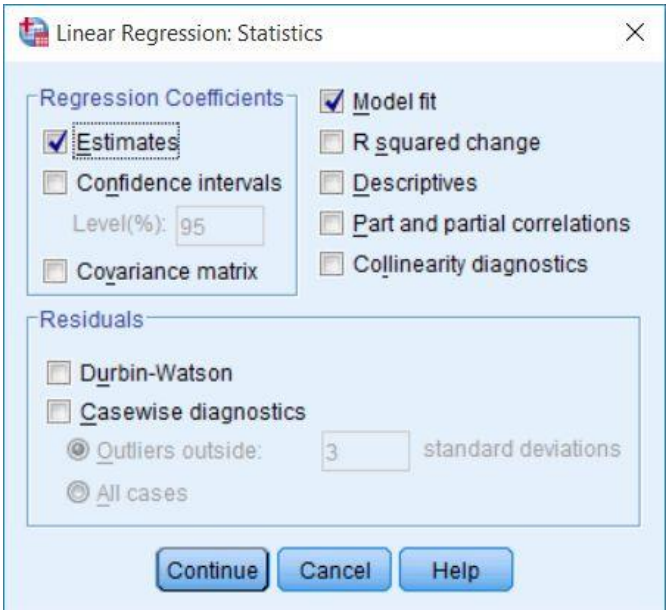
(ก)

ภาพที่ 10.7 คำสั่ง Regression > Linear ในเมนู Analyze (ก)

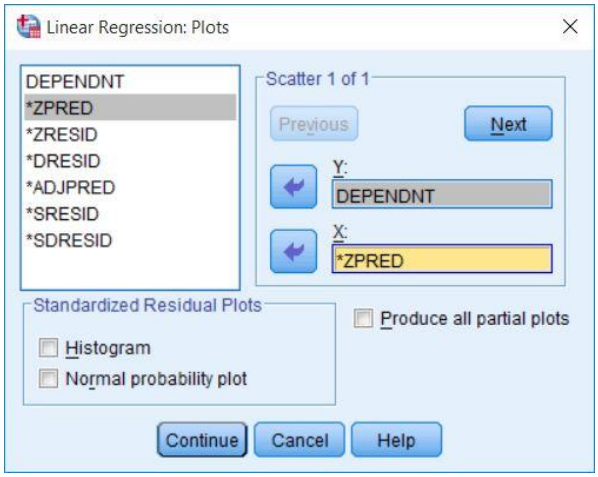
(ข)



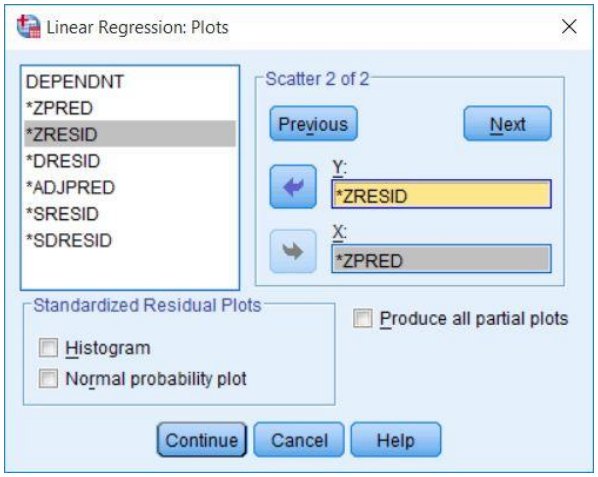
(ค)



ภาพที่ 10.7 (ต่อ) กำหนดตัวแปร (ข) เลือกค่าสถิติ (ค)

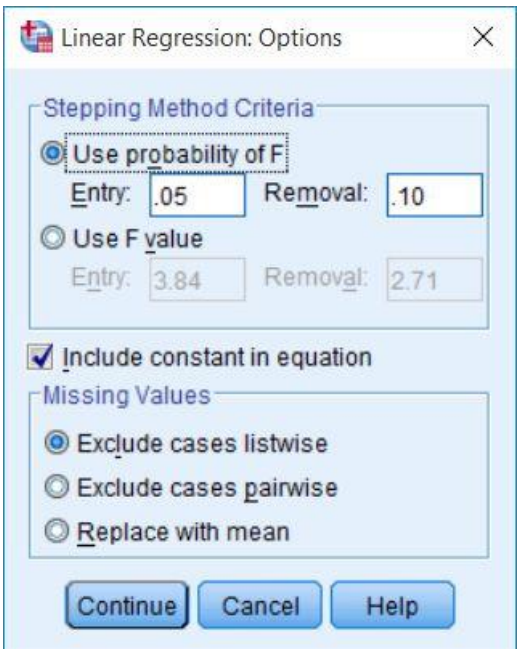


(ง)



(จ)

(ฉ)



