

การวางแผนการวิจัยและการวิเคราะห์ข้อมูล

ผศ.ดร. บุญเสฐียร บุญสูง

ภาควิชาสัตววิทยา คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

การศึกษาวิจัยมีขั้นตอนดำเนินการสรุปได้เป็น 7 ขั้นตอน ดังนี้คือ

1. การตั้งปัญหา
2. การตั้งสมมติฐาน
3. การวางแผนการวิจัย
4. การดำเนินการการวิจัย
5. การบันทึกผลข้อมูล
6. การวิเคราะห์ข้อมูล
7. การสรุปผลการศึกษา และหาความเชื่อถือได้ของผลการวิจัย

จากขั้นตอนเหล่านี้ จะเห็นได้ว่าสถิติเข้ามามีบทบาทในขั้นตอนของ

- ☒ การวางแผนการวิจัย
- ☒ การวิเคราะห์ข้อมูล
- ☒ การสรุปผลการศึกษา และหาความเชื่อถือได้ของผลการวิจัย

การวางแผนงานวิจัย

ก่อนที่ผู้ทำวิจัยจะเริ่มลงมือทำงานวิจัย นักวิจัยต้องมีการวางแผนงานวิจัยก่อน เริ่มแรกสุดคือการตอบคำถามก่อนว่า

- ☒ จะทำอะไร
- ☒ จะทำเมื่อไร
- ☒ จะทำที่ไหน
- ☒ จะทำอย่างไร

ทั้งนี้เพื่อมิให้เกิดการหลงทางหรือทำวิจัยโดยไม่ทราบเป้าหมายที่แน่นอน การวางแผนการวิจัยจะทำให้ผู้วิจัยมีกรอบว่า จะทำอะไร เมื่อไร ที่ไหน อย่างไร จะวิเคราะห์ข้อมูลและแสดงผลอย่างไร รวมทั้งต้องเพิ่มพูนทักษะด้านใดบ้างก่อนเริ่มลงมือทำจริง เพื่อให้ง่ายแก่การเข้าใจ ในที่นี้จะขอแบ่งการวางแผนงานวิจัยออกเป็น 2 กลุ่มใหญ่ๆ คือ การวางแผนงานสำรวจ และการวางแผนงานทดลอง

การวางแผนงานสำรวจ

งานสำรวจ เป็นการศึกษาเกี่ยวกับข้อมูลหรือปัจจัยบางอย่างที่มีอยู่ในธรรมชาติ โดยผู้วิจัยมีได้ไปเปลี่ยนแปลงแก้ไขสิ่งที่มีในธรรมชาตินั้น ส่วนมากเป็นความพยายามที่จะตอบคำถามเกี่ยวกับ

1. ต้องการทราบว่ามีสิ่งมีชีวิต กี่ชนิด อะไรบ้าง รวมถึงการคาดคะเนขนาดประชากร และหาความหนาแน่นของประชากร
2. การเปรียบเทียบจำนวนชนิด และจำนวนตัวของสิ่งมีชีวิตที่พบในสถานที่เดียวกันแต่ต่างเวลากัน
3. การเปรียบเทียบจำนวนชนิด และจำนวนตัวของสิ่งมีชีวิตที่พบในสถานที่แตกต่างกัน
4. การศึกษาเกี่ยวกับปัจจัยกายภาพต่างๆ
5. การหาความสัมพันธ์ระหว่างชนิดและจำนวนตัวของสิ่งมีชีวิตกับปัจจัยทางกายภาพและปัจจัยชีวภาพ เป็นต้น

หากผู้วิจัยต้องการตอบคำถามว่ามีสิ่งมีชีวิต กี่ชนิด อะไรบ้าง เช่น ผู้วิจัยต้องการทราบว่ามีไม้วงศ์ถั่วกี่ชนิด อะไรบ้าง ที่ขึ้นอยู่ภายในมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตบางเขน โดยไม่สนใจว่าจะมีจำนวนชนิดละเท่าไร แสดงว่าผู้วิจัยสนใจ**ศึกษาเชิงคุณภาพ (qualitative)** ผู้วิจัยจะใช้วิธีการออกเก็บตัวอย่างพรรณไม้วงศ์ถั่วที่ขึ้นภายในมหาวิทยาลัยทุกบริเวณ แล้วรายงานผลว่า มีไม้วงศ์ถั่วกี่ชนิด อะไรบ้าง

แต่ถ้าผู้วิจัยต้องการตอบคำถามอื่นๆ ซึ่งมีเรื่องของจำนวนเข้ามาเกี่ยวข้องจะเป็นการ**ศึกษาเชิงปริมาณ (quantitative)** การสำรวจมีแบบแผนมากกว่าในการศึกษาจริงผู้วิจัยไม่สามารถสำรวจประชากรพรรณไม้วงศ์ถั่วได้หมดทุกต้น ดังนั้นต้องใช้วิธีการสุ่มตัวแทน (sample) ของประชากรแทน

ในการสุ่มตัวอย่างเชิงปริมาณ มักจะกำหนดขนาดพื้นที่ คือใช้ **quadrat** คำถามที่เกิดขึ้นคือ

- ☒ เราจะทราบได้อย่างไรว่า quadrat ขนาดใดเป็นขนาดที่เหมาะสมสำหรับงานหนึ่งๆ
- ☒ เราจะต้องเก็บตัวอย่างจำนวนกี่ชิ้น เพื่อให้ได้ข้อมูลที่จะใช้เป็นตัวแทนของประชากรสำหรับนำไปวิเคราะห์ทางสถิติต่อไปได้

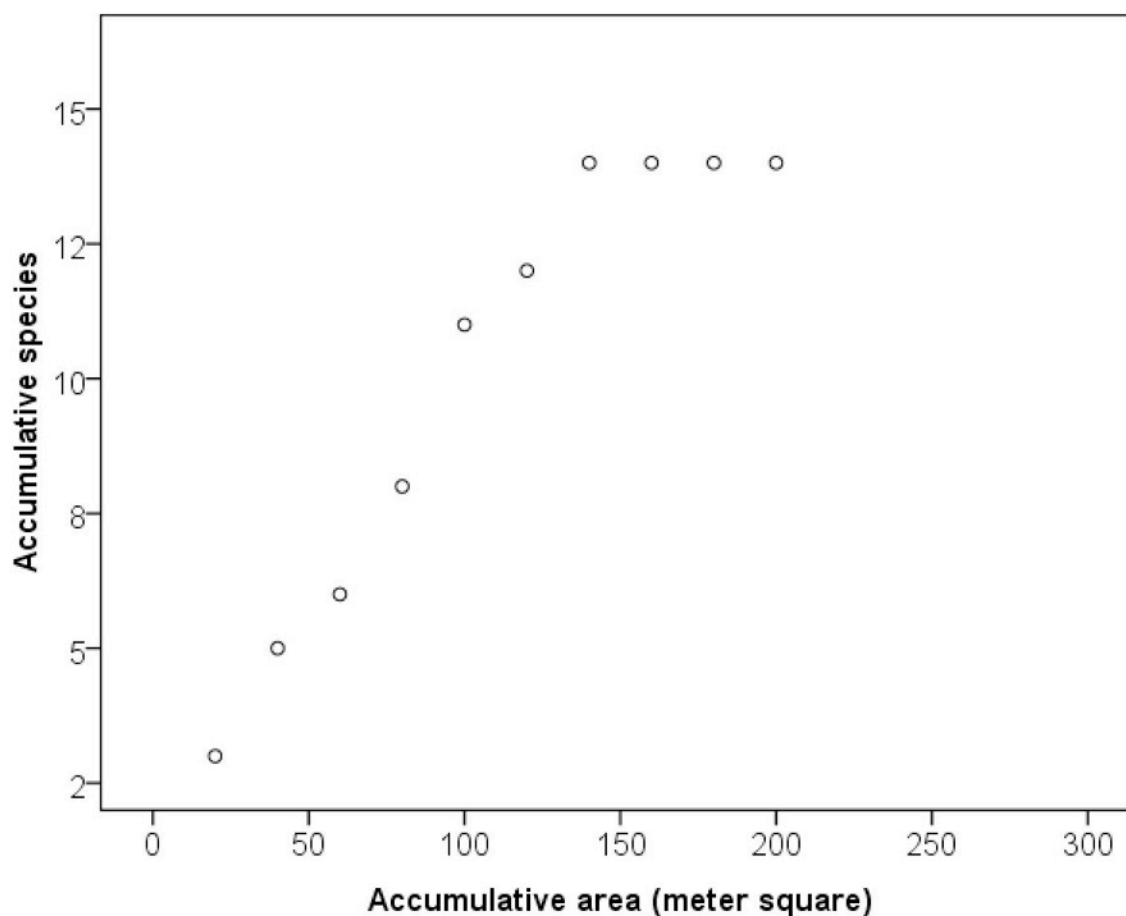
การหาขนาดที่เหมาะสมของ quadrat

ปัญหาที่ผู้วิจัยต้องตอบคือ จะสุ่มตัวแทนจาก พื้นที่เท่าใด จึงจะครอบคลุมจำนวนชนิดของสิ่งมีชีวิตที่เราต้องการศึกษาที่กระจายอยู่ในประชากรนั้น

