เอกสารประกอบการสอน

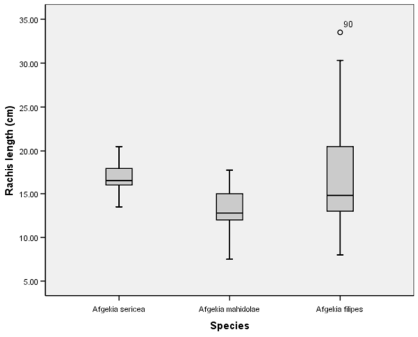
รายวิชา 2305 384 Biological Statistics Laboratory

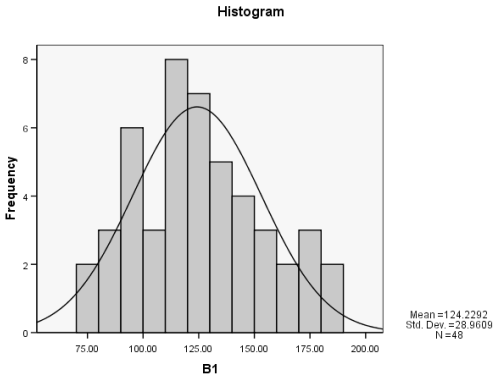
ผศ.ดร.ต่อศักดิ์ สีลานันท์

ภาควิชาพฤกษศาสตร์

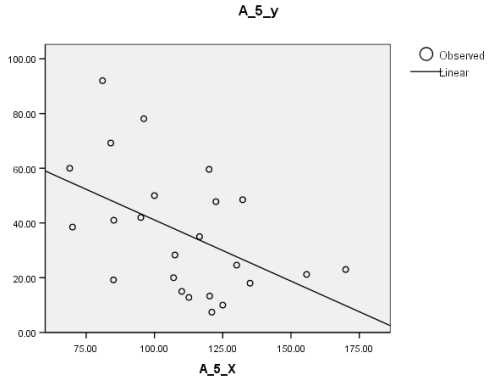
คณะวิทยาศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย





ภาคปลาย ปีการศึกษา 2556



สารบัญ

ปฏิบัติการที่ 3 Descriptive Statistics I 1
การสร้างตารางแจกแจงความถี่และ Histogram ด้วยโปรแกรม Microsoft® Excel 1
การสร้างตารางแจกแจงความถี่และ Histogram ด้วยโปรแกรม SPSS 4
การนำเสนอข้อมูลแบบกราฟ 11
การหาค่า Percentile และ Percentile Rank 15
ปฏิบัติการที่ 4 Descriptive Statistics II 18
การหาค่าสถิติเบื้องต้นด้วยโปรแกรม Microsoft® Excel 18
การหาค่าสถิติเบื้องต้นด้วยโปรแกรม SPSS 19
ปฏิบัติการที่ 5 ความน่าจะเป็น 22
คะแนนมาตราฐาน หรือ z score 22
การหาค่า Continuous probability 24
การหาค่า Binomial distribution 24
ปฏิบัติการที่ 6 Hypothesis Testing: One population/sample 27
การป้อนข้อมูลสำหรับการวิเคราะห์ทางสถิติด้วยโปรแกรม SPSS 27
One sample t-test 28
One sample z test 30
ปฏิบัติการที่ 7 Hypothesis Testing: Two or more populations/samples 34
Paired-samples t test 34
Independent two-samples t test 36
Three or more samples test 39
ปฏิบัติการที่ 8 General Linear Model for CRD and RCBD 46
GLM สำหรับ CRD 46
GLM สำหรับ RCBD 51
ปฏิบัติการที่ 9 GLM for Factorial Design 58
GLM สำหรับ Factorial Analysis in CRD 58
GLM สำหรับ Factorial in RCBD 68
ปฏิบัติการที่ 10 การวิเคราะห์ถดถอยเชิงเส้นและการวิเคราะห์สหสัมพันธ์ 75
การสร้างกราฟ 75
การวิเคราะห์ถดถอยเชิงเส้นตรง 77
การประมาณเส้นโค้ง 81
การวิเคราะห์ความสัมพันธ์เชิงเส้นตรง 83
การใช้โปรแกรม Microsoft® Excel ในการสร้างกราฟและการประมาณเส้นโค้ง 86
การใช้โปรแกรม Microsoft® Excel ในการวิเคราะห์ถดถอยเชิงเส้นตรง 90
การใช้โปรแกรม Microsoft® Excel ในการวิเคราะห์ความสัมพันธ์เชิงเส้นตรง 91
ปฏิบัติการที่ 11 การวิเคราะห์ทางสถิติแบบนอนพาราเมตริกซ์ 95
X^2 Test for Single variable 95
Mann-Whitney U test 109
Wilcoxon Matched-Pairs Signed Rank (WSR) Test 110
Kruskal-Wallis Test 112
Spearman's rank order correlation coefficient 113

ประมวลรายวิชา (Course Syllabus)

1.	รหัสวิชา	2305 384	
2.	จำนวนหน่วยกิต	1 หน่วยกิต	
3.	ชื่อวิชา	ปฏิบัติการชีวสถิติ / Biological Statistics Laboratory	
4.	คณะ / ภาควิชา	วิทยาศาสตร์ / พฤกษศาสตร์	
5.	ภาคการศึกษา	ปลาย	
6.	ปีการศึกษา	2556	
7.	ชื่อผู้สอน	อ.ดร.ธีรดา หวังสมบรูณ์ดี (ผู้ประสานงาน)	ตอนเรียนที่ 1, 2 (พฤ, ศุกร์) (90 ชม.)
		อ.ดร.วราลักษณ์ เกษตรานันท์	ตอนเรียนที่ 1, 2 (พฤ, ศุกร์) (90 ชม.)
		ผศ.ดร.ศศิธร พ่วงปาน	ตอนเรียนที่ 1 (พฤ) (45 ชม.)
		อ.ดร.อัญชลี ใจดี	ตอนเรียนที่ 1 (พฤ) (45 ชม.)
		ผศ.ดร.ต่อศักดิ์ สีลานันท์	ตอนเรียนที่ 2 (ศุกร์) (45 ชม.)

8. เงื่อนไขรายวิชา
   - 8.1 วิชาที่ต้องเรียนมาก่อน: -
   - 8.2 วิชาบังคับร่วม: 2305383 Biological Statistics
   - 8.3 วิชาควบ: -
9. สถานภาพของวิชา: วิชาบังคับในภาควิชา / วิชาเลือกในคณะ
10. ชื่อหลักสูตร: วท.บ. (พฤกษศาสตร์) / วท.บ. (พันธุศาสตร์) / วท.บ. (ชีววิทยา) / วท.บ. (จุลชีววิทยา)
11. วิชาระดับ: ปริญญาตรี
12. จำนวนชั่วโมงที่สอน: 3 ชั่วโมง / สัปดาห์
13. เนื้อหารายวิชา (Course description)

    การสาธิตและการฝึกหัดปฏิบัติการการใช้วิธีทางสถิติในงานวิจัยทางชีววิทยา และวิธีวิเคราะห์ในแง่สถิติ
14. ประมวลการเรียนวิชา (Course Outline)
    - 14.1 วัตถุประสงค์ทั่วไป / หรือวัตถุประสงค์เชิงพฤติกรรม นิสิตสามารถ
      - 14.1.1 คำนวณและวิเคราะห์ผลทางสถิติในงานวิจัยทางชีววิทยาได้
      - 14.1.2 ใช้โปรแกรมคอมพิวเตอร์ช่วยวางแผนงานวิจัยและวิเคราะห์ผล
      - 14.1.3 แปลผลที่ได้จากการวิเคราะห์และสรุปงานวิจัยทางชีววิทยาและพันธุศาสตร์ได้
    - 14.2 เนื้อหารายวิชาต่อสัปดาห์ (เอกสารแนบ)
    - 14.3 วิธีจัดการเรียนการสอน
      - สอนปฏิบัติการในชั้นเรียนโดยใช้โปรแกรมคอมพิวเตอร์
    - 14.4 สื่อการสอน
      - คอมพิวเตอร์และเครื่องฉายแอลซีดี
    - 14.5 การมอบหมายงานผ่านระบบเครือข่าย - **ไม่มี**
    - 14.6 การวัดผลการเรียน
      - 14.6.1 การประเมินความรู้ทางวิชาการ ร้อยละ 90
      - 14.6.2 การประเมินผลงานที่ได้รับมอบหมาย ร้อยละ 5
      - 14.6.3 การประเมินการเข้าชั้นเรียนและความตั้งใจ ร้อยละ 5

15. รายชื่อหนังสืออ่านประกอบ

15.1 หนังสือบังคับ – **ไม่มี**

15.2 หนังสืออ่านเพิ่มเติม – หนังสือการใช้โปรแกรม SPSS และ Excel

16. การประเมินผลการเรียนการสอน

16.1 ใช้แบบประเมินการสอนรูปแบบ 01 การสอนแบบปฏิบัติการ

16.2 การปรับปรุงจากผลการประเมินการสอนครั้งที่ผ่านมา

ปรับปรุงวิธีการประเมินผลการวัดผลการเรียนการสอน

16.3 การอภิปรายหรือวิเคราะห์ที่เสริมสร้างคุณสมบัติที่พึงประสงค์ของบัณฑิตจุฬาลงกรณ์มหาวิทยาลัย

เป็นรายวิชาที่เสริมสร้างคุณลักษณะด้านสติปัญญาและวิชาการ ส่งเสริมให้นิสิตรู้จักคิด และวิเคราะห์ผลที่ได้ทางสถิติ และนิสิตสามารถนำไปใช้เป็นพื้นฐานในการประกอบอาชีพได้

2305384 Biostatistics Lab เวลาเรียน – วันพฤหัสบดี (ตอนเรียนที่ 1) และศุกร์ (ตอนเรียนที่ 2)
เวลา 13.00-16.00 น. ห้องคอมพิวเตอร์ 3 (อาคารวชิรุณหิศ ชั้น 5)

	หัวข้อ	วันที่สอน	
		พฤหัสบดี	ศุกร์
1. 2.	แนะนำปฏิบัติการ Familiarization with computer programs used in biostatistics 2.1 Excel 2.2 SPSS	31 ต.ค. 56	1 พ.ย. 56
3.	Descriptive statistics I 3.1 Frequency distribution 3.1.1 Frequency distribution, Relative frequency distribution, Cumulative frequency distribution 3.1.2 Percentile, Percentile rank 3.2 Stem-and-leaf plot, Histogram, Box plot	7 พ.ย. 56	8 พ.ย. 56
4	Descriptive statistics II 4.1 Mean, Median, Mode 4.2 Standard error, Standard deviation, Variance, Range	14 พ.ย. 56	15 พ.ย. 56
. 5	Probability: Discrete and Continuous probability 5.1 Binomial probability distribution 5.2 Normal probability distribution, *z* Scores	21 พ.ย. 56	22 พ.ย. 56
. 6.	Hypothesis testing I: One and Two populations/samples test: 6.1 z test and *t* test for normal distribution 6.2 Test for binomial distribution 6.3 Two independent samples test (*t* test) and *F* Test for Equality of Variance 6.4 Paired samples test (paired *t* test)	28 พ.ย. 56	29 พ.ย. 56
	หยุดวันเฉลิมพระชนมพรรษา	5 ธ.ค. 56	6 ธ.ค. 56
7.	Hypothesis testing II: Three or more populations/samples tests	12 ธ.ค. 56	13 ธ.ค. 56
	สอบกลางภาค	19 ธ.ค. 56	20 ธ.ค. 56
8.	ANOVA I: Univariate experiment 8.1 Complete Random Design with fixed effect 8.2 Randomized Complete Block Design	2 ม.ค. 57	3 ม.ค. 57
9.	ANOVA II: Factorial experiment	9 ม.ค. 57	10 ม.ค. 57
	สอบปลายภาคครั้งที่ 1	23 ม.ค. 57	24 ม.ค. 57
10.	Linear regression and correlation	30 ม.ค. 57	31 ม.ค. 57
11.	Non-parametric statistics 11.1 Contingency table: 11.1.1 Test of Independence 11.1.2 Test of Homogeneity	6 ก.พ. 57	7 ก.พ. 57
	หยุดวันมาฆะบูชา	13 ก.พ. 57	14 ก.พ. 57
12.	Non-parametric statistics 12.1 Mann-Whitney U test 12.2 Wilcoxon Signed-Rank test 12.3 Kruskal-Wallis test 12.4 Spearman's rank correlation coefficient	20 ก.พ. 57	21 ก.พ. 57
	สอบปลายภาคครั้งที่ 2 (สอบถาม)	27 ก.พ. 57	28 ก.พ. 57

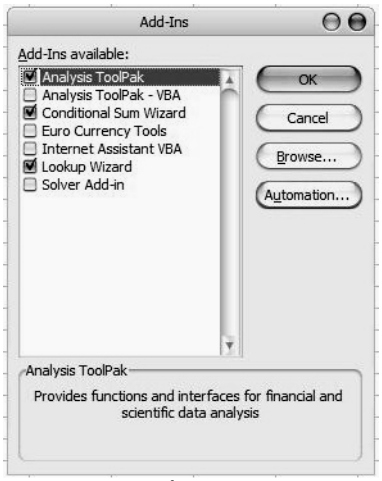
ปฏิบัติการที่ 3 Descriptive Statistics I

ในการอธิบายข้อมูลดิบที่รวบรวมได้จากตัวอย่างหรือประชากร สามารถทำได้หลายวิธี ในบทปฏิบัติการนี้ จะกล่าวถึงการสร้างตารางแจกแจงความถี่และการนำเสนอในรูปกราฟด้วยโปรแกรม Microsoft® Excel และ SPSS

1. การสร้างตารางแจกแจงความถี่และ Histogram ด้วยโปรแกรม Microsoft® Excel

การสร้างตารางแจกแจงความถี่และ Histogram ด้วยโปรแกรม Microsoft® Excel จะใช้ Add-Ins ชื่อ Analysis ToolPack โดยทั่วไป add-ins ชุดนี้จะไม่ถูกติดตั้งใน Excel แต่เมื่อมีการติดตั้งแล้ว จะอยู่ในเมนู Tools, Data Analyses...

การติดตั้ง add-ins ชุดนี้ใน Microsoft Excel® 2003 ทำได้โดยเลือก Tools, Add-Ins... ซึ่งจะมีตัวเลือกดังภาพที่ 3.1 (ก) ให้เลือก Analysis ToolPack (และ add-ins ตัวอื่นที่ต้องการ) แล้วกดที่ OK และตรวจสอบว่ามีการติดตั้ง add-ins แล้วโดยดูที่ Tools จะมี Data Analysis... ปรากฏเพิ่มมา ดังรูปที่ 3.1 (ข)

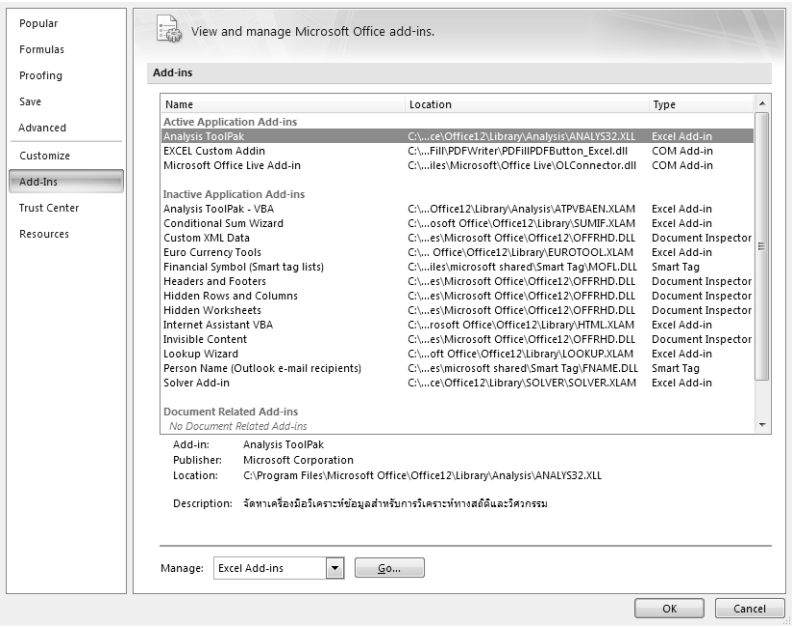


รูปที่ 3.1 (ก)

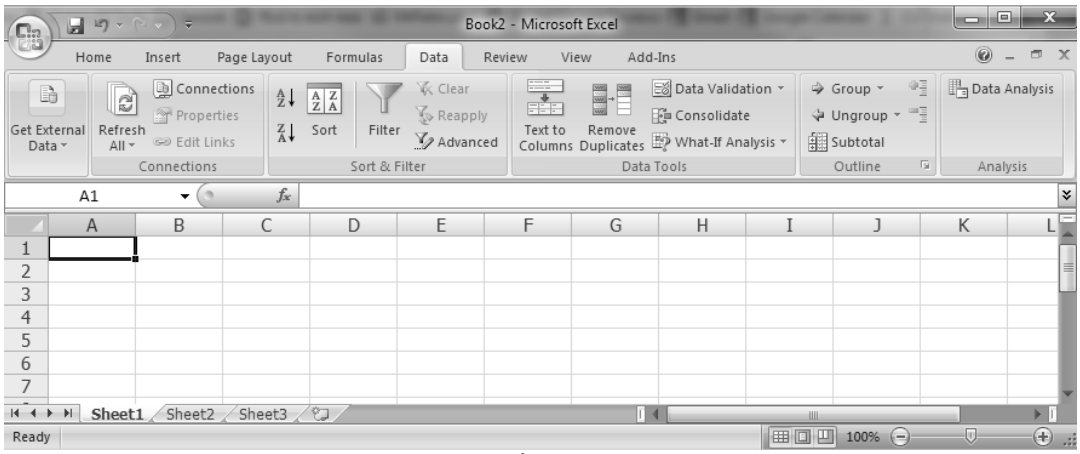


รูปที่ 3.1 (ข)

การติดตั้ง add-ins ชุดนี้ใน Microsoft Excel® 2007 ทำได้กดที่ Office Button (ที่มุมซ้ายบนของเมนู) แล้วเลือก Excel Options ซึ่งจะมีตัวเลือกดังภาพที่ 3.2 (ก) ให้เลือก Add-Ins ซึ่งเป็นตัวเลือกในช่องซ้าย แล้วเลือก Analysis Toolpack ซึ่งเป็นตัวเลือกอยู่ในช่องขวา แล้วกดที่ OK จากนั้นจึงตรวจสอบว่ามีการติดตั้ง add-ins แล้วหรือไม่ โดยเลือก Data ที่เมนู จะเห็น Data Analysis ปรากฏเพิ่มมาในตัวเลือก (ปุ่มขวาสุด) ดังรูปที่ 3.2 (ข)



รูปที่ 3.2 (ก)



รูปที่ 3.2 (ข)

การป้อนข้อมูลดิบใน Microsoft® Excel ควรป้อนข้อมูลดิบเป็นคอลัมน์หรือแถว โดยหากมีข้อมูลดิบหลายประชากร ให้ป้อนข้อมูลดิบแต่ละประชากรในแต่ละคอลัมน์หรือแถว เพื่อสามารถนำไปใช้ในการวิเคราะห์แบบอื่น ๆ ได้ด้วย

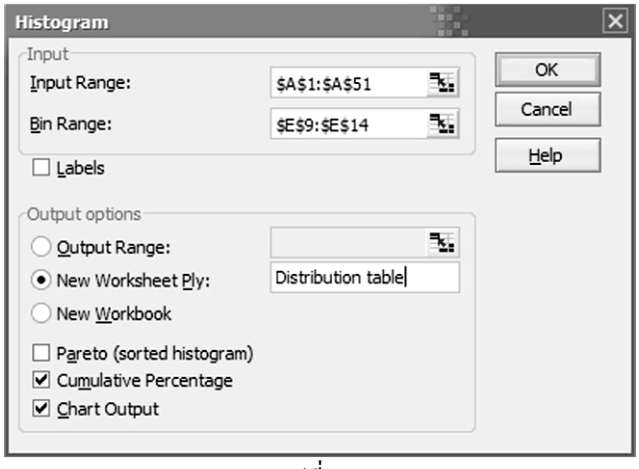
ก่อนจะทำตารางแจกแจงความถี่และ Histogram จะต้องสร้างอันตรภาคชั้น (Class interval) ขึ้นมาก่อน โดยต้องใช้ฟังก์ชั่นต่าง ๆ ได้แก่

ฟังก์ชั่น	คำอธิบาย	ค่าพารามิเตอร์
min(number1,number2,…)	หาค่าที่น้อยที่สุดของชุดข้อมูลดิบ	number1, numer2,… คือ ข้อมูลดิบ
max(number1,number2,…)	หาค่าที่มากที่สุดของชุดข้อมูลดิบ	number1, numer2,… คือ ข้อมูลดิบ

ขั้นตอนในการกำหนดอันตรภาคชั้นโดยสังเขปมีดังนี้

1. หาค่า min() และ max() แล้วหาค่าช่วงข้อมูล (range) จาก max() – min()
2. กำหนดจำนวนอันตรภาคชั้นโดยให้คิดว่าแต่ละชั้นน่าจะมีข้อมูลดิบเท่า ๆ กัน แต่ไม่ควรน้อยกว่า 10 ค่า จำนวนชั้นนี้สามารถปรับเปลี่ยนได้
3. หาค่าความกว้างของแต่ละชั้นจาก ค่าช่วงข้อมูล ÷ จำนวนชั้น (ปรับค่าตามความเหมาะสม)
4. กำหนดค่าต่ำสุดและค่าสูงสุดของแต่ละชั้น โดยให้ค่าต่ำสุดของแต่ละชั้นอยู่คอลัมน์ด้านซ้าย และค่าสูงสุดของแต่ละชั้นอยู่คอลัมน์ด้านขวา
    a. **หมายเหตุ** ทั้งนี้ข้อมูลดิบค่าที่ต่ำที่สุดจะต้องอยู่ในอันตรภาคชั้นที่ 1 และข้อมูลดิบค่าที่สูงที่สุดจะต้องอยู่ในอันตรภาคชั้นสุดท้าย หากไม่เป็นเช่นนี้ ให้เพิ่มจำนวนชั้นหรือขยายความกว้างของชั้นออกไปอีก
5. คอลัมน์ที่มีค่าสูงสุดของแต่ละชั้นนี้ จะใช้ในการแจกแจงความถี่ อนึ่ง Microsoft® Excel จะเรียกแต่ละอันตรภาคชั้นว่า Bin

การสร้าง ตารางแจกแจงความถี่และ Histogram ทำโดยเลือก Tools, Data Analysis... จะมี dialog box ดังรูปที่ 3.3



รูปที่ 3.3

รายละเอียดที่ต้องเพิ่มใน dialog box นี้ประกอบด้วย

ช่องที่ต้องเติม	คำอธิบาย
Input	
Input Range	ข้อมูลดิบ – ต้องระบุ
Bin Range	Class interval ○ หากไม่ระบุ โปรแกรมจะสร้าง class interval ให้ ○ หากระบุค่า ให้เรียงค่าจากน้อยไปหามาก (จากบนลงล่าง) โดยทั่วไป มักระบุค่า upper bound limit ของแต่ละ class interval

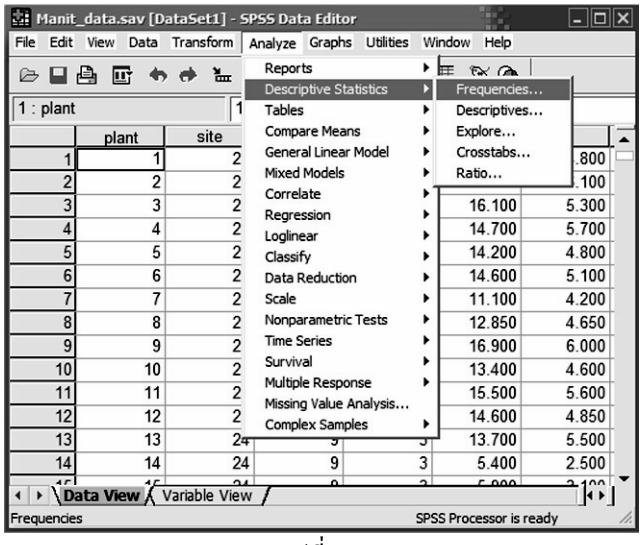
ช่องที่ต้องเติม	คำอธิบาย
Label	ถ้าเลือก (มีเครื่องหมายถูก) หมายความว่า แถวแรก (หรือคอลัมน์แรก) ของข้อมูลดิบเป็นข้อความ
Output Options	
○ Output Range	ระบุผลใน worksheet เดียวกัน หรือใน worksheet ใหม่ หรือใน workbook ใหม่
○ New Worksheet Ply	
○ New Workbook	
Pareto	เรียงผลโดยใช้ค่าความถี่ แทนที่จะเรียงตามค่า Bin ที่กำหนดให้
Cumulative Percentage	แสดงผลเปอร์เซ็นต์ความถี่สะสม
Chart Output	สร้างกราฟ Histogram

2. การสร้างตารางแจกแจงความถี่และ Histogram ด้วยโปรแกรม SPSS

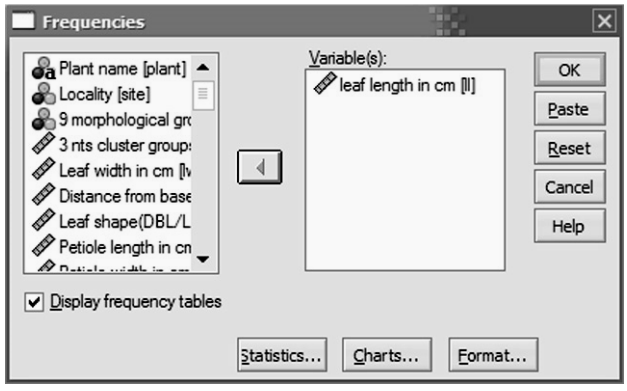
การสร้างตารางแจกแจงความถี่ในโปรแกรม SPSS นั้น ไม่สามารถกำหนดอันตรภาคชั้นได้ดังเช่นโปรแกรม Microsoft® Excel แต่โปรแกรม SPSS สามารถสร้าง Histogram ให้ได้อย่างเหมาะสมและดูดีกว่า อีกทั้งยังมีค่าสถิติอื่น ๆ ที่อาจได้มาพร้อมกันกับการสร้างตารางแจกแจงความถี่และ Histogram ดังนั้น จึงควรรู้จักวิธีการใน SPSS เพื่อการนี้บ้าง

ในตารางแจกแจงความถี่และ Histogram นั้น จะทำดังนี้

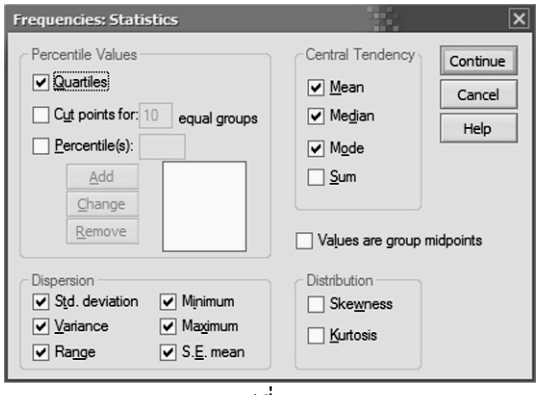
1. เลือก Analyze, Descriptive Statistics, Frequencies... (รูปที่ 3.4)
2. ใน dialog box (รูปที่ 3.5)
   a. เลือกตัวแปรในรายการด้านซ้าย กดปุ่ม ▶ ของช่อง Variable(s):
   b. กดปุ่ม Statistics... มี dialog box ใหม่ (รูปที่ 3.6) เลือกค่าสถิติต่าง ๆ แล้วกดปุ่ม Continue
   c. กดปุ่ม Charts... มี dialog box ใหม่ (รูปที่ 3.7) เลือกกราฟ Histogram แล้วกดปุ่ม Continue
3. กดปุ่ม OK จะมีผลการวิเคราะห์เป็นตารางค่าสถิติ (รูปที่ 3.8) ตารางแจกแจงความถี่ และกราฟ (รูปที่ 3.9)



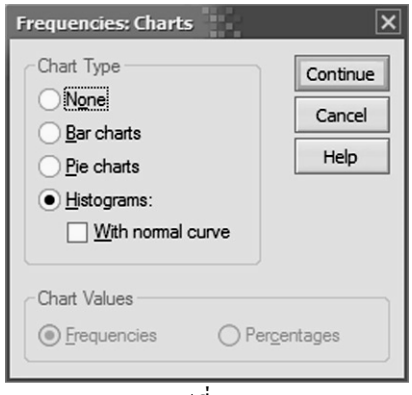
รูปที่ 3.4



รูปที่ 3.5



รูปที่ 3.6



รูปที่ 3.7

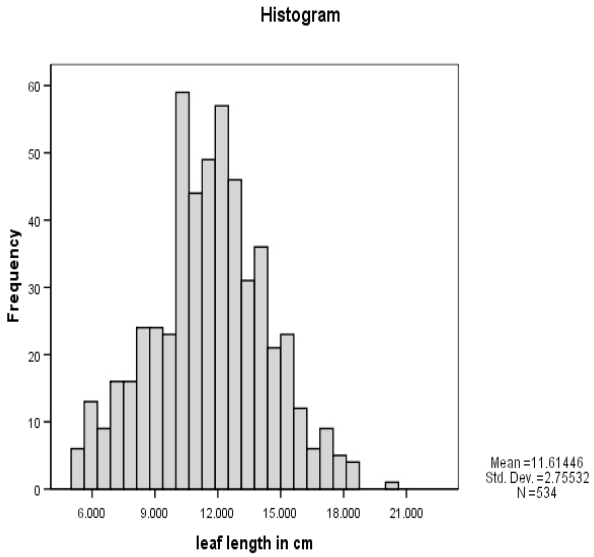
Statistics

leaf length in cm

N	Valid	53
	M ss ng	0
Mean		11.61446
Std. Error of Mean		.119234
Median		11.62500
Mode		11.500(a)
Std. Deviation		2.755320
Variance		7.592
Range		15.400
Minimum		5.000
Maximum		20.400
Percentiles	25	10.00000
	50	11.62500
	75	13.41250

a Multiple modes exist. The smallest value is shown

รูปที่ 3.8



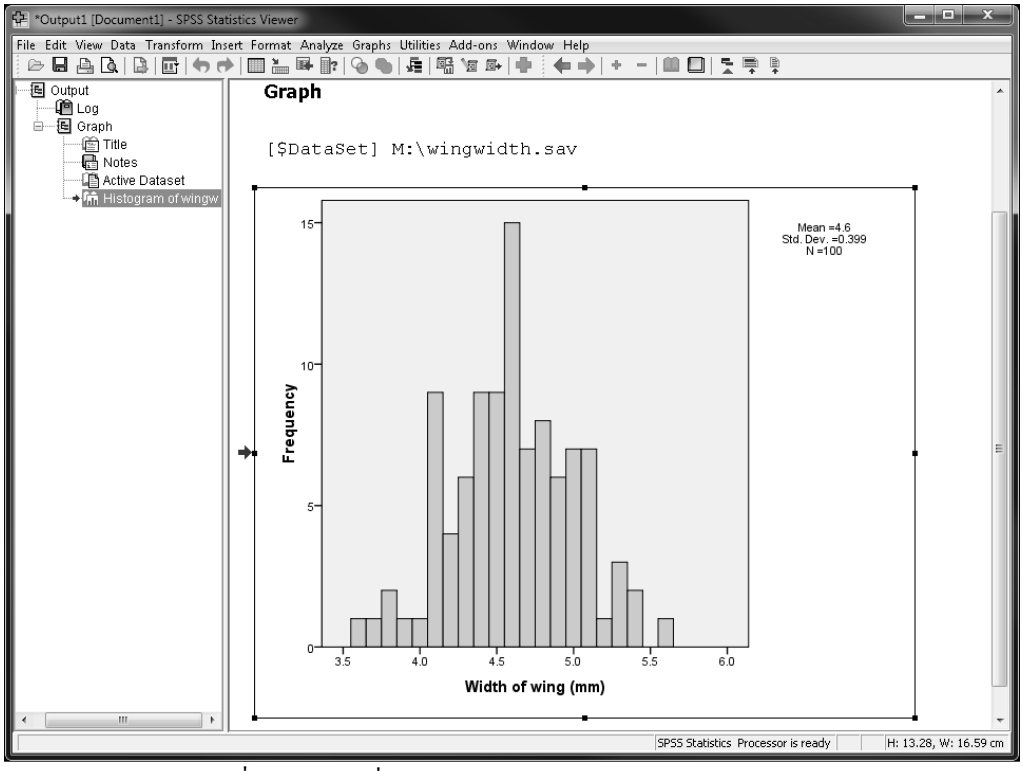
รูปที่ 3.9

การปรับแก้ Histogram ใน SPSS

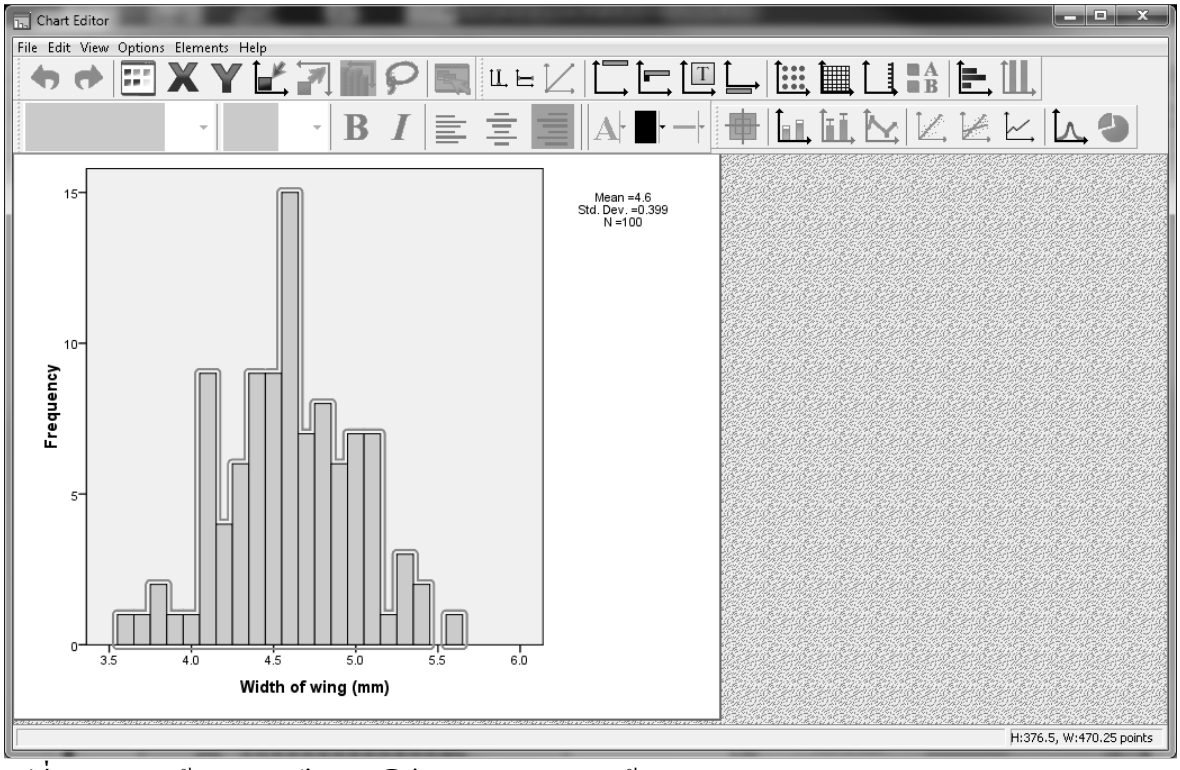
เมื่อนิสิตสร้าง Histogram ใน SPSS จะได้ตารางความถี่และกราฟ Histogram (รูปที่ 3.9-ผ1) เมื่อพิจารณากราฟดังกล่าวโดยละเอียดจะเห็นว่าเป็นกราฟแสดงความถี่ของค่าข้อมูลแต่ละค่า ไม่ได้เป็นอันตรภาคชั้นแต่อย่างใด อีกทั้งมีจำนวน “อันตรภาคชั้น” มากเกินไป ดังนั้น หากต้องการกำหนดจำนวนอันตรภาคชั้น ความกว้างของอันตรภาคชั้น และค่าเริ่มต้นของอันตรภาคชั้นที่ 1 สามารถทำได้โดยการปรับแก้กราฟ ดังขั้นตอนต่อไปนี้

1. ในหน้าต่าง **Output** เลือกกราฟ Histogram ที่ต้องการแก้ไข แล้วกดแป้นซ้ายของเมาส์ 2 ครั้งที่กราฟนี้ จะมีหน้าต่าง **Chart Edit** ปรากฏขึ้นมา (รูปที่ 3.9-ผ2) กดแป้นซ้ายของเมาส์ที่แท่งกราฟ 1 ครั้ง เพื่อเลือกของกราฟ สังเกตว่าจะมีกรอบสีปรากฏรอบกลุ่มแท่งกราฟ
2. กดแป้นซ้ายของเมาส์ 2 ครั้งที่กลุ่มแท่งกราฟเพื่อเรียกหน้าต่าง **Properties** ขึ้นมา (รูปที่ 3.9-ผ3) สังเกตว่าหน้าต่างนี้มี 4 แถบคือ

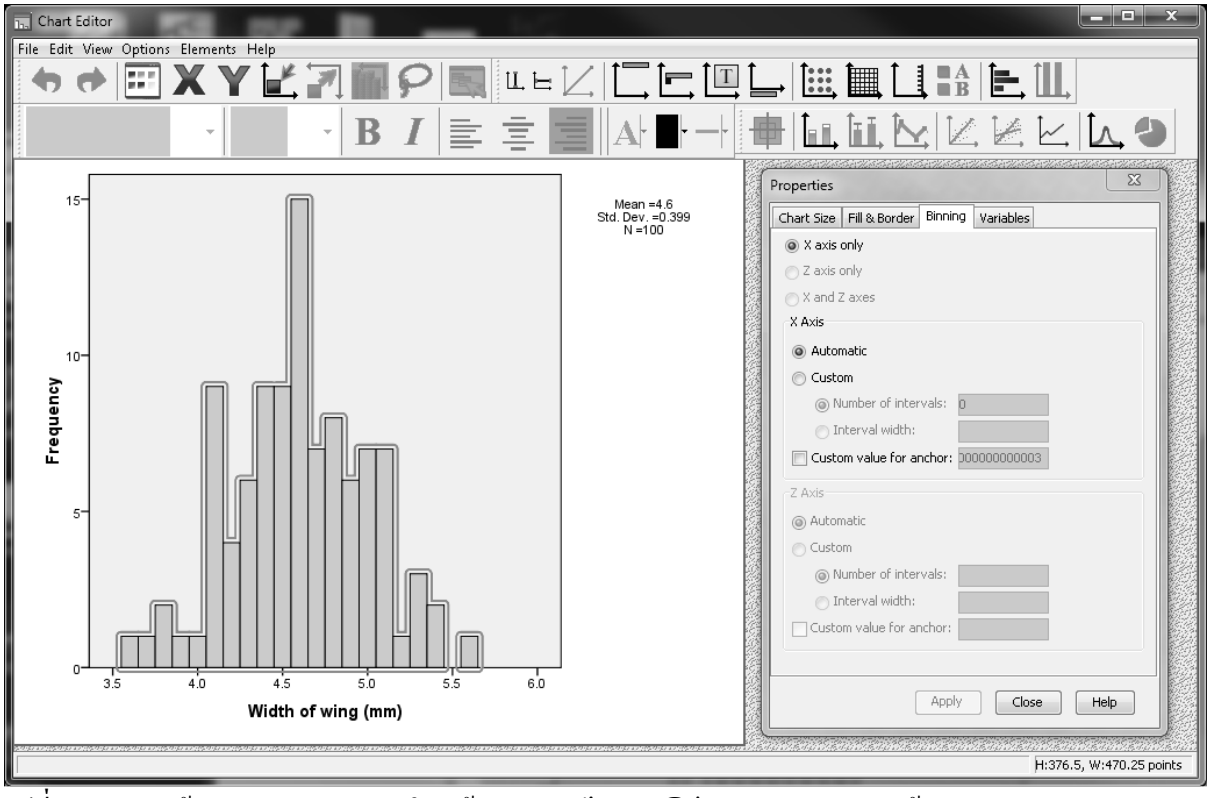
   [Chart Size] [Fill & Border] [Binning] [Variables]
3. การปรับเปลี่ยนกราฟ histogram นั้นทำโดยเลือกแถบ **Binning** เพื่อปรับแก้ค่า จากนั้นจึงเลือกว่าจะแก้ไขจำนวนอันตรภาคชั้นหรือความกว้างของอันตรภาคชั้น (เลือกแก้ได้อย่างใดอย่างหนึ่ง) ดังนี้
   a. *กำหนดจำนวนอันตรภาคชั้น* – ที่กรอบ **X Axis** เลือก **Custom** แล้วเลือก **Number of intervals:** ในตัวอย่างนี้ใส่ค่า 12 จากนั้นกดที่ปุ่ม **Apply** สังเกตว่ากราฟ Histogram มีจำนวนแท่งกราฟ 12 แท่ง และค่าของแต่ละแท่งได้เปลี่ยนไปด้วย (รูปที่ 3.9-ผ4)
   b. *กำหนดความกว้างของอันตรภาคชั้น* – ที่กรอบ **X Axis** เลือก **Custom** แล้วเลือก **Interval width:** ในตัวอย่างนี้ใส่ค่า 0.18 จากนั้นกดที่ปุ่ม **Apply** สังเกตว่ากราฟ Histogram มีจำนวนแท่งกราฟ 12 แท่ง และค่าของแต่ละแท่งได้เปลี่ยนไปด้วย (รูปที่ 3.9-ผ5) เปรียบเทียบกับกราฟในรูปที่ 3.9-ผ4 สังเกตว่าค่าของแต่ละแท่งกราฟและค่าเริ่มต้นต่างกันอย่างไร
4. นอกจากนี้ ยังสามารถกำหนดค่าเริ่มต้นของอันตรภาคชั้นที่ 1 ได้โดยเลือกที่ ☐ **Custom value for anchor:** ในตัวอย่างนี้ใส่ค่า 3.55 จากนั้นกดที่ปุ่ม **Apply** สังเกตว่ากราฟ Histogram มีจำนวนแท่งกราฟ 12 แท่ง และค่าของแต่ละแท่งได้เปลี่ยนไปด้วย (รูปที่ 3.9-ผ6) เปรียบเทียบกับกราฟในรูปที่ 3.9-ผ4 และ ผ-5 สังเกตว่าค่าของแต่ละแท่งกราฟและค่าเริ่มต้นต่างกันอย่างไร
5. เมื่อแก้ไขเสร็จแล้ว เลือก **File, Close** เพื่อปิดหน้าต่าง **Chart Editor** และกลับมาที่หน้าต่าง **Output** สังเกตว่ามีกราฟ Histogram ที่ปรับแก้แล้วมาแทนกราฟเดิม (รูปที่ 3.9-ผ7) เปรียบกับกราฟในรูปที่ 3.9-ผ1



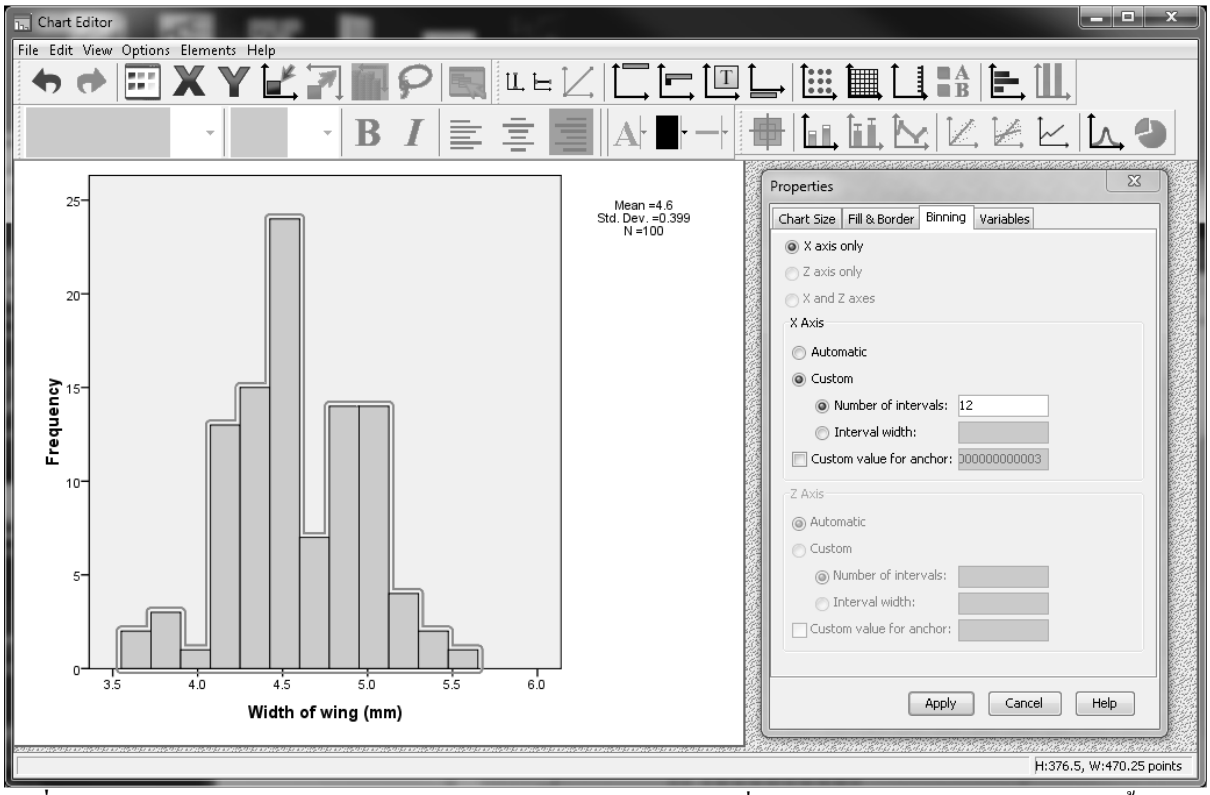
รูปที่ 3.9-ผ1 กราฟ Histogram ที่สร้างจากคำสั่ง Graphs, Legacy Dialogs, Histogram... หรือจากคำสั่ง Analyze, Descriptive Statistics, Frequencies... หรือจากคำสั่ง Analyze, Descriptive Statistics, Explore...