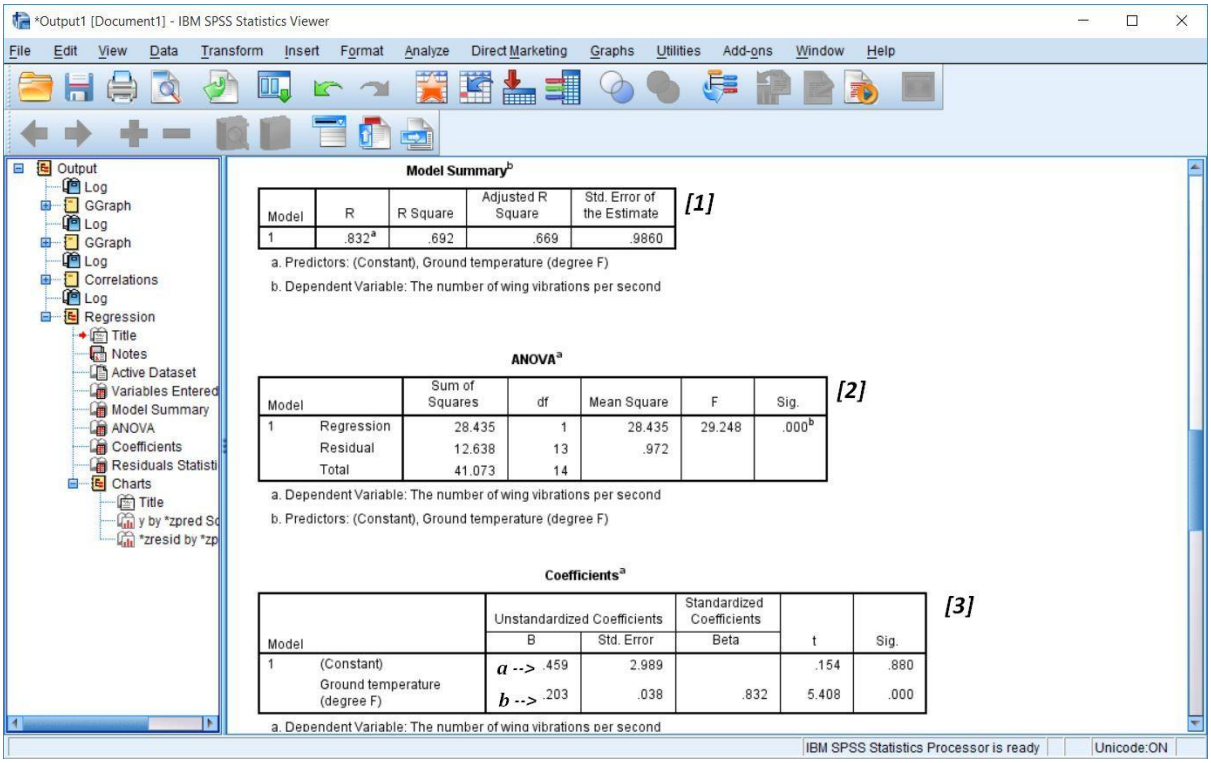
ภาพที่ 10.7 (ต่อ) กำหนดตัวแปรในการสร้างกราฟ (ง-จ) กำหนดให้มี intercept ในแบบจำลอง (ฉ)

ผลการวิเคราะห์ประกอบด้วยตาราง Model Summary (ภาพที่ 10.8 ก *[1]*) ซึ่งจะมีค่า R^2 หรือ coefficient of determination เท่ากับ 0.692 ค่านี้มักใช้เป็นดัชนีที่บ่งบอกว่า regression line ที่สร้างขึ้นนี้สามารถอธิบายเปอร์เซ็นต์การแปรของข้อมูลดิบได้มากน้อยเพียงใด หรืออีกนัยหนึ่ง ค่า R^2 นี้ เป็นค่าที่แสดงถึงความพอดีของสมการถดถอยเชิงเส้นกับข้อมูลดิบ (คือ goodness of fit)

ตารางถัดไป คือ ตาราง ANOVA (ภาพที่ 10.8 ก *[2]*) ซึ่งจะทดสอบสมมุติฐานที่ตั้งไว้ ค่า sig. = 0.000 ซึ่งน้อยกว่า 0.05 จึงปฏิเสธ Ho แล้วยอมรับ H1 นั้นคือ ค่าความชันของ regression line ที่ดีที่สุดต่างจาก 0 อย่างมีนัยสำคัญทางสถิติที่ความเชื่อมั่น 95 เปอร์เซ็นต์ หรือกล่าวอีกนัยหนึ่งว่าตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์เชิงเส้นตรงอย่างมีนัยสำคัญทางสถิติ

ตารางถัดต่อไป คือ ตารางค่า **coefficient** (ภาพที่ 10.8 ก *[3]*) โดยในสมการถดถอยเชิงเชิงเส้นตรง **Y = a + bX** จะใช้ค่าในช่อง **Unstandardized coefficients**[1] ซึ่งค่า **a** เป็น 0.459 และ **b** เป็น 0.203 (บรรทัด ***a*** และ ***b*** ตามลำดับ) แต่ก่อนจะนำค่าดังกล่าวไปแทนค่าในสมการ ต้องพิจารณาก่อนว่าค่าของ **a** และ **b** มีค่าต่างจาก 0 หรือไม่[2] โดยพิจารณาจากค่า **Sig.** ของค่า **a** และ **b** ในตาราง ซึ่งในตัวอย่างข้อมูลนี้ ค่า **a** มีค่า **Sig.=0.880** ซึ่งมีค่ามากกว่า 0.05 จึงไม่ปฏิเสธ **Ho: a = 0** แสดงว่าค่าคงที่ **a=0.459** ไม่ต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ และค่า **b** มีค่า **Sig.=0.000** ซึ่งมีค่าน้อยกว่า 0.05 จึงปฏิเสธ **Ho: b = 0** แล้วยอมรับ **H1: b ≠ 0** แสดงว่าค่าความชันหรือค่าสัมประสิทธิ์ **b=0.203** ต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ ดังนั้น สมการถดถอยเชิงเส้นตรงอาจเขียนได้เป็น **Y = 0.459 + 0.203X**