วิธีการคือ กำหนดพื้นที่โดยใช้ quadrat ขนาดพื้นที่แตกต่างกัน เริ่มจากพื้นที่น้อย สุ่มตัวอย่างในบริเวณที่ต้องการศึกษา นับจำนวนชนิด (species) ที่พบภายใน quadrat สุ่มตัวอย่างอีกโดยเพิ่มขนาดของพื้นที่ quadrat เรื่อยๆ เช่น จาก 1 ตารางเมตร เป็น 2, 3, 4, ตารางเมตร เป็นต้น จนกระทั่งพบว่า **จำนวนสะสมของชนิดใหม่ (cumulative number of new species)** คงที่ แสดงว่าเราได้ขนาดของพื้นที่ที่เล็กที่สุดที่ครอบคลุมจำนวนชนิดที่ต้องการศึกษา ตารางที่ 1 แสดงข้อมูลการสำรวจไม้วงศ์หนึ่ง และรูปที่ 1 แสดง species area curve โดยใช้ข้อมูลจากตารางที่ 1

ตารางที่ 1 ข้อมูลการสำรวจไม้วงศ์หนึ่ง โดยได้ทำการสุ่มแต่ละครั้งพื้นที่ละ 20 ตารางเมตร

จำนวนครั้งของการสุ่ม	พื้นที่สุ่มสะสม (ม ²)	จำนวนชนิด	จำนวนชนิดใหม่	จำนวนสะสมชนิดใหม่
1	20	3	3	3
2	40	4	2	5
3	60	5	1	6
4	80	3	2	8
5	100	4	3	11
6	120	4	1	12
7	140	4	2	14
8	160	3	0	14
9	180	5	0	14
10	200	3	0	14

**ภาพที่ 1** แสดง species area curve

จากกราฟ แสดงว่าพื้นที่น้อยที่สุดที่ครอบคลุมจำนวนสิ่งมีชีวิตที่ต้องการศึกษาหรือขนาดที่เหมาะสมของ quadrat คือ 140 ตารางเมตร

การหาจำนวนซ้ำที่เหมาะสม

เมื่อหาขนาดของพื้นที่ที่เหมาะสมในการสำรวจ (ขนาด quadrat) ได้แล้วตั้งข้างต้น คำถามต่อไปคือ จะต้องเก็บตัวอย่างจำนวนกี่ซ้ำ จึงจะได้ข้อมูลที่เป็นตัวแทนของประชากร นั่นคือ จำนวน quadrat ที่น้อยที่สุดที่จะสามารถนำข้อมูลที่ได้มาใช้ในการวิเคราะห์ด้านปริมาณได้ วิธีที่นิยม มี 2 วิธีคือ

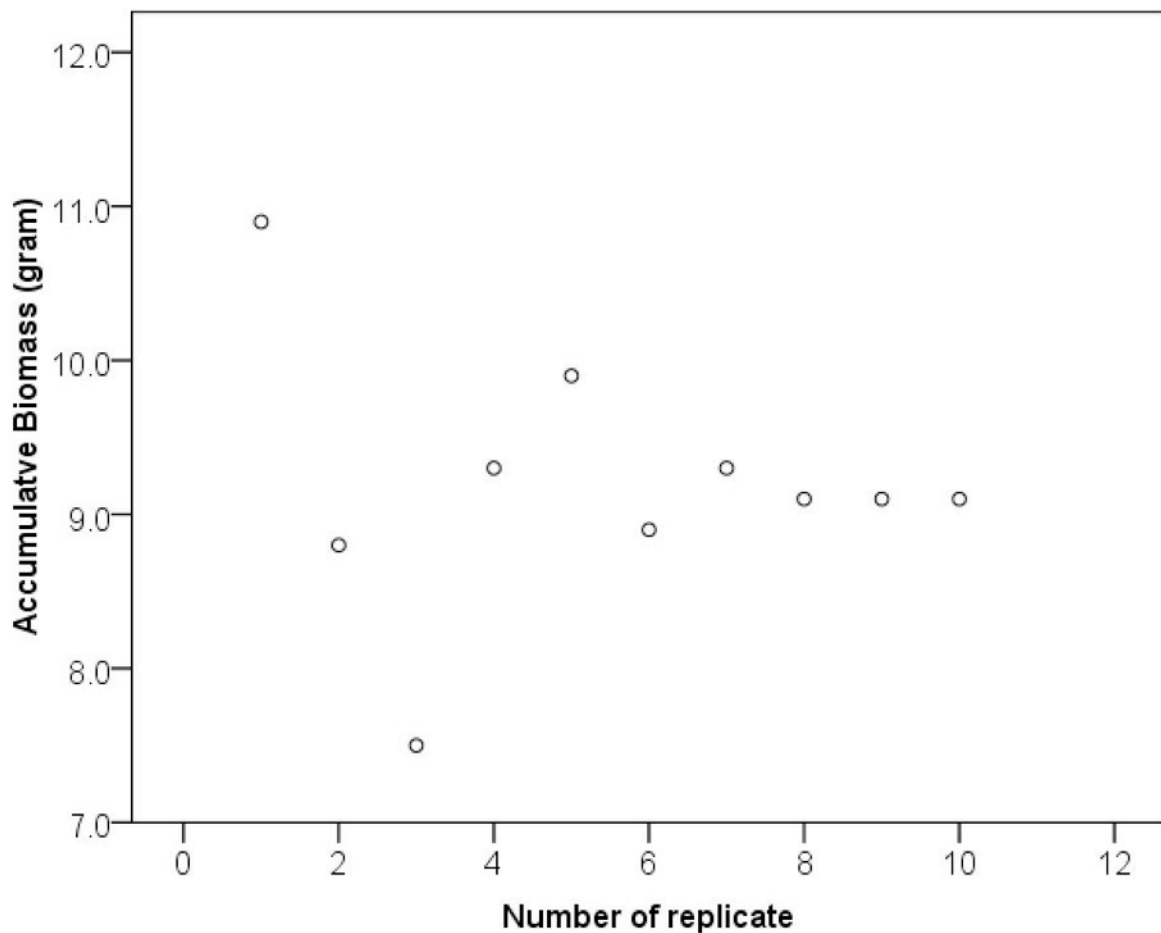
- ☒ ทำ performance curve
- ☒ การคำนวณจากสูตร

การทำ performance curve

ตัวอย่าง จากการสุ่มตัวอย่างตัวอ่อนแมลงซีปะขาวชนิดหนึ่งด้วย quadrat ขนาด 30*30 ซม. จำนวน 10 ครั้ง ได้มวลชีวภาพของแต่ละตัวอย่าง แสดงดังตารางที่ 2 หาค่าสะสมของมวลชีวภาพ ค่าสะสมของมวลชีวภาพคงที่ที่การสุ่มตัวอย่างครั้งที่เท่าใด แสดงว่าต้องสุ่มตัวอย่างเป็นจำนวนเท่านั้นซ้ำ จึงจะได้ข้อมูลที่เป็นตัวแทนของประชากรสามารถนำมาวิเคราะห์ทางสถิติได้ รูปที่ 2 เป็นตัวอย่างของการทำ performance curve จากข้อมูลตารางที่ 2

ตารางที่ 2 แสดงข้อมูลมวลชีวภาพของตัวอ่อนแมลงซีปะขาว จากการสุ่มตัวอย่างจำนวน 10 ครั้ง

การสุ่มตัวอย่างครั้งที่	มวลชีวภาพ (กรัม)	ค่าสะสมของมวลชีวภาพ (กรัม)
1	10.9	10.9
2	6.7	8.8
3	4.9	7.5
4	14.7	9.3
5	12.3	9.9
6	3.9	8.9
7	11.7	9.3
8	7.7	9.1
9	9	9.1
10	11	12



ภาพที่ 2 แสดง performance curve
จากกราฟ จำนวนซ้ำที่เหมาะสมคือ 8 ครั้ง

การคำนวณจากสูตร

การหาจำนวนซ้ำที่เหมาะสมอีกวิธีหนึ่งคือ วิธีคำนวณโดยใช้สูตร

$$N = 25\text{variance}/\text{mean}^2$$

วิธีนี้ต้องมีการสำรวจเบื้องต้นก่อน โดยเก็บตัวอย่างประมาณ 5-6 ครั้งด้วยขนาดของ quadrat ที่เหมาะสม นำมาหาค่าเฉลี่ย และความแปรปรวน นำมาเข้าสูตรข้างต้น จะได้จำนวนซ้ำที่เหมาะสม สูตรนี้ใช้ได้เมื่อผู้วิจัยยอมรับความคลาดเคลื่อนของเฉลี่ยเท่ากับ 20 เปอร์เซนต์ เมื่อได้จำนวนซ้ำที่เหมาะสมแล้ว จึงเริ่มดำเนินการเก็บตัวอย่าง หรือเริ่มงานวิจัยได้

การวางแผนงานทดลอง (experimental design)

งานทดลอง หมายถึง การจำลองสภาพการทดลองจากความเป็นจริงให้มาอยู่ในสภาพที่ผู้วิจัยสามารถควบคุมปัจจัยบางอย่างได้