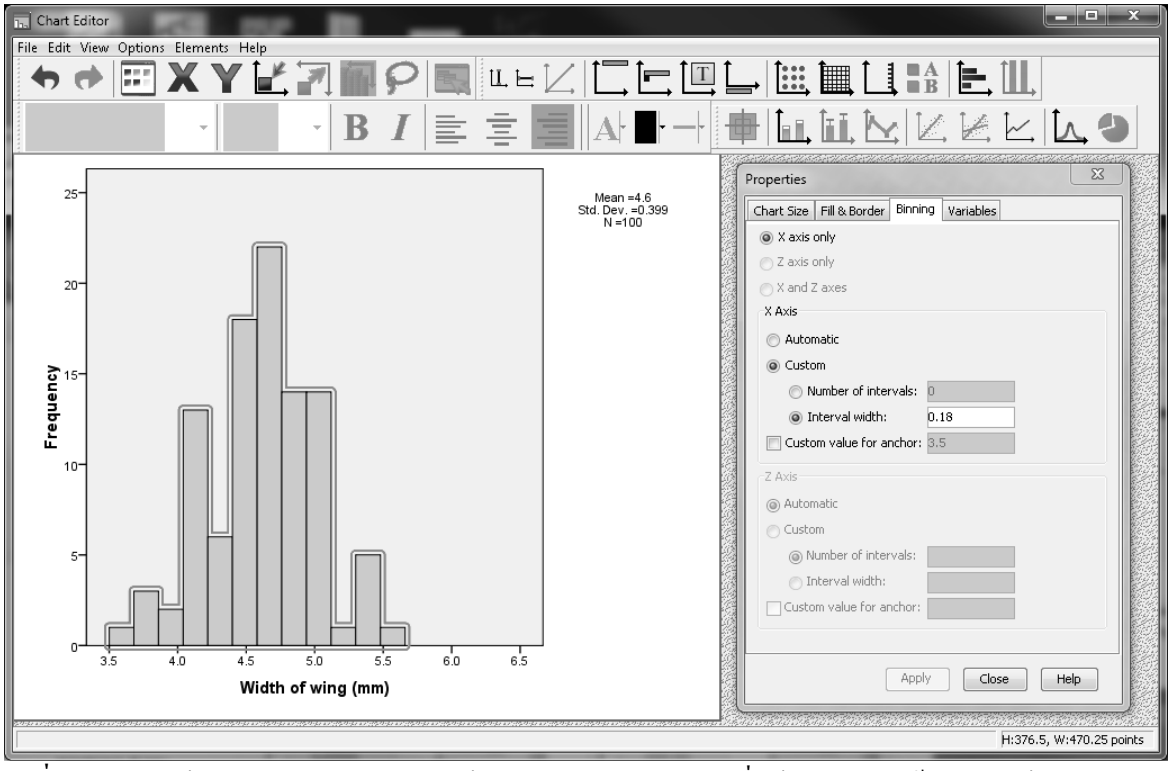
รูปที่ 3.9-ผ2 หน้าต่างของ Chart Edit สำหรับการปรับแก้ค่าต่าง ๆ ของกราฟ



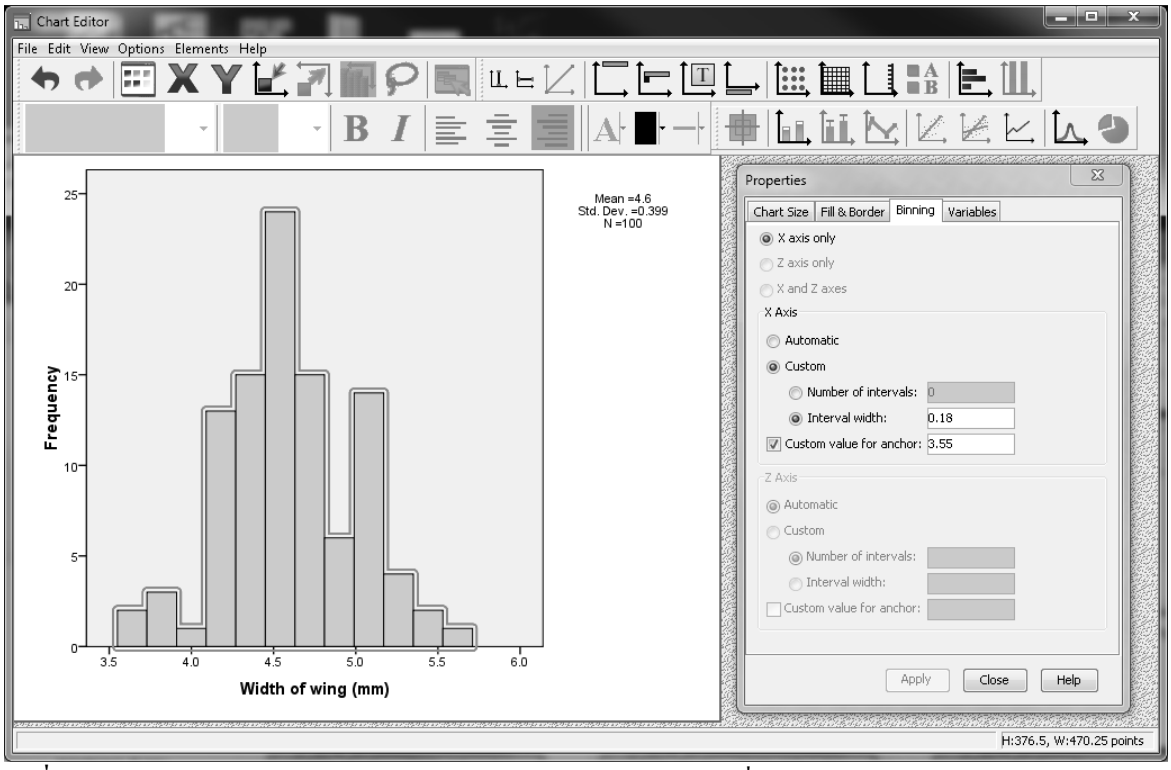
รูปที่ 3.9-ผ3 หน้าต่าง Properties ในหน้าต่างของ Chart Edit สำหรับการปรับแก้ค่าต่าง ๆ ของกราฟ



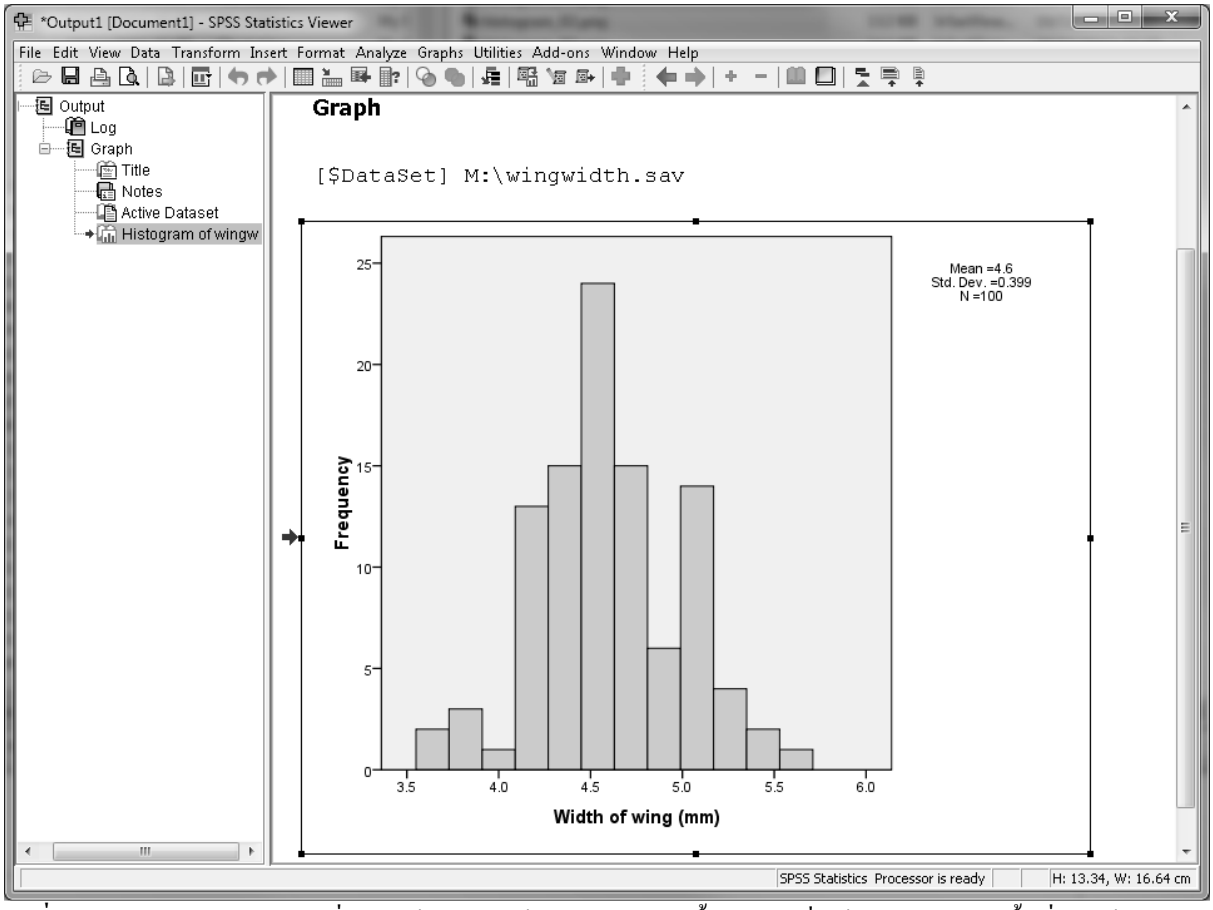
รูปที่ 3.9-ผ4 หน้าต่าง Properties ในหน้าต่างของ Chart Edit เมื่อต้องการปรับแก้ค่าจำนวนอันตรภาคชั้นของกราฟ



รูปที่ 3.9-ผ5 หน้าต่าง **Properties** ในหน้าต่างของ **Chart Edit** เมื่อต้องการปรับแก้ค่าความกว้างของอันตรภาคชั้นของกราฟ



รูปที่ 3.9-ผ6 หน้าต่าง **Properties** ในหน้าต่างของ **Chart Edit** เมื่อต้องการปรับแก้ค่าความกว้างของอันตรภาคชั้น และค่าเริ่มต้นของอัตรภาคชั้นที่ 1 ของกราฟ



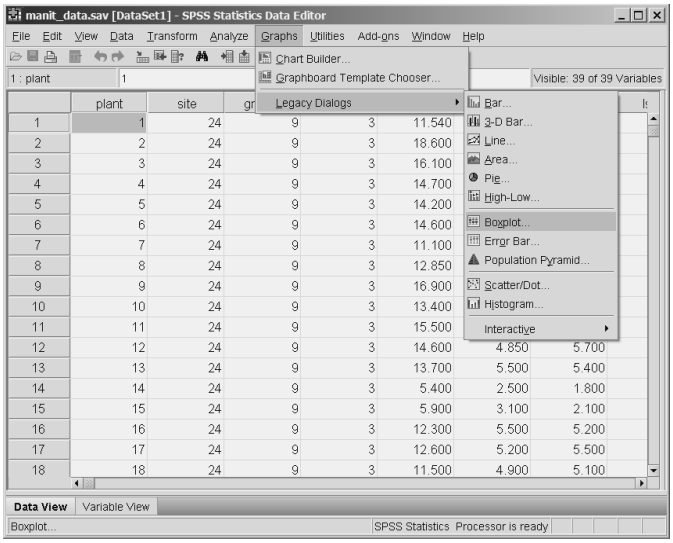
รูปที่ 3.9-ผ7 กราฟ Histogram ที่ปรับแก้ค่าความกว้างของอันตรภาคชั้น และค่าเริ่มต้นของอันตรภาคชั้นที่ 1 แล้ว

3. การนำเสนอข้อมูลแบบกราฟ

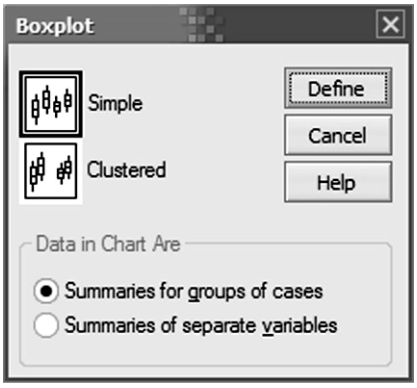
การนำเสนอข้อมูล (หรือผลการวิเคราะห์) เป็นกราฟ มักจะช่วยให้เข้าใจได้ง่ายขึ้น ในหัวข้อที่แล้วได้กล่าวถึง Histogram ซึ่งเป็นการนำเสนอในรูปแบบกราฟอย่างหนึ่ง ในที่นี้จะขอกล่าวถึงกราฟอีก 2 แบบ คือ Box plot และ Stem-and-leaf plot โดยโปรแกรมที่จะใช้ในการสร้างกราฟทั้งสองแบบคือ SPSS

3.1) Box plot เป็นกราฟที่แสดงการกระจายตัวของค่าสังเกตได้อย่างรวดเร็ว ช่วยให้สามารถทราบถึงค่าสังเกตที่เป็น outlier ในกรณีที่มีกลุ่มตัวอย่างหลายกลุ่ม ก็สามารถใช้กราฟ Box plot เปรียบเทียบการกระจายของข้อมูลระหว่างกลุ่มได้ ในการสร้างกราฟแบบนี้ ทำได้ดังนี้

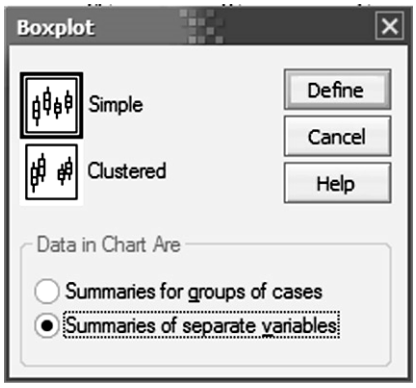
1. เลือก Graph, Legacy Dialogs, Boxplot... (รูปที่ 3.10)
2. ใน dialog box (รูปที่ 3.11 ก) เลือก Simple และใน Data in Chart Area ให้เลือกโดยพิจารณาดังนี้
   a. หากต้องการดูการกระจายข้อมูลของตัวแปร 1 ตัวแปร แยกตามกลุ่มของข้อมูลที่มีในตัวแปรนี้ ให้เลือก Summaries for Groups of Cases (รูปที่ 3.11 ก) แล้วกดปุ่ม Define
      i. เลือกตัวแปรจากรายการในช่องด้านซ้าย กดปุ่ม ▶ ของช่อง Variable:
      ii. เลือกตัวแปรระบุกลุ่มจากรายการในช่องด้านซ้าย กดปุ่ม ▶ ของช่อง Category Axis: (รูปที่ 3.12 ก)
      iii. กดปุ่ม OK จะได้กราฟดังรูปที่ 3.13 ก
   b. หากต้องการดูการกระจายข้อมูลทั้งหมดของตัวแปร 1 ตัวแปรหรือหลายตัวแปรในคราวเดียวกัน (ไม่มีกลุ่มข้อมูลในตัวแปร) ให้เลือก Summaries for Separate Variables (รูปที่ 3.11 ข) แล้วกดปุ่ม Define
      i. เลือกตัวแปรจากรายการในช่องด้านซ้าย กดปุ่ม ▶ ของช่อง Boxes Represent: (รูปที่ 3.12 ข)
      ii. กดปุ่ม OK จะได้กราฟดังรูปที่ 3.13 ข



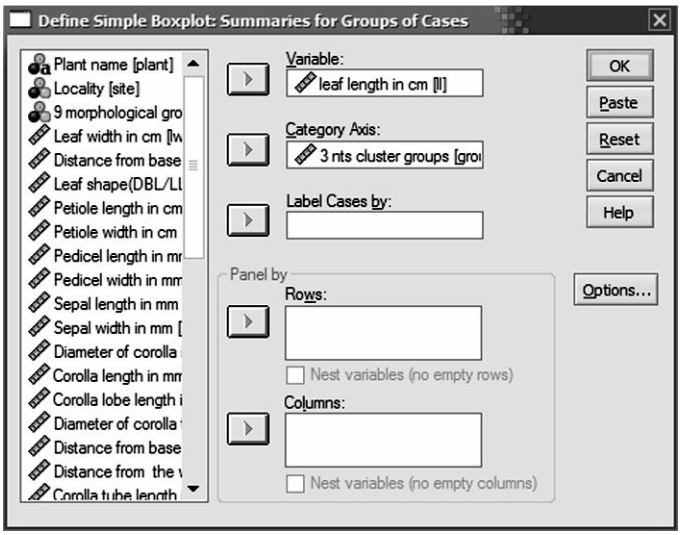
รูปที่ 3.10



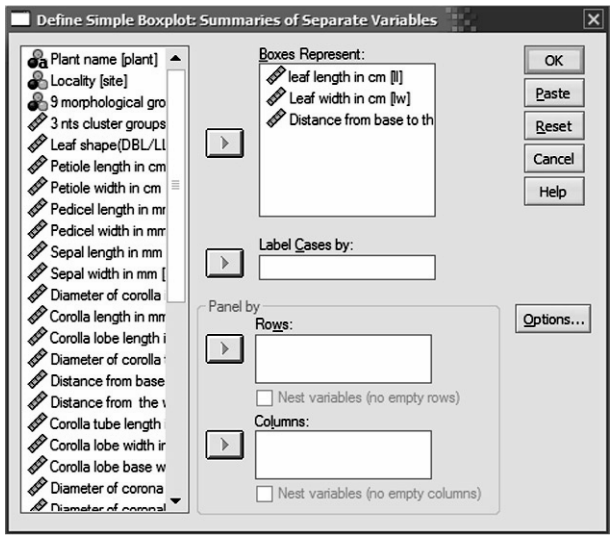
รูปที่ 3.11 ก



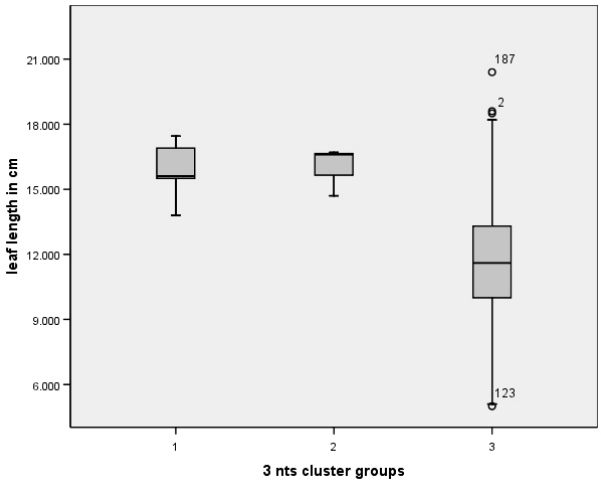
รูปที่ 3.11 ข



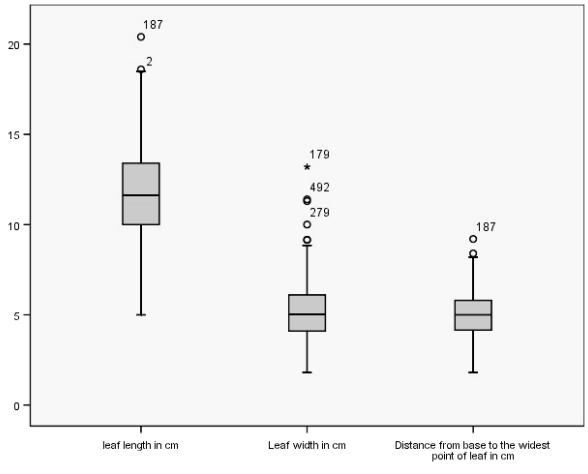
รูปที่ 3.12 ก



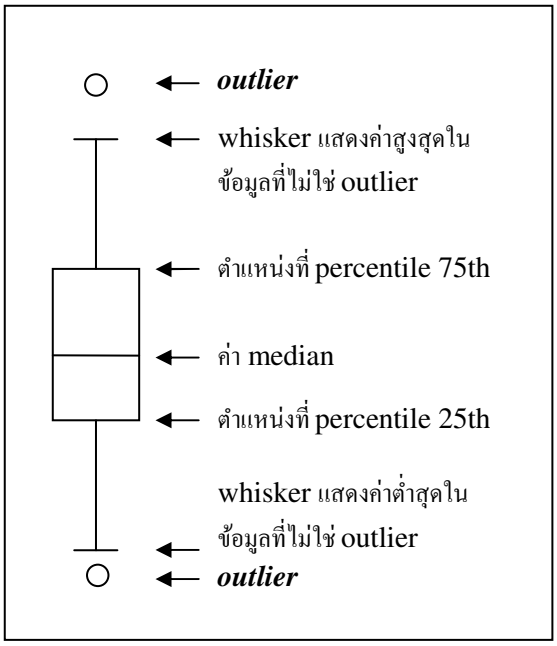
รูปที่ 3.12 ข



รูปที่ 3.13 ก



รูปที่ 3.13 ข



<u>รายละเอียดและข้อพึงทราบเกี่ยวกับ boxplot เป็นดังนี้</u>

- Box แสดงช่วงข้อมูลระหว่าง P75 และ P25 เรียกว่า interquartile range (IQR, หรือมีค่าเท่ากับ P75-P25)
- ถ้าข้อมูลมีการกระจายแบบปกติ ค่า median จะอยู่กึ่งกลางระหว่าง P75 และ P25
- เส้นในแนวนอนเหนือ P75 และใต้ P25 เรียกว่า whisker แสดงข้อมูลที่มีค่าสูงสุดและค่าต่ำสุดที่ไม่ใช่ outlier ตามลำดับ และจะลากเส้นเชื่อม whisker กับ box
- ข้อมูลที่จัดเป็น outlier คือข้อมูลที่มีค่าน้อยกว่าค่า P25 เกิน 1.5 เท่าของ IQR หรือมีค่ามากกว่า P75 เกิน 1.5 เท่าของ IQR
- ข้อมูลที่มีค่าน้อยกว่า P25 ระหว่าง 1.5 – 3 เท่าของ IQR หรือข้อมูลที่มีค่ามากกว่า P75 ระหว่าง 1.5 – 3 เท่าของ IQR จัดเป็น mild outlier โดยทั่วไปให้สัญญลักษณ์ o และหากว่าในข้อมูลมีค่าที่น้อยกว่า P25 เกิน 3 เท่าของ IQR หรือมีค่ามากกว่า P75 เกิน 3 เท่าของ IQR จัดเป็น extreme outlier จะให้สัญญลักษณ์ที่แตกต่างไป เช่น * เป็นต้น
- ค่า outlier คือ ค่าที่ตกอยู่นอกเขตการกระจายของข้อมูลส่วนใหญ่ มักเป็นค่าที่สูงหรือต่ำเกินไป และควรพิจารณาเป็นพิเศษ
- ควรพิจารณา Histogram ของข้อมูลควบคู่ไปด้วย

3.2) Stem-and-leaf plot เป็นกราฟที่สามารถสร้างได้อย่างรวดเร็วโดยไม่จำเป็นต้องใช้เครื่องคอมพิวเตอร์หรือโปรแกรมใด ๆ โดยหลักการง่าย ๆ ดังนี้

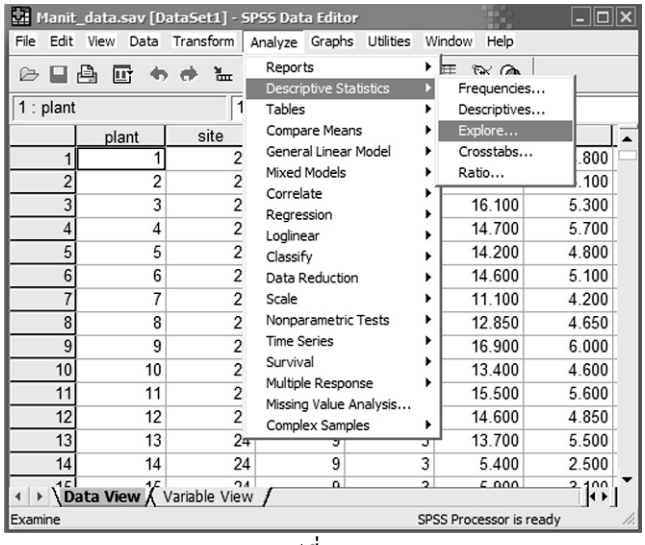
1. ข้อมูลแต่ละค่า กำหนดให้ตัวเลขด้านซ้ายของข้อมูลเป็น stem และตัวเลขด้านขวาเป็น leaf เช่น ถ้าหากว่าเป็นเลข 2 หลัก หลักสิบเป็น stem ส่วนหลักหน่วยเป็น leaf แต่หากเป็นเลข 3 หลัก ให้เอา 2 หลักซ้ายเป็น stem และหลักขวาสุด 1 หลัก เป็น leaf
2. เมื่อได้ stem แล้ว ให้เรียง stem ในแนวตั้งจากค่าน้อยลงมาหาค่ามาก (อาจทำตาราง 2 คอลัมน์ ด้านซ้ายใส่ค่า stem)
3. ที่ stem ค่าเดียวกัน ให้เอา leaf มาใส่
4. ทำจนครบทุกค่าของข้อมูล จะได้ตารางที่ดูคล้ายกับกราฟ Histogram ตะแคงข้าง

ข้อดีของ stem-and-leaf plot คือ นอกจากจะแสดงการกระจายตัวของข้อมูลคล้ายกับ Histogram แล้ว ยังสามารถทราบได้ว่าในแต่ stem มีข้อมูลดิบอะไรบ้าง ซึ่ง Histogram ไม่สามารถบอกได้ (บอกได้เฉพาะจำนวนข้อมูล) การทำ Stem-and-leaf plot ใน SPSS นั้น มีขั้นตอนดังนี้

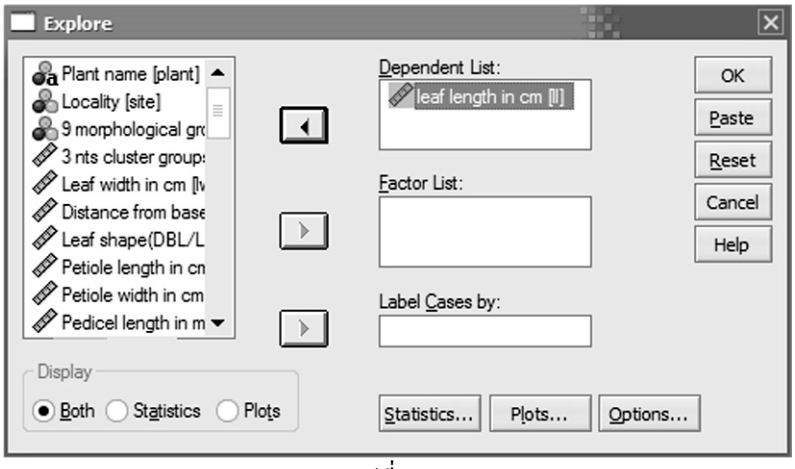
1. เลือก Analyze, Descriptive Statistics, Explore... (รูปที่ 3.14)
2. ใน dialog box (รูปที่ 3.15)
    a. เลือกตัวแปรจากรายการด้านซ้าย กดปุ่ม ▶ ของช่อง Dependent List:
    b. ที่ช่อง Display เลือก Both
    c. กดปุ่ม Statistics... ใน dialog box ใหม่ (รูปที่ 3.16) เลือกค่าสถิติต่าง ๆ แล้วกดปุ่ม Continue

d. กดปุ่ม **Plot**... ใน dialog box ใหม่ (รูปที่ 3.17) ที่ช่อง **Descriptive** เลือกกราฟ **Stem-and-leaf** แล้วกดปุ่ม **Continue**

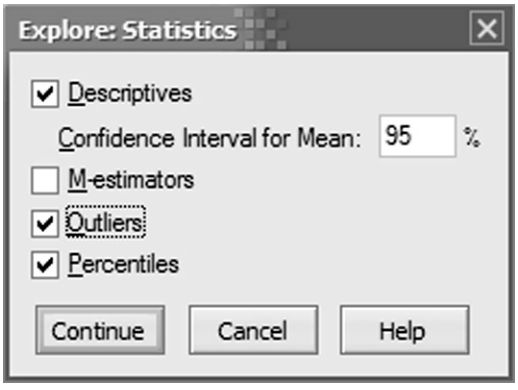
3. กดปุ่ม OK จะได้ผลการวิเคราะห์เป็นตารางค่าสถิติต่าง ๆ และกราฟ Stem-and-leaf (รูปที่ 3.18)



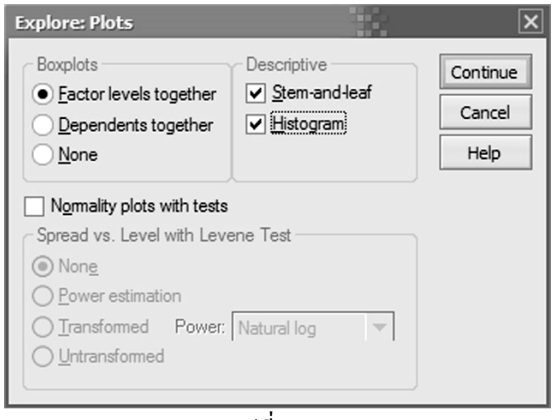
รูปที่ 3.14



รูปที่ 3.15



รูปที่ 3.16



รูปที่ 3.17

```
leaf length in cm Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

    13.00        5 .  0144567779999
    18.00        6 .  001222555666688999
    24.00        7 .  000011112233356677888999
    37.00        8 .  00111222334455555666667777777778899999
    39.00        9 .  000001111122223344455556666678888899999
    73.00       10 .  0000001111111111122222222222333333334444455555555555666666677777888889999
    86.00       11 .  00000000001111111122222222222233333334444444455555555555555666666777777888888888899999999
    78.00       12 .  000000000000011111111112222222222222333333334444445555555566666677888888899999999
    60.00       13 .  00000000000011111122333333344444445555555555667777888888999999
    46.00       14 .  000000011111111222223333333555566666666777888899
    29.00       15 .  000111123333445555555666777899
    15.00       16 .  111222666777999
    11.00       17 .  01124455579
     3.00       18 .  225
     2.00 Extremes    (>=18.6)

 Stem width:     1.000
 Each leaf:       1 case(s)
```

รูปที่ 3.18

4. การหาค่า Percentile และ Percentile Rank

ในการวิเคราะห์หาค่า Percentile และ Percentile Rank นั้น หากมีข้อมูลดิบอยู่ สามารถใช้โปรแกรม Microsoft® Excel หรือ SPSS วิเคราะห์ให้ได้ โดยในโปรแกรม Excel จะมีโมดุลชื่อ Rank and Percentile ใน Data Analysis... หรือจะใช้ฟังก์ชั่น percentile() และ percentrank() ก็ได้ ส่วนโปรแกรม SPSS จะอยู่ใน Statistics... ของ Analyze, Descriptive Statistics, Frequencies... ทั้งนี้ จะหาค่าได้เฉพาะ percentile เท่านั้น

ในกรณีที่ไม่มีข้อมูลดิบ จะมีเพียงโปรแกรม Microsoft® Excel ที่พอจะใช้หาค่า Percentile และ Percentile Rank ได้ ทั้งนี้ จะต้องอยู่บนพื้นฐานที่ว่าข้อมูลดิบนี้มาจากประชากรที่มีการกระจายแบบปกติ มิฉะนั้น ค่าที่คำนวณมาได้จะผิดจากค่าจริงมาก (ไม่อาจบอกได้ว่ามากเท่าใด) โดยใช้ฟังก์ชั่น ดังนี้

ฟังก์ชั่น	คำอธิบาย	ค่าพารามิเตอร์
normdist(x,mean,sd,TRUE)	หาค่า percentile ที่มีค่าต่ำกว่า x เมื่อทราบเพียงค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน	x : ค่าข้อมูลดิบ mean : ค่าเฉลี่ย sd : ค่าเบี่ยงเบนมาตรฐาน
norminv(k-th,mean,sd)	หาค่า x ที่ k-th percentile เมื่อทราบเพียงค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน	k-th : percentile rank มีค่าระหว่าง 0-1
<u>กรณีที่มีข้อมูลดิบ</u>		
percentile(array,k)	หาค่าข้อมูล (x) ที่ percentile k-th	array : ข้อมูลดิบ k : percentile มีค่าระหว่าง 0-1
percentrank(array,x,[sign.])	หาค่า percentile rank ของ x	x : ค่าข้อมูลดิบ [sign.] : ไม่ต้องระบุค่า

แบบฝึกหัดที่ 3

1. อัตราการตายจากโรคหัวใจ (ต่อประชากร 100,000 คน) ในมลรัฐทั้งหมด 50 มลรัฐ และเมืองหลวงของประเทศสหรัฐอเมริกาเป็นดังนี้

445.4	417.4	377.7	351.6	322.4	396.8	361.4	332.7	310.6	265.4
442.7	414.4	376.2	349.4	319.1	395.2	358.2	330.2	306.8	236.9
427.3	407.6	369.8	340.3	315.9	394.0	356.4	329.1	300.6	214.2
422.9	406.7	369.2	338.7	311.8	391.7	353.3	328.4	297.4	194.0
420.8	396.9	368.9	334.2	311.2	391.0	351.6	327.1	274.6	169.0

ก) จงสร้างตารางแจกแจงความถี่ ความถี่สัมพัทธ์ และความถี่สะสม โดยกำหนดให้จำนวนชั้น (category) ของอัตราการตายจากโรคหัวใจ มี 6 ชั้น

ข) จงสร้าง histogram, box plot และ stem-and-leaf โดยระบุด้วยว่ามี outlier ในข้อมูลหรือไม่ คือค่าใด

ค) จงหาค่า percentile ที่ 20^{th} 40^{th} 60^{th} และ 80^{th}

ง) จงหาค่า percentile rank ของค่า 400 300 200 และ 100

2. ความดันเลือดในคนมีความแตกต่างกันขึ้นอยู่กับปัจจัยเรื่องอายุและเพศ กล่าวคือ ที่อายุ 3 ปี จะมีความดันเลือด 95 มม.ปรอท และจะเพิ่มขึ้น 1.5 มม.ปรอทต่อปี จนกระทั่งอายุ 13 ปี หลังจากนั้น ความดันเลือดจะเพิ่มขึ้น 2 มม.ปรอทต่อปี ในเพศชาย และ 1 มม.ปรอทต่อปี ในเพศหญิง จนกระทั่งอายุ 18 ปี ถ้าหากว่าความดันเลือดมีการกระจายแบบปกติโดยมีค่าเบี่ยงเบนมาตรฐาน 12 มม.ปรอท ในทุกช่วงอายุและเพศแล้ว

ก) จงหาค่าความดันเลือดที่ percentile 80^{th} ในเด็กชายอายุ 7 ปี

ข) จงหาค่าความดันเลือดที่ percentile 70^{th} ในเด็กหญิงอายุ 12 ปี

การบ้านปฏิบัติการที่ 3

1. นักวิทยาศาสตร์สนใจจำนวนไข่ของปูชนิดหนึ่งที่จะวางในช่วงฤดูผสมพันธุ์ จึงทำการเก็บข้อมูลจำนวนไข่ที่วางจากปูจำนวน 45 ตัว ดังนี้

1,959	4,534	7,020	6,725	6,964	7,428	9,359	9,166
2,802	2,462	4,000	3,378	7,343	4,189	8,973	4,327
2,412	7,624	1,548	4,801	737	5,321	849	5,749
6,837	8,639	417	6,082	10,241	962	3,894	1,801
5,099	6,627	4,484	5,633	4,148	6,588	5,847	4,632
6,472	8,372	8,225	6,142	12,130			

ก) จงสร้างตารางแจกแจงความถี่ ความถี่สัมพัทธ์ และความถี่สะสม โดยกำหนดให้จำนวนชั้น (category) ของจำนวนไข่ มี 6 ชั้น

ข) จงสร้าง histogram, box plot และ stem-and-leaf โดยระบุด้วยว่ามี outlier ในข้อมูลหรือไม่ คือค่าใด

ค) จงหาค่า percentile ที่ 15^{th} 30^{th} 55^{th} และ 85^{th}

ง) จงหาค่า percentile rank ของค่า 500 1,500 5,000 และ 10,000

2. ข้อมูลต่อไปนี้เป็นเปอร์เซ็นต์เถ้าในพีทมอสส์จากตัวอย่าง 50 ตัวอย่าง

0.5	1.8	4.0	1.0	2.0	2.6	3.1	2.5	1.7	5.0
1.1	1.6	2.3	3.5	2.2	1.3	3.0	2.7	1.2	1.5
2.0	3.8	3.0	2.3	1.8	3.2	2.4	2.5	1.9	3.1
3.6	2.4	0.8	3.4	1.4	2.4	2.8	2.7	4.5	2.1
1.9	2.3	1.2	1.9	2.3	1.5	0.7	3.7	1.8	1.7

a) จงสร้างตารางแจกแจงความถี่ ความถี่สัมพัทธ์ และความถี่สะสม โดยกำหนดให้จำนวนชั้น (category) ของเปอร์เซ็นต์เถ้ามี 6 ชั้น

b) จงเขียน histogram, box plot, และ stem-and-leaf โดยระบุด้วยว่ามี outlier ในข้อมูลหรือไม่ คือค่าใด

c) จงหา Percentile (P) และ Percentile Rank (PR) ดังนี้

   1. $P\text{-}10^{th}$, $P\text{-}25^{th}$, $P\text{-}45^{th}$, $P\text{-}80^{th}$
   2. PR-1.75, PR-3.55, PR-0.75, PR-2.25

ปฏิบัติการที่ 4 Descriptive Statistics II

ในการวิเคราะห์ทางสถิตินั้น ในเบื้องต้น มักต้องการทราบค่าพื้นฐานของกลุ่มตัวอย่างหรือประชากร ได้แก่ ค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน ค่ามัธยฐาน ค่าฐานนิยม ฯลฯ ดังนั้น จึงต้องทราบว่าจะต้องใช้ฟังก์ชั่นหรือคำสั่งใดในการประมวลค่าดังกล่าว ซึ่งทั้งโปรแกรม Microsoft® Excel และ SPSS สามารถวิเคราะห์หาค่าดังกล่าวให้ได้

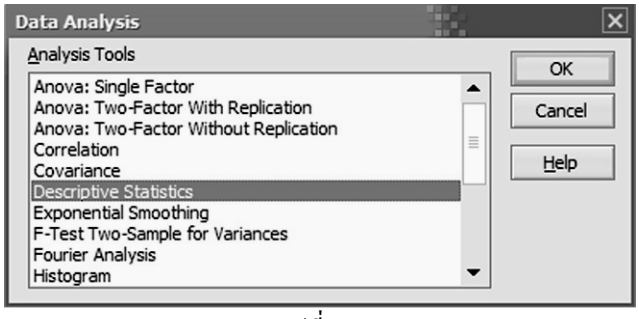
1. การใช้โปรแกรม Microsoft® Excel เพื่อหาค่า Mean, Median, Mode, Range, Standard deviation, Variance

ฟังก์ชันที่สามารถใช้หาค่าทางสถิติเบื้องต้น ได้แก่

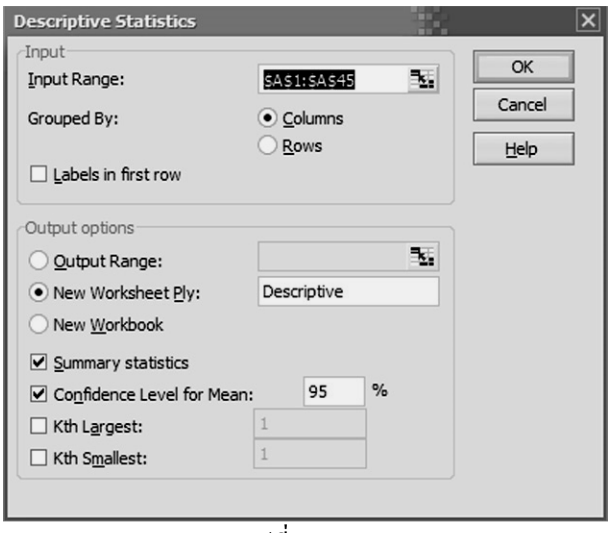
ค่าสถิติ	ฟังก์ชัน	ตัวแปรต่างๆ
Mean	average(x1, x2, …)	x1, x2, … คือค่าข้อมูลดิบ
Median	median(x1, x2, …)	x1, x2, … คือค่าข้อมูลดิบ
Mode	mode(x1, x2, …)	x1, x2, … คือค่าข้อมูลดิบ
Maximum	max(x1, x2, …)	x1, x2, … คือค่าข้อมูลดิบ
Minimum	min(x1, x2, …)	x1, x2, … คือค่าข้อมูลดิบ
Standard deviation	stdev(x1, x2, …)	x1, x2, … คือค่าข้อมูลดิบ
Variance	var(x1, x2, ...)	x1, x2, … คือค่าข้อมูลดิบ
Range	max(x1, x2, …)- min(x1, x2, …)	x1, x2, … คือค่าข้อมูลดิบ

นอกจากนี้ โปรแกรม Microsoft® Excel ยังมี Data Analysis… ซึ่งเป็นโมดุลหนึ่งใน Add-Ins ที่ช่วยในการวิเคราะห์และแสดงผลข้อมูล ในการป้อนข้อมูลเพื่อการวิเคราะห์ หากว่ามีข้อมูลเพียงประชากรเดียว ให้ป้อนข้อมูลให้อยู่คอลัมน์เดียว (หรือเพียงแถวเดียว) แต่หากว่ามีข้อมูลหลายประชากร ให้ป้อนข้อมูลแยกคอลัมน์ (หรือแถว) โดยแต่ละคอลัมน์ (หรือแถว) คือข้อมูล 1 ประชากร ในการวิเคราะห์หาค่าสถิติเบื้องต้นนั้น จะทำดังนี้

1. เลือก Tools, Data Analysis…
   - สำหรับ Excel 2007 เลือก Data, Data Analysis แทน
2. จาก dialog box เลือก Descriptive Statistics (รูปที่ 4.1)
3. ใน dialog box ใหม่ ที่ช่อง Input (รูปที่ 4.2)
   a. ระบุช่วงข้อมูลในช่อง Input Range:
   b. หากมีข้อมูลหลายชุด ให้ระบุว่าข้อมูลแต่ละชุดป้อนตามคอลัมน์หรือแถว ที่ช่อง Grouped By:
4. ใน dialog box เดียวกัน ที่ช่อง Output (รูปที่ 4.2)
   a. ให้ระบุว่าต้องการให้ผลการวิเคราะห์อยู่ใน sheet เดียวกัน หรือ sheet ใหม่ หรือ workbook ใหม่
   b. เลือก Summary Statistics เพื่อแสดงค่าสถิติเบื้องต้น หรือเลือกตัวเลือกอื่นเพิ่มเติม
5. กดปุ่ม OK จะได้ผลดังรูปที่ 4.3
6. หากว่ามีข้อผิดพลาด ดังในรูปที่ 4.3 นี้ Mode ไม่สามารถหาค่าได้ จึงมี #N/A



รูปที่ 4.1



รูปที่ 4.2

Column1	
Mean	5398.711111
Standard Error	409.4198774
Median	5633
Mode	#N/A
Standard Deviation	2746.4 2031
Sample Variance	7543108.619
Kurtosis	-0.371897288
kewness	0.076566774
Range	11713
Minimum	417
Maximum	12130
Sum	242942
Count	45
Confidence Level(95.0%)	825.1315338

รูปที่ 4.3

2. การใช้โปรแกรม SPSS เพื่อหาค่าสถิติเบื้องต้น

ในโปรแกรม SPSS นี้ ค่าสถิติเบื้องต้นสามารถหาได้จาก Analyze, Descriptive Statistics, Frequencies... หรือ Analyze, Descriptive Statistics, Descriptive... หรือ Analyze, Descriptive Statistics, Explore... ดังได้กล่าวมาในปฏิบัติการที่ 3 มาแล้ว ในบทนี้จะกล่าวถึงอีกคำสั่งหนึ่ง ซึ่งสามารถนำมาใช้วิเคราะห์หาค่าสถิติเบื้องต้น โดยปฏิบัติดังนี้

1. เลือก Analyze, Report, Case Summaries... (รูปที่ 4.4)
2. ใน dialog box (รูปที่ 4.5)
   a. เลือกตัวแปรที่ต้องการหาค่าสถิติจากช่องรายการตัวแปรด้านซ้าย กดปุ่ม ▶ ของช่อง Variables: จะมีตัวแปรที่เลือกมาปรากฏทางด้านขวาบน
   b. หากว่าต้องการวิเคราะห์หาค่าสถิติเบื้องต้นของประชากรหลาย ๆ กลุ่ม ให้เลือกตัวแปรระบุกลุ่ม แล้วกดปุ่ม ▶ ของช่อง Grouping Variable(s):

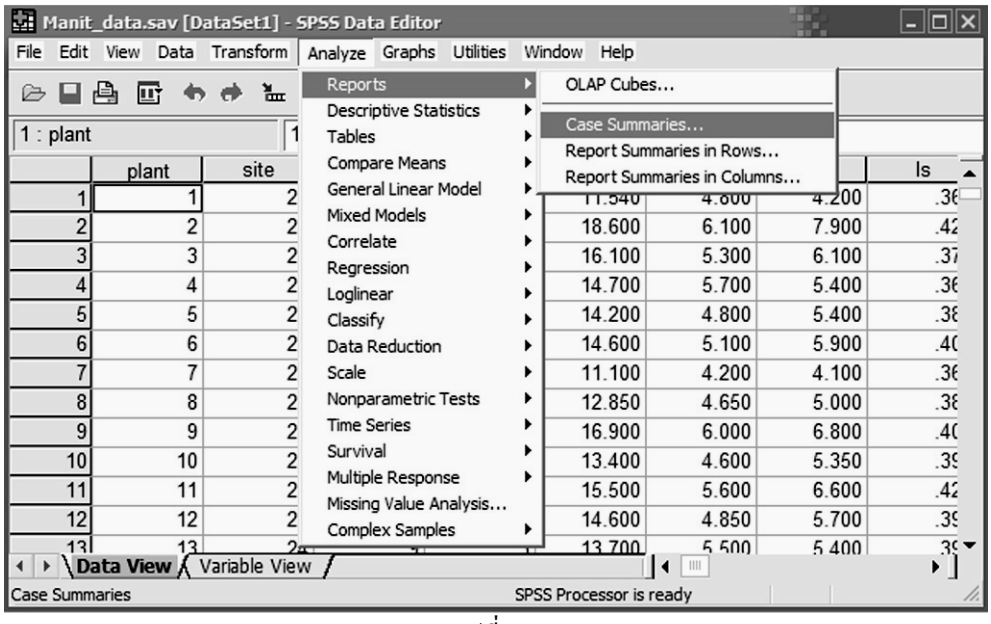
c. กดปุ่ม **Statistics...** จะมี dialog box ใหม่ (รูปที่ 4.6) เลือกค่าสถิติที่ต้องการจากรายการในช่องด้านซ้าย แล้วกดปุ่ม ▶ เมื่อเลือกได้ครบแล้ว ให้กดปุ่ม **Continue**

d. หากต้องพิมพ์ข้อความบรรยายผลการวิเคราะห์หรือจัดการกับ missing value ให้กดปุ่ม **Options...** เพื่อพิมพ์ข้อความดังกล่าวหรือเลือกไม่คิดค่า missing value

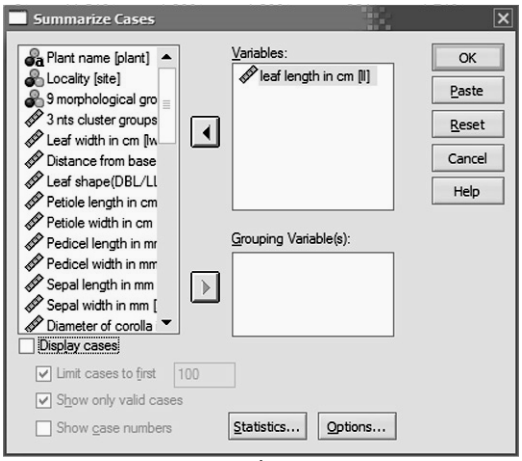
e. หากไม่ต้องการให้แสดงค่าของข้อมูลดิบ ให้เอา ✓ ออกจาก ☐ Display Cases (ดังรูปที่ 4.5)

3. เมื่อกำหนดต่าง ๆ ได้ครบถ้วน ให้กดปุ่ม **OK** (ใน dialog box รูปที่ 4.5)

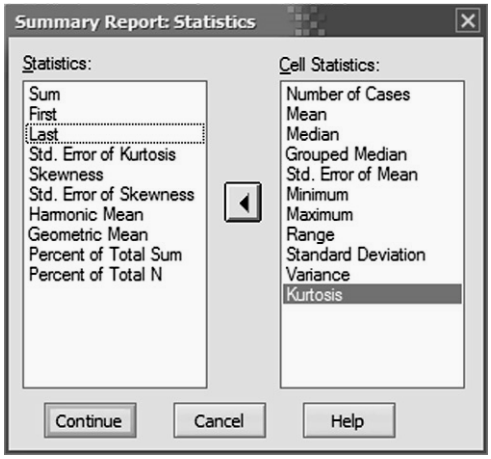
4. ผลการวิเคราะห์แสดงดังรูปที่ 4.7



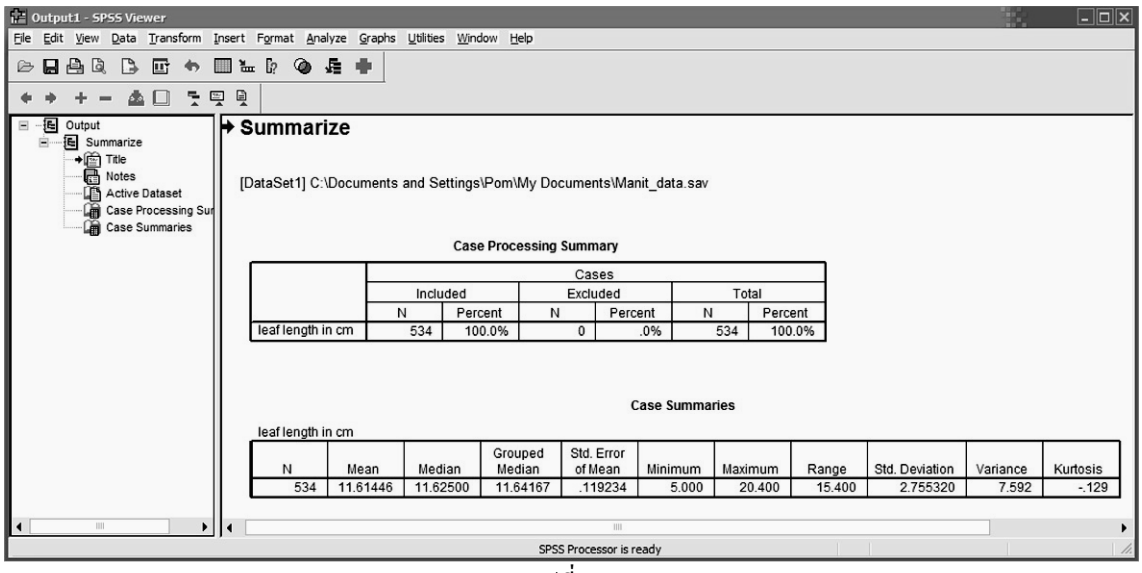
รูปที่ 4.4



รูปที่ 4.5



รูปที่ 4.6



รูปที่ 4.7

แบบฝึกหัดปฏิบัติการที่ 4

1. นักวิทยาศาสตร์สนใจจำนวนไข่ของปูชนิดหนึ่งที่จะวางในช่วงฤดูผสมพันธุ์ จึงทำการเก็บข้อมูลจำนวนไข่ที่วางจากปู จำนวน 45 ตัว ดังนี้

1,959	4,534	7,020	6,725	6,964	7,428	9,359	9,166
2,802	2,462	4,000	3,378	7,343	4,189	8,973	4,327
2,412	7,624	1,548	4,801	737	5,321	849	5,749
6,837	8,639	417	6,082	10,241	962	3,894	1,801
5,099	6,627	4,484	5,633	4,148	6,588	5,847	4,632
6,472	8,372	8,225	6,142	12,130			

ก) จงหาค่าเฉลี่ย (mean) ค่ามัธยฐาน (median) และค่าฐานนิยม (mode) ค่าความแปรปรวน (variance) ค่าเบี่ยงเบนมาตรฐาน (standard deviation) และค่าพิสัย (range) ของจำนวนไข่ของปู โดยใช้โปรแกรม Microsoft® Excel

ข) ให้ใช้โปรแกรม SPSS เพื่อหาค่าในข้อ ก.

2. ข้อมูลต่อไปนี้เป็นเปอร์เซ็นต์เถ้าในพีทมอสส์จากตัวอย่าง 50 ตัวอย่าง

0.5	1.8	4.0	1.0	2.0	2.6	3.1	2.5	1.7	5.0
1.1	1.6	2.3	3.5	2.2	1.3	3.0	2.7	1.2	1.5
2.0	3.8	3.0	2.3	1.8	3.2	2.4	2.5	1.9	3.1
3.6	2.4	0.8	3.4	1.4	2.4	2.8	2.7	4.5	2.1
1.9	2.3	1.2	1.9	2.3	1.5	0.7	3.7	1.8	1.7

ก) จงหาค่า Mean, Median, Mode, Range, Standard deviation, Variance, Standard error โดยใช้โปรแกรม Microsoft® Excel

ข) ให้ใช้โปรแกรม SPSS เพื่อหาค่าในข้อ ก.

ปฏิบัติการที่ 5 ความน่าจะเป็น

ในบทปฏิบัติการ จะใช้โปรแกรม Microsoft® Excel และ SPSS เพื่อหาค่าความน่าจะเป็นโดยจะแบ่งเป็น Discrete probability และ Continuous probability และคะแนนมาตรฐาน (z score) เนื่องจากในการหาค่าทั้งสอง จะต้องใช้ค่าสถิติอื่นที่เกี่ยวข้อง คือค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน ซึ่งจะกล่าวถึงในปฏิบัติการต่อไป โดยในบทปฏิบัติการนี้ จะกล่าวถึงเฉพาะเท่าที่จำเป็นเท่านั้น

1. คะแนนมาตรฐาน หรือ z score

คะแนนมาตรฐานของค่าหนึ่ง ๆ เป็นค่าที่บ่งบอกความห่างและทิศทาง (บวกหรือลบ ซึ่งคือมากกว่าหรือน้อยกว่า) ของค่านั้นว่าเบี่ยงเบนจากค่าเฉลี่ยเท่าใด โดยใช้ค่าเบี่ยงเบนมาตรฐาน (Standard deviation) เป็นตัวชี้ ในการแปลงค่าข้อมูลดิบไปเป็นคะแนนมาตรฐาน จะใช้สมการ $z = \frac{x-\mu}{\sigma}$ หากทราบค่าเฉลี่ยของประชากร (μ) และค่าเบี่ยงเบนมาตรฐานของประชากร (σ) หากไม่ทราบค่าดังกล่าว จะใช้ค่า $\bar{x}$ และค่า s.d. ของตัวอย่างแทนค่าทั้งสองตามลำดับ

คำสั่งที่เกี่ยวข้องในโปรแกรม Microsoft® Excel มีดังนี้

ฟังก์ชัน	คำอธิบาย	ตัวแปร
average(number1,number2,…)	หาค่าเฉลี่ยของชุดข้อมูลดิบ number1, number2, …	number1, number2, … : ข้อมูลดิบ
stdev(number1,number2,…)	หาค่าเบี่ยงเบนมาตรฐานของชุดข้อมูลดิบ number1, number2, …	number1, number2, … : ข้อมูลดิบ
standardize(x,mean,std)	หาค่าคะแนนมาตรฐาน (z score) ของค่าข้อมูลดิบ x	x : ค่าข้อมูลดิบ mean : ค่าเฉลี่ย std : ค่าเบี่ยงเบนมาตรฐาน

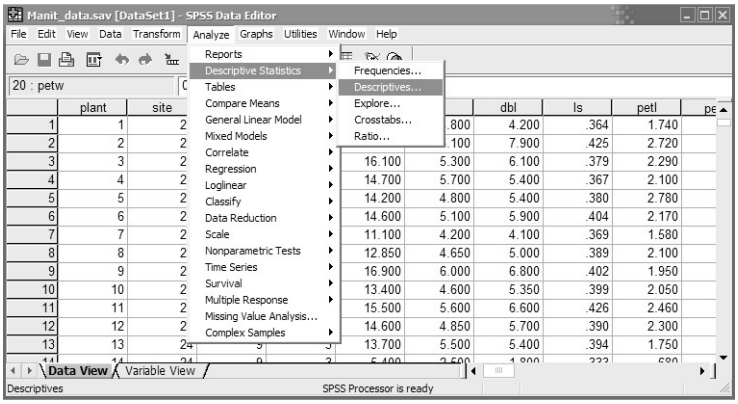
การใช้ฟังก์ชั่นในโปรแกรม Microsoft® Excel เพื่อค่า z score นั้น จำเป็นต้องทราบค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของประชากร/กลุ่มตัวอย่างก่อน มิฉะนั้น จะไม่สามารถหาค่า z score ได้ ดังนั้น หากมีชุดข้อมูลดิบ ก็สามารถแปลงค่าข้อมูลดิบแต่ละค่าไปเป็นค่า z score ได้ โดยใช้ฟังก์ชั่น average(…) และ stdev(…) เพื่อหาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานก่อน แล้วจึงใช้ฟังก์ชั่น standardize(…) เพื่อค่า z score ต่อไป

อนึ่ง หากไม่มีชุดข้อมูลดิบ แต่ทราบค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของประชากร/กลุ่มตัวอย่างแล้ว ก็สามารถใช้ฟังก์ชั่น standardize(…) เพื่อค่า z score ของข้อมูลดิบค่านั้นได้

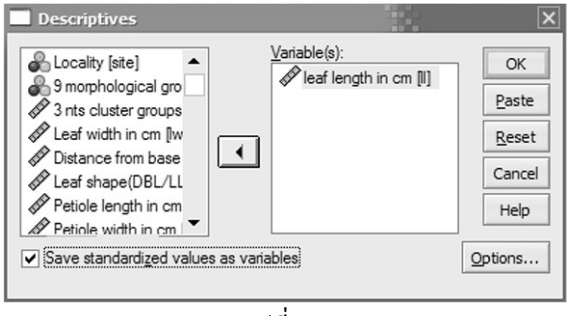
สำหรับในโปรแกรม SPSS นั้น การให้โปรแกรมคำนวณค่าคะแนนมาตรฐานของตัวแปรต่าง ๆ สามารถทำได้ดังนี้

1. เลือก Analyze, Descriptive Statistics, Descriptives... (รูปที่ 5.1)
2. จะมี dialog box (รูปที่ 5.2) เลือกตัวแปรที่ต้องการจากรายการตัวแปรทางซ้าย แล้วกด ▶ จะมีตัวแปรดังกล่าวปรากฏที่ช่อง Variables(s):
3. กดที่ ☐ Save standardized values as variables (รูปที่ 5.2)
4. กดที่ Options จะมี dialog box ดังรูปที่ 5.3 เลือกค่าสถิติที่ต้องการแล้วกดปุ่ม Continue
5. กดปุ่ม OK

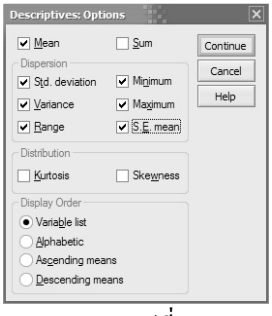
6. ผลการคำนวนในส่วนของค่าสถิติจะปรากฏใน SPSS Viewer (รูปที่ 5.4) ส่วนค่าคะแนนมาตรฐานจะเป็นตัวแปรใหม่ใน SPSS Data Editor ซึ่งตัวแปรใหม่นี้จะมีตัวอักษร “Z” นำหน้าชื่อตัวแปรเดิม ดังรูปที่ 5.5 ตัวแปร Zll ค่าตัวแปรค่าคะแนนมาตรฐานของตัวแปร ll



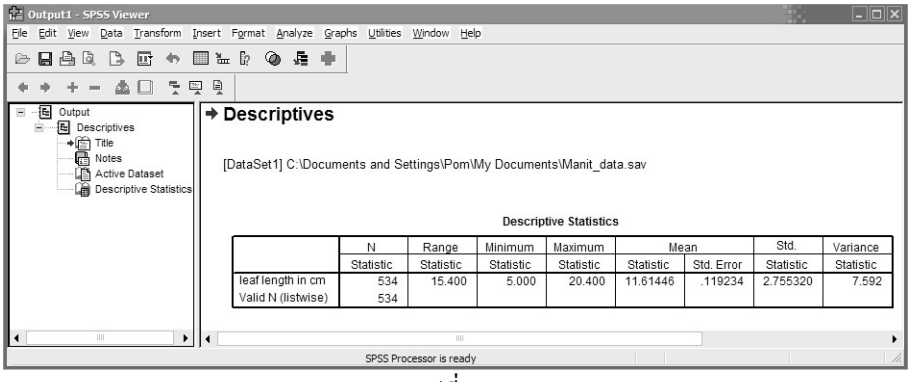
รูปที่ 5.1



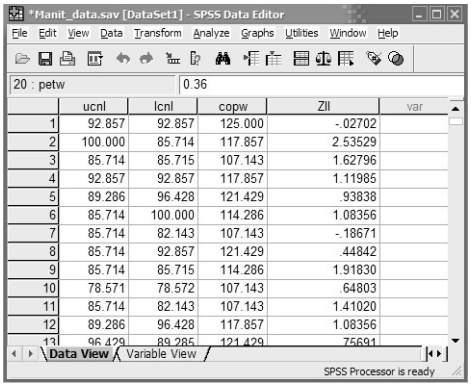
รูปที่ 5.2



รูปที่ 5.3



รูปที่ 5.4



รูปที่ 5.5

2. การหาค่า Continuous probability

ในปฏิบัติการนี้ จะกล่าวถึงเฉพาะ Normal probability distribution ซึ่งการหาค่าความเป็นไปได้แบบนี้เป็นการหาพื้นที่ใต้ Normal curve ของข้อมูลที่มีการกระจายแบบปกติ โดยจะใช้ฟังก์ชันในโปรแกรม Microso® Excel ต่าง ๆ ดังนี้

ฟังก์ชัน	คำอธิบาย	ตัวแปร
normdist(x, mean, s.d., cumulative)	หาค่าความเป็นไปได้ของข้อมูลดิบ x เมื่อกำหนดค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐาน	x : ค่าข้อมูลดิบ mean : ค่าเฉลี่ย s.d. : ค่าเบี่ยงเบนมาตรฐาน cumulative : มี 2 ค่า คือ True – ค่าสะสมจาก -∝ ถึง x False – ค่าที่ x
norminv(prob, mean, s.d.)	หาค่าข้อมูลดิบ เมื่อทราบค่าความเป็นไปได้ เมื่อกำหนดค่าเฉลี่ย และค่าเบี่ยงเบนมาตรฐาน	prob : ค่าความเป็นไปได้ mean : ค่าเฉลี่ย s.d. : ค่าเบี่ยงเบนมาตรฐาน
normsdist(z)	หาค่าความเป็นไปได้ของค่าคะแนนมาตรฐาน z	z : ค่าคะแนนมาตรฐาน
normsinv(prob)	หาค่าคะแนนมาตรฐาน z จากค่าความเป็นไปได้	prob : ค่าความเป็นไปได้

3. การหาค่า Binomial distribution

ในปฏิบัติการนี้ จะกล่าวถึงเฉพาะ Binomial probability distribution ซึ่งการหาค่าความเป็นไปได้แบบนี้เป็นการหาพื้นที่ใต้กราฟของข้อมูลที่มีการกระจายแบบ Binomial โดยจะใช้ฟังก์ชันในโปรแกรม Microso® Excel ต่าง ๆ ดังนี้

ฟังก์ชัน	คำอธิบาย	ตัวแปร
binomdist(number_s, trials, prob_s, cumulative)	หาค่าความเป็นไปได้ของครั้ง ของการเกิดเหตุการณ์ (s) จากจำนวนทดลองทั้งหมด (trial)	number_s : จำนวนครั้งที่เกิดเหตุการณ์ s trials : จำนวนการทดลองทั้งหมด prob_s : ค่าความเป็นไปได้ที่จะเกิดเหตุการณ์ s cumulative : มี 2 ค่า คือ True – ค่าสะสมจาก 0 ถึง number_s False – ค่าที่ number_s

แบบฝึกหัดที่ 5

1. อัตราการตายจากโรคหัวใจ (ต่อประชากร 100,000 คน) ในมลรัฐทั้งหมด 50 มลรัฐ และเมืองหลวงของประเทศสหรัฐอเมริกาเป็นดังนี้

417.4	445.4	377.7	396.8	361.4	351.6	332.7	322.4	310.6	265.4
414.4	442.7	376.2	395.2	358.2	349.4	330.2	319.1	306.8	236.9
407.6	427.3	369.8	394.0	356.4	340.3	329.1	315.9	300.6	214.2
406.7	422.9	369.2	391.7	353.3	338.7	328.4	311.8	297.4	194.0
396.9	420.8	368.9	391.0	351.6	334.2	327.1	311.2	274.6	169.0

   ก) จงหาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐาน

   ข) จงหาค่าคะแนนมาตรฐานของข้อมูลดิบแต่ละค่า

2. ผู้บริหารโรงพยาบาลแห่งหนึ่ง มีความประสงค์จะลดจำนวนห้องดูแลพิเศษทารกคลอดก่อนกำหนด ที่มีน้ำหนักตัวน้อยกว่า 2,500 กรัม เพื่อนำห้องที่ต้องสำรองไว้ไปรองรับทารกแรกคลอดที่คลอดตามกำหนด ทั้งนี้ จากการตรวจสอบข้อมูลย้อนกลับไปพบว่าระยะเวลาที่ทารกคลอดก่อนกำหนดที่มีน้ำหนักตัวน้อยกว่า 2,500 กรัม จะอยู่ในห้องดูแลพิเศษนี้นานที่สุดเพียง 4 วัน ถ้าหากทารกคลอดก่อนกำหนดในข่ายนี้มีการกระจายแบบ Binomial

   ก) จากจำนวนทารก 25 คน และ p = 0.1 จงหาค่าความเป็นไปได้ที่จะมีทารกคลอดก่อนกำหนดที่มีน้ำหนักน้อยกว่า 2,500 กรัม ซึ่งต้องอยู่ในห้องดูพิเศษนานที่สุด 4 วัน จำนวน 0 1 2 3 4 5 6 และ 7 คน

   ข) ผู้บริหารโรงพยาบาลจะต้องการจัดเตรียมเตียงนอนทารกคลอดก่อนกำหนดมากที่สุดกี่เตียง ถ้าหากว่าทารกคลอดก่อนกำหนดที่มีน้ำหนักน้อยกว่า 2,500 กรัม มีโอกาสเกิดขึ้นไม่เกิน 5% และ 1% ตามลำดับ

3. ในประชากรนกพิราบประชากรหนึ่ง นักวิจัยพบนกที่มีปรอทสะสมในร่างกายโดยผ่านเข้าสู่ร่างกายทางการกินเมล็ดพืชที่อยู่ในพื้นที่ถิ่นอาศัยของนกพิราบ ถ้าหากปริมาณปรอทที่สะสมในร่างกายมีการกระจายแบบปกติ โดยมีค่า $\mu = 0.25$ และ $\sigma = 0.08$ แล้ว จงหาค่าความเป็นไปได้เมื่อ

   a. $x \leq 0.3$

   b. $x \geq 0.17$

   c. $0.2 \leq x \leq 0.4$

   d. $0.01 \leq x \leq 0.49$

4. ในการทดลองยาที่ควบคุมโฮอร์โมนในหนู พบว่า ในหนูทดลอง 20 ตัว จะมีหนู 4 ตัวที่รอดตายจนถึงสิ้นสุดการทดลอง ถ้านักวิจัยท่านหนึ่งทดลองยาในหนู 10 ตัว จะมีความน่าจะเป็นเท่าใดที่จะมีหนูรอดตายจนสิ้นสุดการทดลอง 8 ตัว?