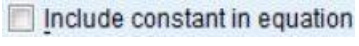
(ก)

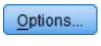


ภาพที่ 10.8 ตารางค่า R *[1]* ตาราง ANOVA *[2]* และตารางค่าสัมประสิทธิ์ *[3]* (ก)

---

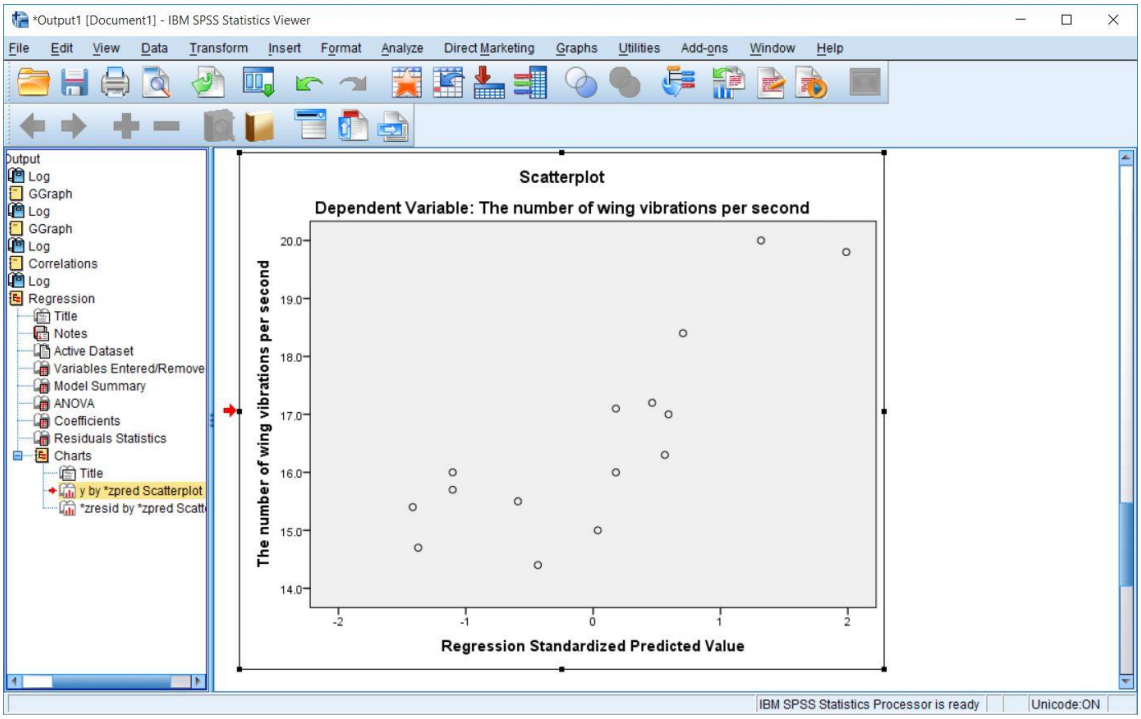
1 หากต้องการใช้กับข้อมูลที่เป็น standardized score แล้ว จะใช้สมการ Y = bX โดยค่า b เป็นค่าจากช่อง Standardized coefficients ในตาราง Coefficient (ภาพที่ 10.8 ก *[3]*) ซึ่ง b เป็น 0.832

2 ในกรณีค่า a จะทดสอบ Ho: a = 0 vs. H1: a ≠ 0 และในกรณีค่า b จะทดสอบ Ho: b = 0 vs. H1: b ≠ 0 อย่างไรก็ตาม หากการทดสอบสมมุติฐานเกี่ยวกับค่า a ได้ผลว่าไม่แตกต่างจาก 0 อย่างมีนัยสคัญทางสถิติแล้ว ก็ไม่ควรนำค่า a นี้ไปใช้ และควรสร้างสมการถดถอยใหม่โดยเลือก 