งานทดลองประกอบด้วยส่วนต่างๆ ดังนี้คือ

1. **ทรีตเมนต์ (treatment)** หมายถึง สิ่งหรือวิธีที่เราให้แก่วัตถุทดลองเพื่อวัดผลเปรียบเทียบ เช่น สูตรอาหาร ระดับความเข้มข้นของฮอร์โมน วิธีการที่แตกต่างกัน เป็นต้น
2. **หน่วยทดลอง (experimental unit)** เป็นหน่วยที่ใช้วัดอิทธิพลของทรีตเมนต์ เช่น การศึกษาสูตรอาหาร 3 สูตร ว่าสูตรใดทำให้สุกรเจริญดีที่สุด เราสามารถทราบได้ว่าอาหารแต่ละสูตรมีส่วนประกอบหรือมีคุณค่าทางโภชนาการอะไรบ้างโดยการวิเคราะห์ส่วนประกอบ แต่เราไม่สามารถทราบได้ว่าสูตรอาหารใดดีกว่ากัน นอกจากนำอาหารแต่ละสูตรนั้นไปให้สุกรกิน เปรียบเทียบการเจริญของสุกรโดยการเปรียบเทียบน้ำหนัก ในที่นี้ สุกร คือ หน่วยทดลอง

หน่วยทดลองเป็น

☒ หน่วยวัดความคลาดเคลื่อนของการทดลอง

☒ หน่วยทดลองมีอิสระแก่กัน อาจเป็นสิ่งเดียวหรือตัวเดียว เช่น สุกรแต่ละตัวถูกเลี้ยง

ในกรงแยกกัน แต่ละตัวเป็น 1 หน่วยทดลอง หรืออาจเป็นกลุ่ม เช่น สุกร 10 ตัว ถูกเลี้ยงอยู่ในแต่ละกรงเรียกว่าแต่ละกรง (สุกร 10 ตัว) เป็น 1 หน่วยทดลอง เนื่องจากสุกรที่อยู่ในกรงเดียวกันไม่เป็นอิสระกัน แต่ละตัวจะได้อิทธิพลของตัวอื่น เช่น เกิดการรังแกกัน ตัวที่สู้เก่งกว่าจะมีน้ำหนักมากกว่า ในขณะที่ตัวที่สู้ไม่ได้จะมีน้ำหนักน้อยกว่า ถ้าคิดว่าสุกรแต่ละตัวในกรงเดียวกันนี้เป็น 1 หน่วยทดลอง ผลจะได้รับความแตกต่างกันของน้ำหนักดังที่กล่าวไปแล้ว ซึ่งอาจไม่ใช่อิทธิพลของสูตรอาหาร

ในพืชหากปลูกพืชเป็นแปลง พืชที่อยู่ริมๆ แปลงมักเจริญดีกว่าพืชที่อยู่แถวใน ดังนั้นในการทดลองจะตัดพืชแถวริมออกไปไม่นำมาใช้ในการทดลอง

3. ความคลาดเคลื่อนของการทดลอง (experimental error)

ความคลาดเคลื่อนของการทดลอง หมายถึง ความแตกต่างระหว่างหน่วยทดลอง 2 หน่วย ที่ได้รับทรีตเมนต์เดียวกัน ความคลาดเคลื่อนของการทดลองเป็นสิ่งที่ใช้ทดสอบความแตกต่างระหว่างทรีตเมนต์ว่าแตกต่างกันอย่างมีนัยสำคัญหรือไม่ คือ ดูว่าความแตกต่างระหว่างทรีตเมนต์มีความผิดแปลกไปจากความแปรปรวนปกติหรือไม่ โดยมีหลักการว่าในทรีตเมนต์เดียวกัน ผลตอบสนองของแต่ละหน่วยทดลอง (หรือค่าสังเกตของหน่วยทดลอง) ควรจะเท่ากันหมด เนื่องจากก่อนการทดลองได้เลือกหน่วยทดลองที่มีความสม่ำเสมอมากที่สุด แต่ในความเป็นจริงแม้จะควบคุมหน่วยทดลองอย่างไรก็ตามก็มีความแปรปรวนเกิดขึ้นซึ่งอาจเนื่องมาจากสาเหตุต่างๆ ดังนี้ คือ ธรรมชาติของหน่วยทดลอง วิธีปฏิบัติในการทดลอง สิ่งแวดล้อมอื่นๆ เป็นต้น สิ่งเหล่านี้เรียกว่าเป็นความแปรปรวนปกติ ใช้ความคลาดเคลื่อนภายในทรีตเมนต์เป็นค่ามาตรฐานในการเปรียบเทียบกับความคลาดเคลื่อนระหว่างทรีตเมนต์หากความแตกต่างภายในทรีตเมนต์มีน้อยแต่ความแตกต่างระหว่างทรีตเมนต์มีมาก กล่าวได้ว่าความแตกต่างนี้เกิดเนื่องมาจากความแตกต่างระหว่างทรีตเมนต์ แต่ถ้าความแตกต่างภายในทรีตเมนต์มีมากทำให้ไม่เห็นความแตกต่างระหว่างทรีตเมนต์จึงกล่าวว่ามี ความคลาดเคลื่อนระหว่างทรีตเมนต์

ข้อมูลในกลุ่มเดียวกันมีความสัมพันธ์กันเช่นแต่ละหน่วยทดลองอยู่ในกรงเดียวกัน ทำให้มีความคลาดเคลื่อนสูง เมื่อนำค่าความคลาดเคลื่อนไปเปรียบเทียบกับหาความแตกต่าง

ระหว่าง ทรีตเมนต์จะได้ผลว่าไม่มีความแตกต่างระหว่างทรีตเมนต์ต่างๆ ที่ความเป็นจริงมีความแตกต่างกัน (เนื่องจากความคลาดเคลื่อนในประเภท 2 (Type II error) มีมาก)

วิธีแก้ไขคือ (i) ในการเลือกหน่วยทดลองต้องคำนึงถึงคือความแปรปรวนที่สามารถมีผลต่อการทดลอง ได้ เช่น เพศ อายุ พันธุ์ สามารถมีผลต่ออัตราการเจริญ ทำให้ไม่สามารถกล่าวได้ว่าผลที่แตกต่างกันนั้นเนื่องมาจากทรีตเมนต์จริง การเลือกหน่วยทดลองจึงต้องคำนึงถึงความสม่ำเสมอของหน่วยทดลอง (ii) มีจำนวนซ้ำของหน่วยทดลอง และ (iii) ลดอคติโดยสุ่มให้แต่ละสมาชิกในตัวแทนหรือในประชากรมีโอกาสเท่าๆ กัน ที่จะอยู่ในกลุ่มใดกลุ่มหนึ่งหรือได้รับทรีตเมนต์ใดทรีตเมนต์หนึ่ง

การวางแผนการทดลอง

หมายถึง วิธีการที่เราจัดหน่วยทดลองให้แก่ทรีตเมนต์

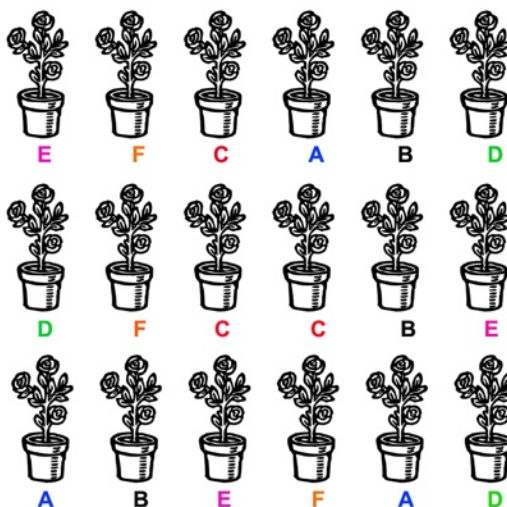
1. แผนทดลองแบบสุ่มตลอด (Completely Randomized Design, CRD)

เหมาะสำหรับสิ่งทดลองที่มีความสม่ำเสมอกันมาก ไม่มีความแตกต่างเนื่องจากอื่นๆ ความแปรปรวนเกิดขึ้นเนื่องจากทรีตเมนต์เท่านั้น เป็นการแจกแจงปัจจัยความแปรปรวนในทิศทางเดียว การสุ่มหรือการจัดสิ่งทดลองให้แก่ทรีตเมนต์จะใช้วิธีสุ่มธรรมดา คือให้ทุกๆ หน่วยทดลองมีโอกาสเท่าๆ กันในการรับทรีตเมนต์ใดทรีตเมนต์หนึ่ง การสุ่มอาจใช้วิธีการจัดสลาก หรืออาจใช้ตารางเลขสุ่ม

ตัวอย่าง 1. ผู้วิจัยต้องการศึกษาอิทธิพลของปุ๋ยไนโตรเจน 6 ระดับ ต่อกุหลาบ จำนวนทรีตเมนต์ละ 3 ข้ว ถ้าให้

- A = หน่วยควบคุม (มีปุ๋ยไนโตรเจนเป็น 0 มก./กระถาง)
- B = ทรีตเมนต์ 2 มีปุ๋ยไนโตรเจนเป็น 20 มก./กระถาง
- C = ทรีตเมนต์ 3 มีปุ๋ยไนโตรเจนเป็น 30 มก./กระถาง
- D = ทรีตเมนต์ 4 มีปุ๋ยไนโตรเจนเป็น 40 มก./กระถาง
- E = ทรีตเมนต์ 5 มีปุ๋ยไนโตรเจนเป็น 50 มก./กระถาง
- F = ทรีตเมนต์ 6 มีปุ๋ยไนโตรเจนเป็น 60 มก./กระถาง