5. ให้นิสิต download ข้อมูลจาก http://www.sc.chula.ac.th/botany/eClass/2305384/biostat_data.sav และ http://www.sc.chula.ac.th/botany/eClass/2305384/biostat_data.xls เพื่อฝึกหัดใช้คำสั่ง/ฟังก์ชั่น หาค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐาน และค่าคะแนนมาตรฐาน

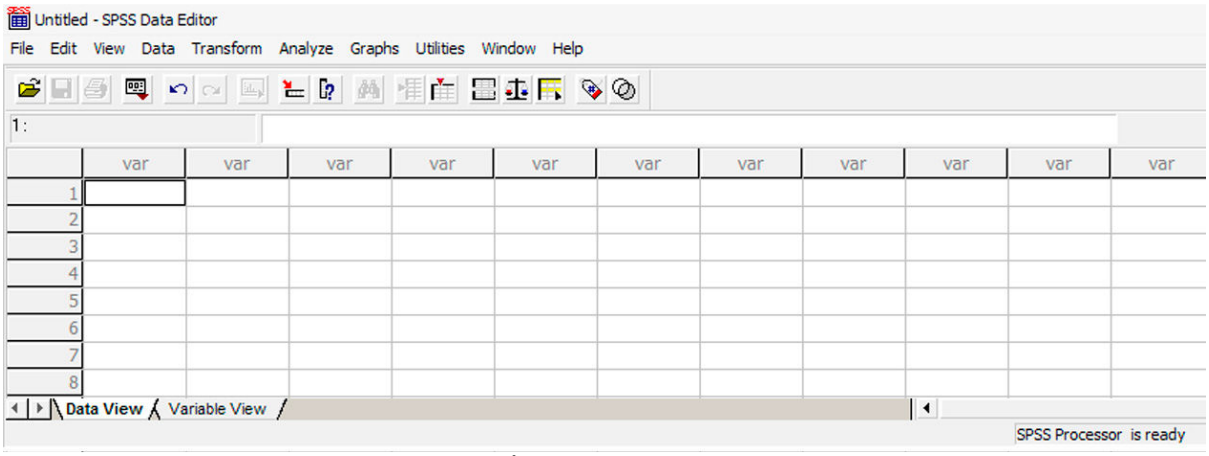
การบ้านปฏิบัติการที่ 5

1. ในประชากรนกประชากรหนึ่งมีค่าเฉลี่ยความยาวของปาก μ = 3.25 และมีค่าเบี่ยงเบนมาตรฐาน σ = 0.75 จงหาค่า z-score เมื่อ
   a. x = 2.50
   b. x = 3.00
   c. x = 4.25
   d. x = 3.15
   e. $2.75 \leq x \leq 3.30$
2. อัตราการตายในโรงพยาบาลจากการเก็บตัวอย่างในโรงพยาบาล 981 แห่ง พบว่า มี 10.5 เปอร์เซ็นต์ สำหรับกรณีการผ่าตัดเปลี่ยนเส้นเลือด และ 5.0 เปอร์เซ็นต์ สำหรับกรณีการเปลี่ยนกระดูกเทียมที่สะโพก ในโรงพยาบาลแห่งหนึ่งพบว่ามีผู้ป่วยที่ผ่าตัดเปลี่ยนเส้นเลือด 20 คน และมีผู้ป่วยที่เปลี่ยนกระดูกเทียมที่สะโพก 20 คน
   ก) มีความเป็นไปได้เท่าใด ที่จะมีผู้ป่วยตาย 5 คนจากการผ่าตัดเปลี่ยนเส้นเลือด?
   ข) มีความเป็นไปได้เท่าใด ที่จะมีผู้ป่วยตายอย่างน้อย 5 คนจากการผ่าตัดเปลี่ยนเส้นเลือด?
   ค) มีความเป็นไปได้เท่าใด ที่จะมีผู้ป่วยตายไม่เกิน 5 คนจากการผ่าตัดเปลี่ยนเส้นเลือด?
   ง) มีความเป็นไปได้เท่าใด ที่จะมีผู้ป่วยตายอย่างน้อย 4 คนจากการเปลี่ยนกระดูกเทียม
   จ) มีความเป็นไปได้เท่าใด ที่จะมีผู้ป่วยตาย 4 คนหรือน้อยกว่าจากการเปลี่ยนกระดูกเทียม
3. ในร่างกายคนจะมีวิตามินอีในน้ำเลือดเฉลี่ย 860 µg/dL และมีค่าเบี่ยงเบนมาตรฐาน 340 µg/dL
   ก) จะมีคนจำนวนเท่าใด (คิดเป็นเปอร์เซ็นต์) ที่จะมีระดับวิตามินอีอยู่ในช่วง 400 – 1000 µg/dL ?
   ข) ระดับวิตามินอีในเลือดของคนอยู่ในระดับที่เป็นพิษ หากว่ามีค่ามากกว่า 2000 µg/dL จะมีคนจำนวนเท่าใด (คิดเป็นเปอร์เซ็นต์) ที่จะถูกวินิจฉัยว่ามีระดับวิตามินอีในเลือดเป็นพิษ ?
4. ให้ X เป็นตัวแปรที่มีการกระจายแบบ binomial ถ้า n = 10 และ p = 0.4 จงหา
   ข. $P[X \leq 1]$
   ค. $P[X \geq 7]$
   ง. $P[X = 7]$
   จ. $P[X \leq 2 \text{ and } X \geq 8]$
   ฉ. $P[4 \leq X \leq 7]$
5. ในประชากรพืชประชากรหนึ่งมีค่าเฉลี่ยความกว้างของดอก μ = 12 และมีค่าเบี่ยงเบนมาตรฐาน σ = 1.5 จงหาค่า x เมื่อ
   a. z = 0.5
   b. z = 1.5
   c. z = -2.1
   d. z = -3.1

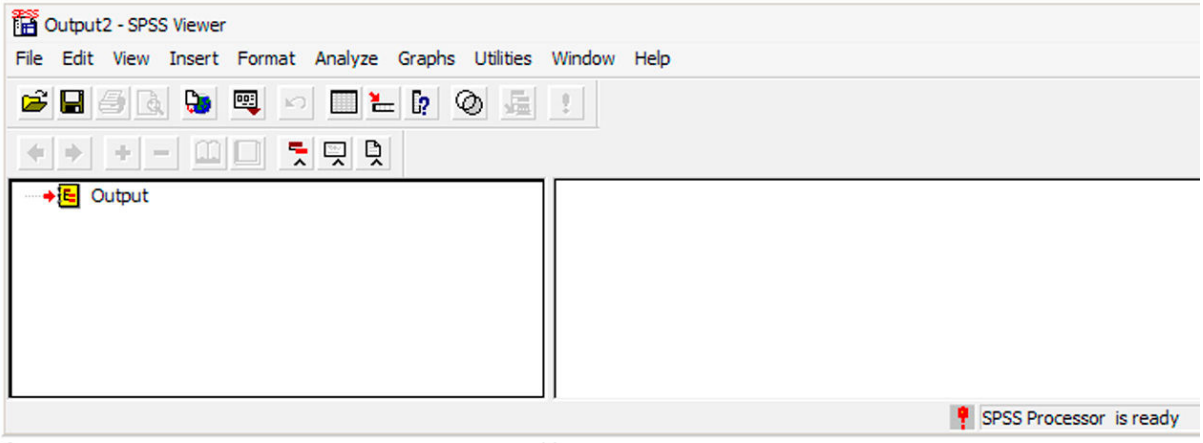
ปฏิบัติการที่ 6 Hypothesis Testing: One population/sample

1. การป้อนข้อมูลสำหรับการวิเคราะห์ทางสถิติด้วยโปรแกรม SPSS

ในโปรแกรม SPSS จะประกอบด้วย window ที่แยกกัน 2 window คือ SPSS Data Editor เพื่อใช้ป้อนข้อมูลและจัดการข้อมูล (รูปที่ 6.1) กับ SPSS Viewer เพื่อดูและจัดการผลการวิเคราะห์ข้อมูล (รูปที่ 6.2) เมื่อเปิดโปรแกรมใช้งาน จะเปิด Data Editor ขึ้นมาก่อนเสมอ แต่ก็สามารถเรียก Viewer เพื่อใช้เปิดดูไฟล์ผลการวิเคราะห์ข้อมูลครั้งก่อนได้

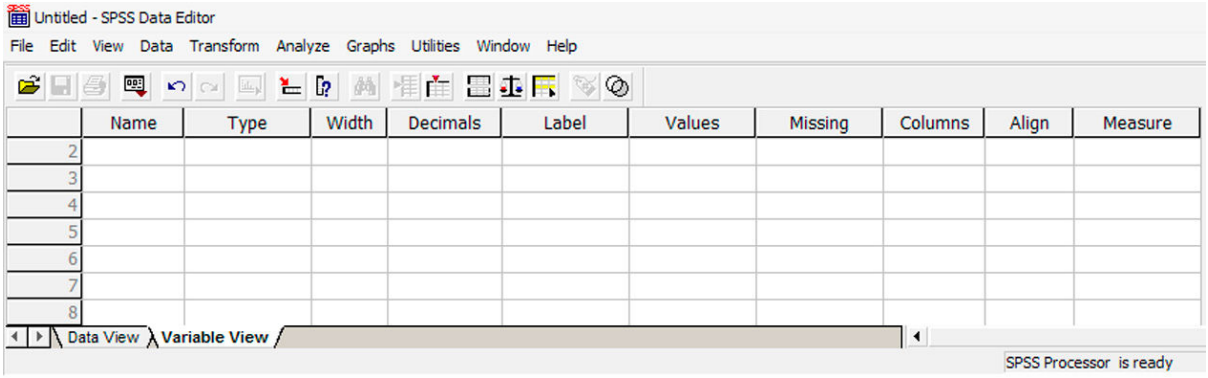


รูปที่ 6.1 SPSS Data Editor



รูปที่ 6.2 SPSS Viewer

ใน Data Editor (ตั้งแต่ ver. 10+) จะประกอบด้วย Data View และ Variable View โดย Data View (รูปที่ 6.1) ใช้ในการป้อนและจัดการเกี่ยวข้อมูลดิบ ส่วน Variable View (รูปที่ 6.3) จะเป็นส่วนที่ใช้ในการกำหนดลักษณะต่างๆ ของตัวแปร เช่น ชื่อ ประเภทข้อมูล ขนาดข้อมูลดิบของตัวแปร ฯลฯ



รูปที่ 6.3 Variable View

ในการป้อนข้อมูล นิสิตควรกำหนดตัวแปรโดยกดที่ Variable View ใน Editor Viewer ก่อน ซึ่งแต่ละแถวจะเป็นตัวแปร 1 ตัวแปร ให้กำหนดค่าต่างๆ ของตัวแปร ดังนี้

คอลัมน์	**คำอธิบาย**
Name	ชื่อตัวแปร (ต้องระบุ)
Type	ประเภทของข้อมูล (ต้องระบุ)
Width	ความกว้างข้อมูลดิบ
Decimal	ตำแหน่งทศนิยม (เป็นส่วนหนึ่งของ Width)
Label	คำอธิบายเพิ่มเติมของตัวแปร
Value	กำหนดคำอธิบายของค่าต่างๆ ของข้อมูลดิบของตัวแปร
Missing	กำหนดค่า missing data
Columns	ความกว้างของคอลัมน์ใน Data View
Align	จัดข้อมูลชิดซ้าย กลาง ขวา
Measure	ข้อมูลวัดแบบ scale, ordinal, nominal

เมื่อกำหนดค่าต่างๆของตัวแปรทุกตัวแล้ว ให้กดที่ Data View ใน Data Editor เพื่อเริ่มบันทึกข้อมูลดิบ ใน Data View นี้ แต่ละคอลัมน์คือ 1 ตัวแปร แต่ละแถวคือค่าสังเกต 1 ค่า ในการกำหนดตัวแปรและการป้อนข้อมูลเพื่อการวิเคราะห์นั้น ผู้วิเคราะห์ต้องทราบก่อนว่าจะเป็นการวิเคราะห์แบบใด เพื่อให้การกำหนดตัวแปรและการป้อนข้อมูลสอดคล้องกัน

2. การทดสอบสมมุติฐาน

2.1 One sample t-test

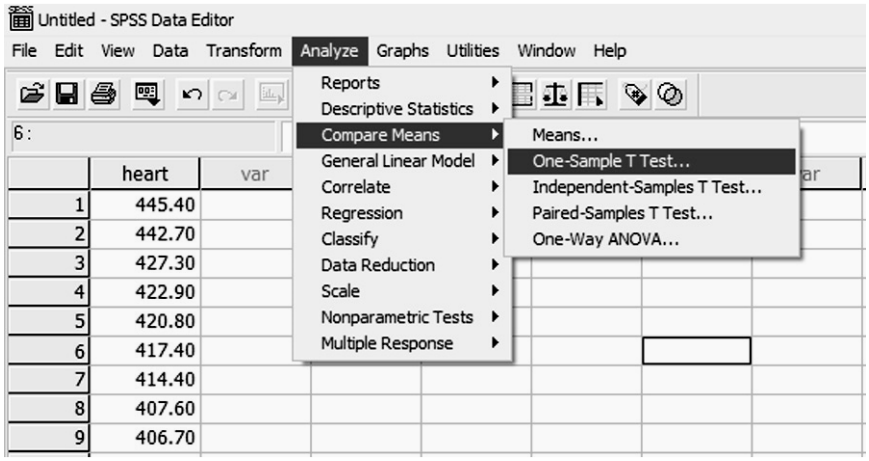
การทดสอบนี้เป็นการทดสอบค่าเฉลี่ยของตัวอย่างกับค่าเฉลี่ยของประชากร แต่ไม่ทราบค่าเบี่ยงเบนมาตรฐานของประชากร ในการป้อนข้อมูลดิบ จะกำหนดตัวแปรให้เท่ากับจำนวนชุดของข้อมูลดิบจากกลุ่มตัวอย่างเดียวกัน เช่น ในกลุ่มตัวอย่าง 30 ตัวอย่าง อาจจะเก็บข้อมูลความสูง และน้ำหนัก ก็จะมีตัวแปร 2 ตัว เป็นต้น การวิเคราะห์ข้อมูล มีขั้นตอนดังนี้

1. เลือก Analyze, Compare Means, One-Sample T Test... (รูปที่ 6.4)
2. ใน dialog box (รูปที่ 6.5)
    a. เลือกตัวแปรที่ต้องการทดสอบจากรายการในช่องด้านซ้าย กดปุ่ม ▶

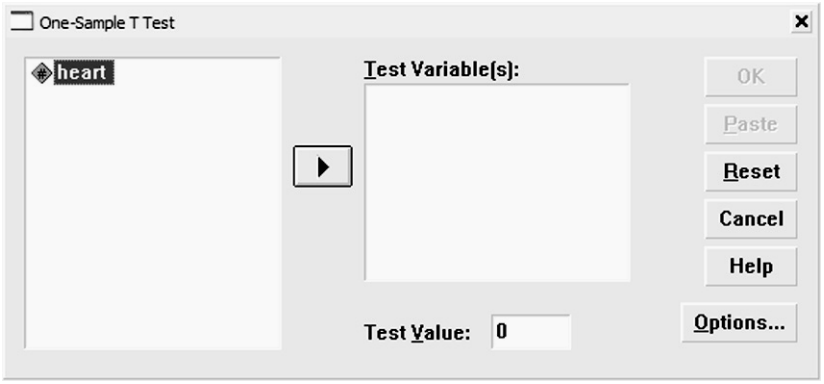
b. กำหนดค่าเฉลี่ยประชากรในช่อง Test Value

c. หากต้องการกำหนดค่านัยสำคัญที่ไม่ใช่ค่า 95% กดที่ปุ่ม Options... ใน dialog box ใหม่ (รูปที่ 6.6) กำหนดค่านัยสำคัญที่ช่อง Confidence Interval แล้วกด Continue

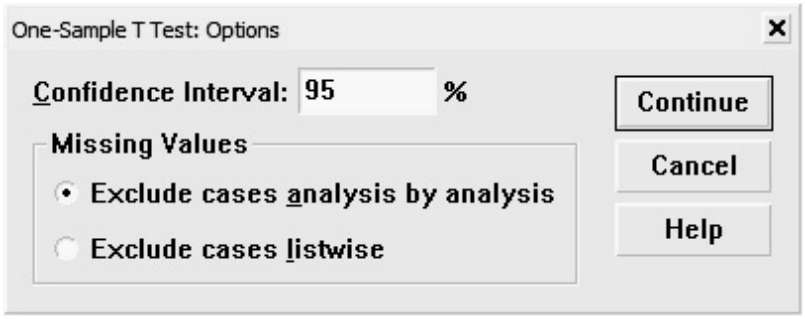
3. แล้วกด OK จะได้ผลดังรูปที่ 6.7



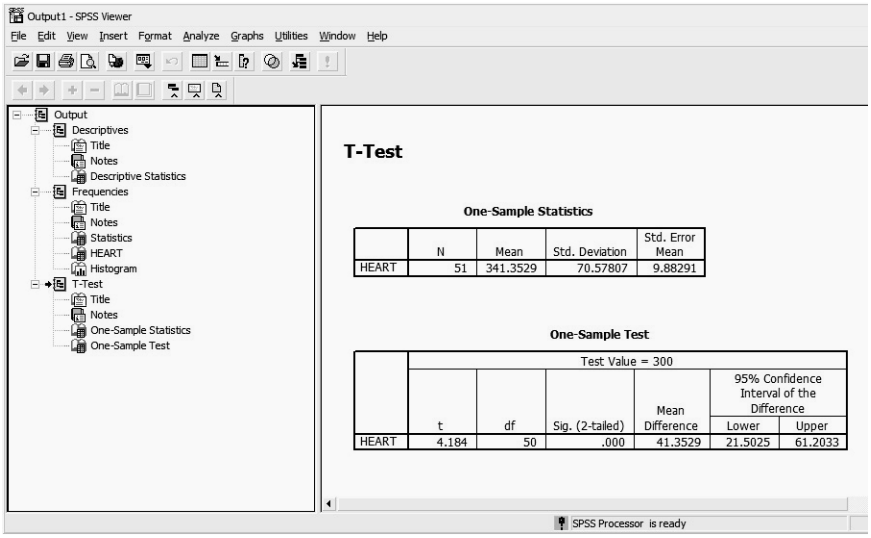
รูปที่ 6.4



รูปที่ 6.5



รูปที่ 6.6



รูปที่ 6.7

จากตัวอย่างผลการวิเคราะห์ จะเห็นว่ามี 2 ตาราง โดยตารางแรกเป็นตารางที่แสดงค่าสถิติต่างๆ ของข้อมูลดิบ ส่วนตารางที่สองแสดงผลการทดสอบโดย

คอลัมน์	**คำอธิบาย**
Test Value	ค่าเฉลี่ยของประชากร
t	ค่า t จากการคำนวณของข้อมูลตัวอย่าง
df	ค่า degree of freedom
Sig. (2-tailed)	ค่าความน่าจะเป็นหรือค่าพื้นที่ใต้กราฟของ t - ในกรณีที่เป็น 2-tailed Hypothesis test ให้ใช้ค่านี้ - ในกรณีที่เป็น 1-tailed Hypothesis test ให้**หาร**ค่านี้ด้วย 2
Mean Difference	ค่าความแตกต่างระหว่างค่าเฉลี่ยของตัวอย่างและค่าเฉลี่ยของประชากร
95% Confidence Interval of the Difference	ค่าต่ำสุดและค่าสูงสุดของ Mean Difference

ในการทดสอบสมมุติฐาน จะแบ่งเป็นกรณีต่างๆ เมื่อ μ_s คือค่าเฉลี่ยของตัวอย่าง และ μ_p คือ ค่าเฉลี่ยของประชากร ดังนี้

กรณีที่ 1 H1: $\mu_s - \mu_p \neq 0$ ปฏิเสธ H0 เมื่อ Sig. $\leq$ 0.05 (หรือค่า α ที่ตั้งไว้)

กรณีที่ 2 H1: $\mu_s - \mu_p < 0$ ปฏิเสธ H0 เมื่อ ½Sig. $\leq$ 0.05 (หรือค่า α ที่ตั้งไว้) และค่า t เป็น ลบ

กรณีที่ 3 H1: $\mu_s - \mu_p > 0$ ปฏิเสธ H0 เมื่อ ½Sig. $\leq$ 0.05 (หรือค่า α ที่ตั้งไว้) และค่า t เป็น บวก

2.2 One sample *z* test

การทดสอบสมมุติฐาน *z* test นี้ จะใช้ทดสอบกับประชากรเดียวเช่นเดียวกับการทดสอบแบบ *t* test แต่ต่างกันคือ หากทราบค่าเบี่ยงเบนมาตรฐานของประชากร (σ) จะใช้ *z* test แทน *t* test โดย $z = \dfrac{\bar{x} - \mu}{\dfrac{\sigma}{\sqrt{n}}}$ และค่า Sig. ที่จะใช้ในการ

ทดสอบสามารถหาได้จากฟังก์ชั่นใน Microsoft® Excel คือ 2×normsdist(*z*) เมื่อ $z < 0$ หรือ 2×(1 -normsdist(*z*)) เมื่อ $z > 0$ สำหรับการทดสอบเป็นดังนี้

กรณีที่ 1 H1: $\mu s - \mu p \neq 0$ ปฏิเสธ H0 เมื่อ Sig. ≤ 0.05 (หรือค่า α ที่ตั้งไว้)

กรณีที่ 2 H1: $\mu s - \mu p < 0$ ปฏิเสธ H0 เมื่อ ½Sig. ≤ 0.05 (หรือค่า α ที่ตั้งไว้) และค่า *z* เป็น ลบ

กรณีที่ 3 H1: $\mu s - \mu p > 0$ ปฏิเสธ H0 เมื่อ ½Sig. ≤ 0.05 (หรือค่า α ที่ตั้งไว้) และค่า *z* เป็น บวก

2.3 การทดสอบสมมุติฐานสำหรับการกระจายแบบ Binomial

การทดสอบสมมุติฐานแบบนี้สามารถแยกได้เป็น 2 กรณี คือ

1. <u>กรณีที่มีข้อมูล (n) 5 ตัวอย่างขึ้นไป</u> จะทดสอบโดยการหาค่า *z*

$$z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o q_o}{n}}}$$

เมื่อ $\hat{p}$ คือ สัดส่วนของเหตุการณ์ที่เกิดขึ้นในกลุ่มตัวอย่าง

p_o คือ สัดส่วนของเหตุการณ์ที่เกิดขึ้นในประชากร

ค่า Sig. สามารถหาได้จากฟังก์ชั่นใน Microsoft® Excel คือ 2×normsdist(*z*) เมื่อ $z < 0$ หรือ 2×(1 - normsdist(*z*)) เมื่อ $z > 0$ สำหรับการทดสอบเป็นดังนี้

กรณีที่ 1 H1: $\mu s \neq \mu p$ ปฏิเสธ H0 เมื่อ Sig. ≤ 0.05 (หรือค่า α ที่ตั้งไว้)

กรณีที่ 2 H1: $\mu s < \mu p$ ปฏิเสธ H0 เมื่อ ½Sig. ≤ 0.05 (หรือค่า α ที่ตั้งไว้) และค่า *z* เป็น ลบ

กรณีที่ 3 H1: $\mu s > \mu p$ ปฏิเสธ H0 เมื่อ ½Sig. ≤ 0.05 (หรือค่า α ที่ตั้งไว้) และค่า *z* เป็น บวก

2. กรณีที่มีข้อมูล (n) น้อยกว่า 5 ตัวอย่าง จะใช้วิธีการหาค่าความเป็นไปได้ หรือค่า Sig. ของเหตุการณ์ที่เกิดในกลุ่มตัวอย่างด้วย Exact method ของการกระจายแบบ Binomial ดังนี้

ถ้า $\hat{p} < p_o$ แล้ว ค่า Sig. หาได้จาก $2 \times \sum_{k=0}^{x} \binom{n}{k} p_o^k q_o^{n-k}$

ถ้า $\hat{p} \geq p_o$ แล้ว ค่า Sig. หาได้จาก $2 \times \sum_{k=x}^{n} \binom{n}{k} p_o^k q_o^{n-k}$

ซึ่งสามารถหาได้จากฟังก์ชั่นใน Microsoft® Excel คือ binomdist(x-1,n,p_o,true) หรือ 1- binomdist(x,n,p_o,true) ตามลำดับ (ทั้งสองกรณี อาจมีเงื่อนไขต่างจากที่กล่าวมาข้างต้นได้ ให้ปรับเปลี่ยนค่าตามความเหมาะสม) โดยการทดสอบจะคล้ายกับที่กล่าวมาในกรณีที่ข้อมูลมีตั้งแต่ 5 ตัวอย่างขึ้นไป

แบบฝึกหัดที่ 6

1. ในอำเภอหนึ่ง ครูสุขศึกษาเชื่อว่านักเรียนหญิงมีความสูงเฉลี่ยเปลี่ยนไปในช่วงเวลา 10 ปีที่ผ่านมา โดยค่าความสูงเฉลี่ยเมื่อ 10 ปีก่อน มีค่าเป็น 160 ซม. แต่ไม่ทราบค่าเบี่ยงเบนมาตรฐาน เพื่อพิสูจน์ความเชื่อดังกล่าว ครูผู้นี้จึงสุ่มตัวอย่างเด็กหญิงจำนวน 10 คน มาวัดความสูง ได้ผลดังนี้

162.6 167.6 172.7 152.4 157.5 165.1 167.6 160.0 170.2 175.3

1.1 ตั้งสมมุติฐานทั้งสอง Null hypothesis และ Alternative hypothesis

1.2 เขียนค่าสถิติที่ได้จากการวิเคราะห์ด้วยโปรแกรม SPSS

1.3 ถ้ากำหนดให้ α=0.05 แล้ว ผลการทดสอบสมมุติฐาน สรุปได้ว่าอย่างไร

2. หิ่งห้อยแต่ละชนิดจะมีรูปแบบการกระพริบแสงแตกต่างกัน โดยหิ่งห้อยชนิดหนึ่งจะมีรูปแบบการกระพริบที่ประกอบด้วยจังหวะที่เปล่งแสงในช่วงเวลาสั้นๆ ตามด้วยระยะพักซึ่งมีค่าเฉลี่ยน้อยกว่า 4 วินาที เพื่อทดสอบการสังเกตนี้ ผู้วิจัยจึงได้สุ่มตัวอย่างหิ่งห้อย 16 ตัว แล้ววัดค่าระยะพักหลังจากเปล่งแสง ได้ผลดังนี้

3.9 3.8 3.5 4.1 3.2 3.5 3.6 3.7 3.8 3.7 4.2 3.4 4.0 4.0 3.6 4.3

2.1 ตั้งสมมุติฐานทั้งสอง Null hypothesis และ Alternative hypothesis

2.2 เขียนค่าสถิติที่ได้จากการวิเคราะห์ด้วยโปรแกรม SPSS

2.3 ถ้ากำหนดให้ α=0.05 แล้ว ผลการทดสอบสมมุติฐาน สรุปได้ว่าอย่างไร

การบ้านปฏิบัติการที่ 6 การทดสอบสมมุติฐาน

1. ระดับเซรั่มครีเอทินีนในเลือดเป็นตัวบ่งชี้ว่าผู้รับการตรวจเป็นหรือไม่เป็นโรคไต จากการตรวจระดับเซรั่มดังกล่าวในผู้รับการตรวจ 15 คน ได้ข้อมูลดังนี้

0.9 1.1 1.6 2.0 0.8 0.7 1.4 1.2 1.5 0.8 1.0 1.1 1.4 2.2 1.4

   ก) ถ้าหากว่าระดับเซรั่มดังกล่าวมีการกระจายแบบปกติโดยมีค่าเฉลี่ยเป็น 1.0 แล้ว นิสิตจะใช้การทดสอบสมมุติฐานแบบใดเพื่อพิจารณาว่าค่าที่วัดได้ไม่แตกต่างจากค่าเฉลี่ย?

   ข) สมมุติฐานเป็น nondirectional หรือ directional?

   ค) Null hypothesis และ Alternative hypothesis เขียนได้ว่าอย่างไร?

   ง) ให้ทำการทดสอบสมมุติฐานโดยแสดงค่าสถิติที่จำเป็น กำหนดให้ α เป็น 0.05 และสรุปด้วาอย่างไร?

2. ในการศึกษาเกี่ยวกับประสิทธิภาพของ echolocation ในค้างคาว ซึ่งเชื่อว่าน่าจะมีประสิทธิภาพดีในการระบุตำแหน่งวัตถุที่อยู่ห่างจากตัวค้างคาวอย่างน้อย 6 เมตรขึ้นไป เพื่อทดสอบข้อสมมุติฐานนี้ จึงได้นำค้างคาวมาทดลองบินในห้องที่มีสิ่งกีดขวาง ได้ผลดังนี้

6.2	6.8	6.1	5.7	6.1	6.3	5.8	6.3
5.9	6.3	6.4	6.0	6.3	6.2	5.9	6.1

   ก) ควรตั้งสมมุติแบบ nondirectional หรือ directional?

   ข) Null hypothesis และ Alternative hypothesis เขียนได้ว่าอย่างไร?

   ค) ควรใช้การทดสอบสมมุติฐานแบบใด?

   ง) ให้ทำการทดสอบสมมุติฐานโดยแสดงค่าสถิติที่จำเป็น กำหนดให้ α เป็น 0.05 และสรุปได้ว่าอย่างไร?

ปฏิบัติการที่ 7 Hypothesis Testing: Two or more populations/samples

ในบทนี้ นิสิตจะได้เรียนรู้วิธีการใช้คำสั่งเพื่อเปรียบเทียบทางสถิติค่าเฉลี่ยจาก 2 ประชากร และมากกว่า 2 ประชากร โดยนิสิตจะต้องเรียนรู้วิธีการป้อนข้อมูลและการกำหนดตัวแปรที่จำเป็น รวมถึงขั้นตอนการวิเคราะห์และการอ่านผลที่ได้ในโปรแกรม SPSS นอกจากนี้ ยังมีวิธีการที่เปรียบเทียบทางสถิติค่าเฉลี่ยด้วยโปรแกรม Microsoft® Excel ด้วย ทั้งนี้ เพื่อให้นิสิตสามารถประยุกต์ใช้โปรแกรมดังกล่าวได้หากไม่มีโปรแกรม SPSS

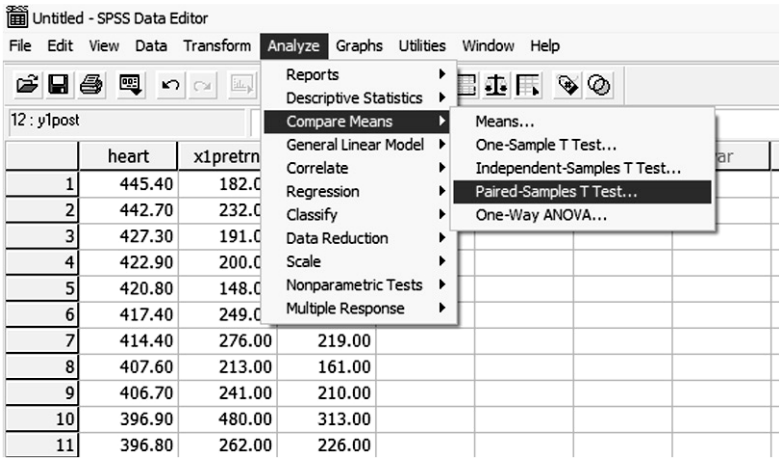
1. การทดสอบสมมุติฐานความแตกต่างของค่าเฉลี่ยกลุ่มตัวอย่าง 2 กลุ่ม

1.1 Paired-samples t test

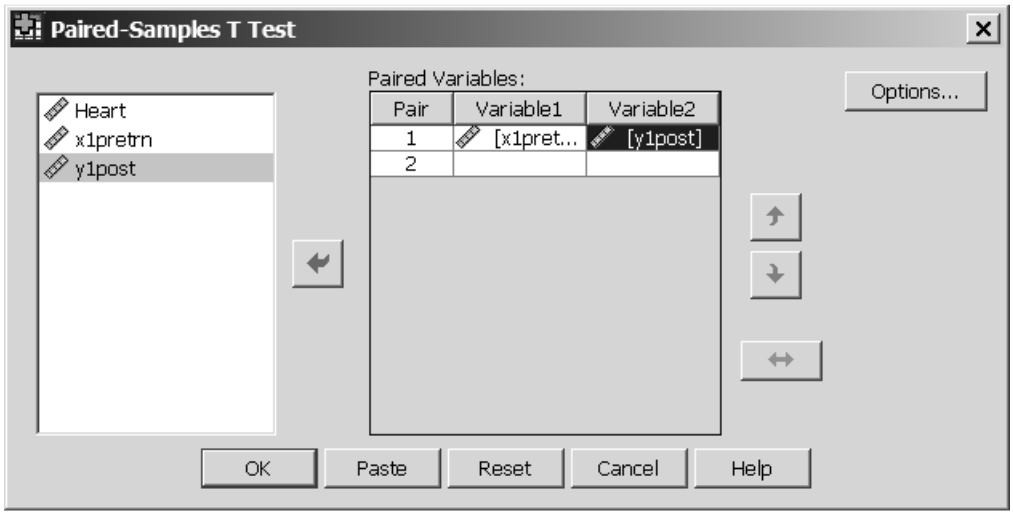
การทดสอบจะใช้ตัวอย่างชุดเดียวกัน แต่จะวัดค่าก่อนการทดลองและหลังการทดลอง การป้อนข้อมูลจึงต้องใช้ตัวแปร 1 คู่ คือ ตัวแปรสำหรับค่าก่อนการทดลอง และตัวแปรสำหรับค่าหลังการทดลอง โดยค่าสังเกตแต่ละคู่ต้องมาจากตัวอย่างเดียวกัน ดังนั้น การป้อนข้อมูลจึงต้องให้ค่าทั้งสองอยู่ในแถวเดียวกัน อนึ่ง เพื่อให้อ่านผลการวิเคราะห์ได้ง่าย ควรตั้งชื่อให้ตัวแปรเก็บค่าก่อนการทดลองอยู่ลำดับแรก และตัวแปรเก็บค่าหลังการทดลองอยู่ลำดับหลัง เช่น มีอักษร x และ y หน้าตัวแปรทั้งสอง ตามลำดับ

การวิเคราะห์ทำได้ดังนี้

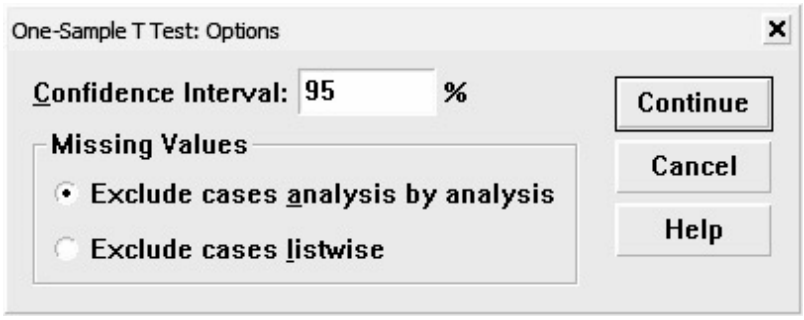
1. เลือก Analyze, Compare Means, Paired-Samples T Test... (ภาพที่ 7.1)
2. จาก dialog box (ภาพที่ 7.2)
   a. เลือกคู่ตัวแปรในช่องด้านซ้าย กดปุ่ม ▸ สังเกตว่าที่ช่อง Current Selections จะมีชื่อตัวแปรที่เลือกไว้ปรากฏ โดย Variable 1: เป็นตัวแปรเก็บค่าก่อนการทดลองและ Variable 2: เป็นตัวแปรเก็บค่าหลังการทดลอง
   b. หากต้องการกำหนดค่านัยสำคัญของ Confidence interval ของค่าความแตกต่างระหว่างค่าเฉลี่ย กดที่ Options... (ภาพที่ 7.3) กำหนดค่าที่ช่อง Confidence Interval: แล้วกด Continue
3. กด OK จะได้ผลดังภาพที่ 7.4



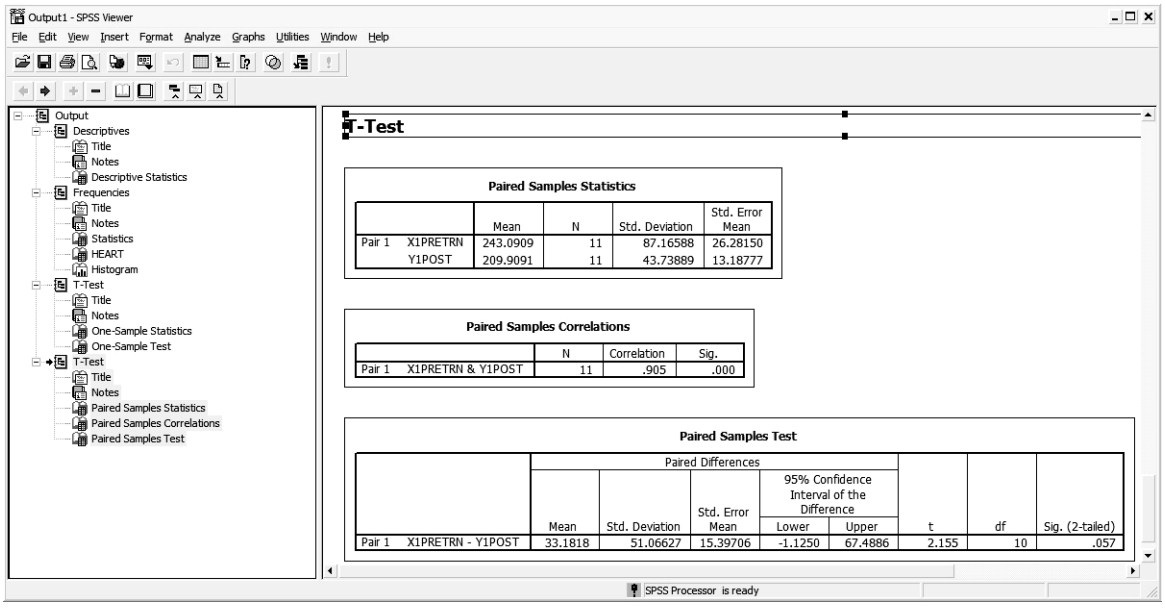
ภาพที่ 7.1



ภาพที่ 7.2



ภาพที่ 7.3



ภาพที่ 7.4

ผลการวิเคราะห์แสดงดังภาพที่ 7.4 ซึ่งมี 3 ตาราง โดยตารางที่ 1 เป็นค่าสถิติพื้นฐานของแต่ละตัวแปร (ในที่นี้คือค่าเฉลี่ยของตัวอย่างก่อนและหลังได้รับ treatment) ตารางที่ 2 เป็นค่าความสัมพันธ์ระหว่างตัวแปรทั้งสอง ส่วนตารางที่ 3 เป็นผลการทดสอบค่าเฉลี่ย โดย

คอลัมน์	คำอธิบาย
Pair 1 … - …	ผลการทดสอบของตัวแปรคู่ใด
Pair Difference	
Mean	ค่าเฉลี่ยของผลต่างระหว่างตัวแปรทั้งสอง
Std. Deviation	ค่าเบี่ยงเบนมาตรฐานของผลต่างระหว่างตัวแปรทั้งสอง
Std. Error Mean	ค่าเบี่ยงเบนของค่าเฉลี่ยของผลต่างระหว่างตัวแปรทั้งสอง
95% Confidence Interval of the Difference	ค่าต่ำสุดและสูงสุดของค่าเฉลี่ยของผลต่างระหว่างตัวแปรทั้งสอง
t	ค่า t จากการคำนวณของข้อมูลตัวอย่าง
df	ค่า degree of freedom
Sig. (2-tailed)	ค่าความน่าจะเป็นหรือค่าพื้นที่ใต้กราฟของ t - ในกรณีที่เป็น 2-tailed Hypothesis test ให้ใช้ค่านี้ - ในกรณีที่เป็น 1-tailed Hypothesis test ให้**หาร**ค่านี้ด้วย 2

ในการทดสอบสมมุติฐาน จะแบ่งเป็นกรณีต่างๆ เมื่อ μx และ μy คือ ค่าเฉลี่ยของตัวอย่างก่อนและหลังได้รับ treatment ตามลำดับ ดังนี้

กรณีที่ 1 H1: $\mu x - \mu y \neq 0$ ปฏิเสธ H0 เมื่อ Sig. ≤ 0.05 (หรือค่า α ที่ตั้งไว้)

กรณีที่ 2 H1: $\mu x - \mu y < 0$ ปฏิเสธ H0 เมื่อ ½Sig. ≤ 0.05 (หรือค่า α ที่ตั้งไว้) และค่า t เป็น ลบ

กรณีที่ 3 H1: $\mu x - \mu y > 0$ ปฏิเสธ H0 เมื่อ ½Sig. ≤ 0.05 (หรือค่า α ที่ตั้งไว้) และค่า t เป็น บวก

1.2 Independent two-samples t test

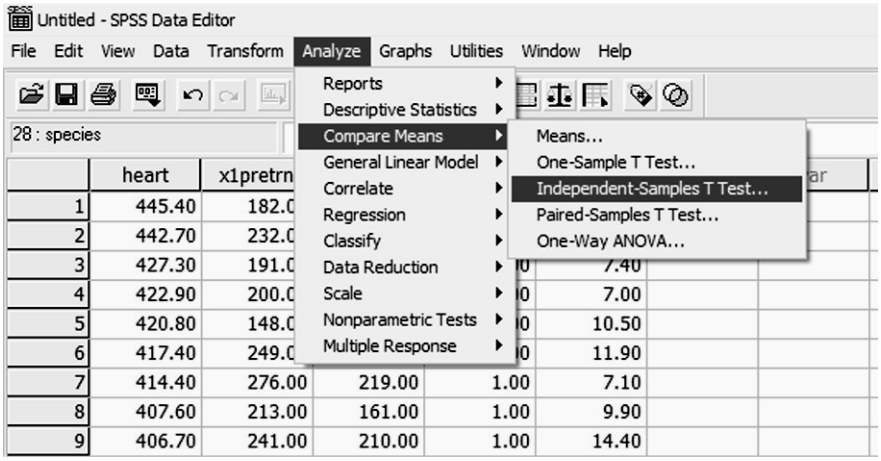
การทดสอบจะใช้ทดสอบค่าเฉลี่ยของตัวอย่างที่แบ่งออกเป็น 2 กลุ่ม แต่ละกลุ่มได้รับ treatment ต่างกัน ในการป้อนข้อมูล ต้องใช้ตัวแปรอย่างน้อย 2 ตัวแปร โดยตัวแปรแรกเป็นตัวแปรระบุกลุ่มตัวอย่าง ซึ่งมีได้เพียง 2 กลุ่มเท่านั้น ส่วนตัวแปรที่สองเป็นตัวแปรสำหรับข้อมูลดิบที่ได้จากการทดลองของแต่ละตัวอย่าง

การวิเคราะห์ทำได้ดังนี้

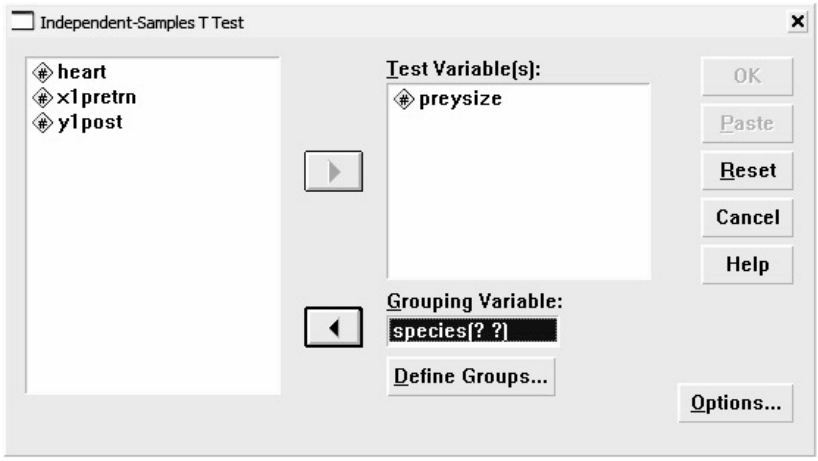
1. เลือก Analyze, Compare Means, Independent-Samples T Test… (ภาพที่ 7.5)
2. จาก dialog box (ภาพที่ 7.6)
   a. เลือกตัวแปรที่ต้องการทดสอบในช่องด้านซ้าย แล้วกดปุ่ม ▸ ที่ช่อง Test Variable(s):
   b. เลือกตัวแปรระบุกลุ่มตัวอย่าง แล้วกดปุ่ม ▸ ที่ช่อง Grouping Variable: ซึ่งจะสังเกตเห็นว่าชื่อตัวแปรระบุกลุ่มตัวอย่างนี้ จะมี (?,?) ต่อท้าย
   c. ระบุกลุ่มตัวอย่างแต่ละกลุ่มโดยกดปุ่ม Define Groups…
   d. จาก dialog box (ภาพที่ 7.7) เลือก Use Specify Values แล้วระบุหมายเลขของกลุ่มตัวอย่างทั้งสองที่ช่องหลัง Group 1: และ Group 2: ตามลำดับ แล้วกดปุ่ม Continue

e. หากต้องการกำหนดค่านัยสำคัญของ confidence interval ของค่าความแตกต่างของค่าเฉลี่ย กดที่ Options... ให้กำหนดค่าในช่อง Confidence Interval: (ภาพที่ 7.8) แล้วกด Continue