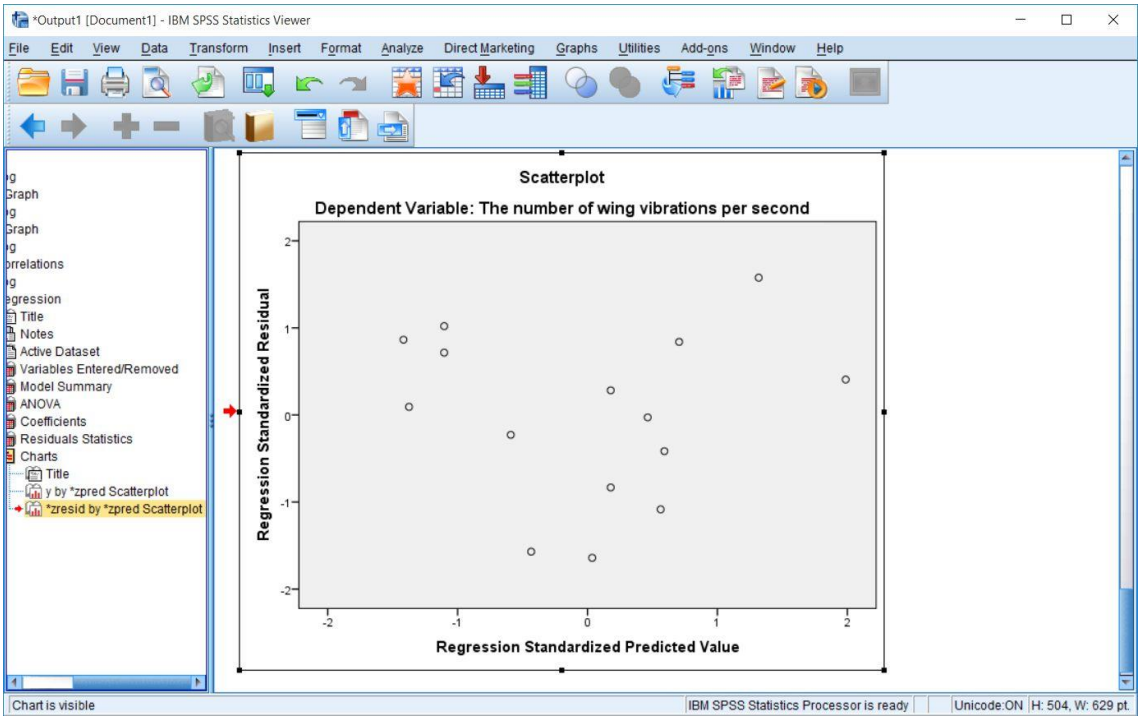
ใน 

อนึ่ง มีข้อแนะนำว่าแม้ว่าจะไม่อาจปฏิเสธ Ho: a = 0 ก็อาจใช้ค่า a ในสมการถดถอยได้ โดยไม่ถือว่าผิดแต่อย่างไร

(ข)



(ค)



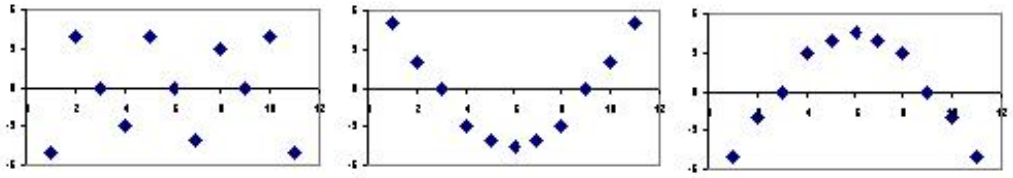
ภาพที่ 10.8 (ต่อ) กราฟ observed values vs. standardized predicted values (ข) กราฟ standardized residuals vs. standardized predicted values (ค)

กราฟ scatter plot ระหว่างตัวแปรตามกับค่าตัวแปรตามที่ได้จากสมการ (observed values vs. standardized predicted values ในภาพ 10.8 ข) เป็นตัวบ่งชี้ตัวหนึ่งว่าสมการถดถอยเชิงเส้นที่ได้นี้ใช้ทำนายค่าตัวแปรตามได้ดีเพียงใด ตามทฤษฎีแล้ว กราฟระหว่างตัวแปรทั้งสองนี้ จะเป็นจุดเรียงตามแนวเส้นทแยงมุม นั้นคือ ค่าตัวแปรตามในข้อมูลดิบกับค่าตัวแปรตามที่ได้จากสมการเป็นค่าดียวกันหรือใกล้เคียงกันมาก แต่จากภาพที่ 10.8 ข เห็นได้ว่าจุด

ส่วนใหญ่ไม่เรียงตามแนวเส้นทะแยงมุม แสดงว่าน่าจะมีการแปรในข้อมูลดิบที่สมการนี้ไม่สามารถอธิบายได้ ซึ่งอาจจะเกิดจากข้อมูลไม่ได้มาจากประชากรเดียวกัน หรือความแปรปรวนของข้อมูลบางส่วนแตกต่างจากข้อมูลอีกส่วนหนึ่ง หรือเหตุผลอื่น ๆ

สำหรับกราฟระหว่าง standardized residuals vs. standardized predicted values (ภาพที่ 10.8 ค) เป็นกราฟที่ประกอบการพิจารณาว่าสมการถดถอยที่ได้นั้นเหมาะสมการข้อมูลดิบนี้หรือไม่ โดย standardized residuals ควรมีการกระจายตัวรอบ ๆ แนวเส้น Y=0 ตลอดค่าของ standardized predicted values ซึ่งจากภาพที่ 10.8 ค นี้ จุดมีการกระจายสม่ำเสมอ แสดงว่าสมการถดถอยเชิงเส้นตรงนี้เหมาะสมกับข้อมูลดิบนี้ อนึ่ง ในกรณีที่การกระจายของจุดไม่สม่ำเสมอ ก็แสดงว่าไม่สามารถใช้สมการถดถอยเชิงเส้นตรงอธิบายข้อมูลดิบได้ ต้องใช้ model อื่น (เช่น S curve, Parabolic, ฯลฯ) ดังตัวอย่างกราฟในภาพที่ 10.9

ข้อควรคำนึงถึงเกี่ยวกับสมการถดถอยเชิงเส้นตรงนี้ คือ ค่าตัวแปรอิสระที่จะใช้ในการทำนายค่าตัวแปรตามจากสมการนี้ **จะต้องอยู่ในช่วงของค่าต่ำสุดและค่าสูงสุดที่ใช้ในการสร้างสมการนี้** ไม่เช่นนั้นก็ไม่สามารถนำสมการนี้ไปใช้ได้ และ**ไม่ควรนำไปใช้**



ภาพที่ 10.9 กราฟ standardized residuals vs. standardized predicted values – กราฟซ้าย-จุดกระจายสม่ำเสมอ เป็น linear modlel กราฟกลางและกราฟขวา-จุดกระจายไม่สม่ำเสมอ เป็น non-linear model (จาก http://stattrek.com/regression/residual-analysis.aspx?Tutorial=AP)