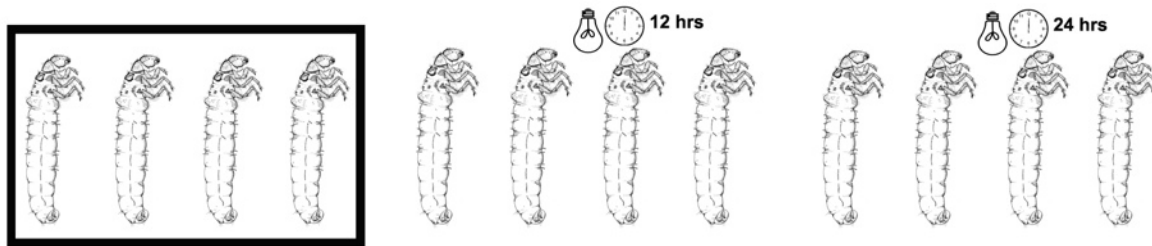
ในการสุ่มทรีตเมนต์ให้แก่ต้นข้าวทั้ง 18 กระถาง ทำได้โดยสุ่มโดยตลอด



แผนผังการทดลอง

2. ผู้วิจัยต้องการทดลองว่าการได้รับแสงแตกต่างกันมีผลต่อการเจริญเป็นดักแด้ และตัวเต็มวัยของตัวอ่อนแมลงหนอนปลอกน้ำชนิดหนึ่งหรือไม่

การทดลองนี้มี 3 ทรีตเมนต์คือ (1) ไม่ได้รับแสงเลย (2) ได้รับแสง 12 ชั่วโมงต่อวัน (3) ได้รับแสง 24 ชั่วโมง



2. แผนทดลองแบบสุ่มในบล็อก (Completely Randomized Blocks Design, RBD)

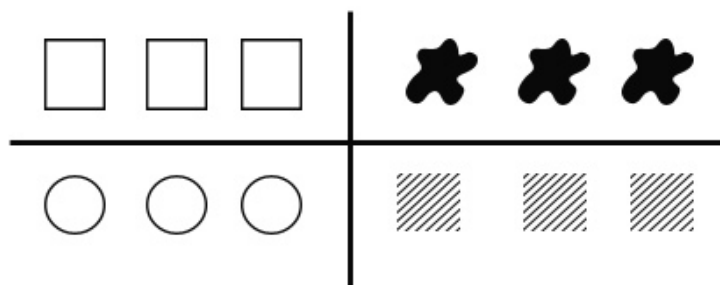
แผนทดลองแบบสุ่มในบล็อกเหมาะสำหรับสิ่งทดลองที่มีความสม่ำเสมอไม่เท่ากันเนื่องจากเพศ พันธุ์ อายุ เป็นต้น หรือปัจจัยอื่นที่ไม่ใช่ทรีตเมนต์ที่สามารถทำให้เกิดความผันแปรขึ้น เช่น ระยะเวลาที่แตกต่างกัน ทำให้สิ่งทดลองนั้นตอบสนองต่อทรีตเมนต์เดียวกันแตกต่างกัน จึงลดความแปรปรวนด้วยการจัดกลุ่ม (block) สิ่งทดลองเช่นเพศเดียวกันไว้ในบล็อกเดียวกัน จากนั้นจึงสุ่มให้ทรีตเมนต์แก่สิ่งทดลองภายในบล็อกอีกทีหนึ่ง วิธีนี้ทำให้ความแตกต่างที่เกิดขึ้นไปรวมอยู่ในบล็อกไม่ไปอยู่ที่ทรีตเมนต์ การแจกแจงความแปรปรวนเป็นไปใน 2 ทิศทาง คือ ความแปรปรวนภายใน (within) บล็อก และความแปรปรวนระหว่าง (between หรือ among) บล็อก

ตัวอย่าง เช่น มีสิ่งทดลองที่แตกต่างกัน 4 แบบ คือ

การทดลองมี 3 ทรีตเมนต์ คือ 1 2 และ 3

วิธีทำ

1. จัดบล็อกสิ่งทดลองเป็น 4 บล็อก
2. ให้ทรีตเมนต์แก่แต่ละบล็อก โดยสุ่ม



หรือถ้าเป็นการทดลองในแปลงที่มีความอุดมสมบูรณ์ของดินต่างกันหรือมีความสูงแตกต่างกัน ซึ่งสิ่งเหล่านี้ทำให้เกิดความแตกต่างได้โดยมิได้มาจากทรีตเมนต์ จึงต้องมีบล็อก เช่น

ต้องการทดสอบการปลูกพืช 2 ชนิด ในพื้นที่มีความสูงแตกต่างกันมาก (ที่สูง และที่ราบ)

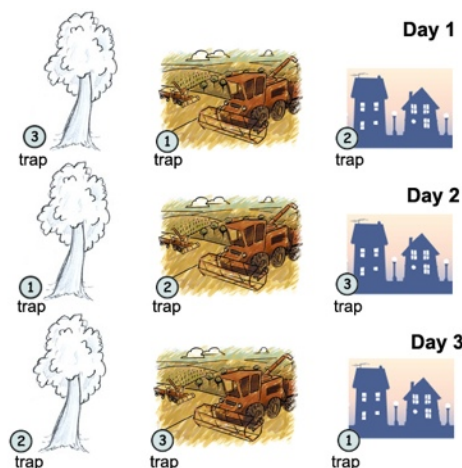
1. แบ่งบล็อก (ที่สูง และที่ราบ)
2. ให้ทรีตเมนต์ คือ พืชทั้ง 2 ชนิด แก่แต่ละบล็อกโดยสุ่ม

3. แผนทดลองแบบละตินสแควร์ (Latin Squares, LS)

บางครั้งแหล่งของความแปรปรวนในหน่วยทดลองอาจมาจาก 2 แหล่ง (หรือ 2 ทิศทาง) ซึ่งเมื่อรวมกับความแปรปรวนที่มาจากทรีตเมนต์ ทำให้มีแหล่งความแปรปรวนรวมทั้งสิ้น 3 แหล่ง (หรือ 3 ทิศทาง) ในการวางแผนการทดลองต้องพยายามหาทางลดแหล่งของความแปรปรวนในหน่วยทดลอง โดยการจัดบล็อกหน่วยทดลองแล้วจึงให้ทรีตเมนต์ (A B C และ D) จะได้แผนงานดังนี้

A	B	C	D
B	C	D	A
C	D	A	B
D	A	B	C

ตัวอย่าง ในการทดสอบประสิทธิภาพของกับดักแมลง 3 ชนิด สำหรับการดักจับแมลงพาหะชนิดหนึ่ง จากการศึกษาเบื้องต้นพบว่า แมลงชนิดนี้มีที่อาศัยทั้งในเขตชุมชน ป่า และเขตเกษตรกรรม และการดักจับแมลงได้จำนวนไม่เท่ากันในแต่ละวันซึ่งไม่ทราบว่าเป็นเพราะเหตุผลใด จะเห็นได้ว่านอกเหนือจากความแปรปรวนที่เกิดจากทรีตเมนต์คือความแตกต่างของกับดักแมลง 3 ชนิดแล้ว ตัวแมลงมีแหล่งของความแปรปรวนในเรื่องที่อาศัยและการตอบสนองต่อกับดักไม่เท่ากันในแต่ละวัน ในการทดลองสามารถลดแหล่งของความแปรปรวนได้โดยบล็อกที่อาศัยออกเป็น 3 แบบ และให้กับดักทั้งสามแบบ (1 2 และ 3)



4. Factorial Experiment

การทดลองที่มีปัจจัยมากกว่า 1 ปัจจัย แต่ปัจจัยมีหลายระดับ เช่น การทดลองสูตรอาหารสำหรับเลี้ยงไก่ ที่มีโปรตีนต่างกัน 3 ระดับ และมีกรดอะมิโน ไลซีน 2 ระดับ การจัดการทดลอง จะจัดให้ทรีตเมนต์ มีส่วนประกอบทุกปัจจัยและทุกระดับ

ทรีตเมนต์	1	โปรตีน ระดับ 1 ไลซีน ระดับ 1
ทรีตเมนต์	2	โปรตีน ระดับ 1 ไลซีน ระดับ 2
ทรีตเมนต์	3	โปรตีน ระดับ 2 ไลซีน ระดับ 1
ทรีตเมนต์	4	โปรตีน ระดับ 2 ไลซีน ระดับ 2

จำนวนทรีตเมนต์ = 2^k เมื่อ k เป็นจำนวนปัจจัยที่ต้องการศึกษา

การวางแผนการทดลองอาจเป็นแบบ CRD หรือ RCRD

การวิเคราะห์ข้อมูล และหาความเชื่อถือได้ของผลการวิจัย

เมื่อได้ทำการศึกษา บันทึกข้อมูลต่างๆ แล้ว งานขั้นต่อไปคือ การวิเคราะห์ข้อมูล การแปลผลข้อมูลและการหาความเชื่อถือได้ของผลการวิจัย สถิติขั้นต้นที่ใช้มากในการวิเคราะห์ข้อมูลทั่วไปคือ

1. สถิติพรรณนา (Descriptive statistics)

1.1 การวัดความโน้มหาคศูนย์ (central tendency measurement) เป็นการแสดงค่าที่เป็นตัวแทนของค่าสังเกตจากประชากรหนึ่งๆ