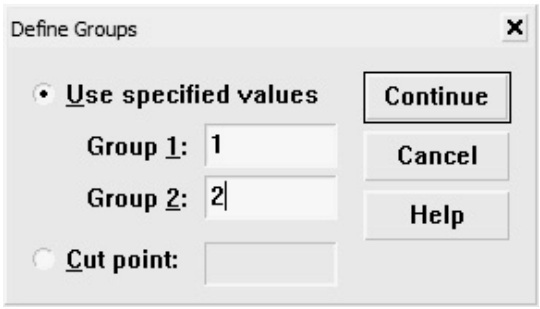
3. กด OK จะได้ผลดังภาพที่ 7.9



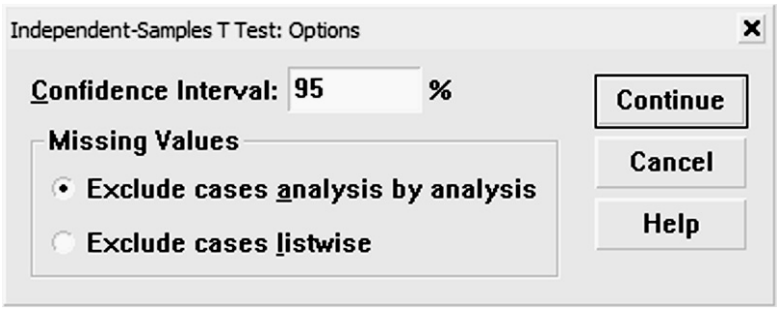
ภาพที่ 7.5



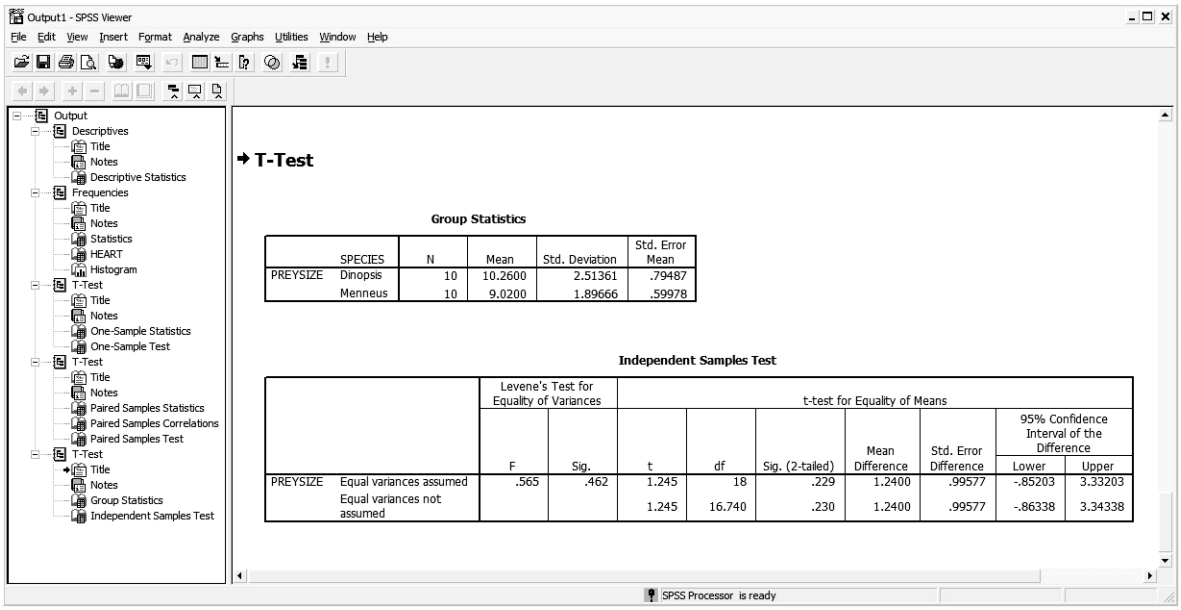
ภาพที่ 7.6



ภาพที่ 7.7

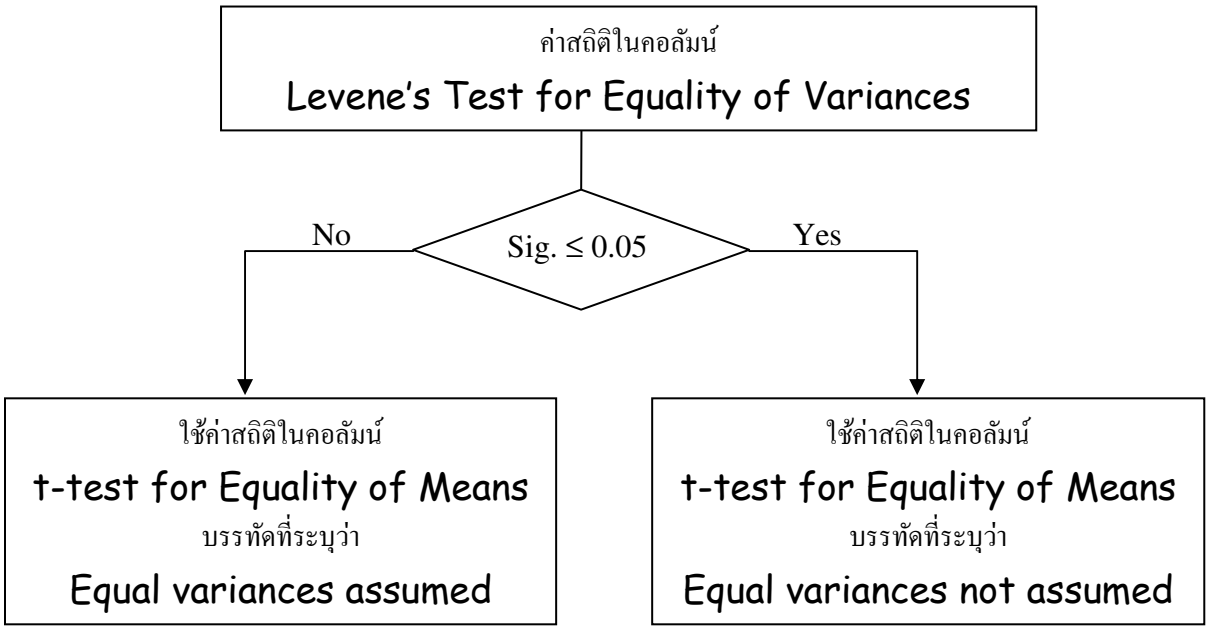


ภาพที่ 7.8



ภาพที่ 7.9

จากตัวอย่างผลการวิเคราะห์ในภาพที่ 7.9 จะเห็นว่ามี 2 ตาราง โดยตารางแรก เป็นค่าสถิติพื้นฐานของกลุ่มตัวอย่างแต่ละกลุ่ม (ในที่นี้คือค่าเฉลี่ยของ Dinopsis และ Menneus) ส่วนตารางที่ 2 เป็นผลการทดสอบ ซึ่งต้องเลือกค่าไปใช้ให้ถูกต้อง ดังนี้



ในการทดสอบสมมุติฐานจะแบ่งเป็นกรณีต่างๆ เมื่อ μ_1 และ μ_2 คือ ค่าเฉลี่ยของตัวอย่างกลุ่มที่ 1 และ 2 ตามลำดับ

กรณีที่ 1 H1: $\mu_1 - \mu_2 \neq 0$ ปฏิเสธ H0 เมื่อ Sig. $\leq$ 0.05 (หรือค่า α ที่ตั้งไว้)

กรณีที่ 2 H1: $\mu_1 - \mu_2 < 0$ ปฏิเสธ H0 เมื่อ ½ Sig. $\leq$ 0.05 (หรือค่า α ที่ตั้งไว้) และค่า t เป็น ลบ

กรณีที่ 3 H1: $\mu_1 - \mu_2 > 0$ ปฏิเสธ H0 เมื่อ ½ Sig. $\leq$ 0.05 (หรือค่า α ที่ตั้งไว้) และค่า t เป็น บวก

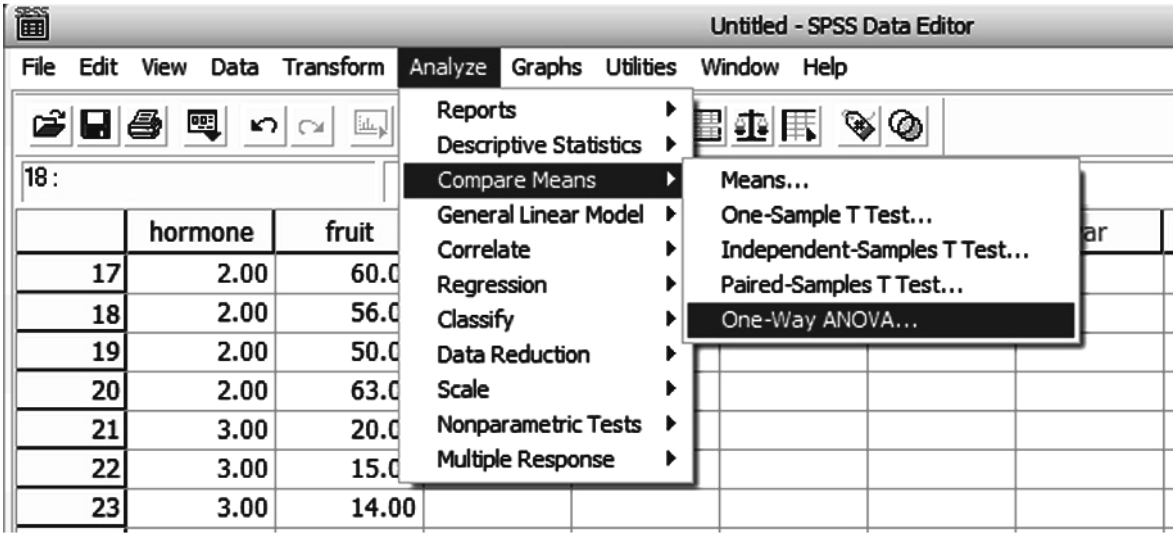
2. การทดสอบสมมุติฐานความแตกต่างของค่าเฉลี่ยของกลุ่มตัวอย่างตั้งแต่ 3 กลุ่มขึ้นไป

ในการเปรียบเทียบค่าเฉลี่ยจากกลุ่มตัวอย่างตั้งแต่ 3 กลุ่มขึ้นไปนั้น จะใช้การวิเคราะห์ที่แตกต่างไปจากที่กล่าวมาข้างต้น โดยจะใช้การวิเคราะห์ความแปรปรวนหรือ ANOVA (**AN**alysis **O**f **VA**riance) ANOVA เป็นการทดสอบค่าเฉลี่ยของกลุ่มตัวอย่างว่าเท่ากันหรือไม่ โดยการเปรียบเทียบความแปรปรวนของตัวอย่างที่ประมาณมาจากค่าเฉลี่ยของกลุ่มกับความแปรปรวนภายในกลุ่มตัวอย่าง อีกประการคือ เมื่อพบว่าในการเปรียบเทียบค่าเฉลี่ย มีค่าเฉลี่ยบางคู่แตกต่างกันอย่างมีนัยสำคัญทางสถิติแล้ว จะต้องหาว่าค่าเฉลี่ยใดที่แตกต่างกันบ้าง

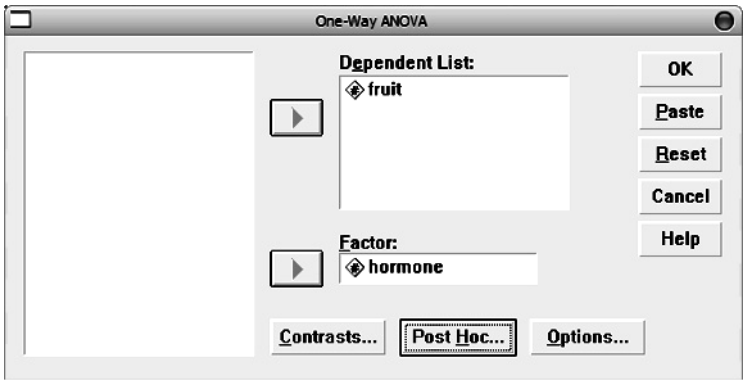
ในการเตรียมข้อมูล จะต้องใช้ตัวแปรอย่างน้อย 2 ตัวแปรเช่นเดียวกับกรณี Independent-samples t test โดยตัวแปรแรกใช้ระบุกลุ่มของค่าข้อมูลดิบ ส่วนตัวแปรที่สองใช้เก็บค่าข้อมูลดิบ

สำหรับขั้นตอนในการวิเคราะห์ จะมีขั้นตอนดังนี้

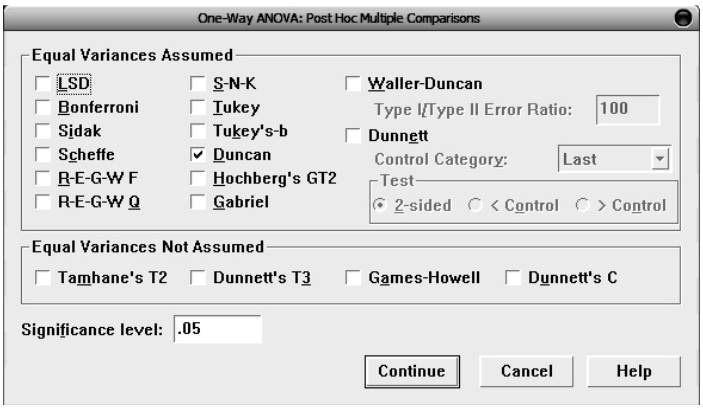
1. เลือก Analyze, Compare Means, One-Way ANOVA... (ภาพที่ 7.10)
2. จาก dialog box (ภาพที่ 7.11)
    a. เลือกตัวแปรเก็บข้อมูลดิบที่จะทดสอบค่าเฉลี่ยจากช่องด้านซ้าย แล้วกดปุ่ม ▶ ของช่อง Dependent List:
    b. เลือกตัวแปรระบุกลุ่มจากช่องด้านซ้าย แล้วกดปุ่ม ▶ ของช่อง Factor:
    c. กดปุ่ม Post Hoc... เพื่อหาค่าเฉลี่ยที่แตกต่างกัน จาก dialog box (ภาพที่ 7.12)
        i. โดยทั่วไปมักจะมีสมมุติฐานว่าความแปรปรวน (variance) ของกลุ่มตัวอย่างแต่ละกลุ่มมีค่าไม่แตกต่างกัน ดังนั้น จึงเลือกวิธีการเปรียบเทียบค่าเฉลี่ยจากตัวเลือกในช่อง Equal Variances Assumed (หากในผลการวิเคราะห์พบว่าความแปรปรวนของกลุ่มตัวอย่างต่างกันแล้ว ให้เลือกวิธีการเปรียบเทียบค่าเฉลี่ยจากตัวเลือกในช่อง Equal Variances Not Assumed แทน)
        ii. กำหนดค่า Significance Level: แล้วกดปุ่ม Continue
    d. หากต้องการค่าทางสถิติพื้นฐานและกราฟค่าเฉลี่ยของกลุ่มตัวอย่าง กดปุ่ม Options... แล้วเลือกค่าสถิติที่ต้องการใน dialog box (ภาพที่ 7.13) แล้วปุ่ม Continue
3. กดปุ่ม OK จะได้ผลการวิเคราะห์ดังภาพที่ 7.14-7.17



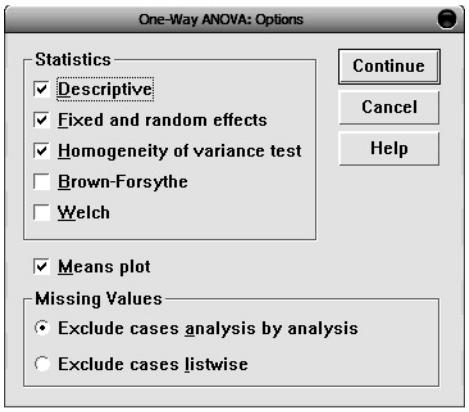
ภาพที่ 7.10



ภาพที่ 7.11



ภาพที่ 7.12



ภาพที่ 7.13

ส่วนที่สำคัญสำหรับการแสดงผลการวิเคราะห์และทดสอบประกอบด้วย

ก.) ตารางค่าสถิติต่างๆ ของกลุ่มตัวอย่าง (ตาราง Descriptives ดังภาพที่ 7.14)

Descriptives

FRUIT

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum	Between-Component Variance
						Lower Bound	Upper Bound			
A		10	89.0000	4.08248	1.29099	86.0796	91.9204	82.00	95.00	
B		10	59.5000	6.29374	1.99025	54.9977	64.0023	50.00	70.00	
C		10	19.9000	5.02107	1.58780	16.3081	23.4919	14.00	28.00	
D		10	51.1000	4.74810	1.50148	47.7034	54.4966	45.00	60.00	
Total		40	54.8750	25.40461	4.01682	46.7502	62.9998	14.00	95.00	
Model	Fixed Effects			5.09984	.80635	53.2396	56.5104			
	Random Effects				14.21093	9.6495	100.1005			805.20167

ภาพที่ 7.14

ข.) ตารางแสดงผลการทดสอบความเท่ากันของค่าความแปรปรวน (variance) ของกลุ่มตัวอย่าง (ตาราง Test of Homogeneity of Variances ดังภาพที่ 7.15) จากตาราง ค่า Sig. > 0.05 หมายถึง ความแปรปรวนของกลุ่มตัวอย่างทุกกลุ่มมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

Test of Homogeneity of Variances

FRUIT

Levene Statistic	df1	df2	Sig.
.761	3	36	.523

ภาพที่ 7.15

ค.) ตาราง ANOVA (ตาราง ANOVA ดังภาพที่ 7.16) จากตาราง ค่า Sig. < 0.05 หมายถึง ค่าเฉลี่ยของกลุ่มตัวอย่างอย่างน้อย 1 คู่ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ANOVA

FRUIT

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	24234.08	3	8078.025	310.594	.000
Within Groups	936.300	36	26.008		
Total	25170.38	39			

ภาพที่ 7.16

ง.) ตารางการเปรียบเทียบค่าเฉลี่ย (Post Hoc Multiple Comparisons) ซึ่งจะแสดงแยกตามวิธีการเปรียบเทียบ ในตัวอย่างนี้ เลือกวิธี Duncan เพียงวิธีเดียว (ภาพที่ 7.12) จึงมีเพียงตารางเดียว ดังภาพที่ 7.17 โดยมีชื่อตัวแปรที่เปรียบเทียบค่าเฉลี่ยอยู่เหนือตาราง (ตารางดังภาพที่ 7.17 ของ FRUIT) จากผลการเปรียบเทียบเห็นได้ว่ามี 4 subset แสดงว่าค่าเฉลี่ยทั้งสี่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

FRUIT

Duncan[a]

		Subset for alpha = .05			
HORMONE	N	1	2	3	4
C	10	19.9000			
D	10		51.1000		
B	10			59.5000	
A	10				89.0000
Sig.		1.000	1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 10.000.

ภาพที่ 7.17

อนึ่ง หากว่าผลจาก ANOVA บ่งชี้ว่าค่าเฉลี่ยของกลุ่มไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ (นั้นคือ Sig. ในตาราง ANOVA มีค่ามากกว่า 0.05) แล้ว จะไม่พิจารณาผลที่ได้จากการเปรียบเทียบค่าเฉลี่ยแต่อย่างใด

ในการนำเสนอผลการทดสอบค่าเฉลี่ยนั้น มักนำเสนอในรูปตารางหรือกราฟ แต่หากเป็นตาราง มักจะเรียงค่าเฉลี่ยจากค่ามากไปหาค่าน้อยหรือจากค่าน้อยไปหาค่ามากแล้วแต่กรณี โดยไม่สนใจว่ากลุ่มหรือ treatment จะเรียงอย่างไร พร้อมกันนั้น จะมีการระบุด้วยตัวอักษรหลังค่าเฉลี่ยเพื่อให้ทราบค่าเฉลี่ยใดแตกต่างกันบ้าง ดังตัวอย่าง

Hormone	Mean	Statistical test
C	19.9	a
D	51.1	b
B	59.5	c
A	89.0	d

หมายเหตุ – ค่าเฉลี่ยที่มีอักษรต่างกัน หมายความว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ความเชื่อมั่น 95 เปอร์เซ็นต์

จ.) สำหรับตารางอื่นๆ นั้น ขึ้นอยู่กับว่าจะกำหนดให้ SPSS วิเคราะห์/เปรียบเทียบค่าใดบ้าง ก็จะมีตารางออกมาเพิ่มเติม รวมทั้งกราฟด้วย

แบบฝึกหัดที่ 7 การทดสอบสมมุติฐาน

7.1 ในการศึกษาเพื่อหาปัจจัยในร่างกายของนักกีฬาหญิงที่ส่งผลต่อนักกีฬาว่ายน้ำและนักกีฬาวิ่งมาราธอน ผู้วิจัยได้รวบรวมข้อมูลปริมาณไขมันในร่างกายซึ่งเป็นปัจจัยหนึ่ง และต้องการทดสอบว่าปริมาณไขมันในร่างกายของนักกีฬาทั้งสองประเภท ต่างกันหรือไม่ จากการเก็บข้อมูลในนักกีฬาทั้งสองประเภท ดังนี้

นักกีฬาวิ่งมาราธอน	11.2	10.1	9.4	9.2	8.3	8.2	7.6	7.3	6.9	5.5	5.0	3.7
นักกีฬาว่ายน้ำ	14.1	15.1	11.4	14.3	9.2	12.7	13.7	11.9	10.7	8.7		

7.1.1 ตั้งสมมุติฐานทั้ง Null hypothesis และ Alternative hypothesis

7.1.2 เขียนค่าสถิติที่ได้จากการวิเคราะห์ด้วยโปรแกรม SPSS กำหนดให้ $\alpha = 0.05$

7.1.3 จากผลการทดสอบสมมุติฐาน สรุปได้ว่าอย่างไร

7.2 ผู้วิจัยต้องการทราบถึงผลของการออกกำลังกายต่อปริมาณ triglyceride ในเลือด จึงได้เลือกอาสาสมัคร 11 คน แล้ววัดปริมาณ triglyceride ก่อนเริ่มการทดลอง จากนั้นให้อาสาสมัครทั้ง 11 คน เข้ารับการออกกำลังอย่างต่อเนื่องเป็นเวลา 3 เดือน แล้ววัดปริมาณ triglyceride อีกครั้งหนึ่ง ได้ผลดังนี้

Triglyceride	อาสาสมัครคนที่										
(mg/100 mL)	1	2	3	4	5	6	7	8	9	10	11
ก่อนเข้า course	68	77	94	73	37	131	77	24	99	629	116
หลังเข้า course	95	90	86	58	47	121	136	65	131	630	104

7.2.1 ตั้งสมมุติฐานทั้ง Null hypothesis และ Alternative hypothesis

7.2.2 เขียนค่าสถิติที่ได้จากการวิเคราะห์ด้วยโปรแกรม SPSS กำหนดให้ $\alpha = 0.05$

7.2.3 จากผลการทดสอบสมมุติฐาน สรุปได้ว่าอย่างไร

ในข้อ 7.3 และ 7.4 ให้นิสิตปฏิบัติดังนี้

ก) จงเขียนสมมุติฐาน H_0 และ H_1

ข) จงหาค่าสถิติที่จำเป็นต้องใช้เพื่อทดสอบสมมุติฐาน กำหนดให้ $\alpha = 0.05$

ค) เขียนตาราง ANOVA และการเปรียบเทียบค่าเฉลี่ย (ถ้ามี) พร้อมสรุปผลการทดสอบ

7.3 ในการทดสอบการให้ฮอร์โมนพืชเพื่อเพิ่มผลผลิตของไม้ผลชนิดหนึ่งด้วยวิธีการ 4 แบบ แล้ววัดผลผลิตที่ได้เมื่อถึงฤดูเก็บเกี่ยว ได้ผลดังนี้

	ผลผลิต (พันกิโลกรัม) จากวิธีการให้ฮอร์โมนแบบต่างๆ			
ต้นที่	A	B	C	D
1	89	59	20	51
2	87	56	15	47
3	84	52	14	45
4	92	67	26	56
5	95	70	28	60
6	90	62	22	53
7	89	60	19	51
8	88	56	17	50
9	82	50	14	45
10	94	63	24	53

7.4 ในการคัดเลือกถ่านหินเพื่อใช้ในการผลิตไฟฟ้า ผู้วิจัยต้องการทราบว่าถ่านหินจากแหล่งต่างๆ มีปริมาณซัลเฟอร์ต่างกันหรือไม่ จึงทดลองวัดค่าปริมาณซัลเฟอร์จากแหล่งต่างๆ ได้ผลดังนี้

Coal core location				
I	II	III	IV	V
1.51	1.69	1.56	1.30	0.73
1.92	0.64	1.22	0.75	0.80
1.08	0.90	1.32	1.26	0.90
2.04	1.41	1.39	0.69	1.24
2.14	1.01	1.33	0.62	0.82
1.76	0.84	1.54	0.90	0.72
1.17	1.28	1.04	1.20	0.57
	1.59	2.25	0.32	1.18
		1.49		0.54
				1.30

การบ้านปฏิบัติการที่ 7 การทดสอบสมมุติฐาน 2

7.1. จักษุแพทย์ผู้หนึ่งสงสัยว่าค่า Spherical refraction ในตาข้างซ้ายและข้างขวา มีความแตกต่างกันหรือไม่ จึงได้วัดค่าดังกล่าวจากอาสาสมัคร 17 คน

คนที่	ตาซ้าย	ตาขวา	คนที่	ตาซ้าย	ตาขวา
1	1.75	2.00	**9**	0.00	0.50
2	-4.00	-4.00	**10**	-1.00	-1.25
3	-1.25	-1.00	**11**	0.50	-1.75
4	1.00	1.00	**12**	-8.50	-5.00
5	-1.00	-1.00	**13**	0.50	0.50
6	-0.75	0.25	**14**	-5.25	-4.75
7	-2.25	-2.25	**15**	-2.25	-2.50
8	0.25	0.25	**16**	-6.50	-6.25
			17	1.75	1.75

ก) นิสิตคิดว่าควรใช้การทดสอบสมมุติฐานแบบใด? one-sample test หรือ two-samples test แบบใด?

ข) ควรตั้งสมมุติแบบ nondirectional หรือ directional?

ค) Null hypothesis และ Alternative hypothesis เขียนได้ว่าอย่างไร?

ง) ให้ทำการทดสอบสมมุติฐานโดยแสดงค่าสถิติที่จำเป็น กำหนดให้ α เป็น 0.05 และสรุปได้ว่าอย่างไร?

7.2. ในการศึกษาเพื่อเปรียบเทียบผลของยา 2 ชนิด ต่อการลดจำนวนสิวที่เกิดขึ้น ได้ผลดังนี้

	จำนวนสิวเฉลี่ย									
ยา A	48.6	49.4	50.1	49.8	50.6	50.8	47.1	52.5	49.0	46.7
ยา B	68.0	67.0	70.1	64.5	68.0	68.3	71.9	71.5	69.9	68.9

ก) นิสิตคิดว่าควรใช้การทดสอบสมมุติฐานแบบใด? one-sample test หรือ two-samples test แบบใด?

ข) ควรตั้งสมมุติแบบ nondirectional หรือ directional?

ค) Null hypothesis และ Alternative hypothesis เขียนได้ว่าอย่างไร?

ง) ให้ทำการทดสอบสมมุติฐานโดยแสดงค่าสถิติที่จำเป็น กำหนดให้ α เป็น 0.05 และสรุปได้ว่าอย่างไร?

ในข้อ 7.3 และ 7.4 ให้นิสิตปฏิบัติดังนี้

ก) จงเขียนสมมุติฐาน H_0 และ H_1

ข) จงหาค่าสถิติที่จำเป็นต้องใช้เพื่อทดสอบสมมุติฐาน กำหนดให้ α=0.05

ค) เขียนตาราง ANOVA และการเปรียบเทียบค่าเฉลี่ย (ถ้ามี) พร้อมสรุปผลการทดสอบ

7.3 ในการศึกษาถึงผลของ CO_2 ต่อการเจริญของแบคทีเรีย *Pseudomanas gragi* ที่เป็นเชื้อสาเหตุที่ทำให้อาหารเป็นพิษ ได้ผลการศึกษาดังนี้

จำนวนโคโลนีเฉลี่ยในอาหารเลี้ยงที่บ่มเลี้ยงในตู้ที่ความเข้มข้นของ CO_2 ต่างๆ				
0	0.083	0.29	0.5	0.86
62.6	50.9	45.5	29.5	24.9
59.6	44.3	41.1	22.8	17.2
64.5	47.5	29.8	19.2	7.8
59.3	49.5	38.3	20.6	10.5
58.6	48.5	40.2	29.2	17.8
64.6	50.4	38.5	24.1	22.1
50.9	35.2	30.2	22.6	22.6
56.2	49.9	27.0	32.7	16.8
52.3	42.6	40.0	24.4	15.9
62.8	41.6	33.9	29.6	8.8

7.4 ในการตรวจสอบตำแหน่งของแหล่งน้ำที่ใช้ในการเลี้ยงปลาว่ามีปริมาณตะกั่วปนเปื้อนต่างกันหรือไม่ หากแหล่งน้ำดังกล่าวอยู่ห่างจากแหล่งโรงงานผลิตธาตุโซเดียมที่ระยะต่างๆ กัน

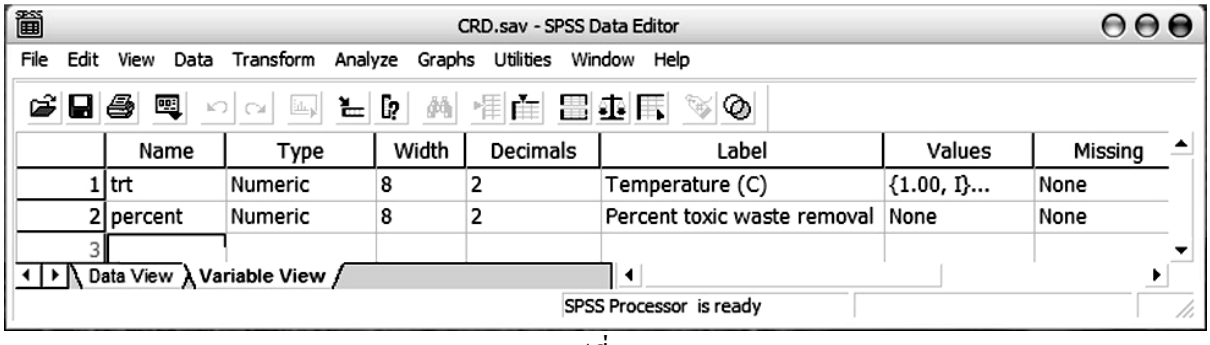
ปริมาณตะกั่วปนเปื้อนในแหล่งน้ำจากตำแหน่งต่าง ๆ			
I	II	III	IV
3.75 กม.เหนือโรงงาน	1.5 กม. เหนือโรงงาน	2.5 กม. ใต้โรงงาน	5.0 กม. ใต้โรงงาน
0.45	1.64	1.56	0.65
0.35	1.67	1.55	0.59
0.32	1.85	1.69	0.69
0.68	1.57	1.67	0.62
0.53	1.59	1.60	0.70
0.34	1.61	1.68	0.64
0.61	1.53	1.65	0.81
0.41		1.59	0.58
0.51		1.75	
		1.49	

ปฏิบัติการที่ 8 General Linear Model for CRD and RCBD

GLM เป็นการวิเคราะห์ทางสถิติที่พิจารณาผล (effect) อันเกิดจากชุดตัวแปรในคราวเดียวกันและร่วมกันเหมือนกับการวิเคราะห์ถดถอย (regression analysis) ในการใช้ GLM จะต้องกำหนดแบบจำลองเพื่อระบุ effect model ว่าเป็นแบบใดเพื่อให้ GLM นำไปหาค่า effect จากชุดตัวแปร (ซึ่งจะเป็น parameter) ของแบบจำลองนั้นๆ ในปฏิบัติการ จะกล่าวถึงการใช้ GLM เพื่อการวิเคราะห์ทางสถิติที่มีการวางแบบการทดลองแบบ CRD และ RCBD

1. GLM สำหรับ CRD

ในการวิเคราะห์ข้อมูลที่ได้จากการทดลองที่มีการวางแบบการทดลองแบบ CRD นั้น จะต้องสร้างตัวแปร 2 ตัว โดยตัวแปรแรกเป็นตัวแปรระบุ treatment ซึ่งอาจเรียก independent variable, fixed factor, grouping variable ก็ได้ ส่วนตัวแปรที่สอง คือตัวแปรที่จะทดสอบ ซึ่งอาจเรียกว่า dependent variable ก็ได้ ในตัวอย่าง (รูปที่ 8.1) ตัวแปร trt คือ independent variable และตัวแปร percent เป็น dependent variable



รูปที่ 8.1

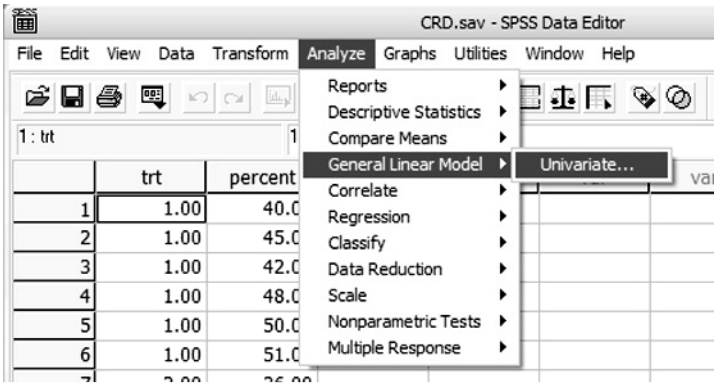
แบบจำลองของการวางแผนการทดลองแบบ CRD สามารถเขียนได้ดังนี้

$$X_{ij} = \mu + \tau_{i.} + \varepsilon_{ij}$$

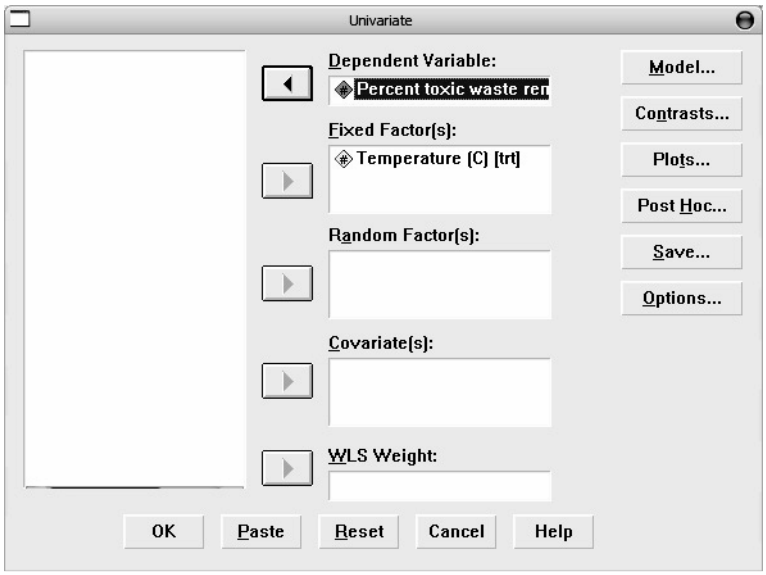
เมื่อ μ คือ ค่าเฉลี่ยโดยรวมของการทดลอง τ คือ ผลจาก treatment และ ε คือ random error

ในการวิเคราะห์ข้อมูล ทำดังนี้

1. เลือก Analyze, General Linear Model, Univariate... (รูปที่ 8.2)
2. จาก dialog box (รูปที่ 8.3)
    a. เลือกตัวแปรที่ระบุกลุ่มในช่องด้านซ้าย กด ▸ ของช่อง Fixed Factor(s):
    b. เลือกตัวแปรที่จะทดสอบค่าในช่องด้านซ้าย กด ▸ ของช่อง Dependent variable(s):

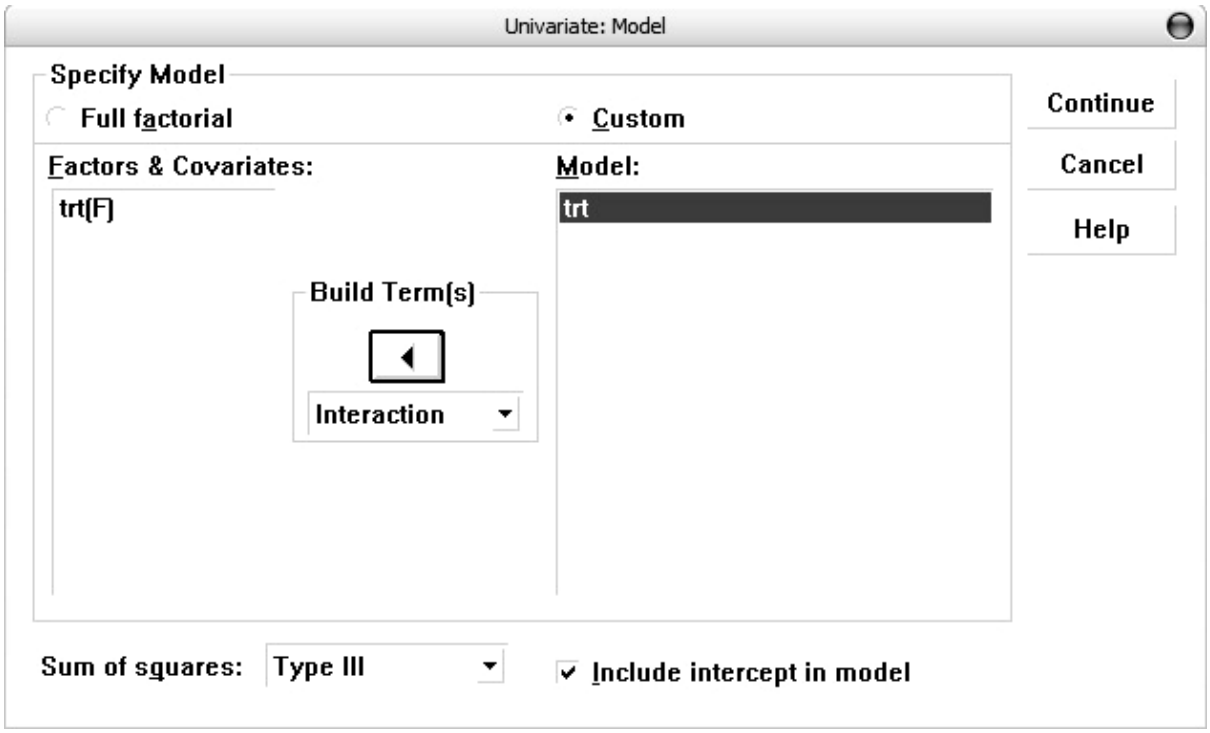


รูปที่ 8.2



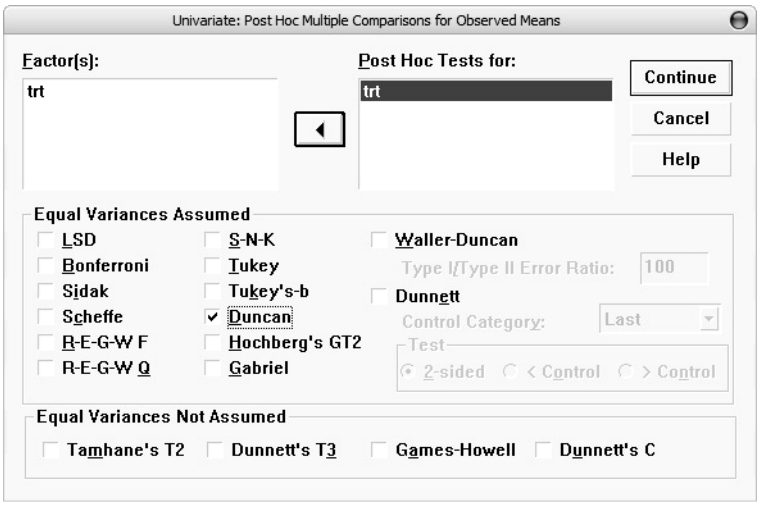
รูปที่ 8.3

c. กดปุ่ม Model... เพื่อกำหนดแบบจำลอง จะมี dialog box ดังรูปที่ 8.4
   i. ที่ช่อง Specify Model เลือก Custom
   ii. ที่ช่อง Factors & Covariates: ให้เลือกตัวแปรระบุกลุ่มในช่องด้านซ้าย แล้วกดปุ่ม ▸ ที่ช่อง Built Term(s): สังเกตว่าตัวแปรมาอยู่ที่ช่อง Model:
   iii. เลือก Sum of Squares: เป็น Type III และเลือก Include intercept in model
   iv. กดปุ่ม Continue

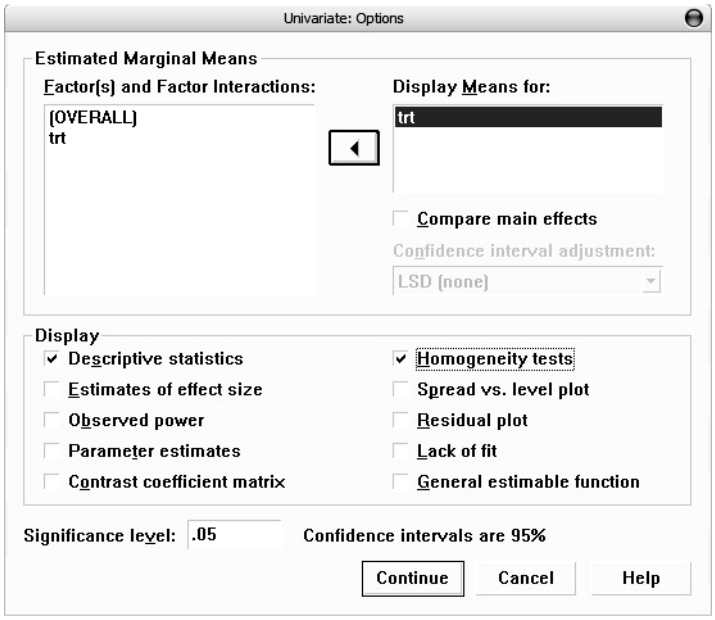


รูปที่ 8.4

d. กดปุ่ม Post Hoc... เพื่อหาค่าเฉลี่ยที่แตกต่างกัน จะมี dialog box ดังรูปที่ 8.5
   i. เลือกตัวแปรระบุกลุ่มในช่อง Factor(s): กดปุ่ม ▸ ของช่อง Post Hoc Test for:
   ii. เลือกแบบการเปรียบเทียบค่าเฉลี่ยในช่อง Equal Variance Assumed ที่ต้องการ อนึ่ง หากผลการวิเคราะห์บ่งชี้ว่าความแปรปรวนระหว่างกลุ่มตัวอย่างไม่เท่ากัน ก็ต้องเลือกแบบการเปรียบเทียบค่าเฉลี่ยในช่อง Equal Variance Not Assumed แทน
   iii. กดปุ่ม Continue
e. กดปุ่ม Options... หากต้องการค่าสถิติอื่น ๆ จะมี dialog box ดังรูปที่ 8.6
   i. เลือกตัวแปรระบุกลุ่มในช่อง Factor(s) and Factor Interaction: กดปุ่ม ▸ ของช่อง Display Means for:
   ii. เลือกค่าสถิติที่ต้องการจากตัวเลือกในช่อง Display
   iii. กำหนดค่า Significant Level: ที่ต้องการ
   iv. กดปุ่ม Continue
f. กดปุ่ม OK เพื่อทำการวิเคราะห์



รูปที่ 8.5



รูปที่ 8.6

ผลการวิเคราะห์ที่ได้จะแตกต่างไปขึ้นอยู่กับตัวเลือกต่าง ๆ ในตัวอย่างนี้ จะแสดงผลที่สำคัญดังตารางในรูปที่ 8.7-8.11 ทั้งนี้ สิ่งที่ต้องพิจารณาจากคือผลการทดสอบ Error variance (ของ dependent varable) ในแต่ละ treatment ว่าเท่ากันหรือไม่ จากตารางในรูปที่ 8.9 จะเห็นว่าค่า Sig. < 0.05 แสดงว่า Error variance มีค่าไม่เท่ากันอย่างน้อย 1 คู่ ซึ่งผลจากตาราง จะช่วยในเลือกใช้แบบการเปรียบเทียบค่าเฉลี่ยใน Post Hoc test

Descriptive Statistics

Dependent Variable: Percent toxic waste removal

Temperature (C)	Mean	Std. Deviation	N
I	46.0000	4.42719	6
II	38.6667	2.16025	6
III	51.3333	1.63299	6
IV	50.0000	1.78885	6
V	60.0000	2.82843	6
Total	49.2000	7.54024	30

รูปที่ 8.7 ค่าสถิติบางส่วนของแต่ละกลุ่ม (treatment)

Temperature (C)

Dependent Variable: Percent toxic waste removal

Temperature (C)	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
I	46.000	1.127	43.678	48.322
II	38.667	1.127	36.345	40.989
III	51.333	1.127	49.011	53.655
IV	50.000	1.127	47.678	52.322
V	60.000	1.127	57.678	62.322

รูปที่ 8.8 ช่วงความเชื่อมั่นของค่าเฉลี่ยแต่ละกลุ่ม (treatment)

Levene's Test of Equality of Error Variances[a]

Dependent Variable: Percent toxic waste removal

F	df1	df2	Sig.
2.951	4	25	.040

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+TRT

รูปที่ 8.9 ผลการทดสอบ Equality of Error Variances

Tests of Between-Subjects Effects

Dependent Variable: Percent toxic waste removal

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1458.133[a]	4	364.533	47.797	.000
Intercept	72619.200	1	72619.200	9521.748	.000
TRT	1458.133	4	364.533	47.797	.000
Error	190.667	25	7.627		
Total	74268.000	30			
Corrected Total	1648.800	29			

a. R Squared = .884 (Adjusted R Squared = .866)

รูปที่ 8.10 ตาราง ANOVA ของ treatment ต่างๆ

Percent toxic waste removal

Duncan[a,b]

Temperature (C)	N	Subset 1	2	3	4
II	6	38.6667			
I	6		46.0000		
IV	6			50.0000	
III	6			51.3333	
V	6				60.0000
Sig.		1.000	1.000	.411	1.000

Means for groups in homogeneous subsets are displayed.
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 7.627.

a. Uses Harmonic Mean Sample Size = 6.000.

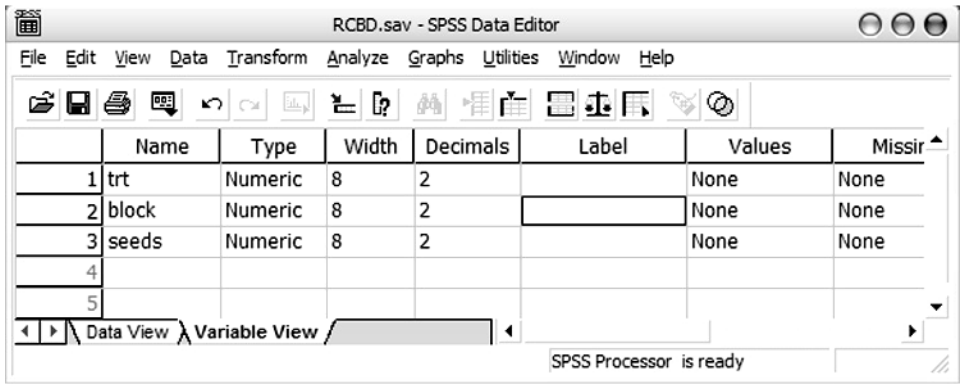
b. Alpha = .05.

รูปที่ 8.11 ผลการเปรียบเทียบค่าเฉลี่ยแบบ DMRT

สิ่งที่ต้องพิจารณาต่อไปคือ ตารางในรูปที่ 8.10 คือ ตาราง ANOVA โดยค่าที่จะนำไปเสนอและพิจารณา คือ ค่าที่ระบายแถบดำ จากผลการวิเคราะห์ จะเห็นว่า ค่า Sig. ของตัวแปรระบุกลุ่ม (ในตัวอย่าง คือ TRT) มีค่าน้อยกว่า 0.05 แสดงว่ามีค่าเฉลี่ยอย่างน้อย 1 คู่ ที่แตกต่างกัน หากต้องการทราบว่าค่าเฉลี่ยคู่ใดต่างกัน ให้ดูจากตารางผลการเปรียบเทียบค่าเฉลี่ย ดังตัวอย่างในรูปที่ 8.11[1]

2. GLM สำหรับ RCBD

ในการวิเคราะห์ข้อมูลที่ได้จากการทดลองที่มีการวางแบบการทดลองแบบ RCBD นั้น จะต้องสร้างตัวแปร 3 ตัว โดยตัวแปรแรกเป็นตัวแปรระบุ treatment ซึ่งอาจเรียก independent variable, fixed factor, grouping variable ก็ได้ ส่วนตัวแปรที่สอง คือ ตัวแปรระบุ block และตัวแปรที่สาม คือตัวแปรที่จะทดสอบ ซึ่งอาจเรียกว่า dependent variable ก็ได้ ในตัวอย่าง (รูปที่ 8.12) ตัวแปร trt คือ independent variable ตัวแปร block คือ block และตัวแปร seeds เป็น dependent variable



รูปที่ 8.12

[1] ในตัวอย่างนี้ ไม่ควรใช้ DMRT เพราะเป็นแบบการเปรียบเทียบค่าเฉลี่ยกรณีที่ Equal Variances Assumed แต่ควรใช้แบบการเปรียบเทียบค่าเฉลี่ยที่ Equal Variances Not Assumed แทน เนื่องจากผลการวิเคราะห์ในตารางของรูปที่ 8.9 ระบุว่า Variance ของตัวแปรที่ต้องการทดสอบค่านั้นต่างกันอย่างน้อย 1 คู่

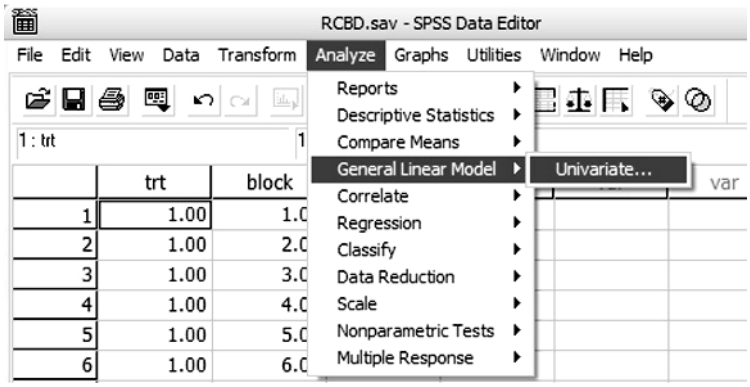
แบบจำลองของการวางแผนการทดลองแบบ RCBD สามารถเขียนได้ดังนี้

$$X_{ij} = \mu + \tau_{i.} + \beta_{.j} + \varepsilon_{ij}$$

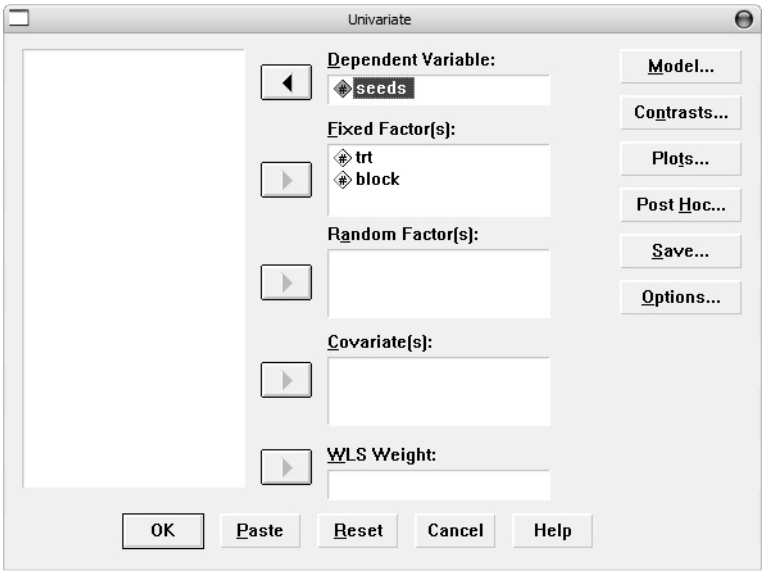
เมื่อ μ คือ ค่าเฉลี่ยโดยรวมของการทดลอง τ คือ ผลจาก treatment ส่วน β คือ ผลจาก Block และ ε คือ random error

ในการวิเคราะห์ข้อมูล มีขั้นตอนดังนี้

1. เลือก Analyze, General Linear Model, Univariate... (รูปที่ 8.13)
2. จาก dialog box ดังรูปที่ 8.14
   a. เลือกตัวแปรที่ระบุกลุ่มและตัวแปรที่เป็น block ในช่องด้านซ้าย กด ▶ ของช่อง Fixed Factor(s):
   b. เลือกตัวแปรที่จะทดสอบค่าในช่องด้านซ้าย กด ▶ ของช่อง Dependent variable(s):

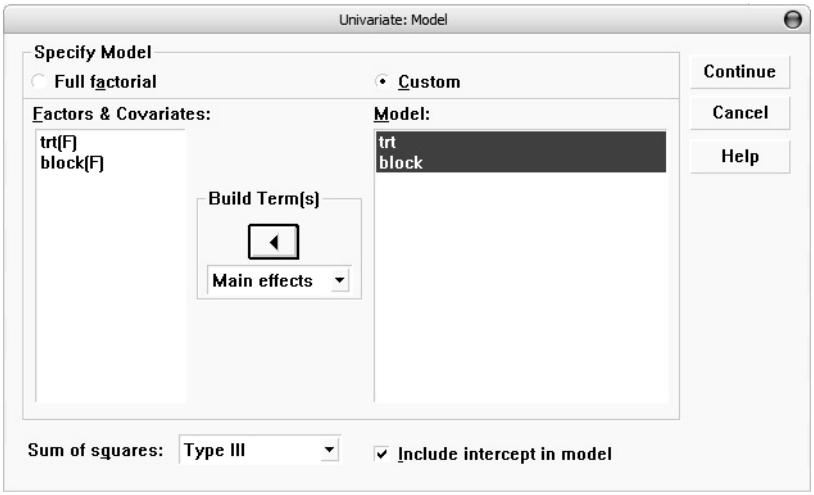


รูปที่ 8.13



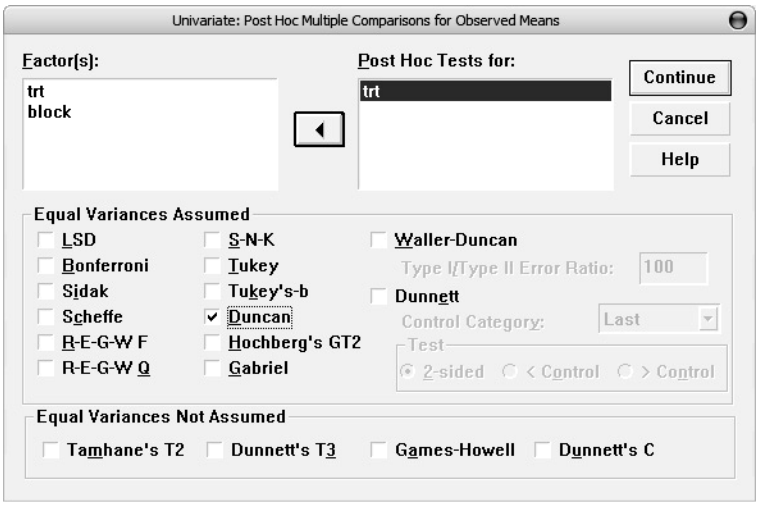
รูปที่ 8.14

c. กดปุ่ม Model... เพื่อกำหนดแบบจำลอง จะมี dialog box ดังรูปที่ 8.15
   i. ที่ช่อง Specify Model เลือก Custom
   ii. ที่ช่อง Factors & Covariates: ให้เลือกตัวแปรระบุกลุ่มในช่องด้านซ้าย แล้วกดปุ่ม ▸ ของช่อง Build Term(s):
   iii. ที่ช่อง Factors & Covariates: ให้เลือกตัวแปรที่เป็น block ในช่องด้านซ้าย แล้วกดปุ่ม ▸ ของช่อง Build Term(s): สังเกตว่าในช่อง Model: จะมีตัวแปรจาก ii. และ iii. อยู่บรรทัดละ 1 ตัวแปร
   iv. เลือก Sum of Squares: เป็น Type III และเลือก Include intercept in model
   v. กดปุ่ม Continue



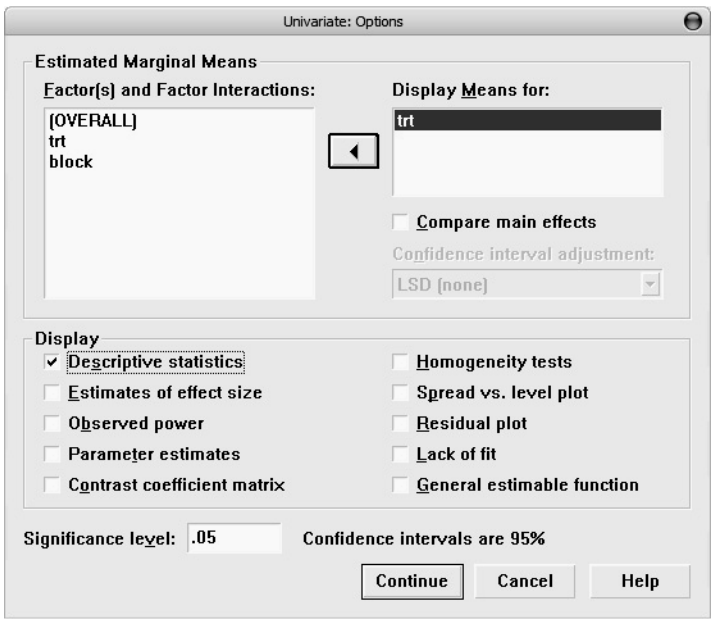
รูปที่ 8.15

d. กดปุ่ม Post Hoc... เพื่อหาค่าเฉลี่ยที่ต่างกัน จะมี dialog box ดังรูปที่ 8.16
   i. เลือกตัวแปรระบุกลุ่มในช่อง Factor(s): กดปุ่ม ▸ ของช่อง Post Hoc Test for:
   ii. เลือกแบบการเปรียบเทียบค่าเฉลี่ยในช่อง Equal Variance Assumed อนึ่ง หากผลการวิเคราะห์ชี้ว่าความแปรปรวนระหว่างกลุ่มตัวอย่างไม่เท่ากัน ก็ต้องเลือกแบบการเปรียบเทียบค่าเฉลี่ยในช่อง Equal Variance Not Assumed แทน
   iii. กดปุ่ม Continue



รูปที่ 8.16

e. กดปุ่ม Options... หากต้องการหาค่าสถิติต่าง ๆ จะมี dialog box ดังรูปที่ 8.17
   i. เลือกตัวแปรระบุกลุ่มในช่อง Factor(s) and Factor Interaction: กดปุ่ม ▸ ของช่อง Display Means for:
   ii. เลือกค่าสถิติต้องการหาในช่อง Display
   iii. กำหนดค่า Significant Level: ที่ต้องการ
   iv. กดปุ่ม Continue
f. กดปุ่ม OK เพื่อทำการวิเคราะห์



รูปที่ 8.17

ผลการวิเคราะห์ จะมีตารางคล้ายกับที่กล่าวมาในหัวข้อ ***GLM สำหรับ CRD*** โดยตาราง Descriptive Statistics จะแสดงค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐานของแต่ละ treatment และ block ส่วนตารางทดสอบค่าเฉลี่ยแต่ละ treatment หรือตาราง ANOVA แสดงดังรูปที่ 8.18 ซึ่งบรรทัดที่จะนำเสนอในตาราง ANOVA คือ บรรทัดที่ระบายสีดำ ในการทดสอบสมมุติฐานนั้น จะพิจารณาค่า Sig. ในบรรทัดของตัวแปรระบุกลุ่ม (ในตัวอย่างนี้คือ TRT) ซึ่งมีค่าน้อยกว่า 0.05 แสดงว่ามีค่าเฉลี่ยอย่างน้อย 1 คู่ ที่มีค่าแตกต่างกัน ตารางที่ 8.19 แสดงผลการเปรียบเทียบค่าเฉลี่ยด้วยวิธี DMRT

Tests of Between-Subjects Effects

Dependent Variable: SEEDS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	25120.700[a]	12	2093.392	1137.827	.000
Intercept	120450.625	1	120450.625	65468.89	.000
TRT	24234.075	3	8078.025	4390.673	.000
BLOCK	886.625	9	98.514	53.546	.000
Error	49.675	27	1.840		
Total	145621.000	40			
Corrected Total	25170.375	39			

a. R Squared = .998 (Adjusted R Squared = .997)

รูปที่ 8.18 ตาราง ANOVA

SEEDS

Duncan[a,b]

TRT	N	Subset 1	2	3	4
3.00	10	19.9000			
4.00	10		51.1000		
2.00	10			59.5000	
1.00	10				89.0000
Sig.		1.000	1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 1.840.

a. Uses Harmonic Mean Sample Size = 10.000.

b. Alpha = .05.