4. การประมาณเส้นโค้ง

เนื่องจากคำสั่ง Regression > Linear... เหมาะสำหรับการวิเคราะห์ถดถอยเชิงเส้นเท่านั้น ไม่สามารถใช้การวิเคราะห์ข้อมูลที่ไม่มีความสัมพันธ์เชิงเส้นตรง จึงต้องใช้การประมาณเส้นโค้งมาวิเคราะห์เพื่อหาสมการถดถอยแบบอื่นที่เหมาะสมกับข้อมูลดิบแทน

สำหรับตัวอย่างข้อมูล จะใช้ขนาดโมเลกุลของดีเอ็นเอในข้อ 2 โดยใช้ข้อมูลดิบที่ยังไม่ได้แปลงค่า ขั้นตอนการวิเคราะห์เป็นดังนี้

1. เลือก Analyze > Regression > Curve Estimation... (ภาพที่ 10.10 ก)
2. กำหนดตัวเลือกในไดอะล็อคดังนี้ (ภาพที่ 10.10 ข)
    a. เลือก Fragment size (kbp)... จากรายการตัวแปรในช่องด้านซ้าย กด ▶ ของช่อง Dependent(s):
    b. เลือก Migration from origin... จากรายการตัวแปรในช่องด้านซ้าย กด ▶ ใต้ ◉ Variable: ของช่อง Independent
    c. เลือก ☑ Include constant in equation
    d. เลือก ☑ Plot models

e. ในช่อง Models เลือกแบบกราฟที่จะประมาณเส้นโค้งจากตัวเลือก ซึ่งหากไม่ทราบ อาจทดลองเลือกกราฟทุกแบบ อย่างไรก็ตาม ในตัวอย่างนี้ จะเลือก ☑ Linear ☑ Logarithmic ☑ Exponential และ ☑ Power

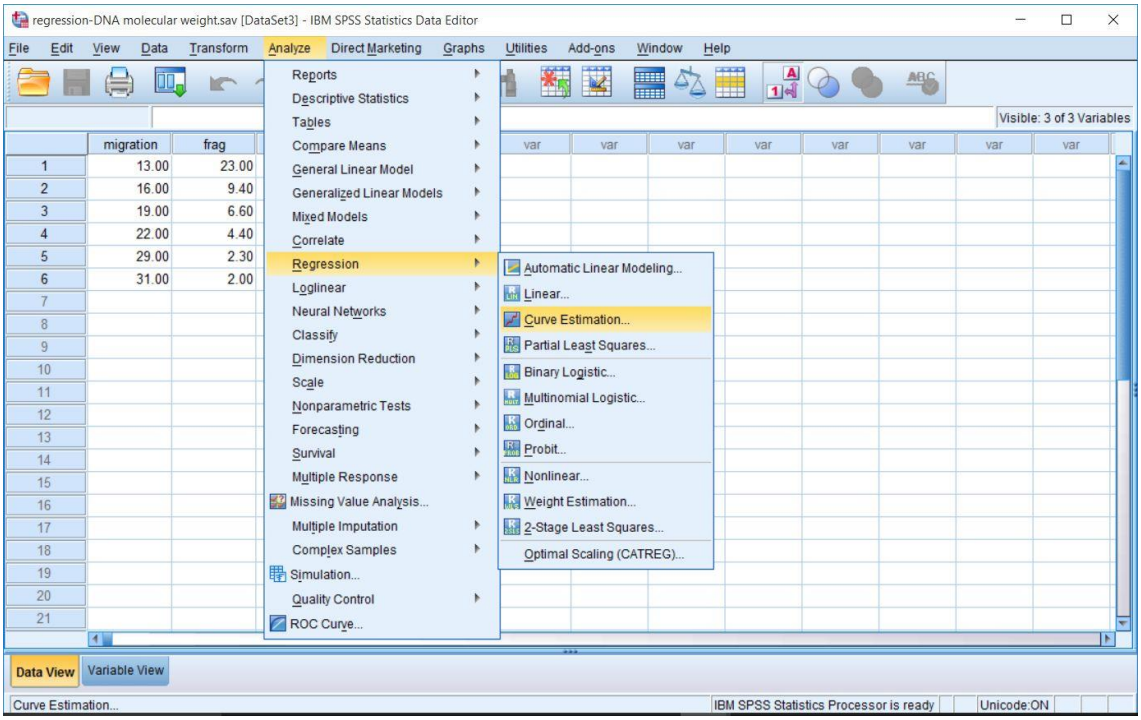
f. เลือก ☑ Display ANOVA table

3. กด OK

ผลการวิเคราะห์แสดงเพียงกราฟในภาพที่ 10.10 ค สำหรับตาราง **Model Summary** ที่ได้จากการวิเคราะห์ถดถอยของแบบจำลองทั้งสี่แบบนั้น พบว่าแบบจำลอง **Power** มีค่า **r=0.992** และ r^2**=0.983** ซึ่งเป็นค่าที่สูงที่สุด และทุกแบบจำลองค่าความชันของสมการถดถอยต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ (ไม่มีตารางนำเสนอในที่นี้) เมื่อพิจารณาเส้นกราฟจากสมการถดถอยพร้อมกับข้อมูลดิบในกราฟในภาพที่ 10.10 ค เส้นกราฟจากแบบจำลอง **Power** ผ่านหรือเข้าใกล้ข้อมูลดิบมากกว่าเส้นกราฟจากแบบจำลองแบบอื่นอีก 3 แบบ เมื่อพิจารณาร่วมกันกับ r^2 แล้ว สรุปได้ว่าเส้นกราฟในแบบจำลอง **Power** เหมาะสมกับข้อมูลดิบมากที่สุด สำหรับสมการถดถอยของข้อมูลขนาดโมเลกุลของดีเอ็นเอนี้ เป็น