ในการศึกษาใดๆ ข้อมูลที่ได้จะประกอบด้วยค่าสังเกตจำนวนหนึ่งซึ่งจะมีค่าหนึ่งซึ่งแสดงให้เห็นว่าค่าสังเกตส่วนมากจะมีค่าใกล้เคียงกับค่านี้ เราเรียกว่า เฉลี่ยของตัวแทน เฉลี่ย มี 3 แบบ คือ

1.1.1 เฉลี่ยเลขคณิต (arithmetic mean) เป็นค่าความโน้มหาคศูนย์ ที่นิยมมากที่สุด คำนวณได้จากผลรวมของค่าสังเกตทั้งหมดหารด้วยจำนวนค่าสังเกต

1.1.2 median เป็นค่าความโน้มหาคศูนย์ที่ได้จาก เมื่อนำค่าสังเกตทั้งหมดมาเรียงกันตามลำดับจากน้อยไปหามากแล้ว median นี้เป็นค่าที่อยู่ตรงกลาง นั่นคือ จะมีค่าสังเกตที่อยู่สูงกว่าและต่ำกว่าค่านี้อยู่อย่างละครึ่ง

1.2.3 mode เป็นค่าความโน้มหาคศูนย์ที่ได้จากการพิจารณาค่าซึ่งปรากฏบ่อยที่สุดในค่าสังเกตชุดนั้น

ตัวอย่าง การสำรวจประชากรผีเสื้อชนิดหนึ่ง พบว่ามีความยาวปีกดังตารางที่ 3 ให้หาเฉลี่ยของปีกตารางที่ 3 แสดงความยาวของปีกผีเสื้อ

ผีเสื้อตัวที่	ความยาวปีก (ซม.)
1	3.3
2	3.5
3	3.6
4	3.6
5	3.7
6	3.8
7	3.8
8	3.8
9	3.8
10	3.9
11	3.9
12	13

$$\begin{aligned}\text{หา mean} \quad \text{mean} &= \sum x/N \\ &= 43.7/12 \\ &= 3.27 \text{ เซนติเมตร}\end{aligned}$$

หา median

3.3 3.5 3.6 3.6 3.7 **3.8** **3.8** 3.8 3.8 3.9 3.9 4.0

ในที่นี้เมื่อเรียงลำดับค่าสังเกตแล้ว ค่าที่อยู่ตำแหน่งตรงกลางมีสองตัว คือ 3.8 และ 3.8 median จะเท่ากับผลรวมของสองค่าสังเกตนี้หารด้วย 2

$$\text{median} = 3.8 + 3.8 = 3.8$$

หา mode ค่าสังเกตที่มีจำนวนซ้ำมากที่สุดคือ 3.9

$$\text{mode} = 3.9$$

1.2 การวัดการกระจายของข้อมูล (dispersion)

การกระจายของข้อมูลพิจารณาได้จาก

1.2.1 พิสัย (range) หมายถึงผลต่างของค่าสังเกตสูงสุดกับค่าสังเกตต่ำสุด

$$\text{range} = \text{maximum} - \text{minimum}$$

1.2.2 ความเบี่ยงเบนของค่าเฉลี่ย (sample mean deviation)

เมื่อได้ค่าเฉลี่ยของตัวแทนแล้ว ควรพิจารณาว่าค่าเฉลี่ยเบี่ยงเบนไปจากค่าสังเกตเท่าไร ซึ่งหาได้จากผลรวมของค่าสัมบูรณ์ความแตกต่างระหว่างค่าเฉลี่ยกับค่าสังเกตแต่ละค่า หารด้วยจำนวนค่าสังเกต

$$\text{sample mean deviation} = \sum |x_i - \bar{x}| / n$$

1.2.3 variance

อีกวิธีหนึ่งที่จะกำจัดเครื่องหมายของความเบี่ยงเบนได้คือ การยกกำลังสอง ผลรวมของกำลังสองของความเบี่ยงเบนจากค่าเฉลี่ยเรียกว่า sum of squares มีตัวอย่าง SS

$$\text{population SS} = \sum (x_i - \mu)^2$$

$$\text{sample SS} = \sum (x_i - \bar{x})^2$$

$$\text{population variance} = \sum (x_i - \mu)^2 / N$$

$$\text{sample variance } s^2 = \sum (x_i - \bar{x})^2 / n-1$$

1.2.4 ส่วนเบี่ยงเบนมาตรฐาน (standard deviation)

ส่วนเบี่ยงเบนมาตรฐาน หมายถึงค่าบอกของรากที่สองของ variance

$$\text{population standard deviation } \sigma = \sqrt{\sum (x_i - \mu)^2 / N}$$

$$\text{sample standard deviation } s = \sqrt{\sum (x_i - \bar{x})^2 / n-1}$$

1.2.5 สัมประสิทธิ์ของความผันแปร (coefficient of variation)

$$CV = s / \bar{x}$$

$$\text{หรือ } CV = (s / \bar{x}) 100 \%$$

1.3 ความเที่ยงตรงของค่าเฉลี่ย (accuracy of a sample mean)

1.3.1 ความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย (Standard error of mean)

เมื่อเรารวบรวมตัวอย่างจากหลายๆ ตัวอย่างในประชากรเดียวกัน เราจะได้ค่าเฉลี่ยที่แตกต่างกัน จากแต่ละตัวอย่าง ความแตกต่างของเฉลี่ยหรืออาจเรียกได้ว่าการกระจายของเฉลี่ย เราสามารถหาความเบี่ยงเบนมาตรฐานของค่าเฉลี่ยได้ซึ่งมักเรียกว่า ความคลาดเคลื่อนมาตรฐาน (standard error) ซึ่งคำนวณได้จาก

$$SE = s / \sqrt{n}$$

1.3.2 ช่วงเชื่อมั่นของค่าเฉลี่ย (confidence limits of mean)

ช่วงเชื่อมั่นมักใช้แทนความคลาดเคลื่อนมาตรฐานและสามารถตีความได้ง่าย ช่วงหมายถึงค่าสูงสุดถึงค่าต่ำสุดของพิสัยที่ค่าเฉลี่ยของประชากรมีค่าอยู่ ตัวอย่างเช่น ช่วงความเชื่อมั่น 95% หมายความว่าในการสุ่มตัวอย่าง 100 ครั้ง จะมีค่าสังเกต 95 ค่า ที่มีค่าอยู่ระหว่างช่วงนี้

$$\text{ดังนั้น} \quad \bar{x} \pm t_{\alpha/2, n-1} SE \quad \text{เมื่อ } \alpha \text{ เท่ากับช่วงความเชื่อมั่น}$$

2. Indices of Diversity

Nominal scale data ไม่มี mean, mode และ median ไว้สำหรับการอธิบายเกี่ยวกับการกระจายของข้อมูล เราสามารถใช้แนวความคิดเกี่ยวกับความหลากหลาย (diversity)

ตัวอย่าง กระจกสร้างรังในสถานที่อยู่แตกต่างกัน 4 แบบ คือ บริเวณเกาะวัลย์ กิ่งไม้ ในโพรงต้นไม้ และในถ้ำ จากการสำรวจรังทั้งสิ้น 20 รัง ถ้าในที่อยู่แต่ละแบบมีกระจกรังอยู่ 5 ตัว เราอาจกล่าวได้ว่ามีความหลากหลายของสถานที่สร้างรังมาก (great diversity) แต่ถ้า กระจก 17 รัง ทำรังในโพรง ที่เหลืออีก 3 รัง ทำรังที่กิ่งไม้ 1 รัง ในถ้ำ 1 รัง และในโพรง 1 รัง เราจะบอกได้ว่ามีความหลากหลายของสถานที่สร้างรังต่ำ

$$H' = - \sum p_i \log p_i$$

$$\text{เมื่อ } p_i = \frac{\text{จำนวนของชนิด } i}{\text{จำนวนทั้งหมด}}$$

3. การเปรียบเทียบตัวแทน (comparison of samples)

การสำรวจบางครั้งผู้สำรวจได้ทำการสำรวจในสถานที่แตกต่างกัน หรือสำรวจในระยะเวลาที่แตกต่างกัน ผู้สำรวจอาจต้องการเปรียบเทียบผลการสำรวจนั้น สามารถทำได้โดยการเปรียบเทียบตัวแทนซึ่งก็คือ **ค่าเฉลี่ย (mean)** ของข้อมูลแต่ละชุดนั่นเอง

ค่าเฉลี่ยของตัวแทน 2 ชุด มักมีค่าแตกต่างกัน การทดสอบทางสถิติจะช่วยให้ตัดสินใจได้ว่า ความแตกต่างของค่าเฉลี่ยนั้นเป็นความแตกต่างอย่างมีนัยสำคัญ หรือความแตกต่างนั้นถูกครอบคลุมด้วยความคลาดเคลื่อนของค่าเฉลี่ย ซึ่งในกรณีหลังนี้ถือว่าไม่แตกต่างกัน

ในการเปรียบเทียบตัวแทน เราต้อง

1. ตั้งสมมติฐาน (null hypothesis, H_0) ก่อน โดยทั่วไปจะตั้ง null hypothesis ว่า samples มาจาก population เดียวกัน ดังนั้นความแตกต่างของ sample mean จะถูกยอมรับได้ในความคลาดเคลื่อนของการประมาณค่า population mean