รูปที่ 8.19 ตารางผลการเปรียบเทียบค่าเฉลี่ย

แบบฝึกหัดปฏิบัติการที่ 8 GLM

ในแต่ละข้อ ให้นิสิตปฏิบัติดังนี้

ก) จะใช้การทดสอบแบบใด จงเขียนสมมุติฐาน H_0 และ H_1

ข) จงหาค่าสถิติที่จำเป็นต้องใช้เพื่อทดสอบสมมุติฐาน กำหนดให้ $\alpha = 0.05$ ด้วย GLM

ค) เขียนตาราง ANOVA และการเปรียบเทียบค่าเฉลี่ย (ถ้ามี) พร้อมสรุปผลการทดสอบ

1. ในการศึกษานกที่อยู่ในธรรมชาติและมีถิ่นอาศัยร่วมกัน ผู้วิจัยเชื่อว่าเสียงเพลงที่นกร้องจะมีรูปแบบที่แตกต่างกัน โดยในการศึกษาผู้วิจัยได้ใช้ความยาวของเพลงที่นกร้อง (เป็นวินาที) เพื่อใช้จำแนกชนิดของนก ในการทดลองผู้วิจัยได้ออกแบบการทดลองแบบ CRD โดยใช้นก 3 ชนิด มาร้องเพลงภายใต้สภาพแวดล้อมเดียวกัน ได้ผลดังนี้

Towhee	Common yellowthroat	Brown thrasher
1.11	2.17	0.42
1.23	1.85	0.93
0.91	1.99	0.77
0.95	1.74	0.37
0.99	1.54	0.50
1.08	1.86	0.48
1.18	1.87	0.68
1.29	2.04	0.62
1.12	1.69	0.39
0.88		0.67
1.34		1.03
		0.79

2. ในการศึกษาความเข้นข้นของโปรตีนที่ขับออกมาจากผู้ป่วยที่เป็นโรค cystic fibrosis ผู้วิจัยได้ออกแบบการทดลองแบบ CRD และเก็บข้อมูลจากกลุ่มผู้ป่วย ซึ่งจัดกลุ่มตามปริมาณทริปซินในเลือดได้ 3 กลุ่ม ได้ผลดังนี้

ปริมาณโปรตีนที่ขับออกมา (µL/(kg/hr))		
<50	51-1000	>1000
1.7	1.4	2.9
2.0	2.4	3.8
2.0	2.4	4.4
2.2	3.3	4.7
4.0	4.4	5.0
4.0	4.7	5.6
5.0	6.7	7.4
6.7	7.6	9.4
7.8	9.5	10.3
	11.7	

3. ในการศึกษาเกี่ยวกับพลังงานของคนที่ต้องใช้ในกิจกรรม 3 อย่าง คือ วิ่ง เดิน และถีบจักรยาน เนื่องจากผู้วิจัยเกรงว่าผลของปริมาณกิโลแคลอรีที่ใช้ต่อระยะทาง 1 กิโลเมตร ในแต่ละคนอาจจะแตกต่างกัน ขึ้นอยู่สภาพร่างกายของอาสาสมัครผู้นั้น ผู้วิจัยจึงออกแบบการทดลองแบบ RCBD เพื่อขจัดอิทธิพลดังกล่าว โดยให้อาสาสมัครเป็น block ได้ผลทดลอง ดังนี้

อาสาสมัคร	กิโลแคลอรีที่ใช้ในกิจกรรม		
คนที่	วิ่ง	เดิน	ถีบจักรยาน
1	1.4	1.1	0.7
2	1.5	1.2	0.8
3	1.8	1.3	0.7
4	1.7	1.3	0.8
5	1.6	0.7	0.1
6	1.5	1.2	0.7
7	1.7	1.1	0.4
8	2.0	1.3	0.6

4. ผู้วิจัยต้องการทราบถึงผลของฮอร์โมน 4 ชนิด ที่สามารถกระตุ้นให้พืชเพิ่มการออกดอกและผล จึงออกแบบการทดลองแบบ RCBD เพื่อขจัดอิทธิพลของความแตกต่างทางพันธุกรรมของพืชต่างต้นกัน โดยในแต่ละต้น จะเลือกกิ่ง 4 กิ่ง และให้ฮอร์โมนต่างชนิดกัน เมื่อสิ้นสุดการทดลอง ได้วัดผลผลิตที่เกิดขึ้น ดังนี้

ต้นไม้ที่	ผลผลิตจากกิ่งที่ใช้ฮอร์โมนชนิด			
	A	B	C	D
1	100	59	20	51
2	87	56	15	47
3	84	52	14	45
4	92	67	26	56
5	105	50	20	60
6	90	62	22	53
7	89	60	19	51
8	88	56	17	50
9	82	50	14	45
10	94	63	24	53

การบ้านปฏิบัติการที่ 8 GLM

ในแต่ละข้อ ให้นิสิตปฏิบัติดังนี้

ก) จะใช้การทดสอบแบบใด จงเขียนสมมุติฐาน H_0 และ H_1

ข) จงหาค่าสถิติที่จำเป็นต้องใช้เพื่อทดสอบสมมุติฐาน กำหนดให้ $\alpha = 0.05$ ด้วย GLM

ค) เขียนตาราง ANOVA และการเปรียบเทียบค่าเฉลี่ย (ถ้ามี) พร้อมสรุปผลการทดสอบ

1. ผู้วิจัยมีความสนใจเกี่ยวกับวิธีการบำบัดของเสียอินทรีย์ที่มีคาร์บอนเป็นองค์ประกอบในแหล่งน้ำ ด้วยวิธีการ 3 วิธี จึงได้ออกแบบการทดลองแบบ CRD และทดลองใช้บำบัดน้ำที่มีของเสียอินทรีย์ที่มีคาร์บอนเป็นองค์ประกอบ ได้ผลดังนี้

ปริมาณที่หายไปของของเสียอินทรีย์ที่มีคาร์บอนเป็นองค์ประกอบเมื่อบำบัดด้วยวิธี		
Air flotation	Foam separation	Ferric-chloride coagulation
34.6	38.8	26.7
35.1	39.0	26.7
35.3	40.1	27.0
35.8	40.9	27.1
36.1	41.0	27.5
36.5	43.2	28.1
36.8	44.9	28.1
37.2	46.9	28.7
37.4	51.6	30.7
37.7	53.6	31.2

2. ผู้วิจัยได้ทำการศึกษาถึงผลของการล่ากวางต่อพฤติกรรมของกวางในแต่ละฤดู จึงได้ออกแบบการทดลองเป็น RCBD โดยได้เลือกเส้นทางที่กวางใช้ในการสัญจร 4 เส้นทาง และกำหนดให้เป็น block ผู้วิจัยได้นับรอยเท้ากวางในระยะทางช่วงหนึ่งในแต่ละเส้นทาง และบันทึกจำนวนรอยเท้ากวางเฉลี่ยต่อสัปดาห์ โดยทำการนับและบันทึก 3 ช่วงเวลา ได้ผลดังนี้

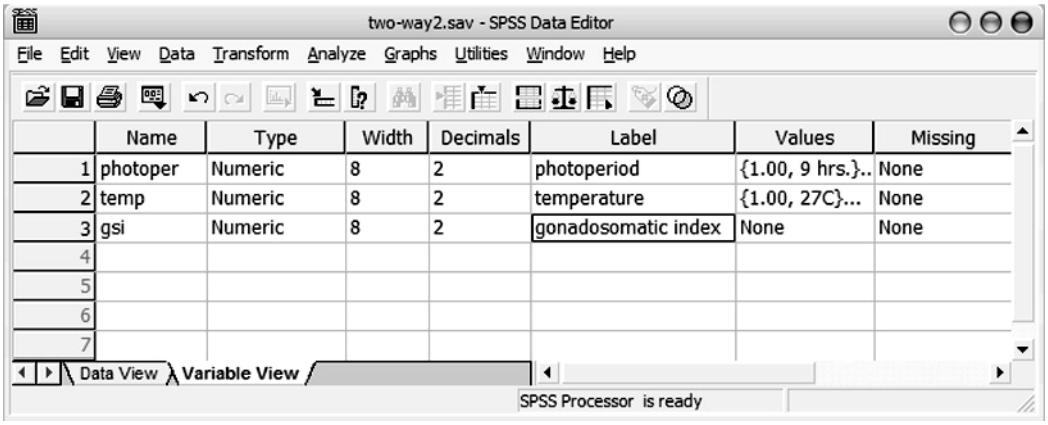
เส้นทางที่	จำนวนรอยเท้าเฉลี่ยต่อสัปดาห์		
	ก่อนฤดูล่า	ระหว่างฤดูล่า	หลังฤดูล่า
1	62.5	57.0	49.0
2	46.5	53.3	50.0
3	45.0	59.3	37.0
4	24.0	35.7	50.0

ปฏิบัติการที่ 9 GLM for Factorial Design

ในปฏิบัติการที่ 8 เป็นการใช้ GLM สำหรับการวิเคราะห์ทางสถิติที่มีตัวแปรระบุกลุ่ม หรือ Fixed factor เพียง 1 ตัวแปร สำหรับในปฏิบัตการนี้ จะเป็นการใช้ GLM เพื่อการวิเคราะห์ทางสถิติที่มีตัวแปรระบุกลุ่ม หรือ Fixed factor 2 ตัวแปร หรือมักเรียกว่า Two-way ANOVA หรือ Factorial analysis โดยจะฝึกหัดใช้ GLM สำหรับ Factorial analysis ที่มีการวางแบบการทดลองแบบ CRD และ RCBD

1. GLM สำหรับ Factorial Analysis in CRD

ในการวิเคราะห์ข้อมูลที่ได้จากการทดลองที่มีการวางแบบการทดลองแบบนี้ จะต้องสร้างตัวแปร 3 ตัว โดยตัวแปรที่ 1 เป็นตัวแปรระบุ treatment สำหรับ fixed factor ตัวแปรแรก ส่วนตัวแปรที่ 2 เป็นตัวแปรระบุ treatment สำหรับ fixed factor ตัวแปรที่ 2 ส่วนตัวแปรที่ 3 คือ ตัวแปรที่จะทดสอบหรือ dependent variable ในตัวอย่าง (รูปที่ 9.1) ตัวแปรที่ 1-3 ตัวแปร photoper เป็น fixed factor A ตัวแปร temp เป็น fixed factor B ซึ่งทั้งสองตัวแปรนี้เป็น independent variable และตัวแปร gsi เป็น dependent variable



รูปที่ 9.1

แบบจำลองของการวางแผนการทดลองแบบ Factorial analysis in CRD สามารถเขียนได้ดังนี้

$$X_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ij}$$

เมื่อ μ คือ ค่าเฉลี่ยโดยรวมของการทดลอง α และ β คือ ผลจาก factor A และ B ตามลำดับ $\alpha\beta$ คือ interaction ระหว่าง factor ทั้งสอง และ ε คือ random error

ในการวิเคราะห์ข้อมูล มีขั้นตอนดังนี้

1. เลือก Analyze, General Linear Model, Univariate... (รูปที่ 9.2)
2. จาก dialog box รูปที่ 9.3
    a. เลือกตัวแปรที่เป็น fixed factor จากช่องด้านซ้าย แล้วกดปุ่ม ▸ ของช่อง Fixed Factor(s):

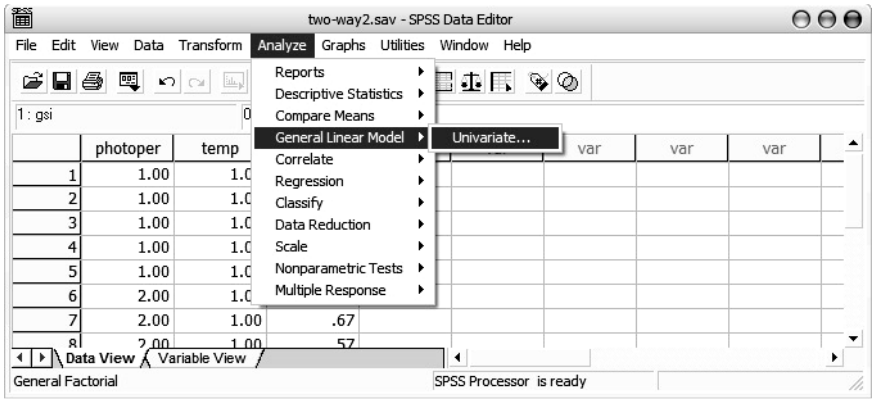
b. เลือกตัวแปรที่จะทดสอบค่าจากช่องด้านซ้าย กด ▸ ของช่อง Dependent variable(s):

c. กดปุ่ม Model... เพื่อกำหนดแบบจำลอง จะมี dialog box ดังรูปที่ 9.4

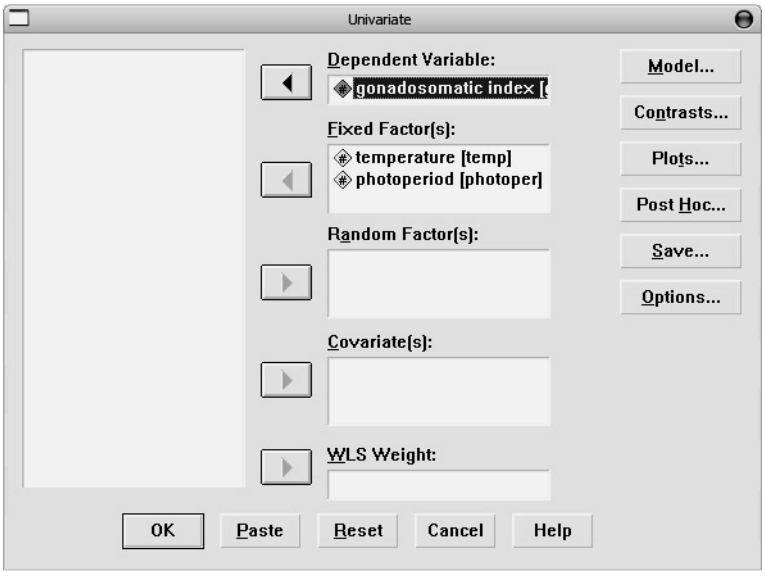
   i. ที่ช่อง Specify Model เลือก Full factorial

   ii. เลือก Sum of Squares: เป็น Type III และเลือก Include intercept in model

   iii. กดปุ่ม Continue



รูปที่ 9.2



รูปที่ 9.3

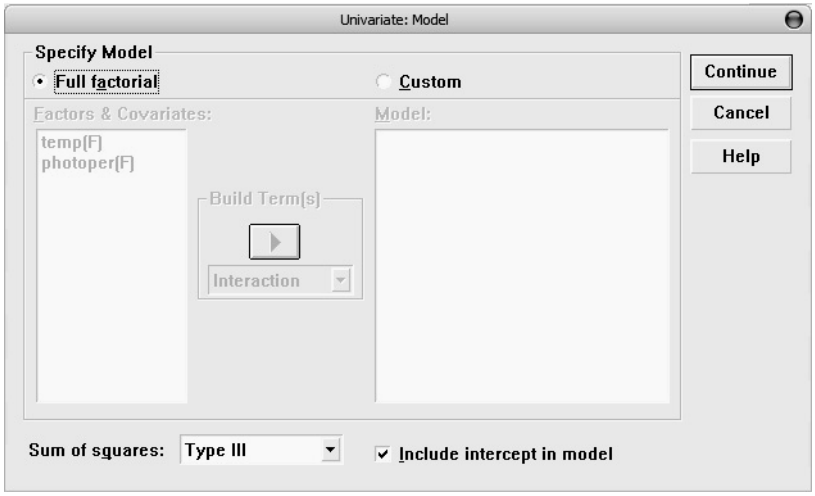
d. กดปุ่ม Plots... เพื่อสร้างกราฟค่าเฉลี่ยสำหรับพิจารณาผลของตัวแปรที่เป็น fixed factor ทั้งสอง (ในตัวอย่างนี้ คือ photoper และ temp) จะมี dialog box ดังรูปที่ 9.5

i. กำหนดกราฟแรก เลือก temp ที่ช่อง **Factors:** แล้วกดปุ่ม ▸ ของช่อง **Horizontal axis:** จากนั้นเลือก photoper แล้วกดปุ่ม ▸ ของช่อง **Separate lines:** แล้วจึงกดปุ่ม **Add** สังเกตว่าที่ช่อง **Plot:** มีพารามิเตอร์ temp*photoper ปรากฏขึ้น

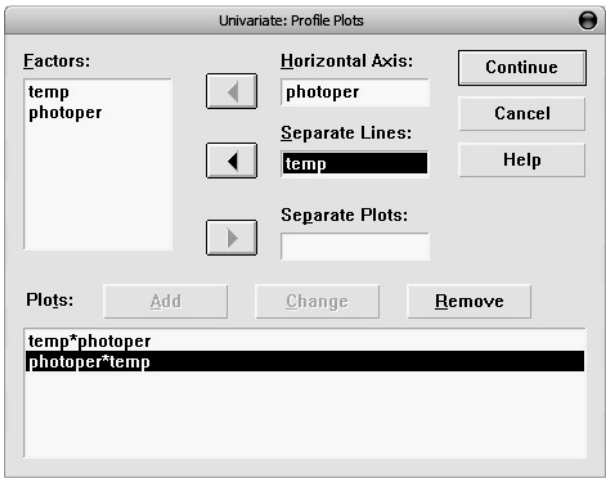
ii. กำหนดกราฟที่สอง เลือก photoper ที่ช่อง **Factors:** แล้วกดปุ่ม ▸ ของช่อง **Horizontal axis:** จากนั้นเลือก temp แล้วกดปุ่ม ▸ ให้มาอยู่ในช่อง **Separate lines:** แล้วจึงกดปุ่ม **Add** สังเกตว่าที่ช่อง **Plot:** มี พารามิเตอร์ photoper*temp เพิ่มขึ้น

iii. หากมีตัวแปรมากกว่านี้ ให้กำหนดกราฟเพิ่มเติมตามวิธีที่กล่าวมาข้างต้น

iv. กดปุ่ม **Continue**

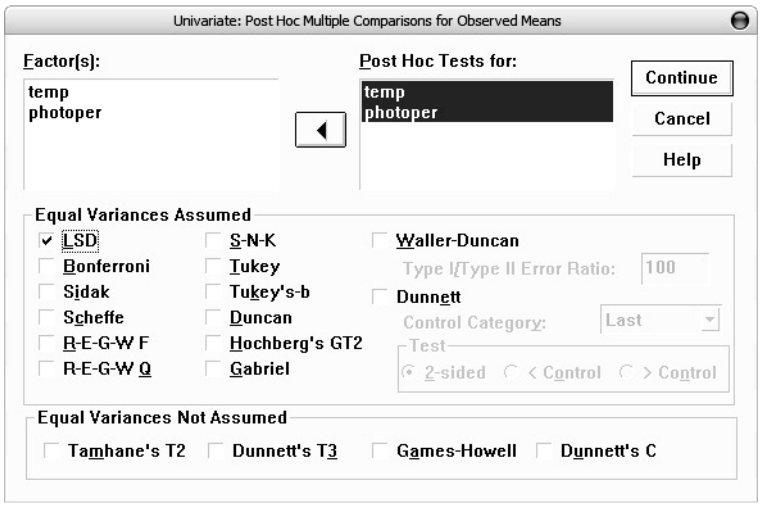


รูปที่ 9.4



รูปที่ 9.5

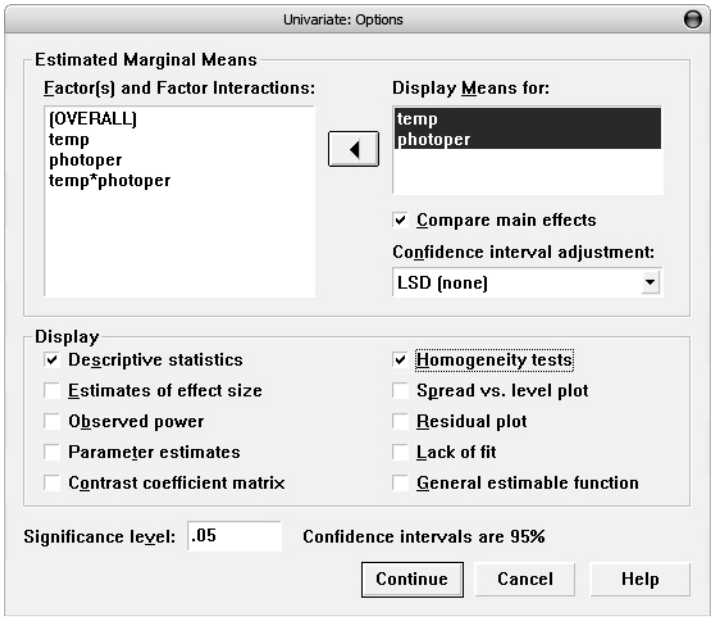
e. กดปุ่ม Post Hoc... เพื่อเปรียบเทียบค่าเฉลี่ย จะมี dialog box ดังรูปที่ 9.6
   i. เลือกตัวแปร fixed factor ในช่อง Factor(s): กดปุ่ม ▸ ของช่อง Post Hoc Test for: ในตัวอย่างนี้ คือ temp และ photoper
   ii. เลือกแบบการเปรียบเทียบค่าเฉลี่ยในช่อง Equal Variance Assumed อนึ่ง หากการวิเคราะห์ชี้ว่าความแปรปรวนระหว่างกลุ่มตัวอย่างไม่เท่ากัน ก็ต้องเลือกแบบการเปรียบเทียบค่าเฉลี่ยในช่อง Equal Variance Not Assumed แทน
   iii. แล้วกดปุ่ม Continue



รูปที่ 9.6

f. กดปุ่ม Options... เพื่อค่าสถิติอื่น จะมี dialog box ดังรูปที่ 9.7
   i. ที่ช่อง Estimated Marginal Means เลือกตัวแปร fixed factor ในช่อง Factor(s) and Factor Interaction: กดปุ่ม ▸ ของช่อง Display Means for:
   ii. เลือก Compare main effects
   iii. เลือกว่าต้องการให้แสดงค่าสถิติใดจากตัวเลือกในช่อง Display
   iv. กำหนดค่า Significant Level: ที่ต้องการ
   v. กดปุ่ม Continue

g. กดปุ่ม OK เพื่อทำการวิเคราะห์



รูปที่ 9.7

ผลการวิเคราะห์ที่ได้จะแตกต่างไปขึ้นอยู่กับตัวเลือกต่าง ๆ ในตัวอย่างนี้ จะแสดงผลที่สำคัญดังตารางและกราฟในรูปที่ 9.8-9.11

สิ่งที่ต้องพิจารณาจากตารางและกราฟในรูปที่ 9.8-9.11 คือ ผลการทดสอบ Error variance (ของ dependent variable) ในแต่ละ treatment ว่าเท่ากันหรือไม่ จากตารางในรูปที่ 9.9 จะเห็นว่าค่า Sig. > 0.05 แสดงว่า Error variance ของทุก treatment ของ fixed factor ทั้งสองมีค่าเท่ากัน ผลสรุปจาก homogeneity test นี้ จะเป็นตัวบ่งชี้ว่าควรจะใช้วิธีการใดในการเปรียบเทียบความแตกต่างระหว่าง treatment ใน fixed factor ทั้งสอง

ตารางที่แสดงผลการวิเคราะห์ที่ต้องใช้พิจารณาผลของ fixed factor คือ ตาราง Tests of Between-Subjects Effects (รูปที่ 9.10) โดยค่าที่ต้องนำเสนอในตาราง ANOVA คือ บรรทัดที่ระบายด้วยสีดำ โดยบรรทัด TEMP และ PHOTOPER เป็นการทดสอบ main effect ส่วนบรรทัด TEMP*PHOTOPER เป็นการทดสอบ interaction effect

Descriptive Statistics

Dependent Variable: gonadosomatic index

temperature	photoperiod	Mean	Std. Deviation	N
27C	9 hrs.	1.0700	.14832	5
	14 hrs.	.6400	.13342	5
	Total	.8550	.26277	10
16C	9 hrs.	2.4400	.66933	5
	14 hrs.	1.3000	.28293	5
	Total	1.8700	.77181	10
Total	9 hrs.	1.7550	.85455	10
	14 hrs.	.9700	.40557	10
	Total	1.3625	.76550	20

รูปที่ 9.8 ค่าสถิติบางส่วนของแต่ละกลุ่ม (treatment)

Levene's Test of Equality of Error Variances[a]

Dependent Variable: gonadosomatic index

F	df1	df2	Sig.
2.808	3	16	.073

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+TEMP+PHOTOPER+TEMP * PHOTOPER

รูปที่ 9.9 ผลการทดสอบ Equality of Error Variances

Tests of Between-Subjects Effects

Dependent Variable: gonadosomatic index

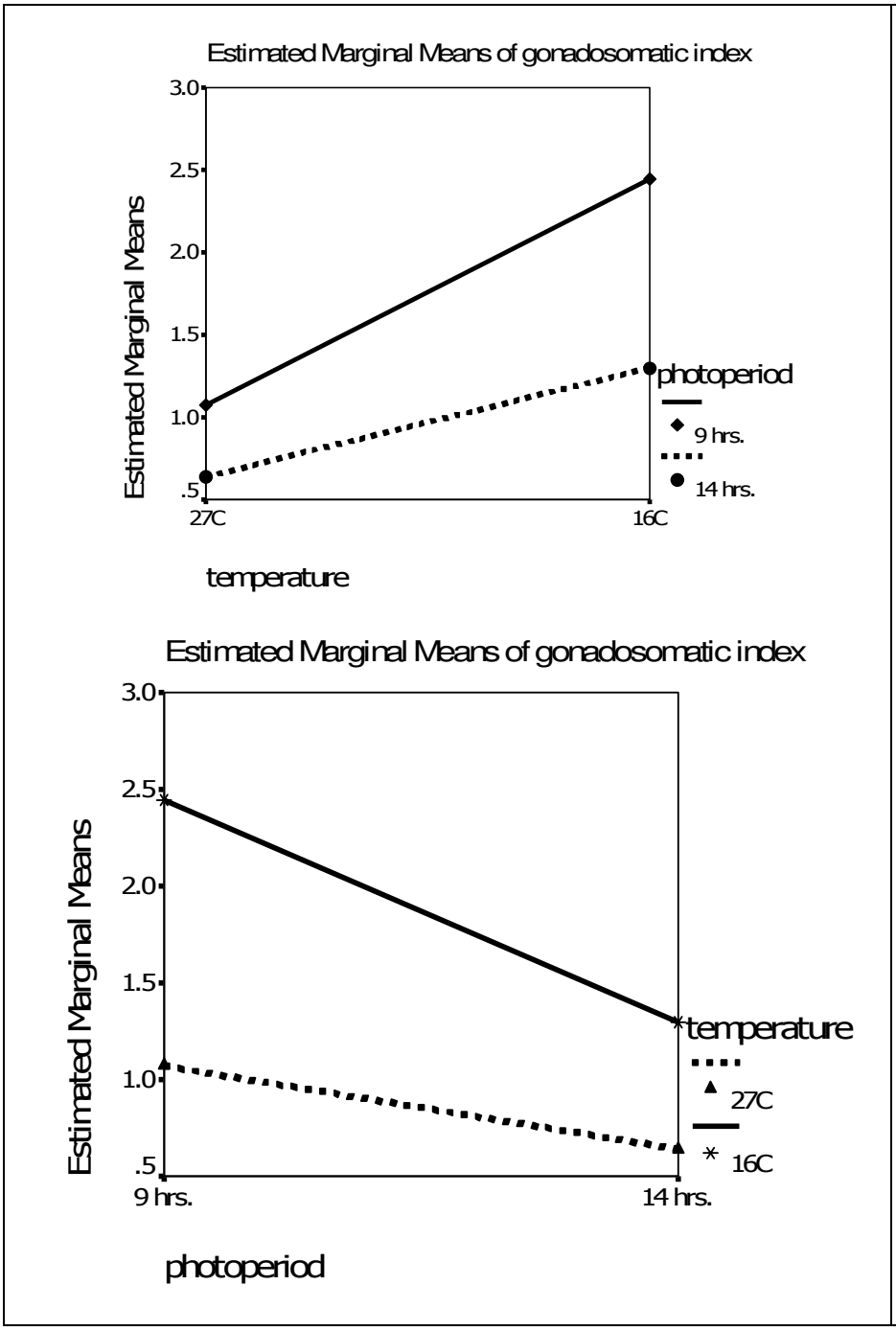
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8.862[a]	3	2.954	20.809	.000
Intercept	37.128	1	37.128	261.535	.000
TEMP	5.151	1	5.151	36.285	.000
PHOTOPER	3.081	1	3.081	21.704	.000
TEMP * PHOTOPER	.630	1	.630	4.439	.051
Error	2.271	16	.142		
Total	48.262	20			
Corrected Total	11.134	19			

a. R Squared = .796 (Adjusted R Squared = .758)

รูปที่ 9.10 ตาราง ANOVA [อาจไม่จำเป็นต้องรายงานบรรทัด Corrected Model]

สำหรับกราฟที่แสดงในรูปที่ 9.11 นั้น จะช่วยให้เห็น interaction ระหว่าง fixed factor ได้ชัดเจนขึ้น ถ้าหากว่าไม่มี interaction แล้ว กราฟของค่าเฉลี่ยจากแต่ละ treatment ใน fixed factor ที่นำค่าเฉลี่ยมาสร้างกราฟบน fixed factor อีกตัวหนึ่งที่เป็นแกนนอน จะเป็นเส้นขนานกัน ในทางตรงกันข้าม ถ้ามี interaction แล้ว กราฟจะไม่ขนานกันหรือตัดกัน ในตัวอย่างกราฟที่แสดงในรูปที่ 9.11 นี้ จะเห็นว่าเส้นกราฟในกราฟทั้งสองไม่ขนานกัน แสดงว่ามี fixed factor ทั้งสองมี interaction กัน หากสังเกตที่บรรทัด TEMP*PHOTOPER จากตารางในรูป 9.10 จะพบว่า Sig. มีค่า 0.051 ซึ่งโดยทั่วไปแล้ว จะไม่ปฏิเสธ Ho: ไม่มี interaction ซึ่งไม่สอดคล้องกับกราฟในรูปที่ 9.11 อย่างไรก็ตาม เนื่องจากค่า Sig. นี้มีค่า

ใกล้เคียงกับค่าที่ตั้งไว้สำหรับ Type I error (คือ 0.05) ประกอบกับกราฟที่แสดงในรูปที่ 9.11 ก็สามารถปฏิเสธ Ho: ไม่มี interaction และยอมรับ H1: มี interaction แทนได้ นอกจากนี้ กราฟนี้ยังแสดงถึงการตอบสนองที่มีต่อ treatment แต่ละ treatment เป็นอย่างไร และแตกต่างกันอย่างไร



รูปที่ 9.11

สำหรับการทดสอบเกี่ยวกับ main effect และ interaction effect นั้น จะพิจารณาจากตารางในรูปที่ 9.10 (คือตาราง ANOVA นั้นเอง) โดยพิจารณาค่า Sig. ของ main effect และ interaction effect ซึ่งแต่ละ effect มีสมมุติฐานดังนี้

Main effect: factor A (มี *i* ระดับ)

$Ho: \bar{\mu}_{i\bullet} = \bar{\mu}_{i'\bullet}$; ไม่มีความแตกต่างเนื่องจาก factor A

H1: มีค่าเฉลี่ยของ factor A อย่างน้อย 1 คู่ ที่แตกต่างกัน

Main effect: factor B (มี *j* ระดับ)

$Ho: \bar{\mu}_{\bullet j} = \bar{\mu}_{\bullet j'}$; ไม่มีความแตกต่างเนื่องจาก factor B

H1: มีค่าเฉลี่ยของ factor B อย่างน้อย 1 คู่ ที่แตกต่างกัน

Interaction effect

$Ho: \mu_{ij} - \bar{\mu}_{i\bullet} = \bar{\mu}_{\bullet j} + \bar{\mu}_{\bullet\bullet} = 0$;

ไม่มี interaction ระหว่าง factor A และ factor B

H1: มี interaction ระหว่าง factor A และ factor B

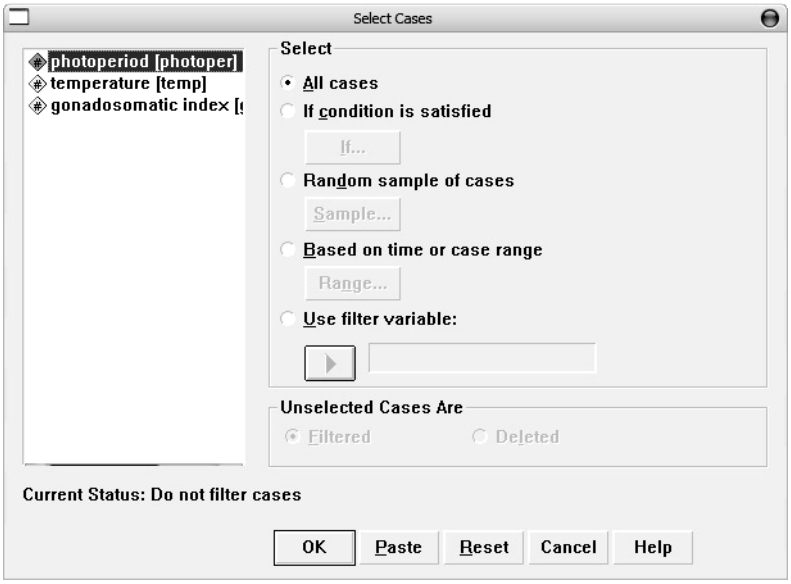
โดยมีแนวทางในการทดสอบสมมุติฐานเป็นขั้นตอนโดยสังเขป ดังนี้

1. พิจารณาว่า ค่า Sig. ของ main effect และ interaction effect ว่าเท่ากับหรือน้อยกว่า 0.05 หรือไม่
2. ถ้ามีค่า Sig. มากกว่า 0.05 ไม่ต้องทำการพิจารณาต่อ สรุปว่าค่าเฉลี่ยไม่แตกต่างกัน
3. ถ้า main effect ตัวใดตัวหนึ่ง มีค่า Sig. เท่ากับหรือน้อยกว่า 0.05 **<u>แต่</u>** interaction effect มีค่า Sig. มากกว่า 0.05 แสดงว่าค่าเฉลี่ยของ main effect แตกต่างกันและไม่มี interaction effect ให้พิจารณา Post Hoc test ของ main effect ที่มีนัยสำคัญ (คือ ปฏิเสธ Ho) เพื่อหาว่าค่าเฉลี่ยใดบ้างที่แตกต่างกัน
4. ถ้า interaction effect มีค่า Sig. เท่ากับหรือน้อยกว่า 0.05 ไม่ควรพิจารณาผลการทดสอบสมมุติฐานของ main effect แต่ให้ดูที่ตาราง Test of Equality of Error Variance (รูปที่ 9.9)
    a. ถ้าหากสรุปได้ว่า variance ไม่เท่ากัน ให้ทดสอบผลของ main effect ด้วย One-way ANOVA ของ fixed factor ตัวหนึ่ง ในแต่ละ treatment (แยกกัน) ของ fixed factor อีกตัวหนึ่ง
    b. ถ้าหากสรุปได้ว่า variance เท่ากัน ให้ทดสอบ main effect ด้วยวิธี simple main effect test และอาจจะทดสอบผลของ interaction effect ได้ด้วยหากต้องการ [*<u>ศึกษาค้นคว้าเอง</u>*]

ในตัวอย่างนี้ มี interaction ระหว่างตัวแปร fixed factor ทั้งสอง จึงแยกการทดสอบ main effect โดยในที่นี้ ผู้วิจัยต้องการทราบว่าผลของ treatment ทั้งสอง (9 ชม. และ 14 ชม.) ของ photoper แตกต่างกันหรือไม่ ในการทดสอบ จะต้องวิเคราะห์ด้วย One-way ANOVA โดยแยกวิเคราะห์กัน คือ One-way ANOVA เมื่อ temp เป็น 16 ซ. และเมื่อ temp เป็น 27 ซ. (ในตัวอย่างนี้ มีเพียง 2 treatment ใน temp)

เนื่องจากว่า จะต้องทำการวิเคราะห์แยกกัน จึงต้องเลือกเอาเฉพาะข้อมูลเพียงบางส่วนมาวิเคราะห์ ซึ่งสามารถทำได้ 2 วิธี คือ วิธีแรก ให้คัดลอกข้อมูลที่ต้องการมาไว้ในตัวแปรใหม่ แล้ววิเคราะห์ด้วย One-way ANOVA ส่วนอีกวิธีหนึ่ง จะใช้คำสั่งในเมนูของ SPSS คือ **Data, Select Cases**... (จะมี dialog box ดังรูปที่ 9.12) แล้วระบุวิธีการในการเลือกข้อมูลในช่อง **Select** โดย

All cases	เลือกข้อมูลทุกค่า
If condition is satisfied	เลือกข้อมูลที่ตรงกับเงื่อนไขที่กำหนด
Random sample of cases	เลือกข้อมูลโดยวิธีการสุ่มเลือก (กำหนดวิธีได้)
Based on time or case range	เลือกข้อมูลโดยกำหนดช่วงเวลาหรือข้อมูล
Use filter variable	เลือกข้อมูลโดยใช้ตัวแปรคัดเลือก

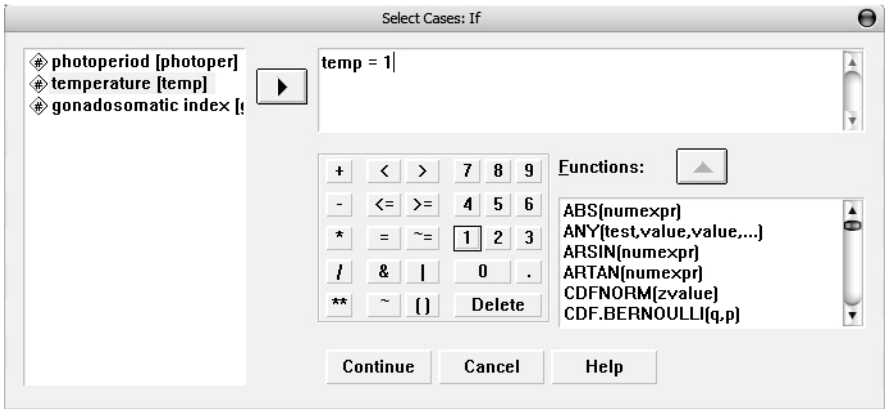


รูปที่ 9.12

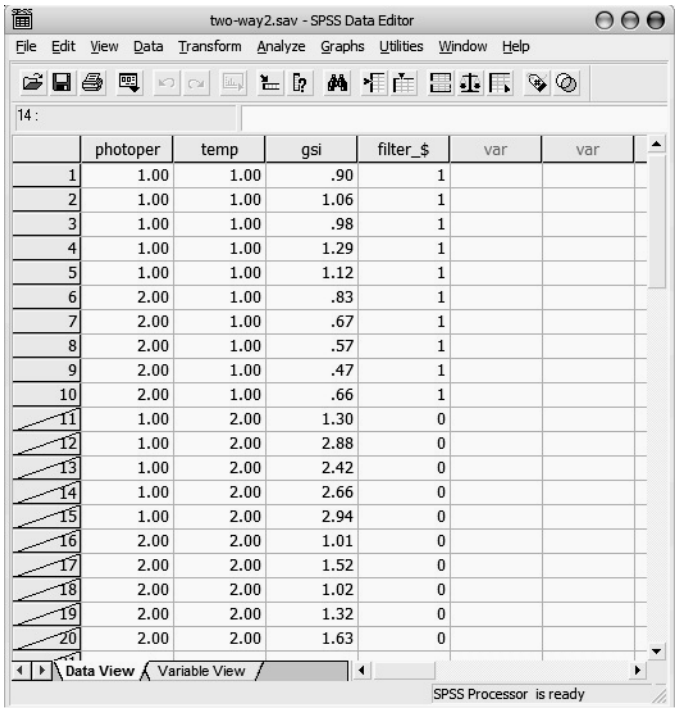
ในตัวอย่างนี้ จะฝึกใช้วิธีการที่ 2 (คือกดปุ่ม ❍ ของ **If condition is satisfied** ในรูปที่ 9.12) และกำหนดเงื่อนไขให้เลือกเฉพาะข้อมูลที่ temp มีค่าเท่ากับ 1 (คือ 27 ซ.) ตามขั้นตอนดังนี้

1. กดปุ่ม **If**... จะมี dialog box ดังรูปที่ 9.13
    a. เลือกตัวแปรจากช่องด้านซ้าย (คือ temp) กดปุ่ม ▸ จะมีตัวแปร temp ปรากฏในช่องด้านซ้าย
    b. กดเลือก mathematic operator จากตัวเลือกด้านล่าง ในตัวอย่างคือ "=" และพิมพ์ 1 เพื่อเลือกเฉพาะข้อมูลที่ temp = 1
    c. แล้วกดปุ่ม **Continue**
2. ในช่อง **Unselected Cases Are** ให้เลือก **Filtered**

3. กดปุ่ม OK ให้สังเกตว่าใน Data View (รูปที่ 9.14) ข้อมูลที่ temp มีค่าเป็น 2 จะมีเครื่องขีดฆ่าและมีตัวแปรเพิ่มมา 1 ตัวแปร คือ filter_$ ซึ่งค่าสังเกตที่มีค่า 1 จะเป็นค่าที่ถูกเลือกไว้สำหรับการวิเคราะห์ ส่วนค่าสังเกตที่มีค่า 0 จะเป็นค่าที่ถูกคัดออกไปจากการวิเคราะห์



รูปที่ 9.13



	photoper	temp	gsi	filter_$	var	var
1	1.00	1.00	.90	1		
2	1.00	1.00	1.06	1		
3	1.00	1.00	.98	1		
4	1.00	1.00	1.29	1		
5	1.00	1.00	1.12	1		
6	2.00	1.00	.83	1		
7	2.00	1.00	.67	1		
8	2.00	1.00	.57	1		
9	2.00	1.00	.47	1		
10	2.00	1.00	.66	1		
11	1.00	2.00	1.30	0		
12	1.00	2.00	2.88	0		
13	1.00	2.00	2.42	0		
14	1.00	2.00	2.66	0		
15	1.00	2.00	2.94	0		
16	2.00	2.00	1.01	0		
17	2.00	2.00	1.52	0		
18	2.00	2.00	1.02	0		
19	2.00	2.00	1.32	0		
20	2.00	2.00	1.63	0		

รูปที่ 9.14

ในการวิเคราะห์ จะใช้ One-way ANOVA โดยระบุตัวแปร photoper ที่ช่อง Factor: และ gsi ที่ช่อง Dependent list: โดยผลการวิเคราะห์ที่จะนำเสนอคือตาราง ANOVA (รูปที่ 9.15) และหากต้องการจะเปลี่ยนไปวิเคราะห์ข้อมูลที่ temp มีค่าเป็น 2 (คือ 16 ซ.) ให้ทำเช่นเดียวกับที่กล่าวมาในย่อหน้าที่แล้ว เพียงแต่ระบุให้ temp = 2 แล้ววิเคราะห์ด้วย One-way ANOVA ซึ่งผลการวิเคราะห์ที่จะนำเสนอ คือ ตาราง ANOVA (รูปที่ 9.16) จากตาราง One-way ANOVA ทั้งสอง เห็นได้ว่าค่า Sig. < 0.05 แสดงว่า treatment ทั้งสองของ photroper มีค่าเฉลี่ย gsi ต่างกันทั้งที่ temp เป็น 16 ซ. และ 27 ซ.

ANOVA

gonadosomatic index

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.462	1	.462	23.229	.001
Within Groups	.159	8	.020		
Total	.621	9			

รูปที่ 9.15

ANOVA

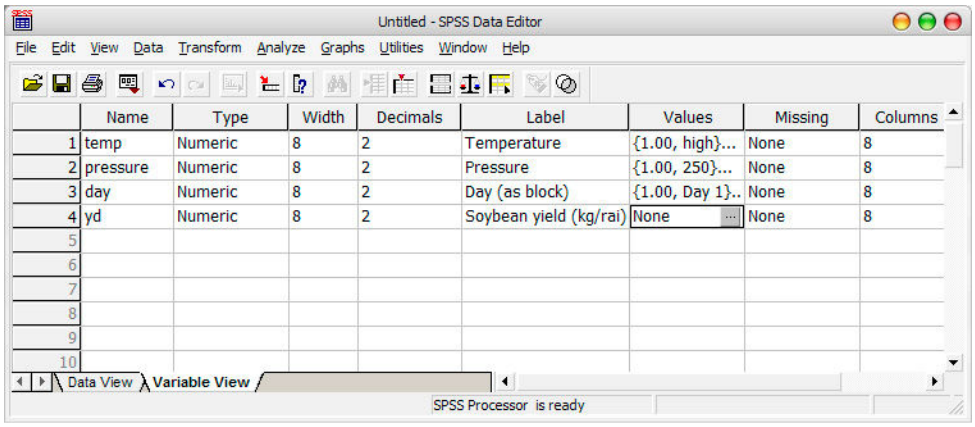
gonadosomatic index

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3.249	1	3.249	12.306	.008
Within Groups	2.112	8	.264		
Total	5.361	9			

รูปที่ 9.16

2. GLM สำหรับ Factorial in RCBD

ในการวิเคราะห์ข้อมูลที่ได้จากการทดลองที่มีการวางแบบการทดลองแบบ RCBD นั้น จะต้องสร้างตัวแปร 4 ตัว คือ ตัวแปรที่ 1 เป็นตัวแปรระบุ treatment สำหรับ fixed factor ตัวแปรแรก ส่วนตัวแปรที่ 2 เป็นตัวแปรระบุ treatment สำหรับ fixed factor ตัวแปรที่ 2 ตัวแปรที่ 3 คือ ตัวแปรระบุ block และตัวแปรที่ 4 คือ ตัวแปรที่จะทดสอบหรือ dependent variable ในตัวอย่าง (รูปที่ 9.17) ตัวแปร temp และ pressure คือ fixed factor ตัวแปร day คือ block และตัวแปร yd เป็น dependent variable



รูปที่ 9.17

แบบจำลองของการวางแผนการทดลองแบบ CRD สามารถเขียนได้ดังนี้

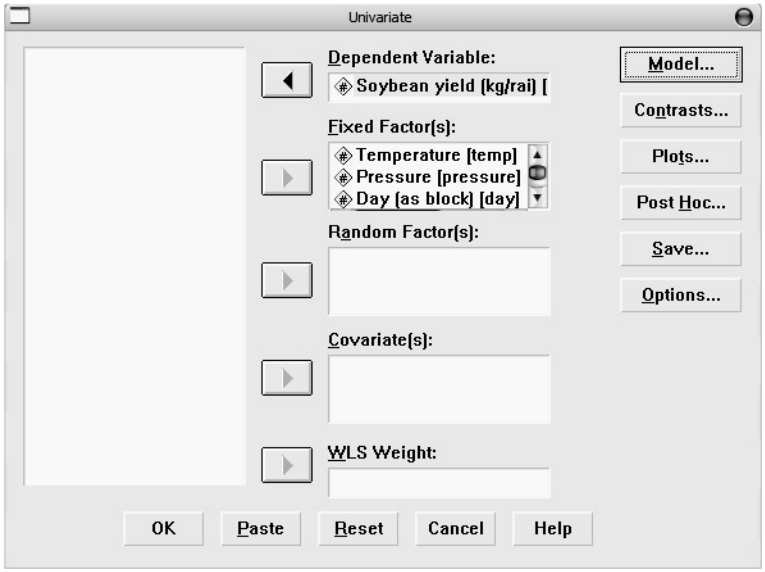
$$X_{ij} = \mu + \alpha_{i.} + \beta_{.j} + \alpha\beta_{ij} + B_k + \varepsilon_{ijk}$$

เมื่อ μ คือ ค่าเฉลี่ยโดยรวมของการทดลอง α และ β คือ ผลจาก fixed factor A และ B ส่วน $\alpha\beta$ คือผลจาก interaction ส่วน B คือ ผลจาก Block และ ε คือ random error

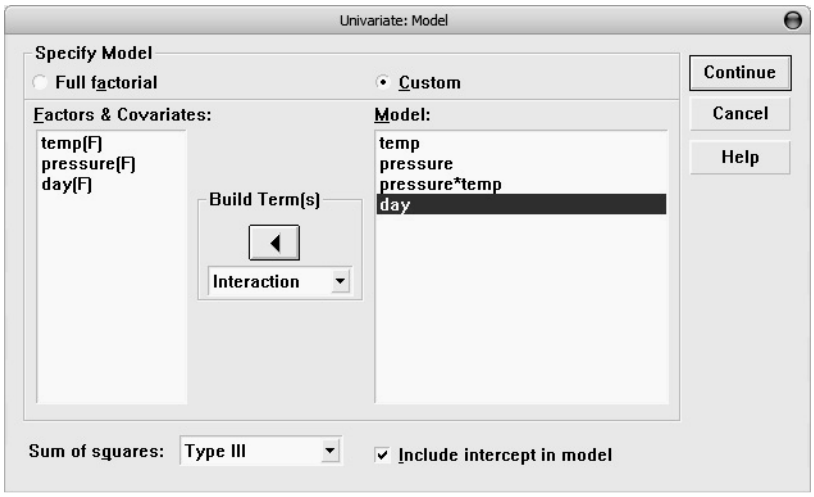
ในการวิเคราะห์ข้อมูล จะมีขั้นตอนดังนี้

1. เลือก Analyze, General Linear Model, Univariate...