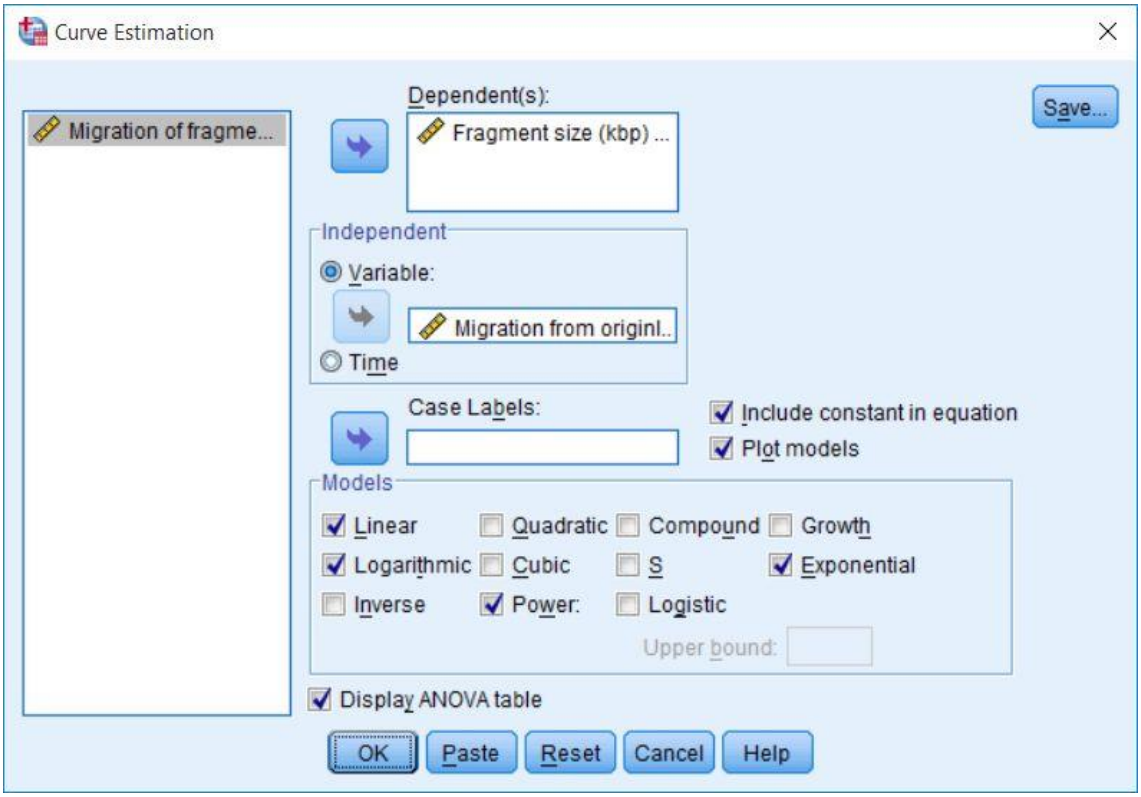
$$\text{Fragment size (kbp)} = 19197.049 \times [\text{Migration (mm)}]^{-2.691}$$

(ก)

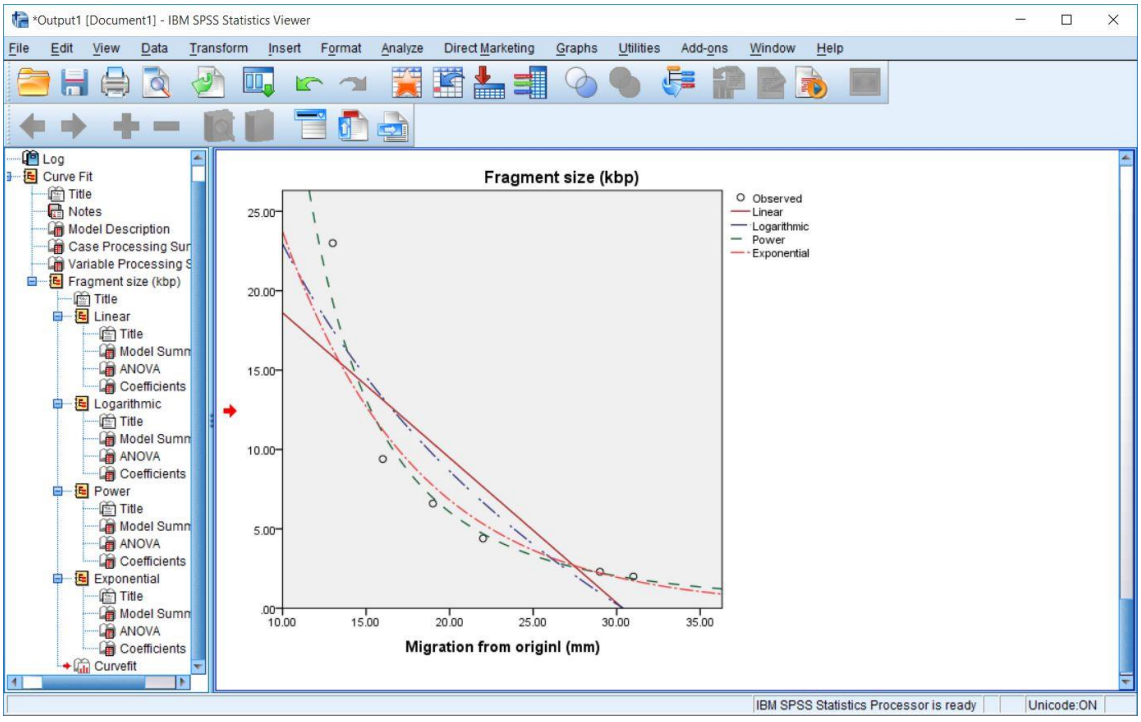


ภาพที่ 10.10 คำสั่ง Regression > Curve Estimation (ก)