2. คำนวณความน่าจะเป็น (probability, P) ที่ค่าเฉลี่ยทั้งสองไม่แตกต่างกัน

ถ้าความน่าจะเป็น (P) สูงแสดงว่าโอกาสที่ sample ทั้งสอง จะไม่แตกต่างกันนั้นมีมาก ดังนั้นจึงยอมรับ null hypothesis ได้

ในทำนองเดียวกัน ถ้าคำนวณความน่าจะเป็น (P) ได้ค่าต่ำ แสดงว่า โอกาสที่ sample ทั้งสอง จะไม่แตกต่างกันมีน้อย นั่นคือ sample mean 2 ค่าแตกต่างกัน จึงไม่ยอมรับ null hypothesis

ระดับนัยสำคัญ (significant level) ที่เท่าไรที่นิยมใช้

งานทางชีววิทยามักใช้ระดับนัยสำคัญที่ความน่าจะเป็น 5% ($P = 0.05$) null hypothesis จะถูกยอมรับเมื่อความน่าจะเป็นสูงกว่า 5% ($P > 0.05$) และถูกปฏิเสธเมื่อความน่าจะเป็นต่ำกว่าหรือเท่ากับ 5% ($P \leq 0.05$)

ที่ระดับนัยสำคัญที่ความน่าจะเป็น 5% หรือ $P = 0.05$ หมายความว่าอย่างไร

หมายความว่า จะมีโอกาสเพียง 1 ใน 20 หรือ 5 ใน 100 ครั้งเท่านั้น ที่ค่าเฉลี่ยทั้งหลายที่นำมาเทียบกันจะไม่แตกต่างกัน ดังนั้นเมื่อ $P > 0.05$ แสดงว่าโอกาสมากขึ้นที่ค่าเฉลี่ยจะไม่แตกต่างกัน จึงยอมรับ null hypothesis ที่ว่า sample mean ทั้งสองค่าไม่แตกต่างกัน

ระดับนัยสำคัญที่ใช้กันมาก มี 3 ระดับ แสดงด้วยเครื่องหมาย ดอกจัน (*) ดังนี้ คือ

* = $P < 0.05$	significant (มีนัยสำคัญ)
** = $P < 0.01$	highly significant (มีนัยสำคัญยิ่ง)
*** = $P < 0.001$	very highly significant (มีนัยสำคัญอย่างยิ่ง)

ความคลาดเคลื่อนที่อาจเกิดขึ้นได้ในการทดสอบทางสถิติ

ความคลาดเคลื่อนที่อาจเกิดขึ้นได้ มี 2 แบบ คือ

1. Type I error คือ H_0 ถูกปฏิเสธ ทั้งๆ ที่ความเป็นจริงจะต้องถูกยอมรับ เช่น

$$H_0: \mu_1 = \mu_2$$

ในความเป็นจริง: μ_1 ไม่ต่างจาก μ_2 แต่เมื่อทดสอบทางสถิติได้ค่า $p < 0.05$ จึงปฏิเสธ H_0

2. Type II error คือ H_0 ถูกยอมรับ ทั้งๆ ที่ความเป็นจริงจะต้องถูกปฏิเสธ เช่น

$$H_0: \mu_1 = \mu_2$$

ในความเป็นจริง: μ_1 ต่างจาก μ_2 แต่เมื่อทดสอบทางสถิติได้ค่า $p > 0.05$ จึงยอมรับ H_0

การเกิด error ทั้งสองแบบนี้ จะทำให้การสรุปผลผิดพลาด

Type I error และ Type II error เกิดขึ้นจากสาเหตุหลายอย่าง เช่น การใช้สถิติที่ไม่เหมาะสม เช่น มีความไวน้อยมาเป็นตัวทดสอบ การมีจำนวนซ้ำน้อยเกินไป การมีความไม่สม่ำเสมอในหน่วยทดลอง การเก็บตัวอย่างแบบไม่สุ่ม การกระจายของข้อมูลไม่เป็นแบบปกติ (แบบปกติ คือมีการกระจายเป็นรูปประฆังคว่ำที่สมมาตร ไม่เบ้ไปทางซ้ายหรือทางขวา) เป็นต้น ซึ่งสิ่งเหล่านี้สามารถควบคุมได้โดยการวางแผนการทดลองที่ดี

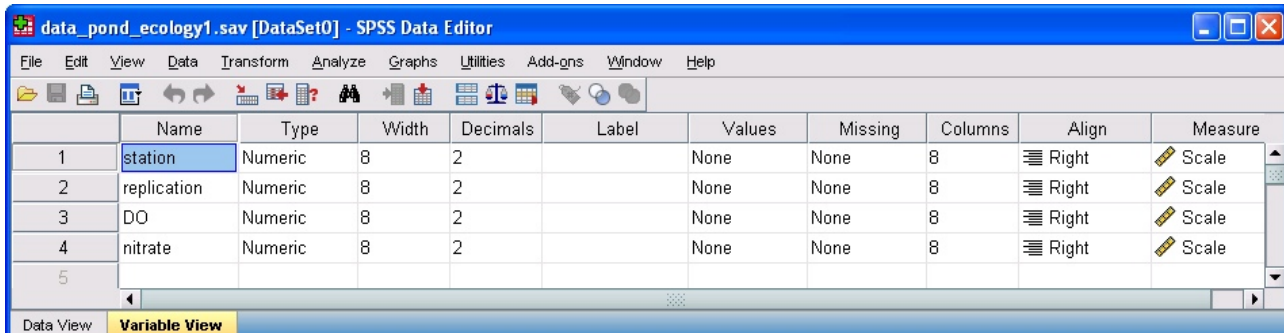
การทดสอบนัยสำคัญโดยใช้ parametric test ซึ่ง assume ว่า ข้อมูลที่นำมาทดสอบนั้นต้องมีการกระจายแบบปกติ (normal distribution) และมี variance เท่ากัน ดังนั้นก่อนการทดสอบด้วย parametric test ต้องทำการทดสอบข้อมูลก่อนว่า มีการกระจายแบบปกติหรือไม่ และมี variance เท่ากันหรือไม่ ถ้าไม่ต้องการแปลง (transform) ข้อมูลก่อน (ด้วย logarithm, square root หรือ arc sine ตามแต่เหมาะสมก่อน) เพื่อให้ข้อมูลมีการกระจายแบบปกติ ส่วนการทดสอบนัยสำคัญโดยใช้ non-parametric test ไม่สนใจการกระจายของข้อมูลว่าจะเป็นแบบใด ในที่นี้จะกล่าวถึงการทดสอบนัยสำคัญโดยใช้ parametric test

การวิเคราะห์ข้อมูลทางนิเวศวิทยาเบื้องต้นโดยใช้โปรแกรมสำเร็จรูป SPSS for Windows

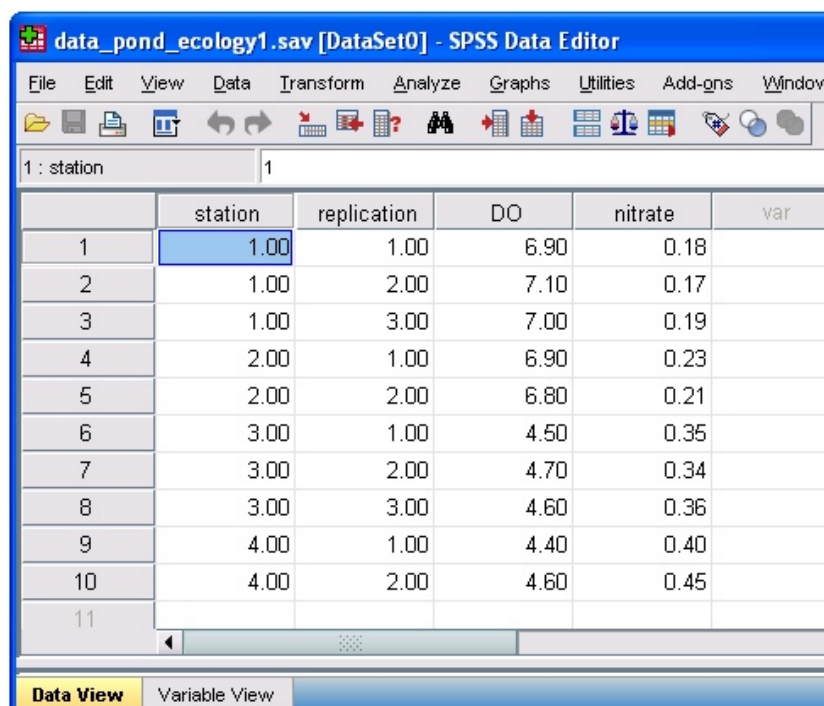
1. การเตรียมข้อมูล

เมื่อเปิดโปรแกรมสำเร็จรูป SPSS for Windows ขึ้นมา จะมีหน้าจอ SPSS Data Editor 2 Tab คือ Data View และ Variable View

Variable View เป็นหน้าจอให้สร้าง หรือแก้ไขตัวแปร โดยต้องกำหนดชื่อ ชนิด ความยาว และความหมายของตัวแปร และใช้ตรวจสอบรายละเอียดของตัวแปร ดังภาพ



Data View เป็นหน้าจอให้พิมพ์ข้อมูลตามตัวแปรที่กำหนดใน Variable View ดังภาพ



2. การหาค่าสถิติพื้นฐาน

สถิติพื้นฐานทางนิเวศวิทยาที่นิยมใช้กันบ่อย คือ สถิติเชิงพรรณนา (descriptive statistics) ได้แก่

2.1 ค่าเฉลี่ย (mean)

ค่าเฉลี่ยเป็นค่าความโน้มหาคำนวณกลางของข้อมูล สามารถหาได้จากผลรวมของค่าทั้งหมดหารด้วยจำนวนค่า สำหรับค่าเฉลี่ยของตัวอย่าง (sample mean) ใช้ $\bar{x}$ เป็นสัญลักษณ์แทน ซึ่ง $\bar{x}$ นี้จะแสดงคุณลักษณะของตัวแทน (sample)

2.2 ความแปรปรวน (variance)

ความแปรปรวนเป็นค่าที่ใช้วัดการกระจายของข้อมูลโดยพิจารณาจากความแตกต่างระหว่างข้อมูลแต่ละค่ากับค่าเฉลี่ย ค่าความแปรปรวนเป็นผลรวมของกำลังสองของความแตกต่างระหว่างแต่ละค่าสังเกตกับค่าเฉลี่ยแล้วหารด้วยจำนวนข้อมูล

2.3 ส่วนเบี่ยงเบนมาตรฐาน (standard deviation, S.D.)