2. จาก dialog box ในรูปที่ 9.18
    a. เลือกตัวแปร fixed factor และตัวแปรที่เป็น block ในช่องด้านซ้าย แล้วกด ▸ ของช่อง Fixed Factor(s):
    b. เลือกตัวแปรที่จะทดสอบค่า แล้วกด ▸ ของช่อง Dependent variable(s):
    c. กดปุ่ม Model... เพื่อกำหนดแบบจำลอง จะมี dialog box ในรูปที่ 9.19
        i. ที่ช่อง Specify Model เลือก Custom
        ii. ที่ช่อง Factors & Covariates:
            1. เลือกตัวแปร fixed factor (โดยเลือกที่ละ 1 ตัวแปร) แล้วกดปุ่ม ▸ ของช่อง Build Terms(s)
            2. เลือกตัวแปร fixed factor ทั้งสอง แล้วกำหนดที่ช่อง Build Term(s) เป็น Interaction แล้วกดปุ่ม ▸ สังเกตว่าที่ช่อง Model: จะมีบรรทัดที่ระบุ interaction ของ fixed factor โดยมี * ระหว่างตัวแปรทั้งสอง
            3. เลือกตัวแปรที่เป็น block แล้วกดปุ่ม ▸ ของช่อง Build Terms(s)
        iii. กำหนด Sum of Squares: เป็น Type III และเลือก Include intercept in model
        iv. กดปุ่ม Continue

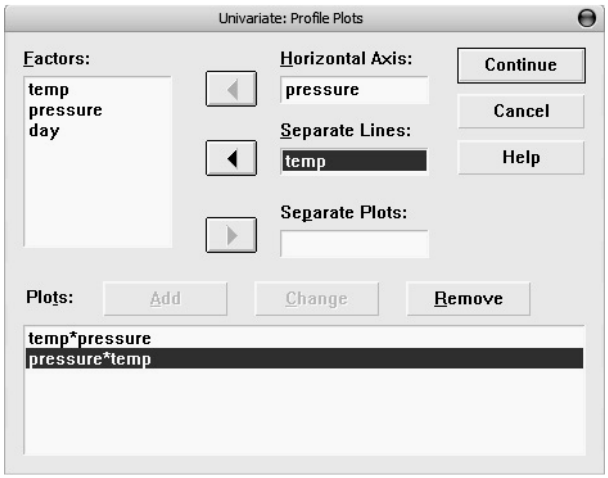


รูปที่ 9.18



รูปที่ 9.19

d. กดปุ่ม **Plots...** เพื่อสร้างกราฟค่าเฉลี่ยสำหรับพิจารณา interaction ของตัวแปรที่เป็น fixed factor ทั้งสอง ในตัวอย่างนี้ คือ temp และ pressure ซึ่งในตัวอย่างจะสร้าง 2 กราฟ จาก dialog box ในรูปที่ 9.20
   i. สร้างกราฟแรก โดยเลือก temp ที่ช่อง **Factors:** แล้วกด ▸ ของช่อง **Horizontal axis:** จากนั้นเลือก pressure แล้วกด ▸ ของช่อง **Separate lines:** แล้วจึงกดปุ่ม **Add** สังเกตว่าที่ช่อง **Plot:** มีพารามิเตอร์ temp*pressure ปรากฏขึ้น
   ii. สร้างกราฟที่สอง เลือก pressure ที่ช่อง **Factors:** แล้วกด ▸ ของช่อง **Horizontal axis:** จากนั้นเลือก temp แล้วกด ▸ ของช่อง **Separate lines:** แล้วจึงกดปุ่ม **Add** สังเกตว่าที่ช่อง **Plot:** มี พารามิเตอร์ pressure*temp เพิ่มขึ้น
   iii. กดปุ่ม **Continue**



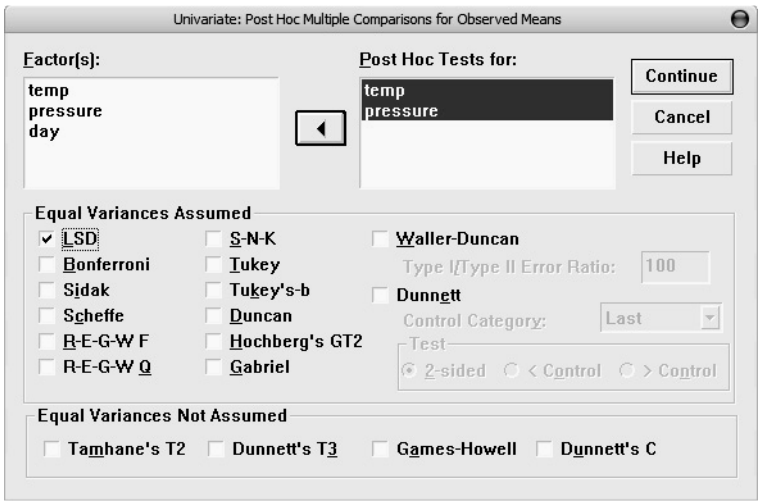
รูปที่ 9.20

e. กดปุ่ม **Post Hoc...** เพื่อกำหนดการเปรียบเทียบค่าเฉลี่ยของ fixed factor จะมี dialog box ในรูปที่ 9.21

i. เลือกตัวแปร fixed factor ในช่อง Factor(s): กดปุ่ม ▸ ของช่อง Post Hoc Test for:

ii. เลือกแบบการเปรียบเทียบค่าเฉลี่ยในช่อง Equal Variance Assumed อนึ่ง หากการวิเคราะห์ชี้ว่าความแปรปรวนระหว่างกลุ่มตัวอย่างไม่เท่ากัน ก็ต้องเลือกแบบการเปรียบเทียบค่าเฉลี่ยในช่อง Equal Variance Not Assumed แทน

iii. กดปุ่ม Continue



รูปที่ 9.21

f. กดปุ่ม Options... เพื่อเลือกแสดงค่าสถิติอื่น ๆ จาก dialog box ในรูปที่ 9.22

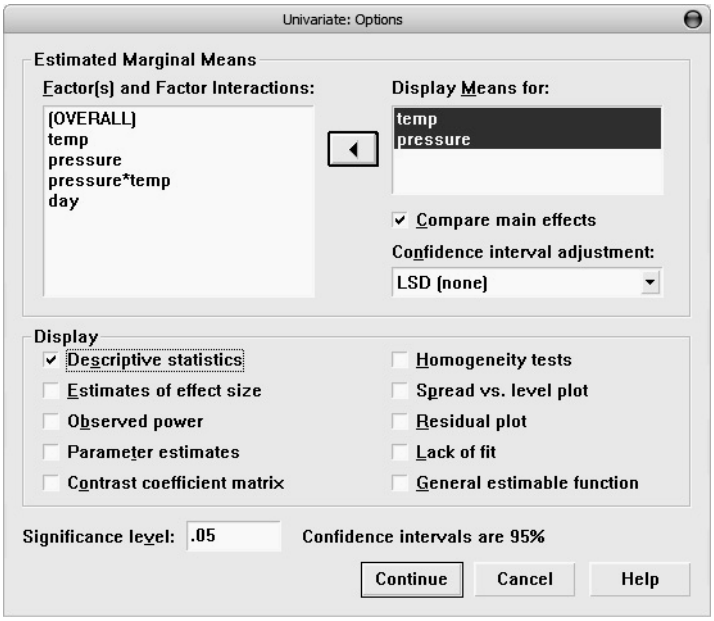
i. ที่ช่อง Estimated Marginal Means เลือกตัวแปร fixed factor ในช่อง Factor(s) and Factor Interaction: กดปุ่ม ▸ ของช่อง Display Means for:

ii. เลือกค่าสถิติที่ต้องการจากตัวเลือกในช่อง Display

iii. กำหนดค่า Significant Level:

iv. กดปุ่ม Continue

g. เมื่อกำหนดค่าต่าง ๆ เรียบร้อยแล้ว กดปุ่ม OK



รูปที่ 9.22

ผลการวิเคราะห์ จะมีตารางคล้ายกับที่กล่าวมาในข้อ 1. โดยตาราง Descriptive Statistics จะแสดงค่าเฉลี่ย ค่าเบี่ยงเบนมาตรฐานของแต่ละ fixed factor และ block ส่วนตารางทดสอบผลของแต่ละ fixed factor แสดงดังรูปที่ 9.23 โดยบรรทัดที่ระบายสีดำนั้นคือส่วนที่จะนำไปเสนอในตาราง ANOVA ส่วนการทดสอบสมมติฐานนั้น จะพิจารณาเช่นเดียวกับที่กล่าวในข้อ 1. อนึ่ง ในการออกแบบการทดลองแบบ factorial analysis in RCBD นั้น ไม่ควรมี interaction ระหว่าง block กับ fixed factor

Tests of Between-Subjects Effects

Dependent Variable: Soybean yield (kg/rai)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1377.617[a]	9	153.069	15.247	.000
Intercept	7072.569	1	7072.569	704.488	.000
TEMP	1220.154	2	610.077	60.769	.000
PRESSURE	96.981	2	48.491	4.830	.042
TEMP * PRESSURE	11.146	4	2.786	.278	.884
DAY	49.336	1	49.336	4.914	.057
Error	80.314	8	10.039		
Total	8530.500	18			
Corrected Total	1457.931	17			

a. R Squared = .945 (Adjusted R Squared = .883)

รูปที่ 9.23ตาราง ANOVA [อาจไม่จำเป็นต้องรายงานบรรทัด Corrected Model]

สำหรับผลการวิเคราะห์อื่นๆ (เช่น กราฟ การเปรียบเทียบผลของ main effect ฯลฯ) นั้น จะแปลความคล้ายกับที่กล่าวมาในข้อ 1.

แบบฝึกหัดปฏิบัติการที่ 9

1. ในการวิจัยเพื่อพิจารณาผลของการให้น้ำและชนิดของพืชต่อการเจริญของลำต้นถั่ว ผู้วิจัยได้ทดลองให้น้ำ 3 ระดับ กับต้นถั่ว 2 แบบ คือ ต้นถั่วที่ไม่มีใบกับต้นถั่วปกติ ในช่วงเวลาที่ทดลองมีต้นถั่วที่ไม่มีใบเพียง 18 ต้น ผู้วิจัยจึงได้สุ่มเลือกโดยแบ่งเป็น 3 กลุ่ม จากนั้นก็สุ่มว่าต้นถั่วแต่ละต้นจะได้รับน้ำระดับใดจาก 3 ระดับ และผู้วิจัยก็ทำเช่นเดียวกันกับต้นถั่วปกติ เมื่อดำเนินการทดลองไปได้ 3 เดือน วัดความสูงของต้นถั่ว (ซม.) ดังนี้

Factor B (Plant type)	Factor A (water level)		
	low	medium	high
leafless	69.0	96.1	121.0
	71.3	102.3	122.9
	73.2	107.5	123.1
	75.1	103.6	125.7
	74.0	100.7	125.2
	74.4	101.8	120.1
conventional	71.1	81.0	101.1
	69.2	85.8	103.2
	70.4	86.0	106.1
	73.2	87.5	109.7
	71.2	88.1	109.0
	70.9	87.6	106.9

ก. จะตั้งสมมุติฐานในทดสอบว่าอย่างไร
ข. ทำการวิเคราะห์ข้อมูล พร้อมนำเสนอตาราง ANOVA และกราฟที่จำเป็นต่อการแปลผล
ค. ผลการทดสอบสามารถสรุปได้ว่าอย่างไร
ง. ถ้าหากว่ามีความแตกต่างระหว่าง treatment ในแต่ละ factor ให้เปรียบเทียบค่าเฉลี่ยด้วยวิธี LSD

2. ในการศึกษาผลของ photoperiod และจีโนไทป์ต่อระยะเวลาที่จะแสดงอาการของโรคราน้ำค้างสายพันธุ์ AB3 ผู้วิจัยได้สุ่มเลือกใบพืชมา 50 ใบ จากแต่ละจีโนไทป์ (4 แบบ) และในแต่ละจีโนไทป์ได้สุ่มแบ่งออกเป็น 5 กลุ่ม ๆ ละ 10 ใบ แต่ละกลุ่มจะถูกปลูกเชื้อ แล้วจึงสุ่มให้ได้รับแสงเป็นเวลาเป็นระยะเวลาต่างกัน (มี 5 ช่วง) จากนั้นจึงนับจำนวนวันรวมที่ใบทั้งสิบใบในแต่ละกลุ่มแสดงอาการของโรคออกมา ได้ผลดังนี้

Factor B (genotype)	Factor A (photoperiod) (hr. darkness in 24-hr cycle)				
	0	2	4	8	16
Armelle Gold	630	610	560	570	590
Promise	640	630	600	620	620
Emir	640	630	650	620	580
Vacla	660	660	620	610	630

ก. จะตั้งสมมุติฐานในทดสอบว่าอย่างไร
ข. ทำการวิเคราะห์ข้อมูล พร้อมนำเสนอตาราง ANOVA และกราฟที่จำเป็นต่อการแปลผล
ค. ผลการทดสอบสามารถสรุปได้ว่าอย่างไร
ง. ถ้าหากว่ามีความแตกต่างระหว่าง treatment ในแต่ละ factor ให้เปรียบเทียบค่าเฉลี่ยด้วยวิธี LSD

3. ในการศึกษาผลของ phytoestrogen ที่สกัดจากพืชต่างชนิด และความเข้มข้นต่าง ๆ ต่อการผุกร่อนของกระดูก (โดยวัดความหนาแน่นของกระดูก) ผู้วิจัยมีลิงที่อยู่ในระยะ menopause 18 ตัว แต่มีอายุแตกต่างกัน จึงแบ่งลิงออกเป็น 2 กลุ่มๆ ละ 9 ตัว โดยพยายามจัดให้แต่ละกลุ่มมีลิงที่อายุและสรีระวิทยาใกล้เคียงกันมากที่สุด ด้วยเหตุนี้ ผู้วิจัยจึงพิจารณาให้กลุ่มลิงแต่ละกลุ่มนี้เป็น block ซึ่งในการทดลอง ลิงแต่ละตัวในกลุ่มจะได้รับ phytoestrogen ที่สุ่มสกัดจากพืช 1 ชนิด และสุ่มระดับที่ได้รับ 1 ระดับ (มีพืช 3 ชนิด ๆ ละ 3 ระดับ) เมื่อเสร็จสิ้นการทดลองได้ผลดังนี้

Conc. (mg/kg BW)	Phytoestrogen (First Block)			Phytoestrogen (Second Block)		
	A	B	C	A	B	C
1	105	125	95	150	175	135
10	150	165	135	200	235	250
100	95	110	100	110	125	120

ก. จะตั้งสมมุติฐานในทดสอบว่าอย่างไร
ข. ทำการวิเคราะห์ข้อมูล พร้อมนำเสนอตาราง ANOVA และกราฟที่จำเป็นต่อการแปลผล
ค. ผลการทดสอบสามารถสรุปได้ว่าอย่างไร
ง. ถ้าหากว่ามีความแตกต่างระหว่าง treatment ในแต่ละ factor ให้เปรียบเทียบด้วยวิธี DMRT

การบ้านปฏิบัติการที่ 9

1. ในการศึกษาผลกระทบของน้ำทิ้งจากโรงงานอุตสาหกรรมที่มีต่อระบบนิเวศวิทยา (โดยใช้จำนวนไดอะตอมเป็นดัชนี) ผู้วิจัยจึงได้เก็บตัวอย่างน้ำจาก 2 แหล่ง ในช่วงเวลา 3 สัปดาห์ หลังจากมีการปล่อยน้ำออกจากโรงงานดังกล่าว แล้วนับจำนวนไดอะตอมที่มีในตัวอย่างน้ำ ได้ผลดังนี้

	week		
Sites	**1**	**2**	**3**
upstream	689	831	558
	756	916	423
downstream	204	56	34
	229	73	78

ก. จะตั้งสมมุติฐานในทดสอบว่าอย่างไร

ข. ทำการวิเคราะห์ข้อมูลโดย Two-way ANOVA นำเสนอตาราง ANOVA และกราฟที่จำเป็นต่อการแปลผล

ค. ผลการทดสอบสามารถสรุปได้ว่าอย่างไร

ง. ถ้าหากว่ามีความแตกต่างระหว่าง treatment ในแต่ละ factor ให้เปรียบเทียบค่าเฉลี่ยด้วยวิธี DMRT

2. ในการศึกษาถึงผลของระยะเวลาในการแช่น้ำร้อน (2 ระดับ) และความเข้มข้นของไคโตซาน (3 ระดับ) ต่อปริมาณน้ำตาลทั้งหมดที่ละลายน้ำได้ในผลกล้วยหอม ในการทดลอง ผู้วิจัยจะแช่ผลกล้วยหอมในน้ำที่อุณหภูมิหนึ่ง แล้วนำไปแช่ต่อในสารละลายไคโตซานที่ความเข้มข้นหนึ่ง จากนั้นจึงเก็บรักษาไว้ที่ 15 องศาเซลเซียส ก่อนจะนำไปวิเคราะห์หาปริมาณน้ำตาลทั้งหมดที่ละลายน้ำได้ในผลกล้วย แต่เนื่องจากผู้วิจัยมีข้อจำกัดเกี่ยวกับตู้เก็บรักษาที่ 15 องศาเซลเซียล โดยสามารถทำการทดลองได้เพียง 6 ชุดต่อครั้ง ทำให้แต่ละครั้งอาจได้กล้วยที่มีสรีรวิทยาต่างกัน ดังนั้น ผู้วิจัยจึงออกแบบให้การทดลองแต่ละครั้งเป็น block โดยจะทดลอง 3 ครั้ง เมื่อสิ้นสุดการทดลองได้ผลดังนี้

First batch				**Second batch**			
	Chitosan				**Chitosan**		
temperature	**0**	**10**	**100**	**temperature**	**0**	**10**	**100**
25	250	200	300	25	175	125	225
40	150	125	100	40	125	100	85

Third batch			
	Chitosan		
temperature	**0**	**10**	**100**
25	300	250	350
40	200	150	125

ก. จะตั้งสมมุติฐานในทดสอบว่าอย่างไร

ข. ทำการวิเคราะห์ข้อมูลโดย Two-way ANOVA นำเสนอตาราง ANOVA และกราฟที่จำเป็นต่อการแปลผล

ค. ผลการทดสอบสามารถสรุปได้ว่าอย่างไร

ง. ถ้าหากว่ามีความแตกต่างระหว่าง treatment ในแต่ละ factor ให้เปรียบเทียบค่าเฉลี่ยด้วยวิธี DMRT

ปฏิบัติการที่ 10 การวิเคราะห์ถดถอยเชิงเส้นและการวิเคราะห์สหสัมพันธ์

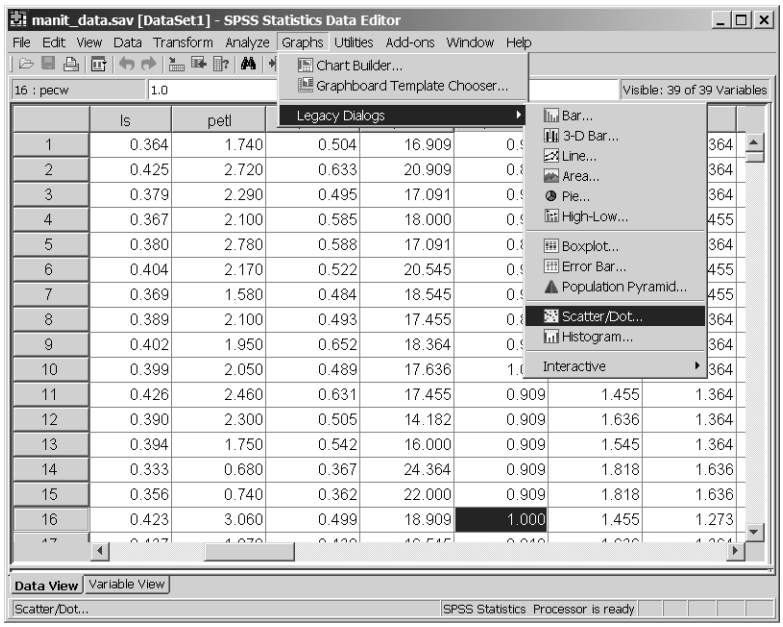
ในปฏิบัติการนี้ จะฝึกหัดการใช้คำสั่งใน SPSS เพื่อใช้สร้างกราฟเบื้องต้น ซึ่งอาจจะนำไปใช้ในการวิเคราะห์หาสมการเชิงเส้นเพื่อใช้ในการทำนายค่าการประมาณเส้นโค้ง (curve estimation) และการวิเคราะห์สหสัมพันธ์ นอกจากนี้ จะฝึกหัดใช้โปรแกรม Excel แทน SPSS ในการวิเคราะห์ที่กล่าวมาข้างต้นด้วย

1. การสร้างกราฟ

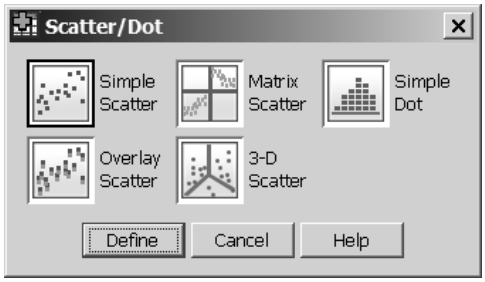
ในปฏิบัติการนี้ จะกล่าวเฉพาะกราฟ 2 มิติเท่านั้น ในกรณีที่มีชุดข้อมูลเป็นคู่ เราอาจจะต้องการใช้ค่าของตัวแปรหนึ่งเพื่อทำนายค่าของอีกตัวแปร หรือต้องการหาว่ามีสหสัมพันธ์กันหรือไม่นั้น ในเบื้องต้นควรสร้างกราฟจากชุดข้อมูลดังกล่าวก่อน เพื่อที่จะตัดสินใจว่าจะวิเคราะห์ต่อไปอย่างไร

ในการสร้างกราฟใน SPSS นั้น จะต้องมีตัวแปรอย่างน้อย 2 ตัว มีขั้นตอนดังนี้

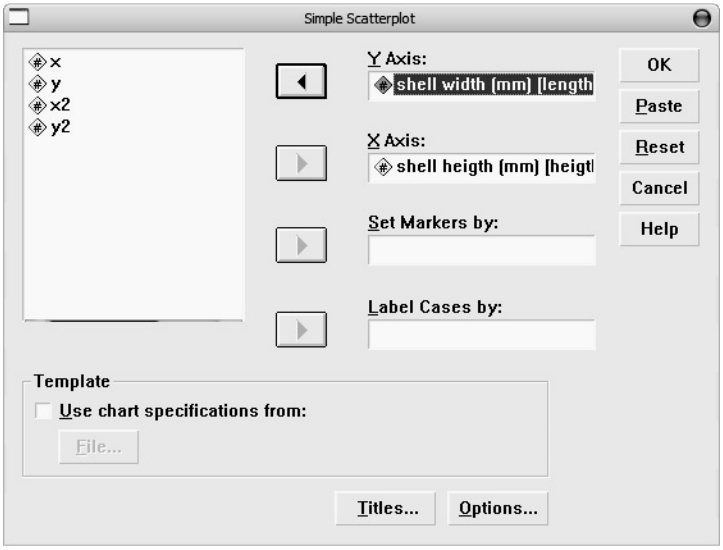
1. เลือก **Graphs**, **Legacy Dialogs**, **Scatter/Dot** (รูปที่ 10.1)
2. จาก dialog box ในรูปที่ 10.2 เลือก **Simple**
3. จาก dialog box ดังรูปที่ 10.3
   a. เลือกตัวแปรที่ต้องการให้เป็นแกนนอนจากช่องด้านซ้าย กดปุ่ม ▸ ของช่อง **X Axis:**
   b. เลือกตัวแปรที่ต้องการให้เป็นแกนตั้งจากช่องด้านซ้าย กดปุ่ม ▸ ของช่อง **Y Axis:**
   c. กด **OK**



รูปที่ 10.1



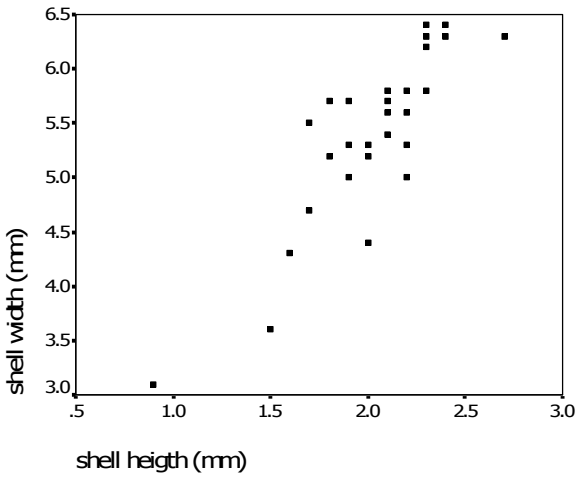
รูปที่ 10.2



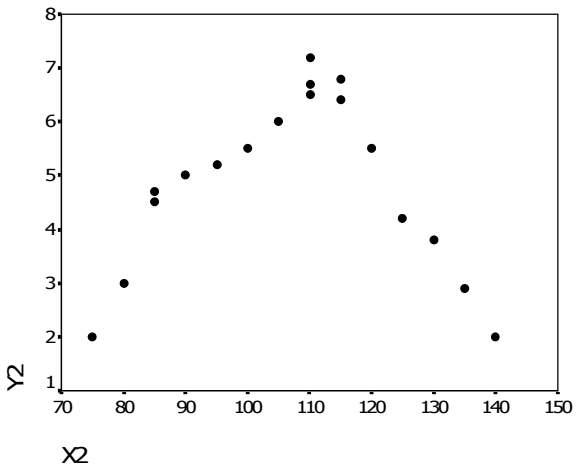
รูปที่ 10.3

จากตัวอย่างกราฟที่แสดงในรูปที่ 10.4 จะเห็นได้ว่าจุดต่าง ๆ ที่ปรากฏนั้นมีแนวโน้มว่าสามารถลากเส้นตรงผ่านจุดส่วนใหญ่ได้ (แม้ว่าบางจุดจะอยู่ห่างจากเส้นตรงที่ลากขึ้น ขณะที่บางจุดจะอยู่ใกล้กว่า) อย่างไรก็ตาม เมื่อพิจารณากราฟอีกตัวอย่างที่แสดงในรูปที่ 10.5 แล้ว จะเห็นว่าหากลากเส้นโค้งในลักษณะคว่ำแล้ว จะผ่านจุดต่างๆส่วนใหญ่ โดยที่ระยะห่างระหว่างจุดต่างๆ ถึงเส้นโค้งจะน้อยกว่ามากเมื่อเทียบว่าหากต้องลากเป็นเส้นตรง ดังนั้น การพิจารณากราฟ จึงช่วยในการตัดสินใจว่าจะวิเคราะห์ข้อมูลต่อไปอย่างไร

.



รูปที่ 10.4



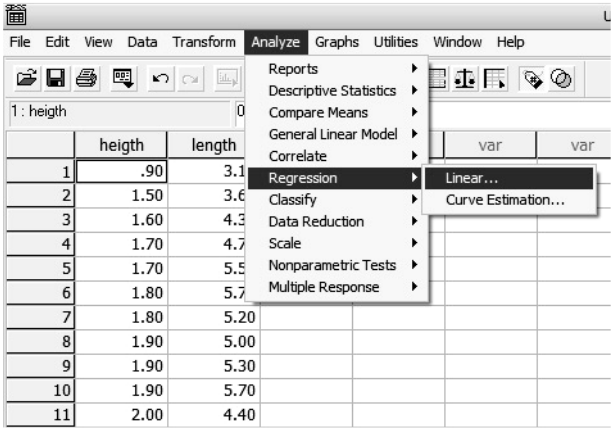
รูปที่ 10.5

2. การวิเคราะห์ถดถอยเชิงเส้นตรง (Linear regression analysis)

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

ในการวิเคราะห์ถดถอยเชิงเส้นตรงนั้น มีขั้นตอนดังนี้

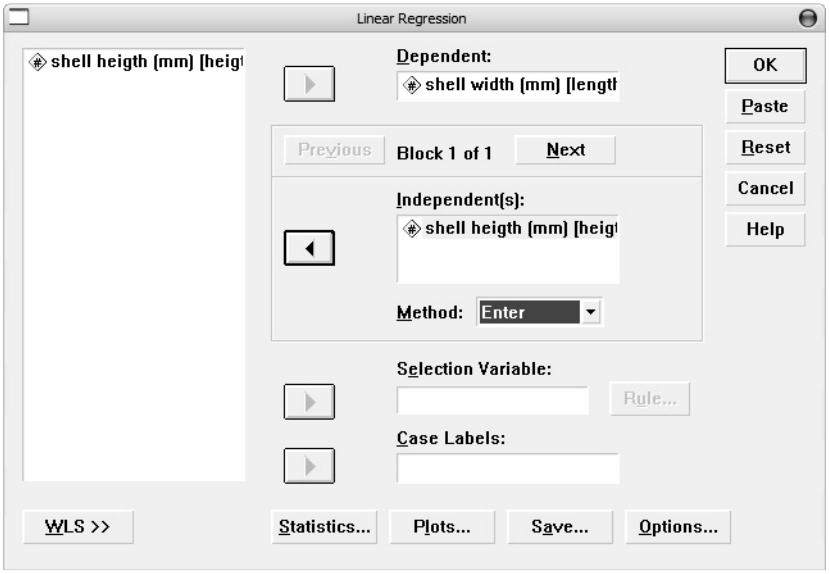
1. เลือก Analyze, Regression, Linear... (รูปที่ 10.6)



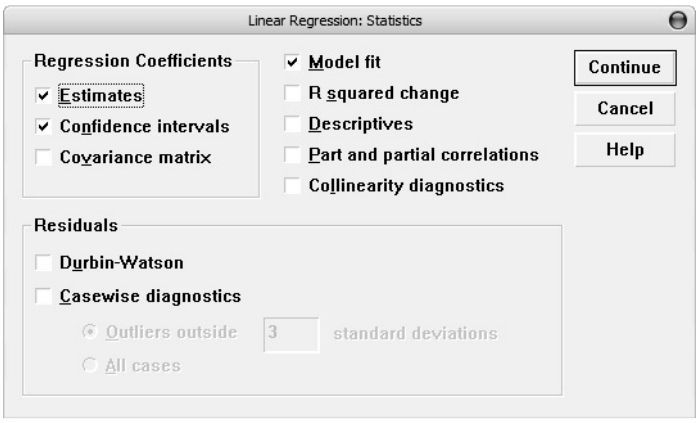
รูปที่ 10.6

2. จาก dialog box ในรูปที่ 10.7
   a. เลือกตัวแปรอิสระจากช่องด้านซ้าย กดปุ่ม ▸ ของช่อง Independent(s):
   b. เลือกตัวแปรตามจากช่องด้านซ้าย กดปุ่ม ▸ ของช่อง Dependent(s):
   c. กดปุ่ม Statistics... เพื่อเลือกค่าสถิติที่ต้องการ จะมี dialog box ในรูปที่ 10.8
      i. ในช่อง Regression Coefficients เลือก Estimates และ Confident Intervals
      ii. เลือก Model Fit
      iii. กดปุ่ม Continue
   d. กดปุ่ม Plots... เพื่อสร้างกราฟ โดยในที่นี้ จะใช้กราฟเพื่อตรวจสอบว่าข้อมูลเป็นไปตาม assumption ของการสร้างสมการถดถอยเชิงเส้นหรือไม่ จาก dialog box ในรูปที่ 10.9
      i. เลือก *ZPRED (=standardized predicted values, $\hat{Y}$) กดปุ่ม ▸ ของช่อง X:
      ii. เลือก *ZRESID (=standardized residuals, $Y-\hat{Y}$) กดปุ่ม ▸ ของช่อง Y:

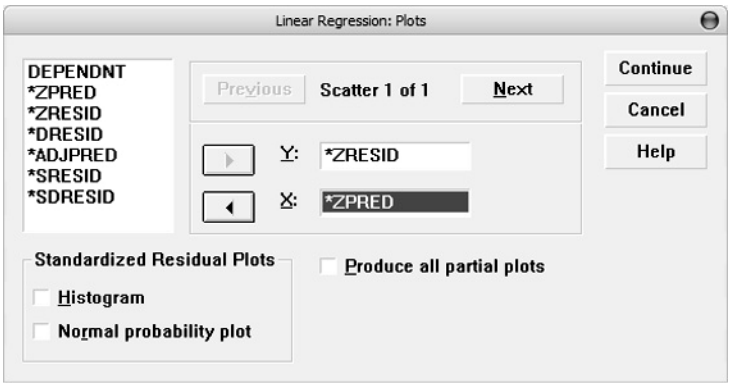
iii. กดปุ่ม Continue



รูปที่ 10.7



รูปที่ 10.8



รูปที่ 10.9

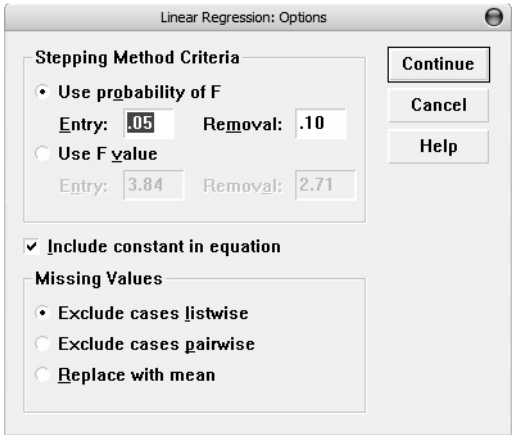
e. กดปุ่ม Option... จะมี dialog box ในรูปที่ 10.10

i. เลือก Include constant in equation[1]

ii. กดปุ่ม Continue

[1] ถ้าหากว่าชุดข้อมูลที่ต้องการวิเคราะห์ เมื่อค่าตัวแปรอิสระเป็น 0 แล้วตัวแปรตามเป็น 0 ด้วย ไม่ต้องเลือก Include constant in equation

f. กดปุ่ม OK



รูปที่ 10.10

ผลการวิเคราะห์ที่สำคัญ คือ ตาราง Model Summary (รูปที่ 10.11) ซึ่งจะมีค่า R^2 (คือ coefficient of determination) ที่จะอธิบายเปอร์เซ็นต์ความแปรปรวนของตัวแปรตามที่สามารถอธิบายได้จากความแปรปรวนของตัวแปรอิสระ หรืออีกนัยหนึ่ง ค่า R^2 นี้ เป็นค่าที่แสดงถึงความพอดีของสมการถดถอยเชิงเส้นกับข้อมูลดิบ (คือ goodness of fit)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.856[a]	.733	.723	.42467

a. Predictors: (Constant), shell heigth (mm)

รูปที่ 10.11

ตารางถัดไป คือ ตาราง ANOVA (รูปที่ 10.12) ซึ่งจะทดสอบว่าตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์เชิงเส้นตรงหรือไม่ (นั่นคือ ทดสอบสมมุติฐาน Ho: b = 0 vs. H1: b ≠ 0) หากผลการทดสอบปฏิเสธ Ho แล้ว แสดงว่า ตัวแปรอิสระและตัวแปรตามมีความสัมพันธ์เชิงเส้นตรงกัน และสามารถนำสมการถดถอยเชิงเส้นตรงนี้ไปใช้ในการทำนายค่า Y ได้เมื่อกำหนดค่า X โดยการทดสอบจะพิจารณาค่า Sig. ว่ามีค่าเท่ากับหรือน้อยกว่า 0.05 แล้ว จะปฏิเสธ Ho

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	12.878	1	12.878	71.407	.000[a]
	Residual	4.689	26	.180		
	Total	17.567	27			

a. Predictors: (Constant), shell heigth (mm)

b. Dependent Variable: shell width (mm)

รูปที่ 10.12

ตารางที่จะพิจารณาต่อไป คือ ตารางค่า coefficient (รูปที่ 10.13) โดยในสมการ Y = a + bX จะใช้ค่าในช่อง Unstandardized coefficients[2] ซึ่งในตัวอย่างที่แสดงในรูปที่ 10.13 นี้ ค่า a และ b จะเป็นค่าในบรรทัด (Constant) และ shell height (mm) ตามลำดับ แต่ก่อนจะนำค่าดังกล่าวไปแทนค่าในสมการต้องพิจารณาก่อนว่าค่า Sig. ของแต่ละค่ามีค่าเท่ากับหรือน้อยกว่า 0.05 หรือไม่[3] ถ้าใช่ แสดงว่าสามารถนำค่าทั้งสองไปใช้ในสมการได้ ในตารางนี้ ยังมีค่า 95% Confidence Interval สำหรับค่าสัมประสิทธิ์ทั้งสองด้วย

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.376	.482		2.857	.008	.386	2.366
	shell heigth (mm)	1.985	.235	.856	8.450	.000	1.502	2.468

a. Dependent Variable: shell width (mm)

รูปที่ 10.13

ในการวิเคราะห์ถดถอยเชิงเส้นตรง มี assumption ที่สำคัญคือ สำหรับค่าของตัวแปรอิสระค่าหนึ่งๆ ค่าตัวแปรตามจะมีการกระจายแบบปกติ โดยค่าตัวแปรตามที่เห็นในข้อมูลดิบเป็นเพียงค่าหนึ่งเท่านั้น ในการทดสอบว่าค่าตัวแปรตามมีการกระจายแบบปกติหรือไม่ รวมถึงความถูกต้องในการทำนายจากสมการถดถอยเชิงเส้นตรง สามารถพิจารณาได้จากกราฟที่แสดงในรูปที่ 10.14 (สร้างจาก Plots... ในรูป 10.9) ถ้าหากว่ามีการกระจายของจุดสม่ำเสมอ หมายความว่าข้อมูลดิบเป็นไปตาม assumption และสมการถดถอยเชิงเส้นตรงมีความเหมาะสมที่จะใช้ในการทำนายค่า แต่หากว่ากราฟมีการกระจายต่างออกไป เช่น ด้านหนึ่งของกราฟมีกระจายของจุดในแนวตั้งเป็นช่วงแคบๆ แต่ที่ปลายอีกหนึ่งของกราฟกลับมีการกระจายของจุดในแนวตั้งเป็นช่วงที่กว้างกว่า แสดงว่าอาจจะต้องแปลงค่าข้อมูลดิบก่อนที่จะวิเคราะห์หาสมการถดถอยเชิงเส้นตรง หรือในกรณีที่การกระจายของจุดเป็นลักษณะโค้ง ก็แสดงว่าไม่สามารถใช้สมการถดถอยเชิงเส้นตรงอธิบายข้อมูลดิบได้ ต้องใช้ model อื่น (เช่น S curve, Parabolic, ฯลฯ) ในตัวอย่างกราฟในรูปที่ 10.14 นี้ อนุโลมได้ว่ามีการกระจายของจุดสม่ำเสมอ แสดงว่า สมการถดถอยเชิงเส้นตรงที่ได้เหมาะสมกับข้อมูลดิบชุดนี้

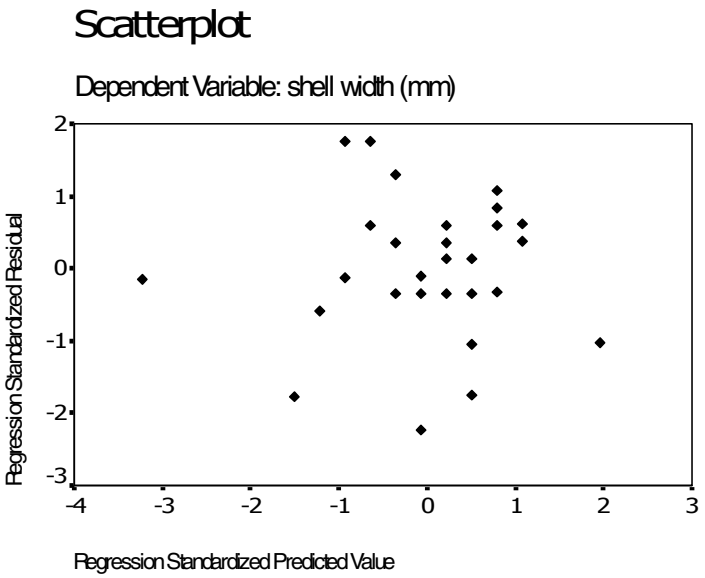
ข้อควรคำนึงถึงเกี่ยวกับสมการถดถอยเชิงเส้นตรงนี้ คือ ค่าตัวแปรอิสระที่จะใช้ในการทำนายค่าตัวแปรตามจากสมการนี้ **จะต้องอยู่ในช่วงของค่าต่ำสุดและค่าสูงสุดที่ใช้ในการสร้างสมการนี้** ไม่เช่นนั้นก็ไม่

---

[2] หากต้องการใช้กับข้อมูลที่เป็น standardized score แล้ว จะใช้สมการ Y = bX โดยค่า b เป็นค่าจากช่อง Standardized coefficients ในตาราง Coefficient (ในรูปที่ 10.13)

[3] 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 และควรสร้างสมการถดถอยใหม่โดยไม่ต้องเลือก Include constant in equation (ใน dialog box รูป 10.10) อนึ่ง มีข้อแนะนำว่าแม้ว่าจะพบว่าไม่อาจปฏิเสธ Ho: a = 0 ก็อาจใช้ค่า a ในสมการถดถอยได้โดยไม่ถือว่าผิดแต่อย่างไร

สามารถนำสมการนี้ไปใช้ได้ และ**ไม่ควรนำไปใช้** นอกจากนี้ หากต้องการทราบช่วงความเชื่อมั่น 95% ของค่าตัวแปรตามที่ทำนายได้จากสมการถดถอยเชิงเส้นตรง ก็สามารถหาได้โดยการแทนค่าสัมประสิทธิ์ a และ b ด้วยค่าที่แสดงในตารางในรูปที่ 10.13 ช่อง 95% Confidence Interval



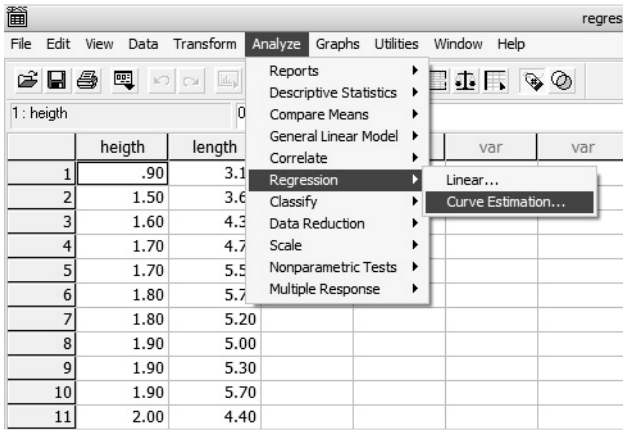
รูปที่ 10.14

3. การประมาณเส้นโค้ง (Curve estimation)

เนื่องจากว่า Regression, Linear... จะไม่สร้างกราฟที่แสดงข้อมูลดิบและเส้นกราฟที่ลากผ่านข้อมูลดิบ จึงจำเป็นต้องใช้คำสั่งอื่นในการสร้างกราฟประกอบการวิเคราะห์ถดถอยเชิงเส้น หรือในกรณีที่ไม่สามารถใช้การวิเคราะห์ถดถอยเชิงเส้นได้ เนื่องมาจากข้อมูลไม่มีความสัมพันธ์เชิงเส้นตรง ก็ยังมีคำสั่งที่ช่วยสร้างกราฟหรือวิเคราะห์หาสมการถดถอยแบบอื่นที่ไม่ใช่เชิงเส้นตรงที่เหมาะสมกับข้อมูลดิบได้

ในการสร้างเส้นกราฟที่ไม่ใช่เส้นตรง มีขั้นตอนดังนี้

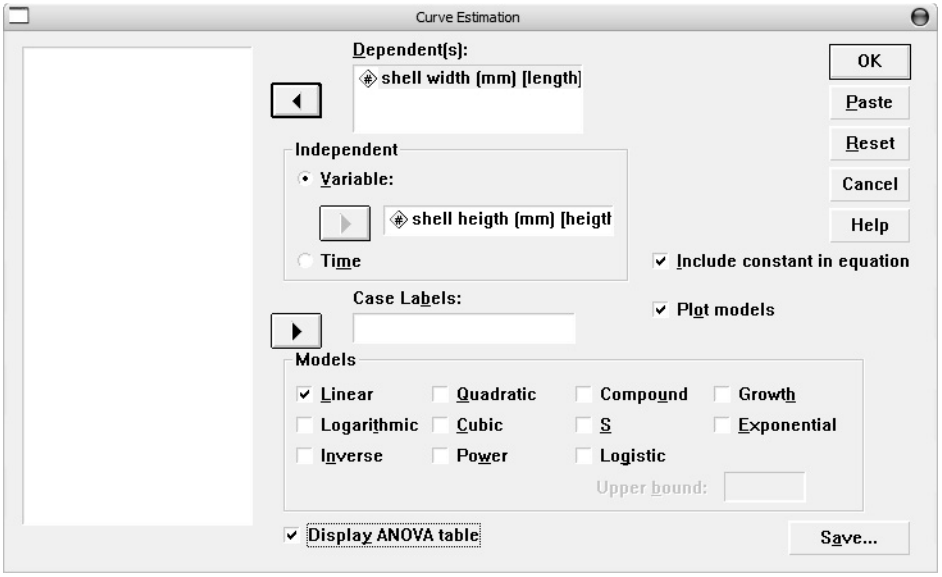
1. เลือก Regression, Curve Estimate... (รูปที่ 10.15)



รูปที่ 10.15

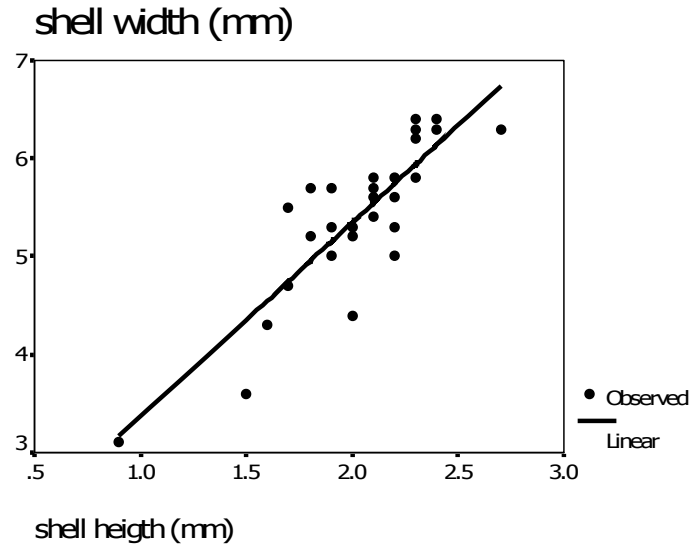
2. จาก dialog box ในรูปที่ 10.16

a. เลือกตัวแปรอิสระจากช่องด้านซ้าย กดปุ่ม ▸ ของช่อง Independent, Variable:
b. เลือกตัวแปรตามจากช่องด้านซ้าย กดปุ่ม ▸ ของช่อง Dependent(s):
c. เลือก Include constant in equation
d. เลือก Plot models
e. ในช่อง Models เลือกแบบกราฟที่จะประมาณจากตัวเลือก ซึ่งหากไม่ทราบ อาจทดลองเลือกกราฟทุกแบบ อย่างไรก็ตาม ควรใช้คำสั่งสร้างกราฟดังที่อธิบายในข้อ 1. เพื่อเลือกแบบกราฟที่เหมาะสม
f. เลือก Display ANOVA table เพื่อให้ทดสอบนัยสำคัญทางสถิติของ model ที่เลือกและหาค่า R^2 ของแต่ละ model

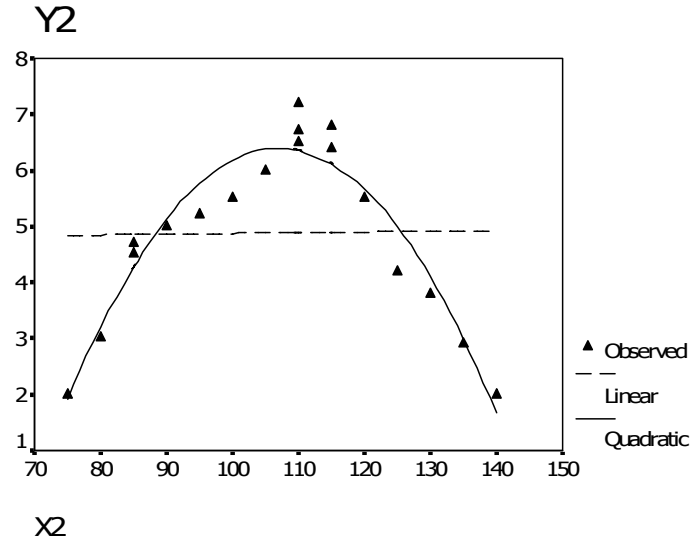


รูปที่ 10.16

ผลการวิเคราะห์ จะได้ตาราง ANOVA ที่ได้จากการวิเคราะห์ถดถอยของ model แต่ละแบบที่เลือกในช่อง Models (ไม่ได้นำมาแสดง) และกราฟที่แสดงจุดข้อมูลดิบพร้อมกราฟเส้น กราฟที่แสดงในรูปที่ 10.17 ได้มาจากการกราฟเลือก Linear ใน Models จึงมีกราฟเส้นเพียง 1 เส้น แต่กราฟที่แสดงในรูปที่ 10.18 ได้มาจากการเลือกกราฟ Linear และ Quadratic ใน Models จึงมีกราฟเส้น 2 เส้น คือ เส้นตรงได้จากตัวเลือก Linear ส่วนเส้นโค้งได้จากตัวเลือก Quadratic ซึ่งจะเห็นได้ว่ากราฟแบบ quadratic เหมาะสมกับข้อมูลดิบ (เมื่อพิจารณาจากกราฟและ R^2)



รูปที่ 10.17



รูปที่ 10.18

4. การวิเคราะห์ความสัมพันธ์เชิงเส้นตรง (สำหรับข้อมูลที่มีการกระจายแบบปกติ)

การวิเคราะห์ความสัมพันธ์เชิงเส้นตรงเป็นการวิเคราะห์ที่ต้องการทราบว่าตัวแปรคู่หนึ่งมีสัมพันธ์เชิงเส้นตรงหรือไม่? ถ้ามี เป็นอย่างไร? ในการเตรียมข้อมูล จะต้องใช้ตัวแปรอย่างน้อย 2 ตัวแปร ซึ่งต้องมีค่าสังเกตเท่ากัน อนึ่ง ในปฏิบัติการที่ 8 ได้เคยกล่าวถึงการวิเคราะห์สหสัมพันธ์แบบนอนพาราเมตริกซ์มาบ้างแล้ว แต่เพื่อให้เกิดความเข้าใจที่ถูกต้องยิ่งขึ้นจะขอเพิ่มเติมรายละเอียดเกี่ยวการตั้งสมมุติฐานและการทดสอบ รวมถึงการแปลผลความสัมพันธ์ระหว่างภายหลังจากได้ค่า r แล้ว ดังนี้

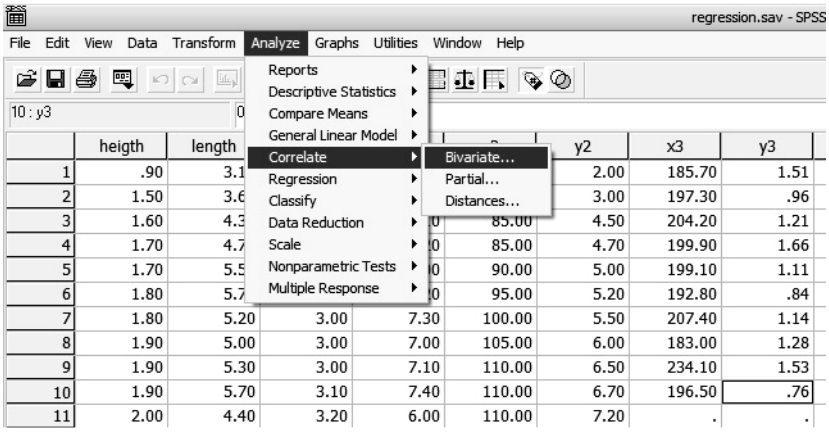
การตั้งสมมุติฐาน	ความหมาย	การทดสอบ
Ho: r = 0 H1: r ≠ 0	Ho: ไม่มีความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรทั้งสอง H1: มีความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรทั้งสอง	2-tailed
Ho: r = 0 H1: r > 0	Ho: ไม่มีความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรทั้งสอง H1: มีความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรทั้งสองแบบบวก (คือ X เพิ่ม Y เพิ่มหรือ X ลด Y ลด)	1-tailed
Ho: r = 0 H1: r < 0	Ho: ไม่มีความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรทั้งสอง H1: มีความสัมพันธ์เชิงเส้นตรงระหว่างตัวแปรทั้งสองแบบลบ (คือ X เพิ่ม Y ลด หรือกลับกัน)	1-tailed

และในการแปลผลจากค่า correlation coefficient หรือค่า r จะเป็นดังนี้ (ค่าเป็นบวกหรือลบ แปลผลเหมือนกัน)

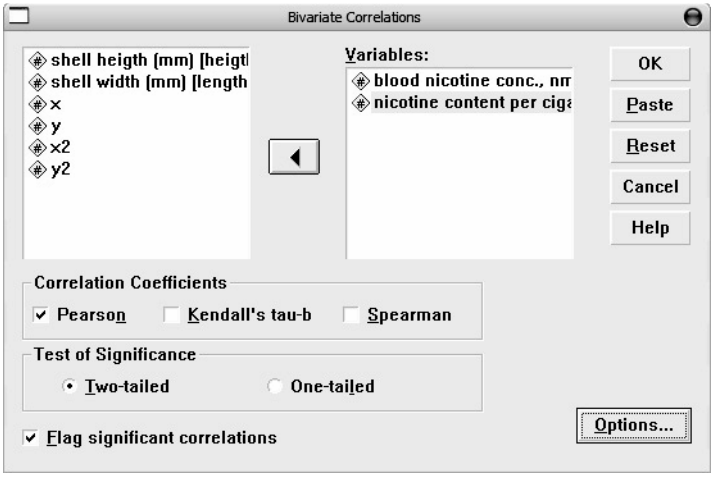
ค่า correlation coefficient	ความหมาย
0.00 - 0.25	no or weak relationship
0.25 - 0.50	fair degree of relationship
0.50 - 0.75	moderate to good relationship
0.75 - 1.00	good to strong relationship

ในการวิเคราะห์สหสัมพันธ์ จะมีขั้นตอนดังนี้

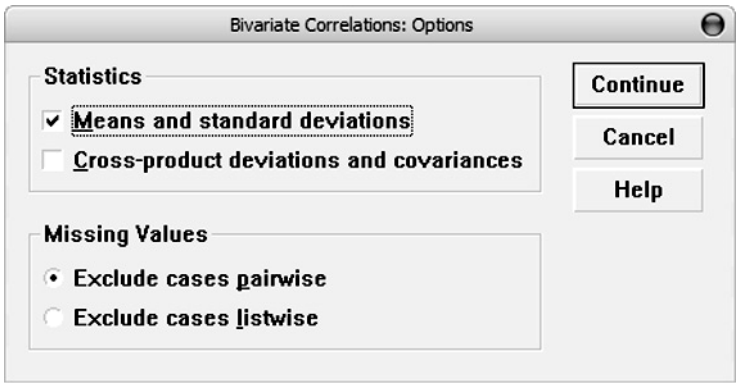
1. เลือก Analyze, Correlate, Bivariate... (รูปที่ 10.19)
2. จาก dialog box ในรูปที่ 10.20
3. เลือกตัวแปรที่ต้องการวิเคราะห์ในช่องด้านซ้าย (2 ตัวแปรหรือมากกว่า) แล้วกดปุ่ม ▸ ของช่อง Variables:
    a. ที่ช่อง Correlation Coefficients เลือก Pearson
    b. ที่ช่อง Test of Significance เลือกแบบ Two-tailed หรือ One-tailed
    c. เลือก Flag significant correlations เพื่อให้โปรแกรมติดสัญญลักษณ์ที่แสดงค่า r ที่มีนัยสำคัญทางสถิติ ซึ่งช่วยให้ง่ายต่อการพิจารณาค่า r ในกรณีที่วิเคราะห์สหสัมพันธ์ตัวแปรหลายคู่
    d. กดปุ่ม Options... แล้วเลือกค่าสถิติที่ต้องการ (ในรูปที่ 10.21)
    e. กดปุ่ม Continue
4. กดปุ่ม OK