(ข)



(ค)



ภาพที่ 10.10 (ต่อ) กำหนดตัวแปรและตัวเลือกอื่น ๆ (ข) กราฟจากแบบจำลอง หลังปรับเปลี่ยนสีเส้นกราฟ (ค)

แบบฝึกหัดปฏิบัติการที่ 10

1. ผู้วิจัยคนหนึ่งมีคำถามว่าพัฒนาการของทารกอายุ 30 สัปดาห์ทั้งสองแบบ มีความสัมพันธ์กันหรือไม่ อย่างไร จึงได้เก็บข้อมูลจากทารก 12 คน ดังนี้

proximal development scores	17	10	14	21	16	21	22	18	18	16	18	20
distal development scores	11	8	13	14	21	19	14	21	16	16	10	14

ก. จงสร้างกราฟ Scatter แล้วดูว่าการกระจายของข้อมูลเป็นอย่างไร

ข. หากวิเคราะห์ว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงหรือไม่ จะตั้งสมมุติฐานอย่างไร ระบุทั้ง Ho และ H1 พร้อมทั้งบอกว่าจะทดสอบแบบ 2-tailed หรือ 1-tailed

ค. จงแสดงค่า coefficient และผลการทดสอบ พร้อมทั้งสรุปผล

2. จากข้อมูลตัวแปร x และตัวแปร y ดังตารางด้านล่าง

x	1.0	1.1	1.5	1.7	2.0	2.5	3.0	3.0	3.0	3.1	3.2	3.5	3.6	4.0
y	3.0	3.2	4.1	4.2	5.0	6.2	7.3	7.0	7.1	7.4	6.0	8.1	8.0	9.0

ก. จงสร้างกราฟ Scatter แล้วดูว่าสามารถสร้างสมการถดถอยเชิงเส้นตรงกับข้อมูลชุดนี้ได้หรือไม่

ข. ถ้าสามารถสร้างได้ ให้สร้างสมการดังกล่าว พร้อมทั้งทดสอบสมมุติฐานที่เกี่ยวข้อง ให้แสดงสมมุติฐาน ตารางสถิติและสรุปผล

ค. ถ้าสามารถใช้สมการถดถอยเชิงเส้นตรงนี้ได้ จงบอกค่า y เมื่อ x มีค่าเป็น 1.3, 2.2, 3.4 และ 3.7

3. ในการศึกษา photoperiodism ในนกเป็ดน้ำชนิดหนึ่ง ผู้วิจัยต้องการสร้างสมการที่สามารถใช้ทำนายระยะเวลาของฤดูผสมพันธุ์จากจำนวนชั่วโมงที่มีแสงใน 1 วัน เมื่อการผสมพันธุ์ได้เริ่มต้นขึ้น จึงออกเก็บข้อมูล ได้ผลดังนี้

จำนวนชั่วโมงที่มีแสงใน 1 วัน	12.8	13.9	14.1	14.7	15.0	15.1	16.0	16.5	16.6	17.2	17.9
ระยะเวลาของฤดูผสมพันธุ์ (วัน)	110	54	98	50	67	58	52	50	43	15	28

ก. จงสร้างกราฟ Scatter แล้วดูว่าสามารถสร้างสมการถดถอยเชิงเส้นตรงกับข้อมูลชุดนี้ได้หรือไม่

ข. ถ้าสามารถสร้างได้ ให้สร้างสมการดังกล่าว พร้อมทั้งทดสอบสมมุติฐานที่เกี่ยวข้อง ให้แสดงสมมุติฐาน ตารางสถิติและสรุปผล

ค. ถ้าสามารถใช้สมการถดถอยเชิงเส้นตรงนี้ได้ จงหาว่าระยะเวลาของฤดูผสมพันธุ์ เมื่อจำนวนชั่วโมงที่มีแสงใน 1 วัน มีค่าเป็น 13.4, 15.5 และ17.5 ชั่วโมง

4. จากข้อมูลการกินอาหารของ Didinium เมื่อเลี้ยงด้วย Paramecium ความหนาแน่นต่างๆ ได้ผลดังนี้

Didinium feedig rate	2.9	2.7	2.7	2.6	2.5	2.2	2.1	1.8	1.7	1.5
Initial Paramecium density	10	5	7	15	20	5	9	10	20	10
Didinium feedig rate	1.4	1.2	1.1	1.1	0.9	0.8	0.8	0.5	0.4	0.5
Initial Paramecium density	20	15	25	30	28	15	10	40	70	80
Didinium feedig rate	0.3	0.3	0.2	0.1	0.2	0.2	0.2	0.3	0.1	0.2
Initial Paramecium density	110	140	142	148	151	178	18	185	190	195

ก. จงสร้างกราฟ Scatter แล้วดูว่าสามารถสร้างสมการถดถอยเชิงเส้นตรงกับข้อมูลชุดนี้ได้หรือไม่

ข. ถ้าสามารถสร้างได้ ให้สร้างสมการดังกล่าว พร้อมทั้งทดสอบสมมุติฐานที่เกี่ยวข้อง ให้แสดงสมมุติฐาน ตารางสถิติและสรุปผล แต่ถ้าไม่ได้ ควรทำอย่างไร?