ส่วนเบี่ยงเบนมาตรฐานเป็นค่าที่บอกถึงความแปรปรวนของชุดข้อมูล โดยการยึดเอาค่าเฉลี่ยเป็นหลัก ถ้าชุดข้อมูลใดมีค่าส่วนเบี่ยงเบนมาตรฐานสูง ก็แสดงให้เห็นว่าค่าต่างๆ ของข้อมูล ชุดนั้นๆ ห่างจากค่าเฉลี่ยมาก ส่วนเบี่ยงเบนมาตรฐานมีค่าได้ทั้งค่าบวก และค่าลบ

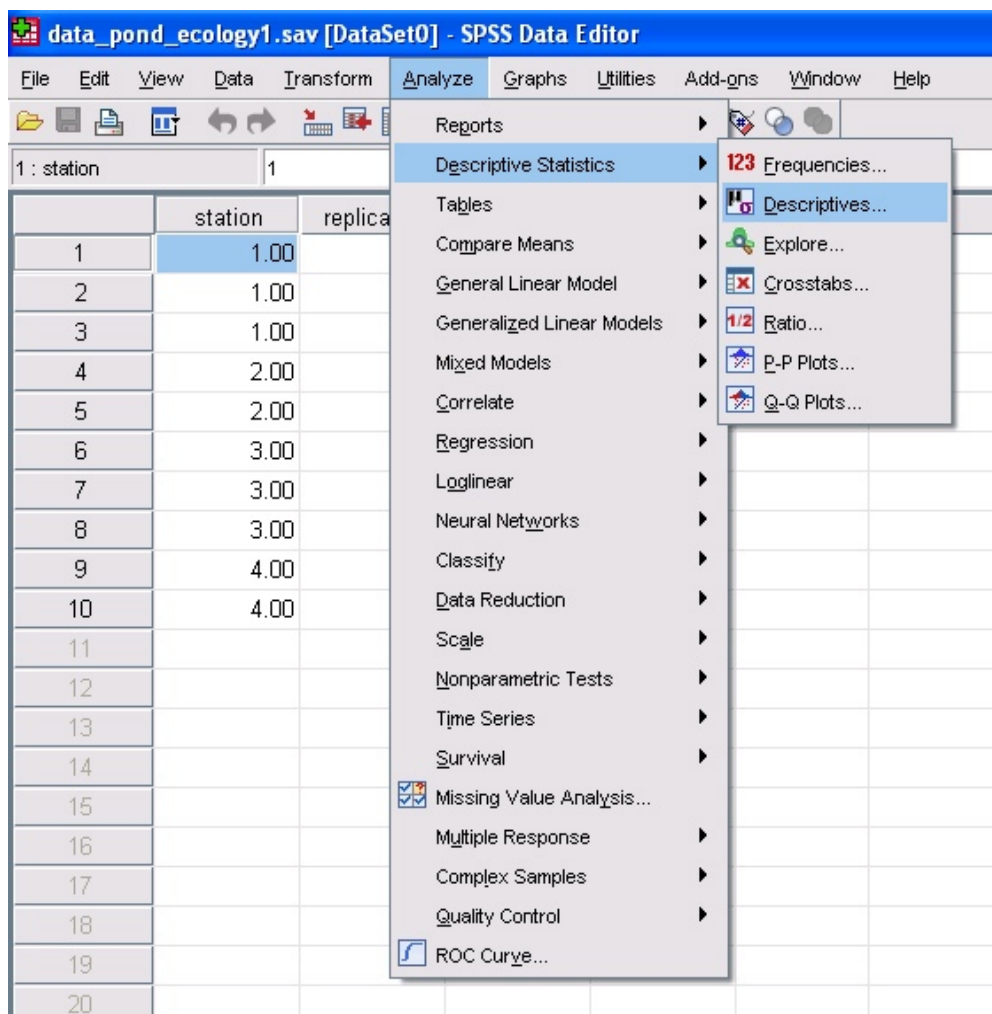
2.4 ความคลาดเคลื่อนมาตรฐาน (standard error, S.E.)

ค่าความคลาดเคลื่อนมาตรฐานสามารถหาได้จาก นำค่าความแปรปรวนหารด้วยขนาดของตัวอย่าง แล้วถอดรากที่สอง

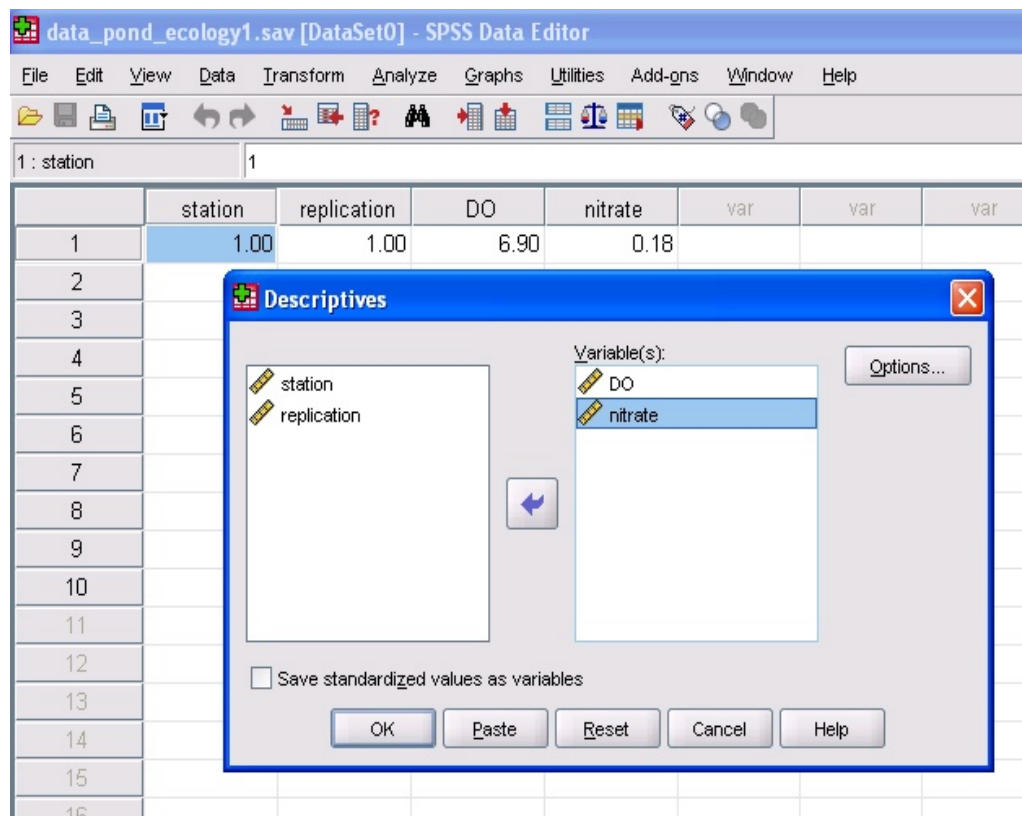
2.5 สัมประสิทธิ์ของความผันแปร (coefficient of variation, C.V.)