รูปที่ 10.19



รูปที่ 10.20



รูปที่ 10.21

ผลการวิเคราะห์ที่สำคัญ คือ ตาราง Correlation ดังแสดงในรูปที่ 10.22 ซึ่งจะมี 2 แถว แต่ละแถวมี 2 คอลัมน์ (จำนวนช่องจะเท่ากับ n×n ช่อง เมื่อ n คือ จำนวนตัวแปรในการวิเคราะห์) โดยช่องในแนวทแยงมุมจะมีค่า coefficient เป็น 1 เนื่องจากเป็นค่าสัมประสิทธิ์สหสัมพันธ์ของคู่ตัวแปร ซึ่งตัวแปรทั้งสองเป็นตัวแปรเดียวกัน ดังนั้น การพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์จึงดูจากค่าที่อยู่ซีกบนหรือซีกล่างของแนวทแยงมุมดังกล่าวก็ได้

ในแต่ละแถว มี 3 บรรทัด คือ บรรทัดแรกเป็นค่า Pearson correlation coefficient บรรทัดที่สองเป็นค่า Sig. ของค่า Pearson correlation coefficient และบรรทัดที่สามเป็นจำนวนข้อมูลดิบ ในการพิจารณา จะดูว่าถ้าค่า Sig. เท่ากับหรือน้อยกว่า 0.05 จะปฏิเสธ Ho แสดงว่าตัวแปรทั้งสองมีความสัมพันธ์เชิงเส้นตรงกัน

Correlations

		blood nicotine conc., nmol/L)	nicotine content per cigarette (mg)
blood nicotine conc., nmol/L)	Pearson Correlation	1	.258
	Sig. (2-tailed)	.	.472
	N	10	10
nicotine content per cigarette (mg)	Pearson Correlation	.258	1
	Sig. (2-tailed)	.472	.
	N	10	10

รูปที่ 10.22

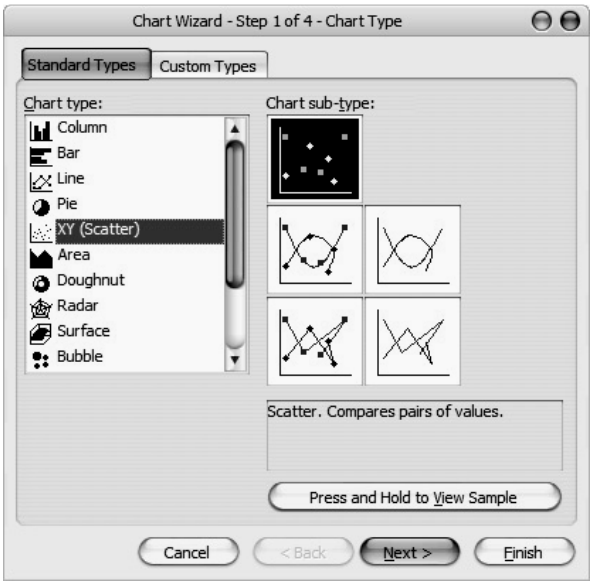
ข้อควรคำนึงถึงเมื่อวิเคราะห์ความสัมพันธ์เชิงเส้นตรง ทั้งที่เป็นแบบพาราเมตริกซ์และนอนพาราเมตริกซ์ คือ ถ้าหากว่าค่าสัมประสิทธิ์สหสัมพันธ์มีนัยสำคัญทางสถิติแล้ว แสดงว่าตัวแปรทั้งสองที่พิจารณามีความสัมพันธ์เชิงเส้นตรงกันเท่านั้น ซึ่งจะมากน้อยขึ้นอยู่กับค่าสัมประสิทธิ์สหสัมพันธ์ แต่ไม่อาจใช้ผลที่ได้ไปสรุปเพิ่มเติมเกี่ยวกับความแตกต่างระหว่างกลุ่มตัวอย่างได้ นอกจากนี้ ยังไม่อาจใช้ผลการวิเคราะห์นี้เพื่ออธิบายเกี่ยวกับเหตุและผลที่เกิดจากตัวแปรหนึ่งต่ออีกตัวแปรหนึ่ง (นั้นคือ ไม่ใช่ causation) ดังนั้น จึงควรระมัดระวังเกี่ยวกับการนำผลการวิเคราะห์ความสัมพันธ์เชิงเส้นตรงไปอธิบาย

5. การใช้โปรแกรม Microsoft® Excel ในการสร้างกราฟและการประมาณเส้นโค้ง (รวมทั้งเส้นตรงและแบบอื่นๆ)

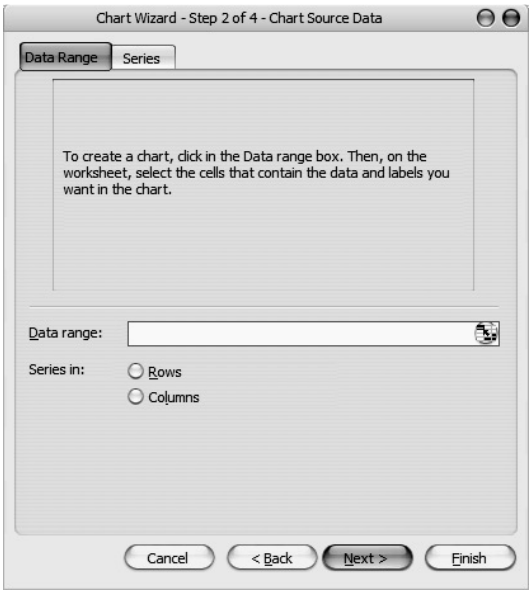
โปรแกรม Microsoft® Excel มีคำสั่งสร้างกราฟที่มีขั้นตอนง่าย ๆ เพื่อใช้สร้างกราฟแบบต่าง ๆ ในการเตรียมข้อมูลนั้น จะพิมพ์ค่าสังเกตของแต่ละตัวแปรให้อยู่ในคอลัมน์แยกกัน (หรือในแถวแยกแถวกันก็ได้) การสร้างกราฟใน Microsoft® Excel มีขั้นตอนดังนี้

1. กดที่ icon  (หรือใช้คำสังจากเมนู Insert, Chart...)
2. จาก dialog box ของ Chart Wizard ดังรูปที่ 10.23
    a. เลือกแบบกราฟโดยกดแถบ Standard Types
        i. ที่ช่อง Chart type: ซึ่งอยู่ด้านซ้าย เลือก XY (Scatter)
        ii. ที่ช่อง Chart sub-type: ซึ่งอยู่ด้านขวา เลือกกราฟที่ไม่มีเส้น
        iii. กดปุ่ม Next >
    b. การระบุชุดข้อมูล ทำได้ 2 วิธี
        i. วิธีแรก เป็นการเลือกขอบเขตของข้อมูล (กดแถบ Data Range ดังรูปที่ 10.24)
            1. กดที่ช่องของ Data range: แล้วระบุขอบเขตข้อมูล

2. ระบุว่าข้อมูล 1 ตัวแปรจัดเรียงแยกอยู่ในแถว (row) หรือคอลัมน์ (column) ข้อจำกัดของการเลือกข้อมูลแบบนี้คือ Microsoft® Excel กำหนดให้ข้อมูลแถวแรก (หรือคอลัมน์แรก) จะอยู่แกน X แล้วข้อมูลถัดไปอยู่แกน Y ดังนั้นจึงต้องป้อนข้อมูลให้สอดคล้องกัน



รูปที่ 10.23



รูปที่ 10.24

ii. วิธีที่สอง เป็นการกำหนดชุดข้อมูลในแต่ละแกนเอง (กดแถบ Series ดังรูปที่ 10.25)

1. กดปุ่ม Add ใต้ช่อง Series:
2. จะมีช่องให้ป้อนข้อมูลทางด้านขวาของช่อง Series:
   a. ที่ช่อง Name: พิมพ์ชื่อชุดข้อมูล
   b. ที่ช่อง X Values: เลือกข้อมูลที่เป็นแกน X
   c. ที่ช่อง Y Values: เลือกข้อมูลที่เป็นแกน Y สังเกตว่าจะมีกราฟตัวอย่างให้ดูด้วย

iii. เมื่อกำหนดชุดข้อมูลตาม 2.b.i หรือ 2.b.ii แล้ว กดปุ่ม Next >

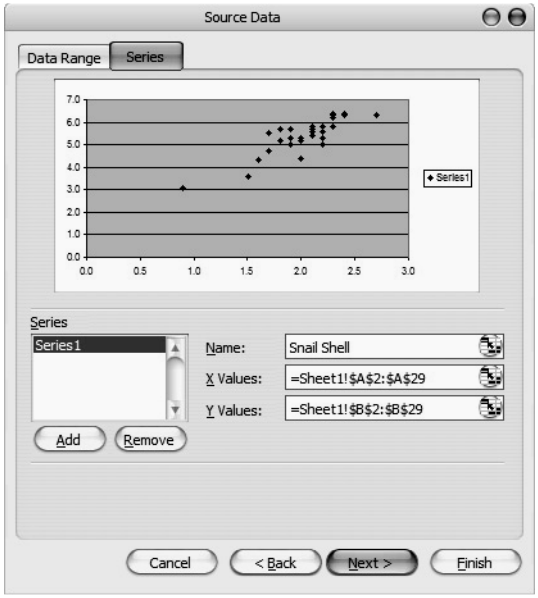
c. การกำหนดองค์ประกอบอื่น ๆ ของกราฟ โดยใช้แถบใน dialog box ดังรูปที่ 10.26

i. แถบ Titles ใช้กำหนดชื่อกราฟ แกน X และ แกนY
ii. แถบ Axes ใช้กำหนดแกน X และแกน Y
iii. แถบ Gridlines ใช้กำหนดเส้น grid ของแกน X และแกน Y
iv. แถบ Legend ใช้กำหนดรูปแบบและตำแหน่งที่วาง legend ของกราฟ

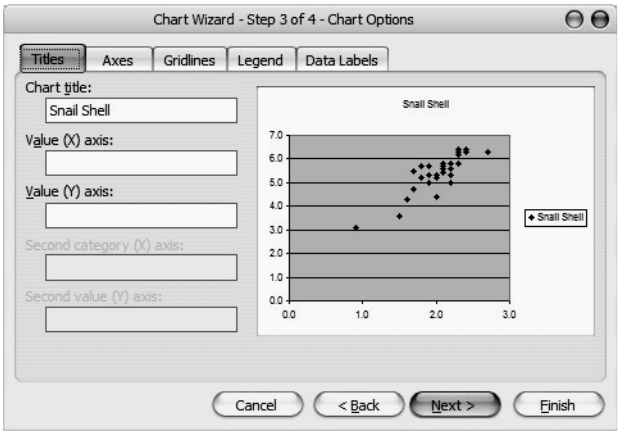
v. แถบ **Data Labels** ใช้กำหนดการแสดงข้อความของชุดข้อมูล

vi. เมื่อกำหนดค่าต่าง ๆ เสร็จแล้ว กดปุ่ม **Next >**

d. กำหนดที่วางกราฟ ซึ่งสามารถกำหนดให้กราฟที่สร้างอยู่ใน sheet ใหม่หรือเป็น chart object อยู่ใน sheet ที่มีข้อมูลของกราฟอยู่ (ดัง dialog box ในรูปที่ 10.27) เมื่อเลือกได้แล้ว กดปุ่ม **Finish** จะได้กราฟดังรูปที่ 10.28



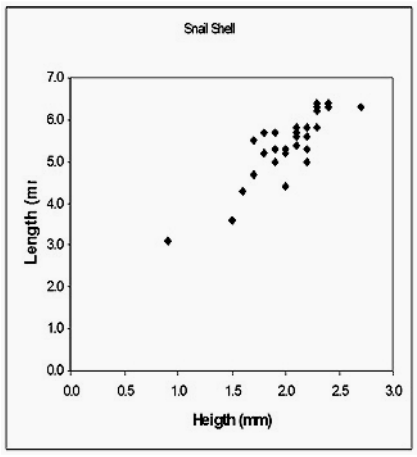
รูปที่ 10.25



รูปที่ 10.26



รูปที่ 10.27



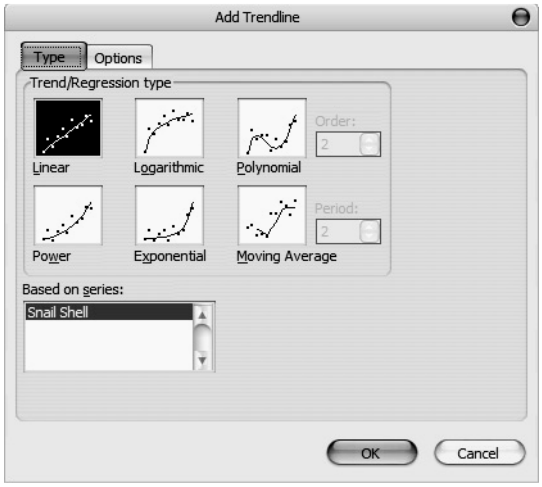
รูปที่ 10.28

เมื่อได้กราฟแล้ว หากต้องการสร้างเส้นกราฟที่ลากผ่านจุดในกราฟ ก็สามารถทำได้ (ไม่จำกัดว่ากราฟจะเป็น scatter plot หรือ line plot) โดยมีขั้นตอนดังนี้

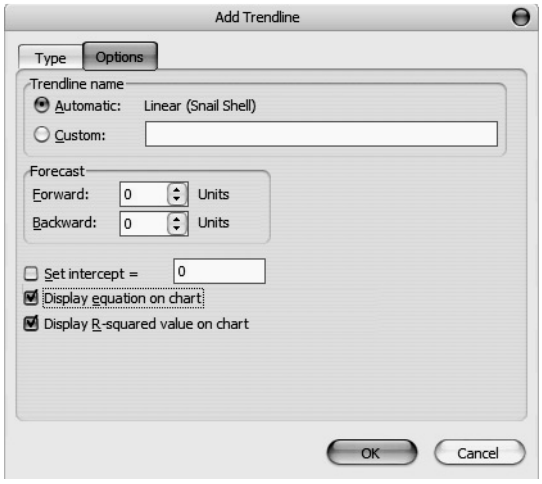
1. กดที่กราฟ จะปรากฏ ■ อยู่รอบกราฟและสังเกตว่าที่เมนูจะมีคำสั่ง **Chart** เพิ่มขึ้นมา
2. เลือก **Chart, Add TrendLine...** จะมี dialog box ดังรูปที่ 10.29

a. กดที่แถบ Type ซึ่งแสดงแบบของการประมาณเส้นกราฟแบบต่างๆ
   i. ที่ช่อง Trend/Regression Type ให้เลือกแบบ trendline ที่สอดคล้องกับข้อมูลในตัวอย่างนี้ การกระจายจุดมีแนวโน้มเป็นเส้นตรง จึงจะเลือก Linear
   ii. ที่ช่อง Based on series:[4] ให้ระบุชื่อกราฟที่ต้องการเพิ่ม trendline ในตัวอย่างนี้ กราฟในรูป 10.28 มีชื่อชุดข้อมูลว่า Snail Shell
b. กดที่แถบ Options เพื่อกำหนดตัวเลือกอื่น ๆ จาก dialog box ในรูป 12.30
   i. ในตัวอย่างนี้ จะเลือก Display equation on chart และ Display R-squared value on chart
c. เมื่อกำหนดเสร็จแล้ว กดปุ่ม OK จะได้กราฟดังรูปที่ 10.31

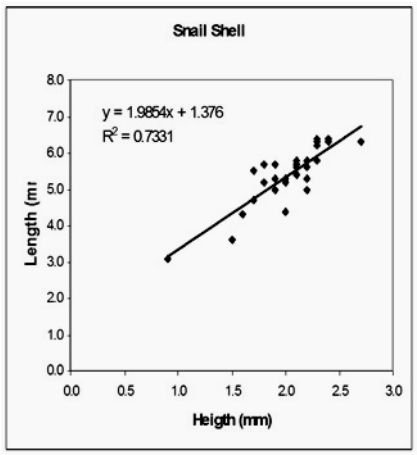
กราฟในรูปที่ 10.32 เป็นตัวอย่างของการใช้คำสั่ง Chart, Add TrendLine... เพื่อสร้างเส้นกราฟแบบ Polynomial และกำหนด Order: เป็น 2



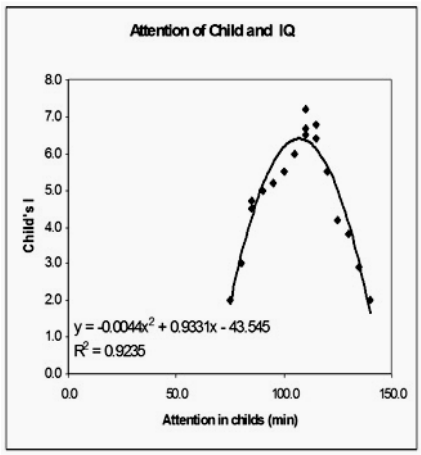
รูปที่ 10.29



รูปที่ 10.30



รูปที่ 10.31



รูปที่ 10.32

[4] ควรระบุชื่อของชุดข้อมูลที่สร้างกราฟทุกๆกราฟ โดยกำหนดใน dialog box ในรูปที่ 10.24

ข้อควรคำนึงถึงในการสร้างกราฟและการประมาณเส้นโค้งด้วยโปรแกรม Microsoft® Excel คือ ค่าสัมประสิทธิ์และค่าคงที่ต่างๆ ที่อยู่ในสมการนั้น ไม่มีการทดสอบทางสถิติแต่อย่างใด เนื่องจากไม่มีค่าสถิติที่จำเป็นต่อการทดสอบสมมุติฐานใดๆ ไม่ว่าจะเป็นสมการถดถอยที่สร้างขึ้น ค่าสัมประสิทธิ์ และค่าคงที่ต่างๆ ในสมการนั้น ดังนั้น หากเป็นไปได้ จึงควรหาวิธีการทดสอบสมมุติฐานที่จำเป็นก่อนที่จะนำผลไปใช้ต่อไป

6. การใช้โปรแกรม Microsoft® Excel ในการวิเคราะห์ถดถอยเชิงเส้นตรง

โปรแกรม Microsoft® Excel มีโมดุล Regression ที่สามารถนำมาใช้สร้างสมการถดถอยและทดสอบทางสถิติด้วย โดยเป็นโมดุลหนึ่งใน Data Analysis add-ins

ในการเตรียมข้อมูล จะป้อนข้อมูลแยกเป็นคอลัมน์ๆ ละ 1 ตัวแปร (หรือเป็นแถวก็ได้) และขั้นตอนการวิเคราะห์มีดังนี้

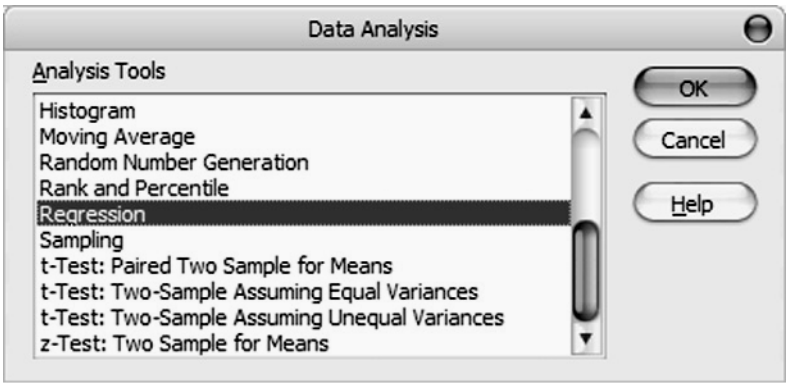
1. ใช้คำสั่ง Tools, Data Analysis…[5]
2. จากรายการตัวเลือกในช่อง Analysis Tools (รูปที่ 10.33) ให้เลือก Regression
3. กดปุ่ม OK
4. จาก dialog box ที่ปรากฏขึ้นมาใหม่ดังรูปที่ 10.34
    a. ที่ช่อง Input
        i. กดที่ช่อง Input Y Range: แล้วเลือกช่วงข้อมูลที่เป็นตัวแปรตาม
        ii. กดที่ช่อง Input X Range: แล้วเลือกช่วงข้อมูลที่เป็นตัวแปรอิสระ
        iii. ถ้าหากว่าค่าคงที่ (ค่า a) เป็น 0 ให้เลือก Constant is Zero
    b. ที่ช่อง Output options
        i. เลือก New Worksheet Ply: แล้วตั้งชื่อ sheet (หรืออาจเลือก Output Range: หรือ New Workbook ก็ได้)
        ii. ที่ช่อง Residuals เลือก Residual Plots และ Line Fit Plot
    c. เมื่อกำหนดค่าต่างๆ แล้ว กดปุ่ม OK

ผลการวิเคราะห์ จะปรากฏใน sheet ที่สร้างขึ้นใหม่ โดยมีตารางต่างๆ และกราฟต่างๆ ดังนี้

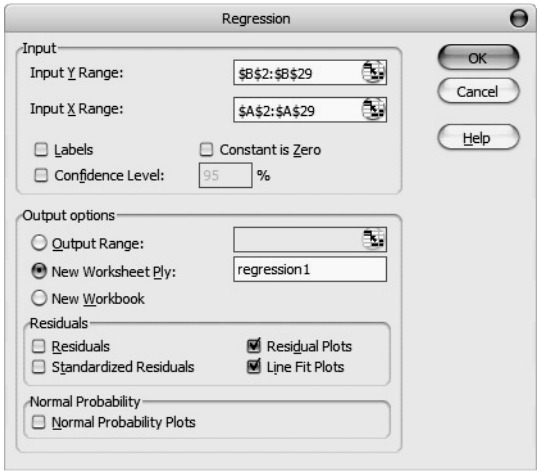
ตารางจากการวิเคราะห์	**ตารางเทียบเคียงจาก SPSS**
SUMMARY OUTPUT	Model Summary รูปที่ 10.11
ANOVA	ANOVA รูปที่ 10.12 และ 10.13
RESIDUAL PLOT	กราฟในรูปที่ 10.14
LINE FIT PLOT	กราฟในรูป 10.17

[5] หากว่าในเมนู Tools ไม่มี Data Analysis แล้ว ให้เพิ่มเข้าไปโดยใช้คำสั่ง Tools, Add-Ins… แล้วเลือก Analysis ToolPak

โดยค่าสถิติต่างๆ สามารถนำไปทดสอบสมมุติฐานที่เกี่ยวข้องได้ เช่นเดียวกับการวิเคราะห์ถดถอยเชิงเส้นตรงด้วย SPSS



รูปที่ 10.33



รูปที่ 10.34

7. การใช้โปรแกรม Microsoft® Excel ในการวิเคราะห์ความสัมพันธ์เชิงเส้นตรง

โปรแกรม Microsoft® Excel มีโมดุล Correlation ที่สามารถนำมาใช้สร้างสมการถดถอยและทดสอบทางสถิติด้วย โดยเป็นโมดุลหนึ่งใน Data Analysis add-ins

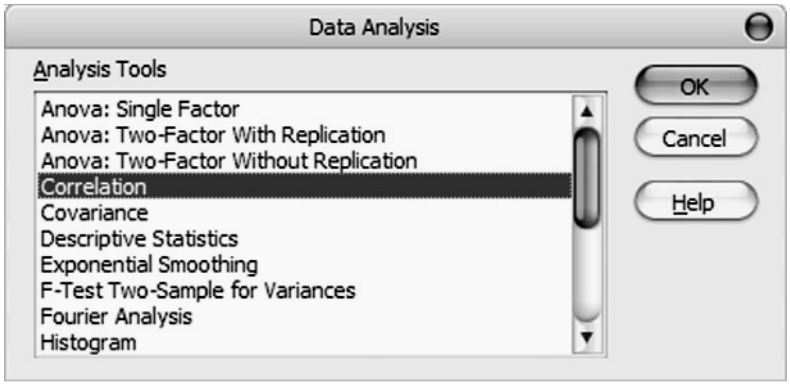
ในการเตรียมข้อมูล จะป้อมข้อมูลแยกเป็นคอลัมน์ๆ ละ 1 ตัวแปร (หรือเป็นแถวก็ได้) และขั้นตอนการวิเคราะห์มีดังนี้

1. ใช้คำสั่ง Tools, Data Analysis...
2. จากรายการตัวเลือกในช่อง Analysis Tools (รูปที่ 10.35) ให้เลือก Correlation
3. กดปุ่ม OK
4. จาก dialog box ที่ปรากฏขึ้นมาใหม่ดังรูปที่ 10.36
    a. ที่ช่อง Input
        i. กดที่ช่อง Input Range: แล้วเลือกช่วงข้อมูลที่เป็น 2 ตัวแปร (หรือมากกว่า)

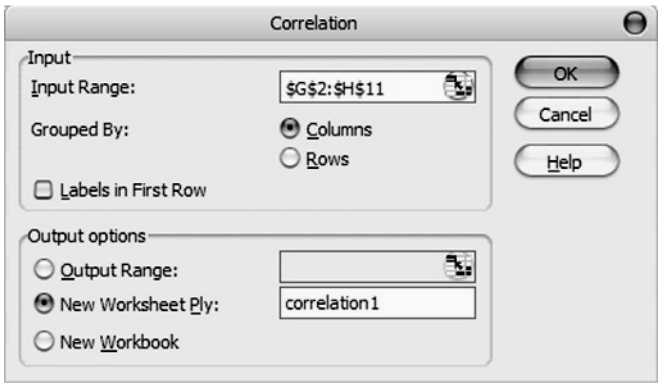
ii. ที่ช่อง **Grouped by:** กดช่อง **Columns** ถ้าแต่ละตัวแปรในช่วงข้อมูลแยกกันตามคอลัมน์ หรือกดช่อง **Rows** ถ้าแต่ละตัวแปรในช่วงข้อมูลแยกกันตามแถว

b. ที่ช่อง **Output options** เลือกที่ช่อง **New Worksheet Ply:** แล้วตั้งชื่อ sheet (หรืออาจเลือกตัวเลือกอื่น)

c. เมื่อกำหนดค่าต่างๆ แล้ว กดปุ่ม **OK**



รูปที่ 10.35



รูปที่ 10.36

	Column 1	*Column 2*
Column 1	1	
Column 2	0.257612	1

รูปที่ 10.37

ผลการวิเคราะห์ได้ผลดังตารางในรูปที่ 10.37 ซึ่งมีเพียงค่าสัมประสิทธิ์สหสัมพันธ์ (r) เท่านั้น [สังเกตว่าค่า r จะแสดงซีกล่างของแนวทแยงมุม] ไม่มีค่าสถิติอื่นๆ ที่จะนำไปทดสอบสมมุติฐานเกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์นี้เลย ดังนั้น จึงต้องใช้ตารางค่าวิกฤติของ r หรือแปลงค่า r เป็นค่าสถิติอื่นแทน เพื่อที่จะสามารถทดสอบสมมุติฐานที่เกี่ยวข้องได้

แบบฝึกหัดปฏิบัติการที่ 10

1. ผู้วิจัยคนหนึ่งมีคำถามว่าพัฒนาการของทารกอายุ 30 สัปดาห์ทั้งสองแบบ มีความสัมพันธ์กันหรือไม่อย่างไร 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. ถ้าสามารถใช้สมการถดถอยเชิงตรงนี้ได้ 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

ค. ถ้าผู้วัดค่าน้ำตาลด้วย stick ได้ค่า 4 mmol/L แล้ว ค่าที่จะวัดในห้องปฏิบัติการจะเป็นเท่าใด

ข. จะอธิบายเกี่ยวกับ coefficient of determination ว่าอย่างไร

ปฏิบัติการที่ 11 การวิเคราะห์ทางสถิติแบบนอนพาราเมตริกซ์

ในปฏิบัติการนี้ เป็นการฝึกวิเคราะห์ข้อมูลที่ไม่สามารถใช้วิธีการวิเคราะห์ทางสถิติแบบพาราเมตริกซ์ได้ โดยในปฏิบัติการนี้จะกล่าวถึง (1) χ^2 Test for Single Variable, Test of Homogeneity & Test of Independence สำหรับการทดสอบข้อมูลที่เป็น nominal scale (2) Mann-Whitney U test สำหรับการทดสอบข้อมูลที่เป็น independent 2-sample test (3) Wilcoxon Signed-Rank test สำหรับการทดสอบข้อมูลที่เป็น dependent 2-sample test (4) Kruskal-Wallis test สำหรับการทดสอบข้อมูลที่มีตั้งแต่ 3 กลุ่มขึ้นไป และ (5) Spearman's rank order correlation coefficient สำหรับการทดสอบความสัมพันธ์ระหว่างตัวแปร

1. χ^2 Test for Single variable, Test of Homogeneity & Test of Independence

การทดสอบทั้งสองนี้จะใช้ในการวิเคราะห์ข้อมูลที่มีมาตราการวัดเป็นแบบ nominal scale ซึ่งในตัวแปรหนึ่ง ๆ จะแบ่งย่อยเป็นหลายหน่วยย่อยหรือหลาย category เช่น ฟีโนไทป์ของลักษณะที่มีการควบคุมด้วยยีนเพียง 1 ยีน ในรุ่น F_2 จะแบ่งเป็น 2 category เป็นต้น สำหรับการทดสอบสมมุติฐานที่มี 1 ตัวแปรนั้น จะใช้ *χ^2 Test for Single Variable* จะทดสอบสมมุติฐาน

Ho: สัดส่วนของค่าสังเกตใน category เป็นไปตามสัดส่วนทางทฤษฎี

H1: สัดส่วนของค่าสังเกตใน category ไม่เป็นไปตามสัดส่วนทางทฤษฎี

ส่วนการทดสอบสมมุติฐานที่มี 2 ตัวแปรนั้น จะแบ่งได้เป็น 2 การทดสอบ คือ

Test of Homogeneity จะกำหนดจำนวนของกลุ่มตัวอย่างของตัวแปรใดตัวแปรหนึ่งก่อน โดยจะระบุด้วยว่าแต่ละ category จะมีจำนวนเท่าใด การทดสอบสมมุติฐานจึงเป็น

Ho: สัดส่วนของค่าสังเกตใน "เซลล์" ไม่มีความแตกต่างกัน[1]

H1: สัดส่วนของค่าสังเกตใน "เซลล์" มีความแตกต่างกัน

Test of Independence จะทดสอบสมมุติฐาน

Ho: ตัวแปรเป็นอิสระต่อกัน

H1: ตัวแปรไม่เป็นอิสระต่อกัน

การทดสอบทั้งสองจะใช้ค่าสถิติ χ^2 ในการทดสอบ โดยจะต้องทราบหรือสามารถหา expected value มาใช้ในการทดสอบ มิฉะนั้น จะไม่สามารถทดสอบทางสถิติใด ๆ ได้

ในปฏิบัติการนี้ จะฝึกใช้ฟังก์ชั่นในโปรแกรม Microsoft® Excel และโปรแกรม SPSS ในการทดสอบ

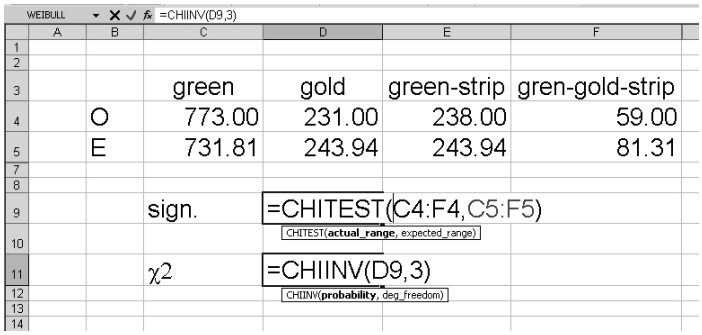
---

[1] "เซลล์" ในที่นี้ หมายถึง category ของตัวแปรที่ศึกษา ถ้าหากว่าตัวแปร A มี 2 category คือ a_1 และ a_2 กับอีกตัวแปร B มี 2 category คือ b_1 และ b_2 และกำหนดจำนวนกลุ่มตัวอย่างของ b_1 และ b_2 ในกรณีนี้ อาจตั้ง Ho: ใน category b_1 สัดส่วนของค่าสังเกตใน $a_1 = a_2$ ถ้า Ho นี้เป็นจริงแล้ว อีกสมมุติฐานหนึ่ง คือ Ho: ใน category b_2 สัดส่วนของค่าสังเกตใน $a_1 = a_2$ ก็จะเป็นจริงด้วย

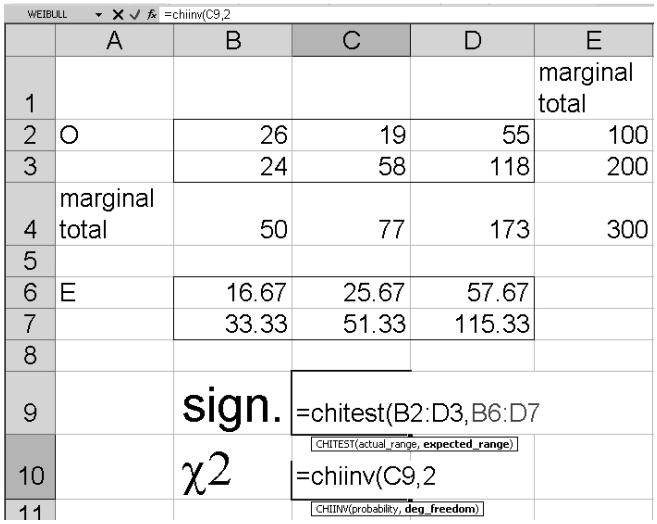
1.1 ฟังก์ชันในโปรแกรม Microsoft® Excel

โดยทั่วไป มักจะนับค่าสังเกตของแต่ละ category ของตัวแปรหนึ่ง ๆ แล้วบันทึกลงในโปรแกรม Microsoft® Excel และทำเช่นเดียวกันกับกรณีที่มีตัวแปรมากกว่า 1 ตัวแปร เพียงแต่ต้องพิจารณาค่าสังเกตที่นับนั้นอยู่ใน category ใดของตัวแปรแรก และอยู่ใน category ใดของอีกตัวแปรหนึ่ง

ค่าสังเกตในแต่ละ category นี้ จะเรียกว่า observed frequency (O) ของแต่ละ category ส่วนค่าคาดคะเนในแต่ละ category จะเรียกว่า expected frequency (E) และเพื่อให้เข้าใจง่าย จึงมักป้อนค่าคาดคะเนให้อยู่ด้านล่างของค่าสังเกต (หรือจะไว้ด้านขวาของค่าสังเกตก็ได้) รูปที่ 11.1 และ 11.2 กรณีที่การทดสอบทางสถิติที่มีตัวแปรเพียง 1 ตัวแปร และ 2 ตัวแปรตามลำดับ



รูปที่ 11.1 การป้อนค่าสังเกตและค่าคาดคะเนในกรณีที่การทดสอบมีเพียง 1 ตัวแปร แบ่งเป็น 4 category



รูปที่ 11.2 การป้อนค่าสังเกตและค่าคาดคะเนในกรณีที่การทดสอบมีตัวแปร 2 ตัวแปร โดยตัวแปรหนึ่งแบ่งเป็น 3 category และอีกตัวแปรหนึ่งแบ่งเป็น 2 category (3×2 Contingency table)

การหาค่าคาดคะเนนี้ อาจจะหาได้จากสัดส่วนทางทฤษฎีในแต่ละ category เช่น สัดส่วนของฟีโนไทป์ของลักษณะที่ถูกควบคุมด้วยยีน 1 ยีน และเป็น complete dominance จะเป็นอัตราส่วน 3:1 (ฟีโนไทป์เด่น : ฟีโนไทป์ด้อย) ถ้าในกรณีที่มีค่าสังเกตทั้งหมด 120 ค่า ค่าคาดคะเนของฟีโนไทป์เด่น เป็น ¾ ×120 เท่ากับ 90 ส่วนค่าคาดคะเนของฟีโนไทป์ด้อย เป็น ¼ ×120 เท่ากับ 30

หากว่าไม่สามารถทราบได้ว่า สัดส่วนทางทฤษฎีในแต่ละ category เป็นเท่าใด เช่น กรณีที่เป็นการทดสอบ Test of Independence ก็อาจสามารถหาได้โดยใช้ row marginal และ column marginal มาคำนวณหาได้ ในรูปที่ 11.2 ค่า row marginal คือ ค่าในช่อง E2 และ E3 ส่วนค่า column

marginal คือ ค่าในช่อง B4 C4 และ D4 ถ้าตัวแปรหนึ่งมี *j* category และอีกตัวแปรหนึ่งมี *i* category ให้ $r_{\cdot j}$ และ $c_{i\cdot}$ เป็น row marginal และ column marginal ตามลำดับ ค่าคาดคะเน E_{ij} หาได้จาก $E_{ij} = \frac{r_{\cdot j} \times c_{i\cdot}}{n}$ เมื่อ n เป็นจำนวนค่าสังเกตทั้งหมด

เมื่อหาค่าคาดคะเนได้แล้ว จะใช้ฟังก์ชั่น 2 ฟังก์ชั่น ในการหาค่าเพื่อทดสอบสมมุติฐาน คือ

ฟังก์ชั่น	ค่าที่หา	ค่าพารามิเตอร์
chitest(*observed* , *expect*)	significance	Observed : ช่วงข้อมูลค่าสังเกต Expect : ช่วงข้อมูลค่าคาดคะเน
chiinv(*sign.* , *df*)	χ^2	Sign. : ค่านัยสำคัญ df : ค่า degree of freedom ถ้ามี 1 ตัวแปร df = i-1 ถ้ามี 2 ตัวแปร df = (i-1)(j-1)

ในกรณีที่ต้องทดสอบว่าตัวแปร 2 ตัวแปร เป็นอิสระต่อกันหรือไม่นั้น หากว่ามีจำนวนค่าสังเกตและหรือจำนวนค่าคาดคะเนน้อยกว่า 5 ค่า ในแต่ละ “เซลล์” ของ contingency table แล้ว ไม่ควรใช้ χ^2 test ในการทดสอบ แต่ควรใช้ Fisher Exact test แทน ในโปรแกรม Microsoft Excel ไม่มีฟังก์ชั่นในการคำนวณ แต่ว่าสามารถใช้ Add-Ins เพิ่มคำนวณได้ โดยสามารถเอามาจาก http://www.obertfamily.com/software/fisherexact.xla แล้วจึงใส่ไว้ใน Add-Ins ของ Excel ดังนี้

สำหรับ Excel 2003 เลือก Tools, Add-Ins จะมีเมนูให้เลือก ดูว่ามีตัวเลือก Fisherexact หรือไม่ ถ้ามีให้กดที่ช่อง ☐ หน้าตัวเลือกนี้ หากว่าไม่พบตัวเลือกนี้ ให้กดปุ่ม Browse แล้วหาไฟล์ที่ชื่อว่า fisherexact.xla แล้วกด OK แล้วกดปุ่ม OK เพื่อปิดเมนูนี้

สำหรับ Excel 2007 เลือก Office Button, Excel Options, Add-Ins, ที่ตัวเลือก Manage: เลือก Excel Add-Ins แล้วกด Go... จะมีเมนูให้เลือก ดูว่ามีตัวเลือก Fisherexact หรือไม่ ถ้ามีให้กดที่ช่อง ☐ หน้าตัวเลือกนี้ หากว่าไม่พบตัวเลือกนี้ ให้กดปุ่ม Browse แล้วหาไฟล์ที่ชื่อว่า fisherexact.xla แล้วกด OK แล้วกดปุ่ม OK เพื่อปิดเมนูนี้

สำหรับการหาค่าความน่าจะเป็นในการทดสอบด้วย Fisher Exact test ด้วย Add-Ins นี้ จะใช้ฟังก์ชั่นดังนี้

ฟังก์ชั่น	คำอธิบาย
FET(g1+,g1-,g2+,g2-,[true or false])	• g1 และ g2 คือ ตัวแปรที่ 1 และตัวแปรที่ 2 โดยแต่ละตัวแปรมีเพียง 2 category (+ และ -) ให้ระบุจำนวนค่าสังเกตในแต่ละ category

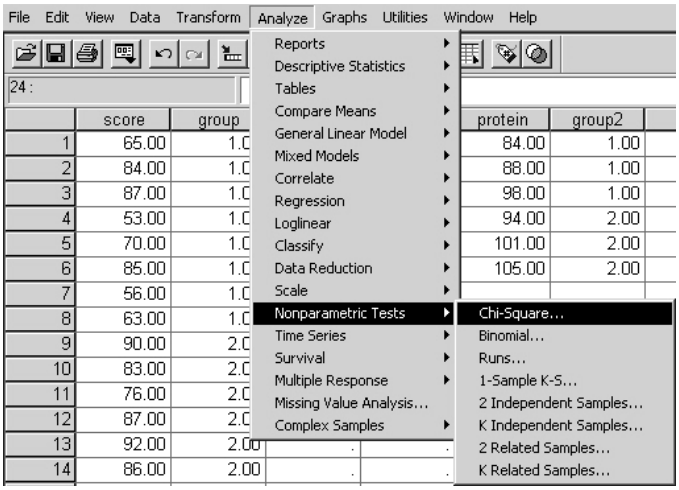
	• ค่าความน่าจะเป็นแบบ 1-tail เมื่อระบุ true และ แบบ 2-tail เมื่อระบุ false
FETL(g1+,g1-,g2+,g2-)	หาค่าความน่าจะเป็นแบบ 1-tail ด้านซ้าย
FETR(g1+,g1-,g2+,g2-)	หาค่าความน่าจะเป็นแบบ 1-tail ด้านขวา

ฟังก์ชั่นนี้ ใช้ได้กับเฉพาะเมื่อแต่ละตัวแปรมีเพียง 2 category เท่านั้น และหากว่าได้คำตอบจากการวิเคราะห์เป็น “NP” แสดงว่าไม่สามารถใช้ฟังก์ชั่นนี้ในการวิเคราะห์หาค่าความน่าจะเป็นจากการทดสอบ Fisher Exact text ได้

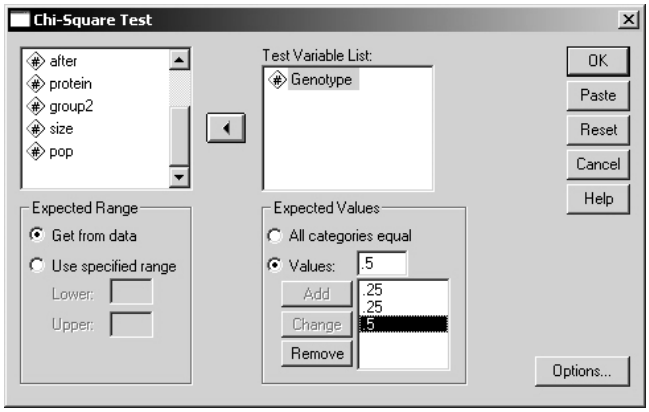
1.2 โปรแกรม SPSS

ในการทดสอบแบบ Goodness of Fit test นั้น หากมีข้อมูลดิบ กล่าวคือ มีค่าสังเกตที่ได้รับการกำหนด category แล้วเท่านั้น การป้อนข้อมูลค่าสังเกต จะใช้เพียง 1 ตัวแปร โดยตัวแปรนี้จะประกอบ *i* category ทั้งอาจให้ 1, 2, 3 … *i* แทนค่าสังเกตใน category ที่ 1, 2, 3 … *i* ตามลำดับ การวิเคราะห์ข้อมูลมีขั้นตอนดังนี้

1. ใช้คำสั่ง **Analyses, Nonparametric tests, Chi-Square**... (รูปที่ 11.3)
2. จาก dialog box ในรูปที่ 11.4
    a. เลือกตัวแปรจากช่องด้านซ้าย แล้วกด ▸ ของช่อง **Test Variables**
    b. ที่ช่อง **Expected Range**: เลือก **Get from data**
    c. ที่ช่อง **Expected Values** กำหนดค่าคาดคะเนโดย
        i. หากว่าสัดส่วนทางทฤษฎีของแต่ละ category เท่ากัน ให้เลือก **All categories equal**
        ii. แต่หากว่าทางทฤษฎีของแต่ละ category ไม่เท่ากัน
            1. ให้เลือก **Value**
            2. ป้อนค่าสัดส่วนของแต่ละ category โดยค่าแรกเป็นของ category แรก แล้วกดปุ่ม **Add** จากนั้นค่าสัดส่วนค่าต่อ ๆ ไป ทั้งนี้ ผลรวมจะต้องมีค่าเป็น 1
        iii. กดปุ่ม **OK**



รูปที่ 11.3



รูปที่ 11.4

ผลการวิเคราะห์จะเป็นตารางดังแสดงในรูปที่ 11.5 และ 11.6 โดยในรูปที่ 11.5 เป็นตารางแสดงจำนวนค่าสังเกต ค่าคาดคะเน และผลต่างระหว่างค่าทั้งสองในแต่ละ category ส่วนตารางในรูปที่ 11.6 เป็นค่าสถิติที่จะนำไปใช้ทดสอบ จากตัวอย่างนี้ ค่า Sig. (=Asymp. Sig.) มีค่าน้อยกว่า 0.05 จึงปฏิเสธสมมุติฐาน Ho ที่ตั้งไว้ว่า Ho: สัดส่วน Genotype 1: Genotype 2: Genotype 3 = 1:1:2

genotype

	Observed N	Expected N	Residual
1.0	7	6.8	.3
2.00	14	6.8	7.3
3.00	6	13.5	-7.5
Total	27		

รูปที่ 11.5

Test Statistics

	genotype
Chi-Square(a)	11.963
f	
Asymp. Sig.	.003

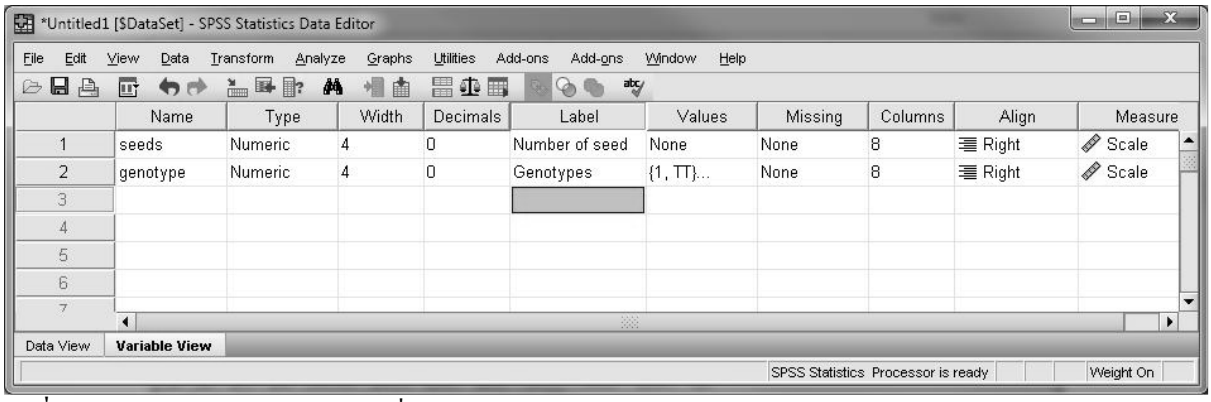
a 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 6.8.

รูปที่ 11.6

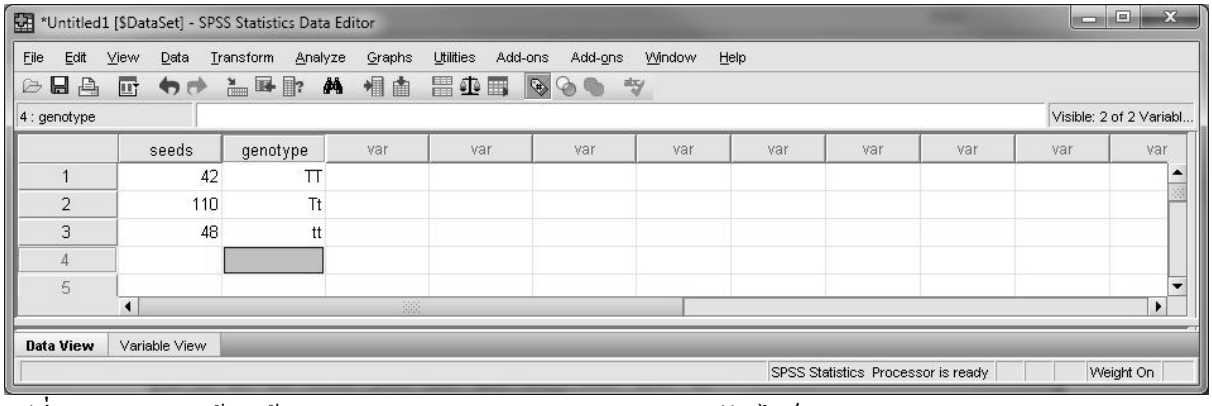
ในกรณีที่ข้อมูลมีการนับจำนวนค่าสังเกตในแต่ละ categories แล้ว การป้อนข้อมูลและขั้นตอนวิเคราะห์ทางสถิติด้วย Chi Squared test ใน SPSS มีขั้นตอนที่แตกต่างไปจากกรณีที่ข้อมูลไม่ได้ถูกนับ (คือเป็นข้อมูลของแต่ละค่าสังเกต) ดังนี้

1. Goodness of Fit test

การป้อนข้อมูลจะใช้ตัวแปร 2 ตัวแปร โดยตัวแปรหนึ่งเป็น nominal variable และอีกตัวแปรหนึ่งเป็นจำนวนค่าสังเกตในแต่ละ categories ของ nominal variable เช่น การทดสอบว่าสัดส่วนของจีโนไทป์ของเมล็ด 3 แบบ (TT, Tt, tt) มีค่าเป็น 1:2:1 หรือไม่ ในการกำหนดตัวแปรจะเป็นดังรูปที่ 11.6-ผ1 โดย genotype เป็น nominal variable ซึ่งมี 3 categories และ seed เป็นตัวแปรเก็บค่าจำนวนเมล็ดในแต่ละจีโนไทป์ ตัวแปรทั้งสองเป็น scale variable ส่วนรูปที่ 11.6-ผ2 เป็นตัวอย่างการป้อนค่า



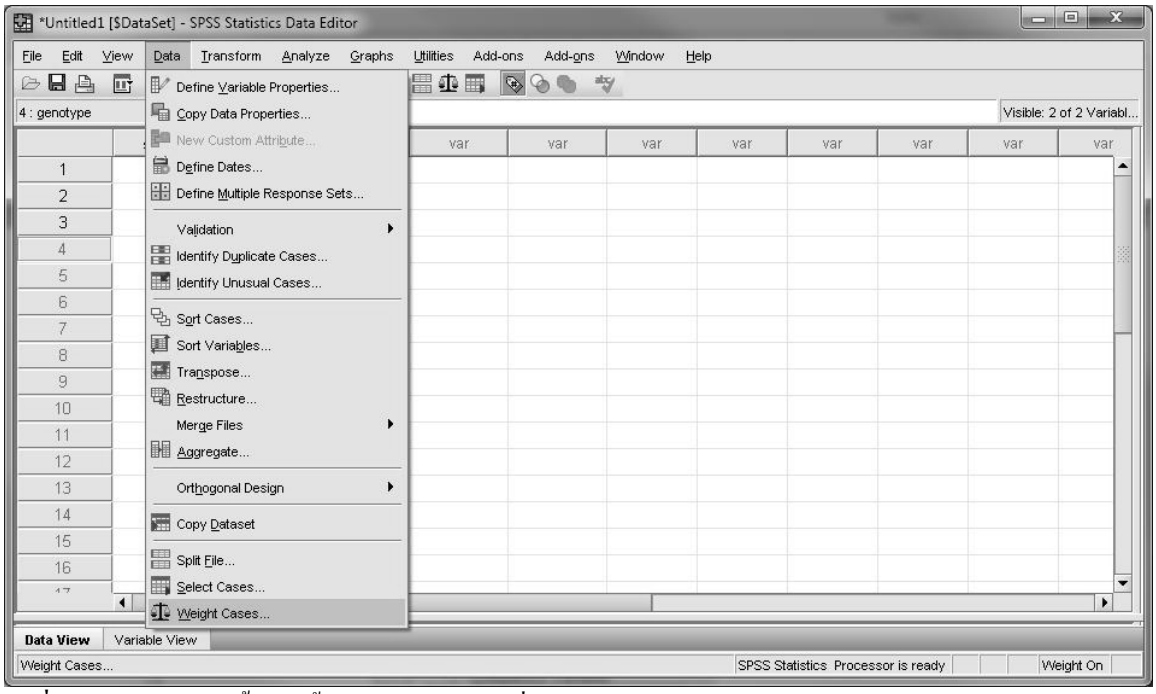
รูปที่ 11.6-ผ1 การกำหนดตัวแปรเพื่อวิเคราะห์ข้อมูลด้วย Chi-square test สำหรับ Goodness of Fit test



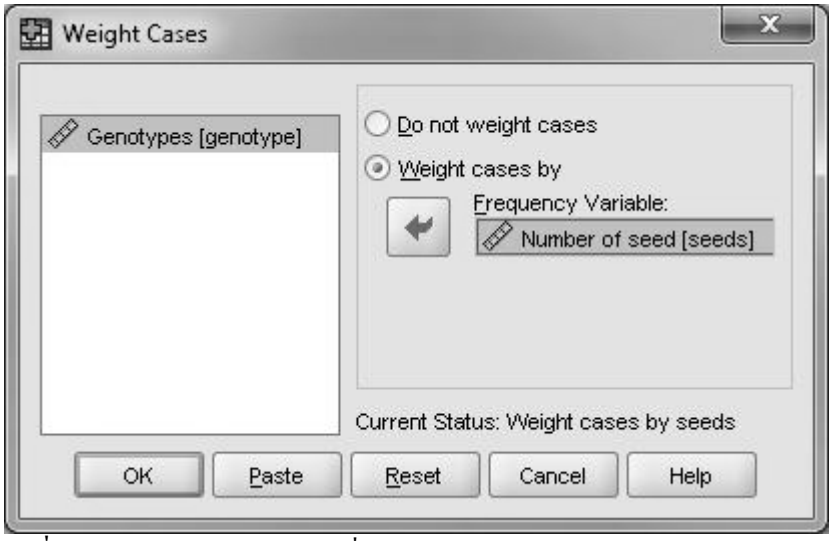
รูปที่ 11.6-ผ2 การป้อนข้อมูล สังเกตว่าตัวแปร genotype แสดงค่าของจีโนไทป์
(โดยเลือก View > Value Labels)

การวิเคราะห์ มีขั้นตอนดังนี้

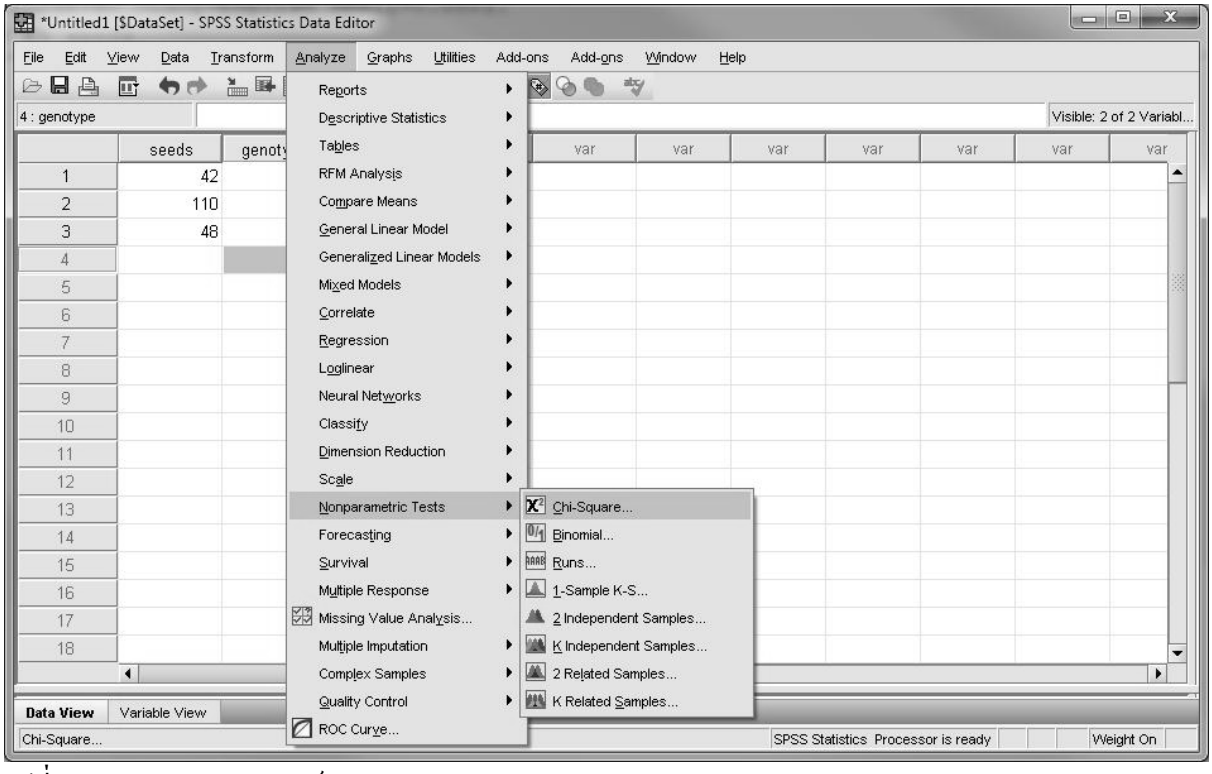
1. เลือก Data > Weight Cases... (รูปที่ 11.6-ผ3)
2. ในหน้าต่างกำหนดค่า (รูปที่ 11.6-ผ4) เลือกตัวแปรที่เก็บจำนวนค่าสังเกตในช่องซ้าย (ในตัวอย่าง คือ seeds) แล้วเลือก ● Weight cases by แล้วจึงกดปุ่ม ➥ สังเกตว่าช่องใต้ข้อความ Frequency variable: จะมีตัวแปรปรากฏขึ้น แล้วจึงกดปุ่ม OK
3. เลือก Statistics > Nonparametric Tests > Chi-Square... (รูปที่ 11.6-ผ5)
4. ในหน้าต่างกำหนดค่า (รูปที่ 11.6-ผ6) กำหนดค่าดังนี้
   a. เลือกตัวแปรที่เป็น nominal variable ในช่องซ้าย แล้วกดปุ่ม ➥ ให้มาอยู่ในช่อง Test Variable List:
   b. ในกล่อง Expected Values: ให้เลือก Values: แล้วป้อนค่าความถี่ของแต่ละ category ของ nominal variable โดยเรียงลำดับตามที่ป้อนค่าไว้ (ในตัวอย่าง คือ 0.25TT 0.50Tt 0.25tt)
   c. กดปุ่ม OK
5. ผลการวิเคราะห์เป็นดังรูปที่ 11.6-ผ7 ซึ่งจะประกอบด้วยตารางจำนวนค่าสังเกตและค่าคาดคะเนในแต่ละ categories ของ nominal variable และตารางแสดงค่า Chi-Square ค่า df และค่า Sig.