การบ้านปฏิบัติการที่ 10

1. เพื่อศึกษาหาสมการที่ใช้ค่าความเข้มข้นของเอสโทรนในน้ำลาย **(x)** ในการทำนายค่าความเข้มข้นของเอสโทรนในน้ำเลือด **(y)** จึงได้รวบรวมข้อมูลจากอาสาสมัครที่แข็งแรง ได้ผลดังนี้ (มีหน่วยเป็น **pg/mL**)

x	7.4	7.5	8.5	9.0	9.0	11.0	13.0	14.0	14.5	16.0	17.0	18.0	20.0	23.0
y	30.0	25.0	31.5	27.5	39.5	38.0	43.0	49.0	55.0	48.5	51.0	64.5	63.0	68.0

ก. จงสร้างกราฟ **scatter** และจากกราฟ สามารถสร้างสมการถดถอยเชิงเส้นตรงได้หรือไม่

ข. หาค่าสัมประสิทธิ์ **a, b** และสมการถดถอยเชิงเส้นตรง พร้อมทั้งการทดสอบทางสถิติ

ค. จงหาว่าความเข้มข้นของเอสโทรนในน้ำเลือดเป็นเท่าใด เมื่อความเข้มข้นของเอสโทรนในน้ำลายมีค่า 17.5 **pg/mL**

2. ในการเฝ้าระวังของเสียในน้ำทิ้งที่ปล่อยลงสู่ทะเลสาบ ผู้วิจัยได้ใช้ปริมาณไนเตรตในน้ำเป็นตัวชี้วัด ซึ่งวิธีการวัดแบบเดิมใช้เวลานานในการวิเคราะห์ จึงมีการพัฒนาวิธีการใหม่ที่ใช้เวลาน้อยลงเพื่อมาทดแทน ถ้าหากว่าปริมาณไนเตรตในน้ำที่หาโดยวิธีการทั้งสองมีความสัมพันธ์กันสูง ผู้วิจัยก็จะใช้วิธีการใหม่แทน ได้ผลการทดสอบดังต่อไปนี้ (มีหน่วยเป็น **µg/L**)

วิธีการวัดแบบเดิม	25	40	125	75	150	300	270	400	450	575
วิธีการวัดแบบใหม่	30	80	150	80	200	350	240	320	470	583

ก. จงวิเคราะห์ว่ามีความสัมพันธ์ระหว่างวิธีการวัดทั้งสองหรือไม่ พร้อมทั้งนำเสนอการทดสอบทางสถิติ

ข. ควรใช้วิธีการวัดปริมาณไนโตรตในน้ำแบบใหม่แทนวิธีการวัดแบบเดิมหรือไม่ อธิบาย

3. การสอนผู้ป่วยโรคเบาหวานให้วัดปริมาณน้ำตาลในเลือดด้วยตนเองโดยใช้ **glucose oxidase stick** ซึ่งมีค่าใช้จ่ายน้อยกว่าการตรวจสอบในห้องปฏิบัติการ จะเป็นประโยชน์อย่างมากต่อผู้ป่วย ผู้วิจัยจึงต้องการทดสอบว่าวิธีการวัดด้วย **glucose oxidase stick** มีความถูกต้องมากน้อยเพียงใด จึงได้ทดสอบในผู้ป่วย 40 คน โดยให้ผู้ป่วยวัดปริมาณน้ำตาลด้วย **glucose oxidase stick** (**x**) และผู้วิจัยวัดปริมาณน้ำตาลของผู้ป่วยด้วยการวิเคราะห์ในห้องปฏิบัติการ (**y**) ได้ผลดังนี้ (มีหน่วยเป็น **mmol/L**)

x	y	x	y	x	y	x	y
1.3	2.4	3.2	4.4	7.0	7.7	15.0	14.9
2.0	3.0	3.6	4.3	8.0	8.0	15.0	13.8
2.4	2.3	3.7	4.3	8.0	10.0	17.5	17.6
2.6	3.0	3.7	5.0	10.0	10.0	18.7	17.5
2.5	2.2	3.8	4.4	10.2	9.5	6.0	6.0
2.6	2.4	4.4	4.5	10.2	11.2	8.7	8.8
2.7	2.5	4.3	5.0	12.5	11.0	5.6	5.7
3.0	3.8	5.0	4.5	11.3	13.0	9.1	9.0
3.7	2.5	5.0	6.2	13.0	13.1	16.2	12.5
3.7	3.5	6.3	6.2	14.5	13.8	9.0	14.0

ก. จงสร้างกราฟ **scatter** และพิจารณาว่าสามารถวิเคราะห์ถดถอยเชิงเส้นตรงได้หรือไม่

ข. ถ้าได้ ให้หาค่า **a, b** และสมการถดถอยเชิงเส้นตรง พร้อมนำเสนอผลการทดสอบทางสถิติ

ก. ถ้าผู้วัดค่าน้ำตาลด้วย **stick** ได้ค่า 4 **mmol/L** แล้ว ค่าที่จะวัดในห้องปฏิบัติการจะเป็นเท่าใด

ข. จะอธิบายเกี่ยวกับ **coefficient of determination** ว่าอย่างไร