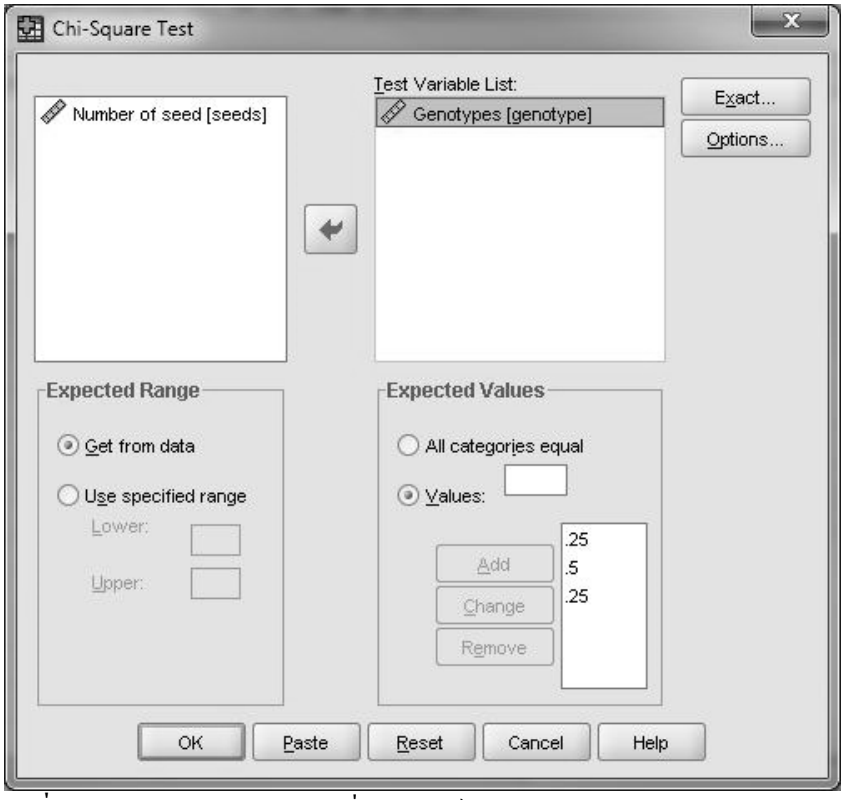
รูปที่ 11.6-ผ3 การตั้งค่า "น้ำหนัก" ของตัวแปรที่เก็บค่าจำนวนค่าสังเกต



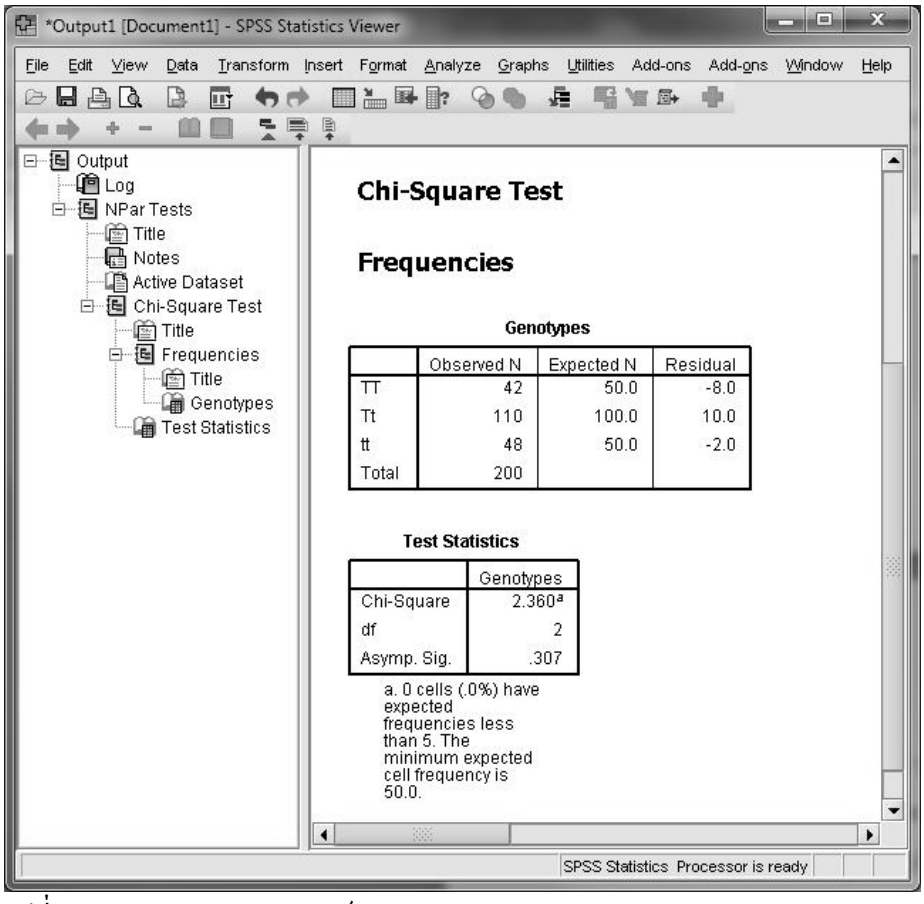
รูปที่ 11.6-ผ4 กำหนดตัวแปรที่ใช้สำหรับ weight



รูปที่ 11.6-ผ5 การวิเคราะห์ Chi Square



รูปที่ 11.6-ผ6 การกำหนดค่าเพื่อวิเคราะห์ Chi Square test

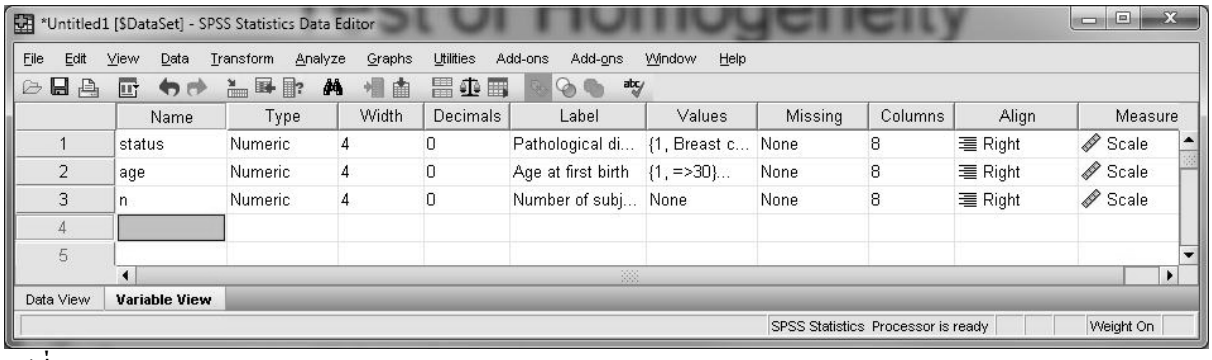


รูปที่ 11.6-ผ7 ผลการวิเคราะห์ Chi Square

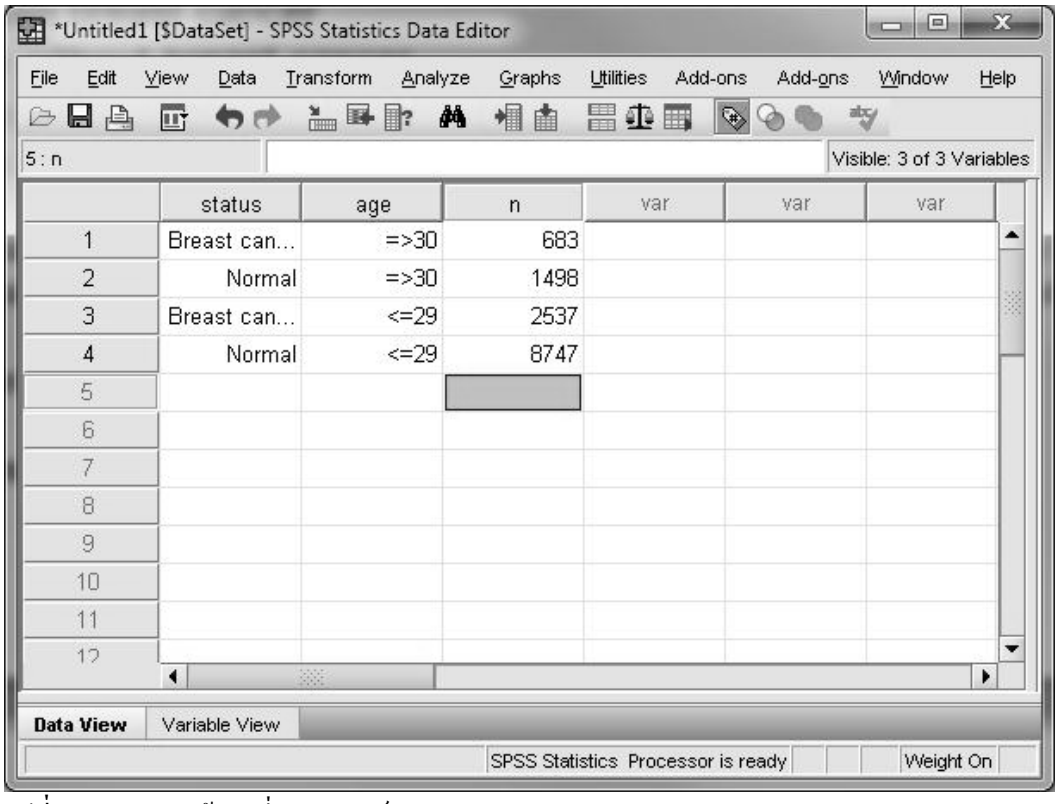
2. การวิเคราะห์ Contingency Table ที่มีเพียง 2 ตัวแปรเท่านั้น

การป้อนข้อมูลและขั้นตอนสำหรับการการวิเคราะห์ด้วย Chi Square test และ Fisher's exact test จะเหมือนกัน เพียงเลือกค่าสถิติที่ถูกต้องจากตารางผลการวิเคราะห์เท่านั้น โดย Chi Squared test จะใช้ค่าสถิติจากบรรทัด Pearson Chi Square ส่วน Fisher's exact test นั้น จะใช้ค่าสถิติในบรรทัด Fisher's exact test (ดูรูปที่ 11.6-ผ15)

ในกรณีที่นับจำนวนค่าสังเกตในแต่ละ categories แล้วนั้น จะต้องกำหนดตัวแปร 3 ตัวแปร คือ nominal variable ตัวที่ 1 nominal variable ตัวที่ 2 และตัวแปรเก็บจำนวนค่าสังเกตในแต่ละ "เซลล์" รูปที่ 11.6-ผ8 เป็นการกำหนดตัวแปร และรูปที่ 11.6-ผ9 เป็นการป้อนค่า



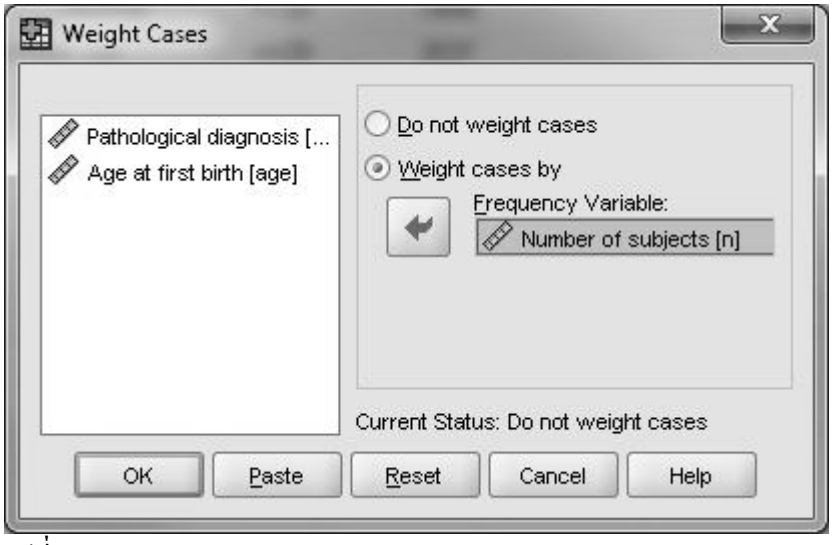
รูปที่ 11.6-ผ8 การกำหนดตัวแปร



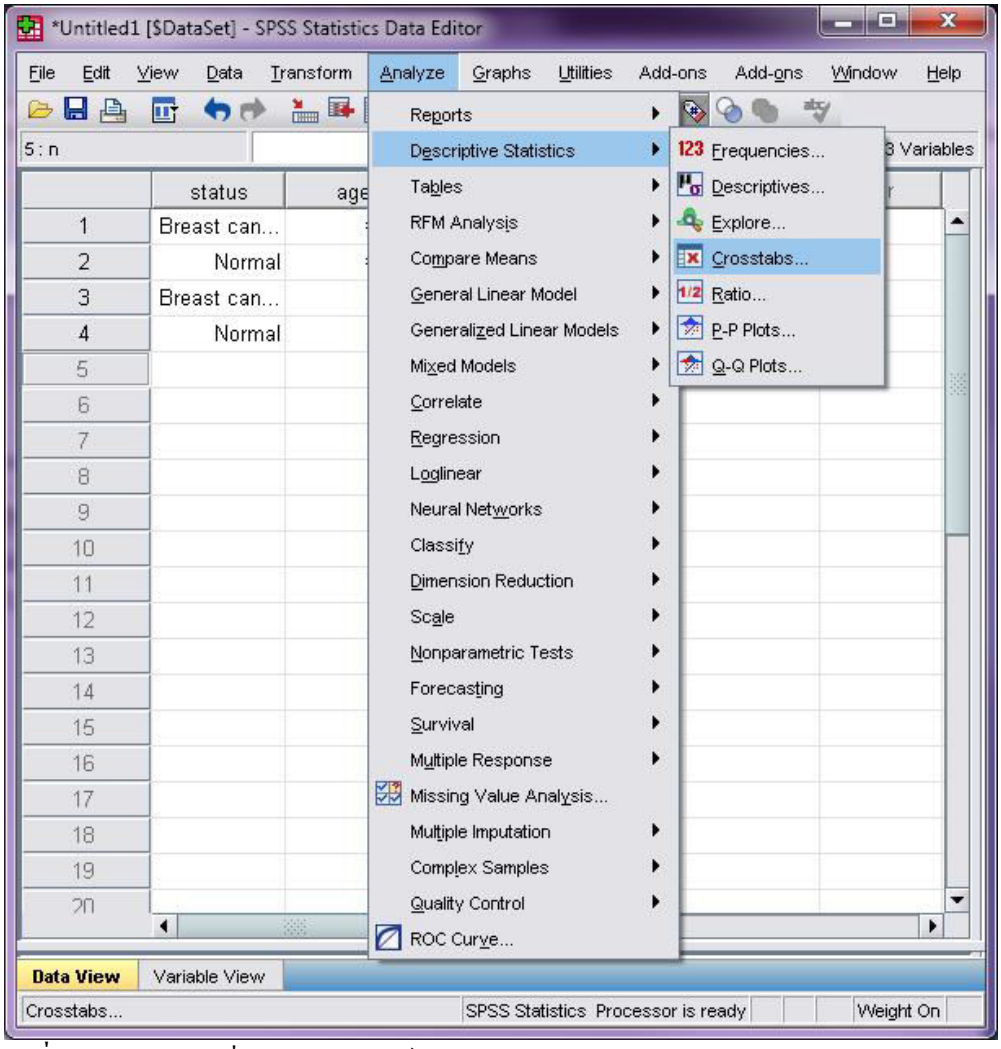
รูปที่ 11.6-ผ9 ข้อมูลที่จะวิเคราะห์

การวิเคราะห์ มีขั้นตอนดังนี้

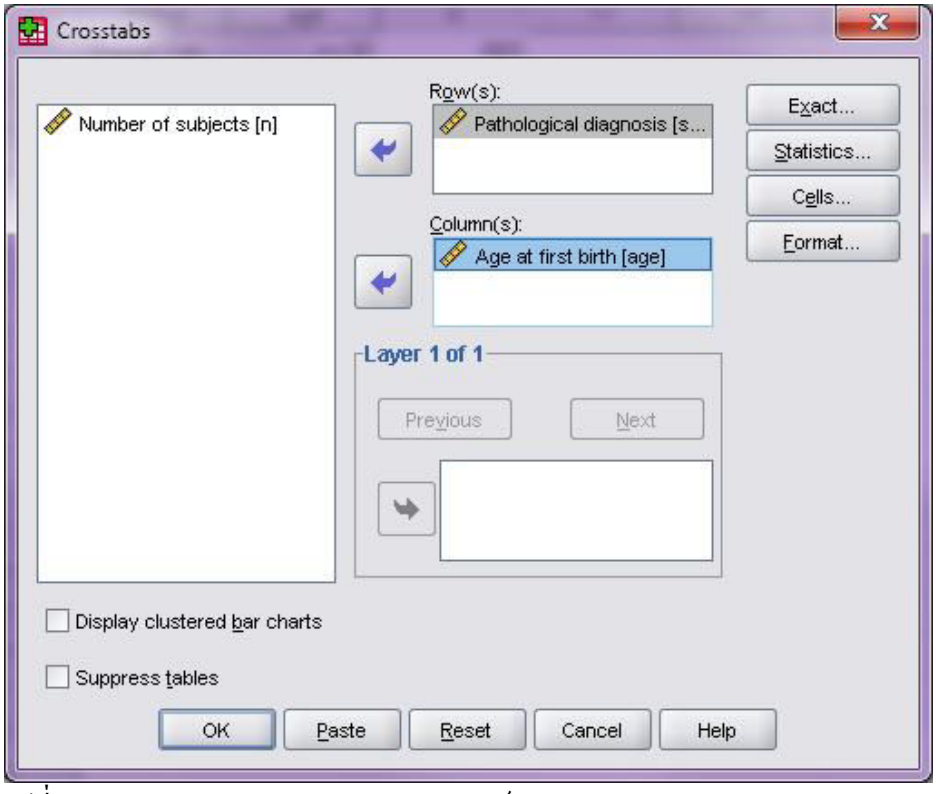
1. เลือก Data > Weight Cases...
2. ในหน้าต่างกำหนดค่า (รูปที่ 11.6-ผ10) เลือกตัวแปรที่เก็บจำนวนค่าสังเกตในช่องซ้าย (ในตัวอย่าง คือ n) แล้วเลือก ⦿ Weight cases by แล้วจึงกดปุ่ม ➥ สังเกตว่าช่องใต้ข้อความ Frequency variable: จะมีตัวแปรปรากฏขึ้น แล้วจึงกดปุ่ม OK
3. เลือก Statistics > Descriptive Statistics > Crosstabs... (รูปที่ 11.6-ผ11)
4. ในหน้าต่างกำหนดค่า (รูปที่ 11.6-ผ12) กำหนดค่าดังนี้
    a. เลือกตัวแปรที่เป็น nominal variable ในช่องซ้ายที่ต้องให้อยู่ในแนวนอน แล้วกดปุ่ม ➥ ให้มาอยู่ในช่อง Row(s):
    b. เลือกตัวแปรที่เป็น nominal variable ในช่องซ้ายที่ต้องให้อยู่ในแนวตั้ง แล้วกดปุ่ม ➥ ให้มาอยู่ในช่อง Column(s):
    c. กดปุ่ม Statistics... แล้วเลือก Chi-Square (รูปที่ 11.6-ผ13) แล้วกดปุ่ม Continue
    d. กดปุ่ม Cell... ที่กล่อง Counts ให้เลือกที่ Observed และ Expected (รูปที่ 11.6-ผ14) แล้วกดปุ่ม Continue
    e. กดปุ่ม OK
5. ผลการวิเคราะห์เป็นดังรูปที่ 11.6-ผ15 ซึ่งจะประกอบด้วยตารางจำนวนค่าสังเกตและค่าคาดคะเนในแต่ละ categories ของ nominal variable ทั้งสอง และตารางแสดงค่า Chi-Square ค่า df และค่า Sig.



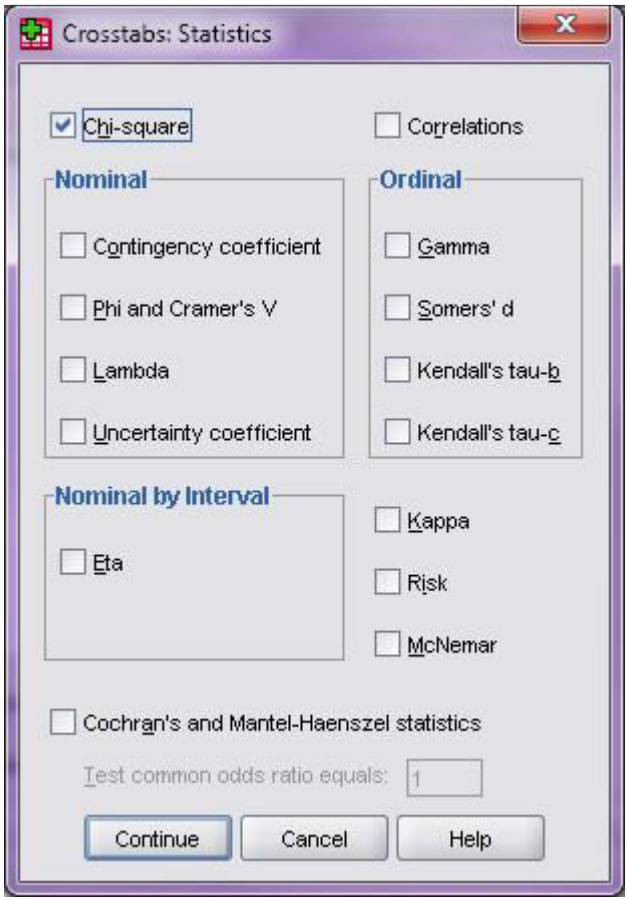
รูปที่ 11.6-ผ10 กำหนดตัวแปรสำหรับ weight



รูปที่ 11.6-ผ11 คำสั่งสำหรับวิเคราะห์ Contingency Table



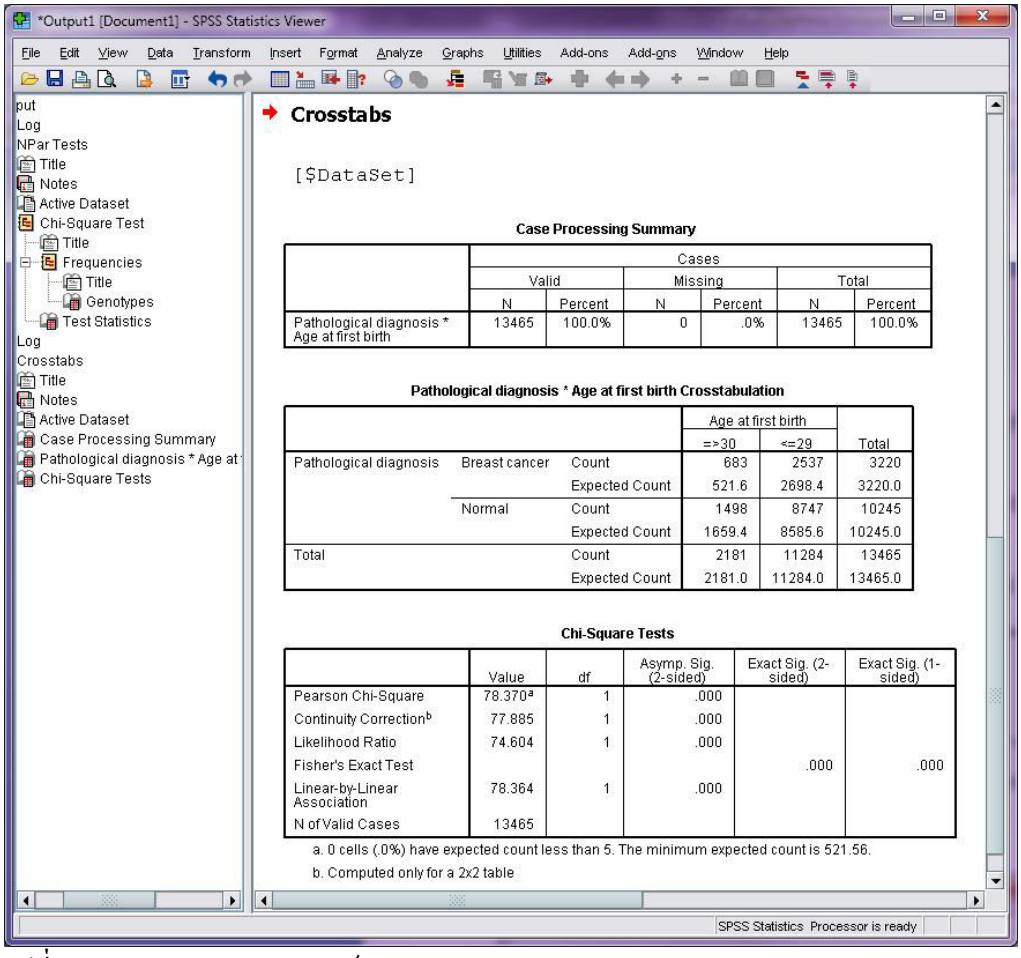
รูปที่ 11.6-ผ12 การกำหนดสำหรับการวิเคราะห์



รูปที่ 11.6-ผ13 กำหนดค่าใน Statistics...



รูปที่ 11.6-ผ14 กำหนดค่าใน Cell...

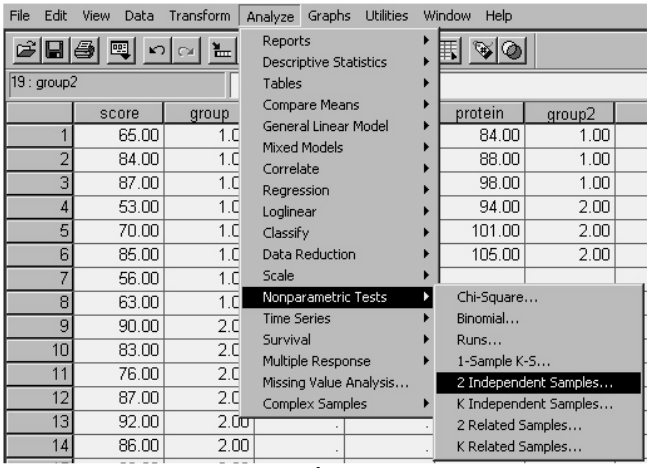


รูปที่ 11.6-ผ15 ผลการวิเคราะห์ Contingency table

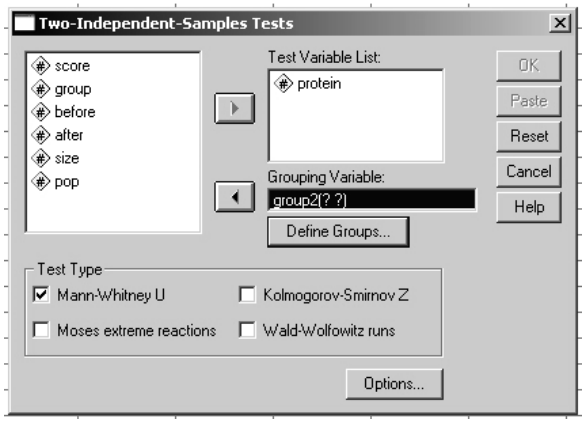
2. Mann-Whitney *U* test

การทดสอบด้วยวิธีนี้ จะใช้แทน Two independent samples *t* test โดยวิธีการป้อนข้อมูลจะเป็นแบบเดียวกัน ขั้นตอนการวิเคราะห์มีดังนี้

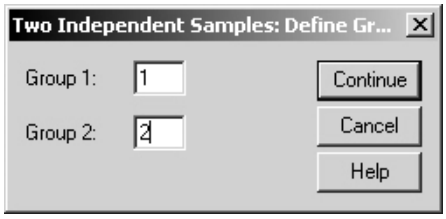
1. เลือก Analyze, Nonparametric, 2 Independent Samples... (รูปที่ 11.7)
2. จาก dialog box ในรูปที่ 11.8
   a. เลือกตัวแปรที่จะทดสอบจากช่องด้านซ้าย แล้วกดปุ่ม ▸ ของช่อง Test Variable List:
   b. เลือกตัวแปรระบุกลุ่มจากช่องด้านซ้าย แล้วกดปุ่ม ▸ ของช่อง Grouping variable: สังเกตว่าจะมี (?,?) หลังชื่อตัวแปรระบุกลุ่ม
   c. กดปุ่ม Define Groups... เพื่อกำหนดค่าที่ใช้ระบุแต่ละกลุ่ม (รูปที่ 11.9) แล้วกดปุ่ม Continue
   d. ที่ช่อง Test Type เลือก Mann-Whitney U
3. กดปุ่ม OK



รูปที่ 11.7



รูปที่ 11.8



รูปที่ 11.9

Ranks

	group2	N	Mean Rank	Sum of Ranks
protein	1.00	3	2.33	7.00
	2.00	3	4.67	14.00
	Total	6		

รูปที่ 11.10

Test Statistics(b)

	protein
Mann-Whitney U	1.000
Wilcoxon W	7.000
Z	-1.528
Asymp. Sig. (2-tailed)	.127
Exact Sig. [2*(1-tailed Sig.)]	.200(a)

a Not corrected for ties.
b Grouping Variable: group2

รูปที่ 11.11

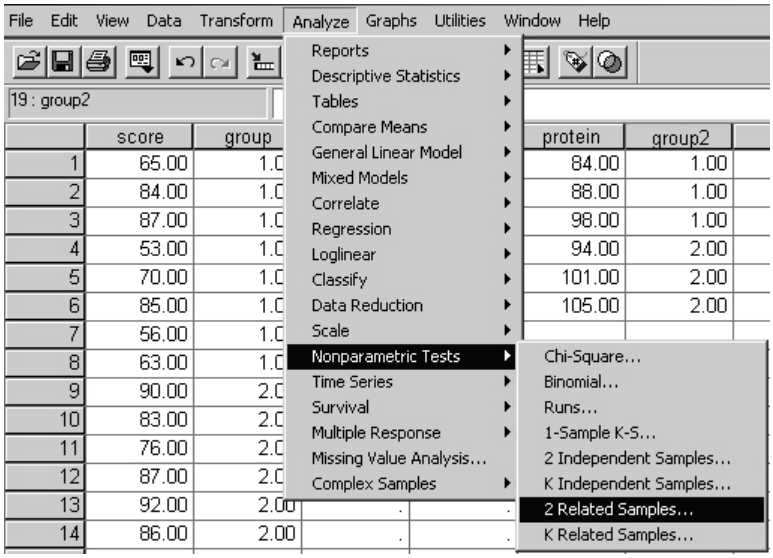
ผลการวิเคราะห์แสดงดังรูปที่ 11.10 และ 11.11 โดยในรูปที่ 11.10 เป็นตารางสรุปค่าสถิติต่าง ๆ ส่วนในรูปที่ 11.11 เป็นตารางที่จะนำไปใช้ในการทดสอบทางสถิติ โดยจะนำค่า Mann-Whitney U และค่า Asymp. Sig. หรือ Exact Sig. ไปทดสอบสมมุติฐาน[2] โดยเปรียบเทียบค่า Sig. กับค่าที่ตั้งไว้ คือ 0.05 นั่นเอง การเลือกจะใช้ Sig. แบบใดนั้น ขึ้นอยู่กับสถานการณ์ของการทดสอบ โดยทั่วไป ทั้งสองค่ามักให้ค่าหรือผลการทดสอบเป็นไปในทิศทางเดียวกัน

3. Wilcoxon Matched-Pairs Signed Rank (WSR) Test

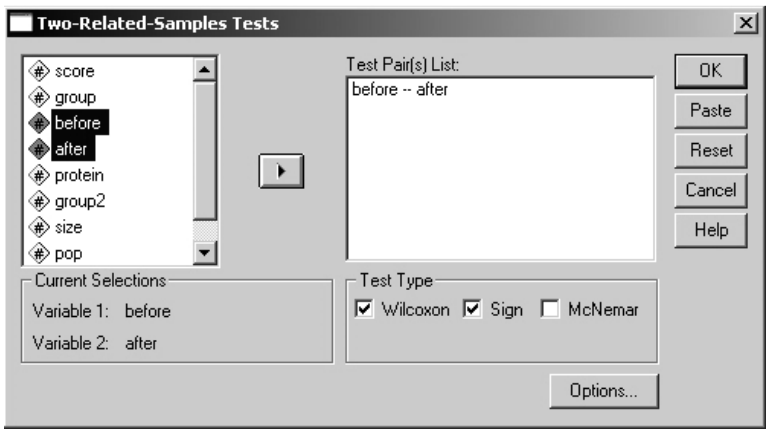
การทดสอบด้วยวิธีนี้ จะใช้แทน Paired samples *t* test โดยวิธีการป้อนข้อมูลจะเป็นแบบเดียวกัน ขั้นตอนการวิเคราะห์มีดังนี้

1. เลือก Analyze, Nonparametric, 2 Related Samples... (รูปที่ 11.12)
2. จาก dialog box ในรูปที่ 11.13
    a. เลือกคู่ตัวแปรที่จะทดสอบจากช่องด้านซ้าย แล้วกดปุ่ม ▸ ของช่อง Test Pair(s) List: สังเกตว่าที่ช่อง Current Selections จะมีตัวแปรหลัง Variable 1 และ Variable 2
    b. ที่ช่อง Test Type เลือก Wilcoxon และ Sign หากต้องการ
3. กดปุ่ม OK

[2] ในการตั้งสมมุติฐาน จะไม่มีการอ้างถึงค่าสถิติที่เป็น parametric statistics แต่อย่างใด ในตัวอย่างนี้ อาจตั้ง Ho: กลุ่มที่แตกต่างกันไม่มีผลต่อระดับโปรตีน vs. H1: กลุ่มที่แตกต่างกันมีผลต่อระดับโปรตีน



รูปที่ 11.12



รูปที่ 11.13

Ranks

		N	Mean Rank	Sum of Ranks
after - before	Negative Ranks	4(a)	6.25	25.00
	Positive Ranks	6(b)	5.00	30.00
	Ties	0(c)		
	Total	10		

a after < before
b after > before
c after = before

รูปที่ 11.14

Test Statistics(b)

	after - before
Z	-.255(a)
Asymp. Sig. (2-tailed)	.798

a Based on negative ranks.
b Wilcoxon Signed Ranks Test

รูปที่ 11.15

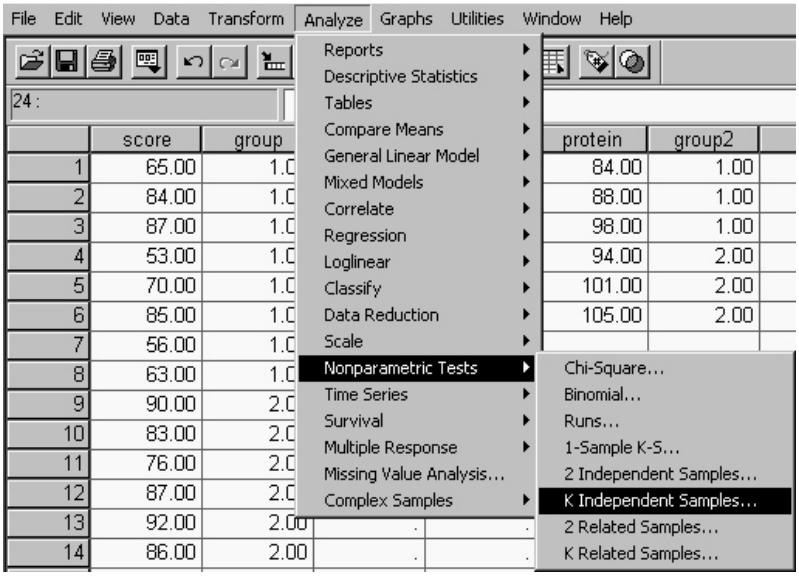
ผลการวิเคราะห์แสดงดังรูปที่ 11.14 และ 11.15 โดยในรูปที่ 11.14 เป็นตารางสรุปค่าสถิติต่าง ๆ ส่วนในรูปที่ 11.15 เป็นตารางที่จะนำไปใช้ในการทดสอบทางสถิติ โดยจะนำค่า Z และค่า Asymp. Sig. ไปทดสอบสมมุติฐาน[3] โดยเปรียบเทียบค่า Sig. กับค่าที่ตั้งไว้ คือ 0.05 นั้นเอง

[3] เช่นเดียวกันกับการตั้งสมมุติฐานของ Mann-Whitney *U* test โดยในตัวอย่างนี้ อาจตั้งได้ดังนี้ Ho: การออกกำลังกายไม่มีผลต่อระดับกรดไขมันในเลือด vs. H1: การออกกำลังกายมีผลต่อระดับกรดไขมันในเลือด

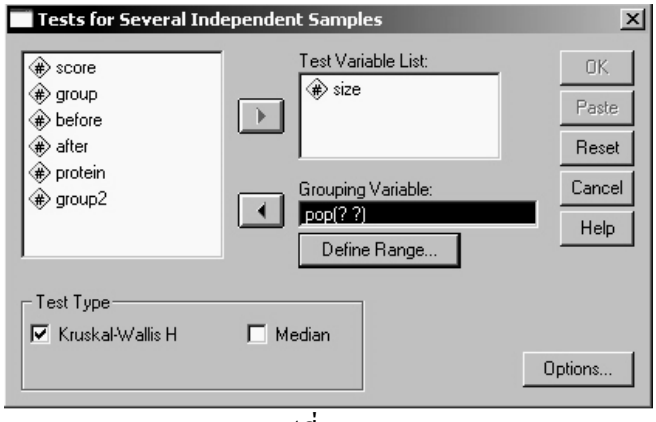
4. Kruskal-Wallis Test

การทดสอบด้วยวิธีนี้ จะใช้แทน One-Way ANOVA หรือกรณีต้องทดสอบกลุ่มตัวอย่างตั้งแต่ 3 กลุ่มขึ้นไป โดยวิธีการป้อนข้อมูลจะเป็นแบบเดียวกับการวิเคราะห์ One-Way ANOVA ขั้นตอนการวิเคราะห์มีดังนี้

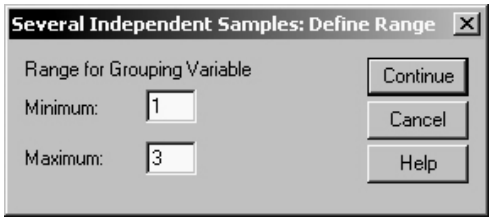
1. เลือก Analyze, Nonparametric, k Independent Samples... (รูปที่ 10.16)
2. จาก dialog box ในรูปที่ 10.17
   a. เลือกตัวแปรที่จะทดสอบจากช่องด้านซ้าย แล้วกดปุ่ม ▸ ของช่อง Test Variable List:
   b. เลือกตัวแปรระบุกลุ่มจากช่องด้านซ้าย แล้วกดปุ่ม ▸ ของช่อง Grouping Variable: สังเกตว่าจะมี (?,?) หลังชื่อตัวแปรระบุกลุ่ม
   c. กดปุ่ม Define Range... เพื่อกำหนดค่าที่ใช้ในการระบุแต่ละกลุ่ม (รูปที่ 10.18) ทั้งนี้ให้ป้อนค่าต่ำสุดและค่าสูงสุดในช่อง Minimum และ Maximum ตามลำดับ แล้วกดปุ่ม Continue
   d. ที่ช่อง Test Type เลือก Kruskal-Wallis
3. กดปุ่ม OK



รูปที่ 10.16



รูปที่ 11.17



รูปที่ 11.18

Ranks

	pop	N	Mean Rank
size	1.00	7	7.71
	2.00	7	7.86
	3.00	10	19.10
	Total	24	

รูปที่ 11.19

Test Statistics(a,b)

	size
Chi-Square	14.936
df	2
Asymp. Sig.	.001

a Kruskal Wallis Test
b Grouping Variable: pop

รูปที่ 11.20

ผลการวิเคราะห์แสดงดังรูปที่ 11.19 และ 11.20 โดยในรูปที่ 11.19 เป็นตารางแสดงค่าสถิติต่าง ๆ ส่วนในรูปที่ 11.20 เป็นตารางที่จะนำไปใช้ในการทดสอบทางสถิติ โดยจะนำค่า Chi-Square (คือค่า *H*) ค่า df และค่า Asymp. Sig. ไปทดสอบสมมติฐาน[4] โดยเปรียบเทียบค่า Sig. กับค่าที่ตั้งไว้ คือ 0.05 นั้นเอง

5. Spearman's rank order correlation coefficient

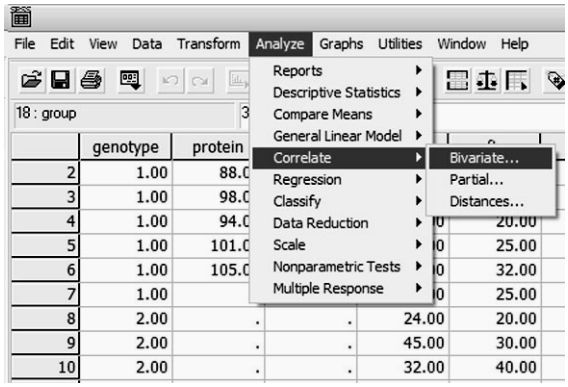
การทดสอบด้วยวิธีนี้ จะใช้กับข้อมูลที่มีมาตรการวัดเป็น ordinal scale เช่น การให้คะแนนการเป็นโรค ซึ่งแบ่งเป็น 10 คะแนน เป็นต้น การวิเคราะห์จะพิจารณาว่าตัวแปรสองตัวมีความสัมพันธ์กันหรือไม่ โดยวิธีการป้อนข้อมูลจะเป็นแบบเดียวกับการวิเคราะห์ Pearson's Correlation Coefficient ขั้นตอนการวิเคราะห์มีดังนี้

1. เลือก Analyze, Correlate, Bivariate... (รูปที่ 10.21)
2. จาก dialog box ในรูปที่ 10.22
   a. เลือกคู่ตัวแปรที่ต้องการหาความสัมพันธ์จากช่องด้านซ้าย แล้วกดปุ่ม ▶ ของช่อง Variables:
   b. ที่ช่อง Correlation Coefficients เลือก Spearman
   c. ที่ช่อง Test of Significance เลือก Two-tailed

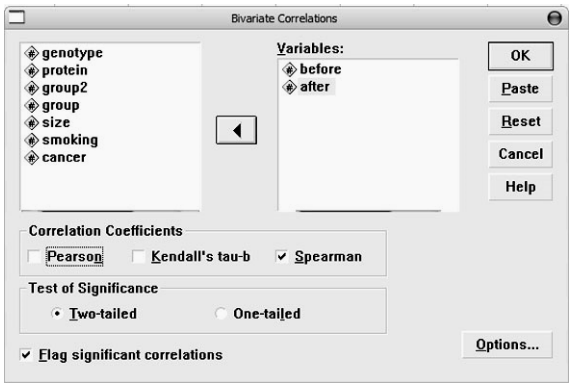
4 การตั้งสมมุติฐาน จะไม่กล่าวถึงค่า parametric statistics โดยตั้ง Ho: ปัจจัยจากกลุ่มต่าง ๆ ไม่มีผลต่อขนาด vs. H1: มีปัจจัยอย่างน้อย 1 ปัจจัยที่มีผลต่อขนาด

d. เลือก Flag Significant correlations

3. กดปุ่ม OK



รูปที่ 11.21



รูปที่ 11.22

ผลการวิเคราะห์แสดงในรูปที่ 11.23 โดยตารางจะมีรายละเอียดเช่นเดียวกันกับผลการวิเคราะห์ความสัมพันธ์เพื่อหาค่า Pearson's correlation coefficient โดยการทดสอบสมมุติฐานจะพิจารณาว่าค่า correlation coefficient มีนัยสำคัญทางสถิติหรือไม่ ถ้ามีนัยสำคัญ ก็จะมี * หรือ ** หลังค่า coefficient

Correlations

			before	after
Spearman's rho	before	Correlation Coefficient	1.000	.854(**)
		Sig. (2-tailed)	.	.002
		N	10	10
	after	Correlation Coefficient	.854(**)	1.000
		Sig. (2-tailed)	.002	.
		N	10	10

** Correlation is significant at the 0.01 level (2-tailed).

รูปที่ 11.23

แบบฝึกหัดที่ 11 การวิเคราะห์ทางสถิติแบบนอนพาราเมตริกซ์

1. ในศึกษาเกี่ยวกับ association ระหว่าง overweight และความสำเร็จทางการศึกษาในโรงเรียนแห่งหนึ่ง ผู้วิจัยได้เลือกนักเรียนมา 500 คน แล้วนำมาแยกกลุ่มได้ผลดังนี้

ความสำเร็จ	overweight	
	✔	✘
✔	162	263
✘	38	37

จงตั้งสมมุติฐาน ทดสอบสมมุติฐานที่ตั้ง และสรุปผลการทดสอบ

2. ในการศึกษาเกี่ยวกับ association ระหว่างหมู่เลือดกับ duodenal ulcer ผู้วิจัยได้สุ่มตัวอย่างผู้ป่วยที่เป็น duodenal ulcer จำนวน 1301 คน และคนปกติจำนวน 6313 แล้วนำตัวอย่างในแต่ละกลุ่มมาแยกตามหมู่เลือด ได้ผลดังนี้

กลุ่ม	หมู่เลือด			
ตัวอย่าง	O	A	B	AB
ผู้ป่วย	698	472	102	29
คนปกติ	2892	2625	570	226

จงตั้งสมมุติฐาน ทดสอบสมมุติฐานที่ตั้ง และสรุปผลการทดสอบ

3. ในการศึกษาเกี่ยวกับการใช้ Captopril ในคนที่จัดว่าเป็น Diuretic-treated hypertensive patient ผู้วิจัยเชื่อว่า สารตัวนี้จะมีผลทำให้ความดันเลือดในผู้ป่วยลดลง เพื่อทดสอบสมมุติฐานนี้ ผู้วิจัยได้ให้สาร Captopril ปริมาณ 6.25 กรัม แก่ผู้ป่วยเหล่านี้ โดยผู้วิจัยบันทึก Systolic blood pressure ของผู้ป่วยทุกคน ก่อนได้รับสาร Captopril (X) และหลังจากนั้น 70 นาที ก็วัดอีกที (Y) ได้ผลดังนี้

No.	X	Y
1	175	140
2	179	143
3	165	135
4	170	133
5	160	162
6	180	150
7	177	182
8	178	139
9	173	140
10	176	141

จงตั้งสมมุติฐาน ทดสอบสมมุติฐานที่ตั้ง และสรุปผลการทดสอบ

4. ในการศึกษาผลของการสูบบุหรี่ต่อการนอนหลับโดยผู้วิจัยได้ให้ความสนใจกับตัวแปรหนึ่ง คือ เวลาที่อาสาสมัครจะหลับ ผู้วิจัยจึงได้สุ่มตัวอย่างกลุ่มคนสูบบุหรี่ 12 คน และกลุ่มคนไม่สูบบุหรี่ 15 คน ได้ผลดังนี้

เวลาที่อาสาสมัครจะหลับ (นาที)			
กลุ่มคนสูบบุหรี่		กลุ่มคนไม่สูบบุหรี่	
69.3	52.7	28.6	30.6
56.0	34.4	25.1	31.8
22.1	60.2	26.4	41.6
47.6	43.8	34.9	21.1
53.2		29.8	36.0
48.1		38.4	37.9
23.2		38.5	13.9
13.8		30.2	

จงตั้งสมมุติฐาน ทดสอบสมมุติฐานที่ตั้ง และสรุปผลการทดสอบ

5. ภาวะขาดวิตามิน A เป็นปัญหาที่พบได้ทั่วไป และเป็นที่ทราบว่าการบริโภคผักจะช่วยให้ความเข้มข้นวิตามิน A ในน้ำเลือดเพิ่มขึ้น ผู้วิจัยจึงหาวิธีที่ให้เด็กได้รับวิตามิน A อย่างเพียงพอ โดยอยากทราบว่าการเติมน้ำมันลงไปในอาหารที่มีผักอยู่แล้ว จะช่วยให้มีความเข้มข้นวิตามิน A ในน้ำเลือดเพิ่มขึ้นหรือไม่ ผู้วิจัยสุ่มตัวอย่างเด็ก 28 คน ได้ผลดังนี้

Vitamin A concentration (mg./100 ml. blood plasma)		
I (No added fat)	II (5 g fat added)	III (10 g fat added)
18.1	29.1	26.6
16.5	15.8	16.1
21.0	20.4	18.8
18.7	23.5	25.0
7.4	18.5	21.8
12.4	21.3	15.4
16.1	23.1	19.9
17.9	23.8	15.5
	20.1	21.1
	11.9	25.5

จงตั้งสมมุติฐาน ทดสอบสมมุติฐานที่ตั้ง และสรุปผลการทดสอบ

6. ในการศึกษาความสัมพันธ์เชิงเส้นตรงระหว่างความสามารถของตาในการติดตามวัตถุที่เคลื่อนที่อย่างช้า ๆ ได้อย่างราบรื่น และระดับปริมาณอัลกอฮอล์ในเลือด ผู้วิจัยได้สุ่มกลุ่มตัวอย่างอาสาสมัคร 12 คน และได้ทำการทดลองโดยวัดความเร็วสูงสุดของตาในการติดตามวัตถุที่เคลื่อนที่ได้อย่างราบรื่น จากนั้นให้อาสาสมัครดื่มจินหรือวิสกี้หรือโทนิคในปริมาณที่มากพอ จนทำให้ไม่สามารถขับรถได้ แล้วจึงวัดความเร็วสูงสุดของตาในการติดตามวัตถุที่เคลื่อนที่ได้อย่างราบรื่นของอาสาสมัครอีกครั้ง และหาเปอร์เซ็นต์ที่ลดลงของการติดตามการเคลื่อนที่อย่างราบรื่น ได้ผลดังนี้

ฮัลกอฮอล์ในเลือด มก/ดล. (X)	เปอร์เซ็นต์ที่ลดลง (Y)
20	2
45	11
68	30
68	50
70	4
75	6
85	19
86	30
108	38
110	10
120	51
150	60

จงตั้งสมมุติฐาน ทดสอบสมมุติฐานที่ตั้ง และสรุปผลการทดสอบ