สัมประสิทธิ์ของความผันแปรคือค่าที่แปลงมาจากการนำส่วนเบี่ยงเบนมาตรฐานไปหารด้วยค่าเฉลี่ย แล้วคูณด้วย 100 มีหน่วยเป็นเปอร์เซ็นต์ (%)

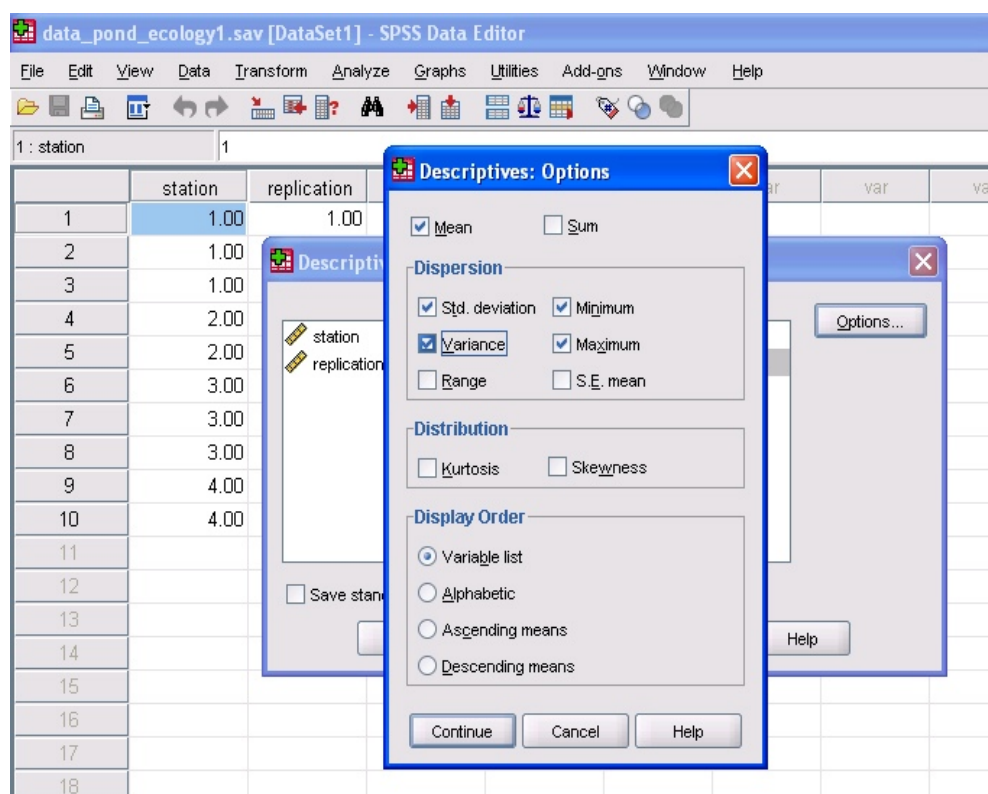
สำหรับการหาค่าสถิติเชิงพรรณนา ในโปรแกรมสำเร็จรูป SPSS ทำได้โดย ไปที่เมนู Analyze --> Descriptive Statistics --> Descriptive ดังภาพ



จะปรากฏหน้าต่างดังภาพ แล้วเลือกตัวแปรที่ต้องการหาค่าไปช่องทางด้านขวา แล้วเลือก Options



และจะปรากฏหน้าต่างดังภาพ จากนั้นคลิก box ในค่าที่ต้องการคำนวณดังภาพ แล้วกด Continue --> OK



จากนั้นจะได้ Output ออกมาในหน้าต่างใหม่ ดังนี้

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
DO	10	4.40	7.10	5.7500	1.25897	1.585
nitrate	10	.17	.45	.2880	.10283	.011
Valid N (listwise)	10					

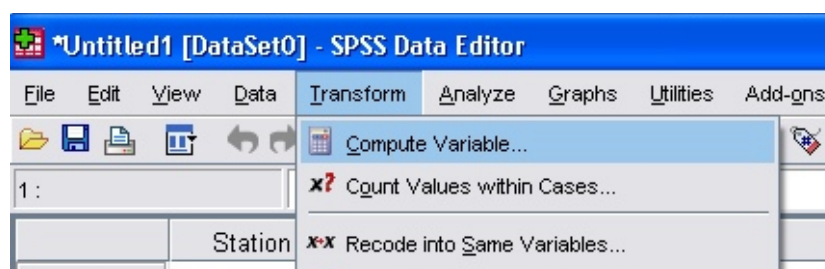
ในการทดสอบสถิติที่ใช้พารามิเตอร์ (parametric test) อาจต้องมีการพิจารณาการกระจายของข้อมูล ก่อนนำไปทดสอบก่อนว่าข้อมูลมีการแจกแจงแบบปกติ (normal distribution) หรือไม่ โดยอาจพิจารณาอย่างง่ายโดยดูจากค่าเฉลี่ยกับค่าความแปรปรวน ถ้าค่าเฉลี่ยมากกว่าค่าความแปรปรวน แสดงว่าข้อมูลมีการแจกแจงแบบปกติ แต่ถ้าค่าเฉลี่ยมีค่าน้อยกว่าค่าความแปรปรวน ข้อมูลมีการแจกแจงแบบไม่ปกติ อาจจะเบ้ขวา หรือเบ้ซ้าย จะต้องมีการแปลงข้อมูล (transform) ก่อนนำมาทดสอบ เช่น การแปลงข้อมูลด้วย $\log(x+1)$ (กรณีที่มีข้อมูลเป็น 0) หรืออาจถอดรากที่สองข้อมูลทั้งหมดก่อน เป็นต้น

3. การแปลงข้อมูล

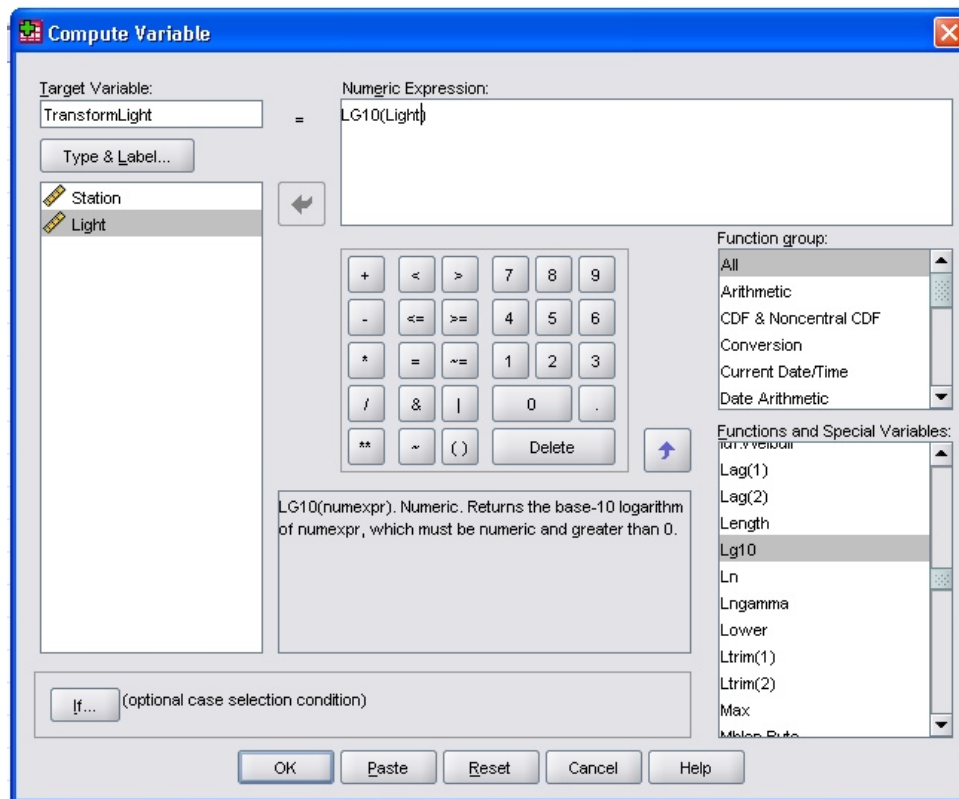
Descriptive Statistics

	N	Mean	Variance
Light	9	1.5056E3	1.059E6
Valid N (listwise)	9		

จากตารางจะเห็นว่าความเข้มแสง (Light) มีค่าเฉลี่ยน้อยกว่าค่าความแปรปรวน แสดงว่าข้อมูลมีการแจกแจงแบบไม่ปกติจะต้องมีการแปลงข้อมูล (transform) ก่อนนำมาทดสอบ โดย ไปที่เมนู Transform --> Compute Variable ดังภาพ



จากนั้น เลือก Function group เป็น All แล้วเลือก Lg10 แล้วกดมาใส่ใน Numeric Expression Box กด Light มาใส่ในวงเล็บ และใส่ชื่อใหม่ใน Target Variable แล้วกด OK ข้อมูลที่ Transform แล้วจะปรากฏใน Data View



4. การเปรียบเทียบสองตัวแทน (Two-samples Test)

การเปรียบเทียบสองตัวแทน เป็นการเปรียบเทียบค่าเฉลี่ยของตัวแทนสองค่า จะใช้สถิติที่เรียกว่า “t-test” ในการทดสอบ ค่าสถิติ t นี้ คำนวณได้จาก

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

เปรียบเทียบค่า P ที่ได้ หากค่า P ต่ำกว่าที่กำหนด จะปฏิเสธ H_0 สรุปผลว่า มีความแตกต่างอย่างมีนัยสำคัญ หากค่า P สูงกว่าที่กำหนด จะยอมรับ H_0 สรุปผลว่า ไม่มีความแตกต่างอย่างมีนัยสำคัญ

ตัวอย่าง ในการทดสอบสมรรถภาพการดักจับแมลงของกับดักจับผีเสื้อกลางคืน 2 ชนิด จำนวนชนิดละ 11 อัน ได้ผลดังตารางที่ 4 อยากทราบว่ากับดักชนิดใดใช้ดักจับผีเสื้อกลางคืนได้ดีกว่า

ตารางที่ 4 จำนวนผีเสื้อกลางคืนที่ถูกจับด้วยกับดัก 2 ชนิด

กับดักชนิดที่ 1	41	34	33	36	40	25	31	37	34	30	38
กับดักชนิดที่ 2	52	57	62	55	64	57	56	55	55	64	67

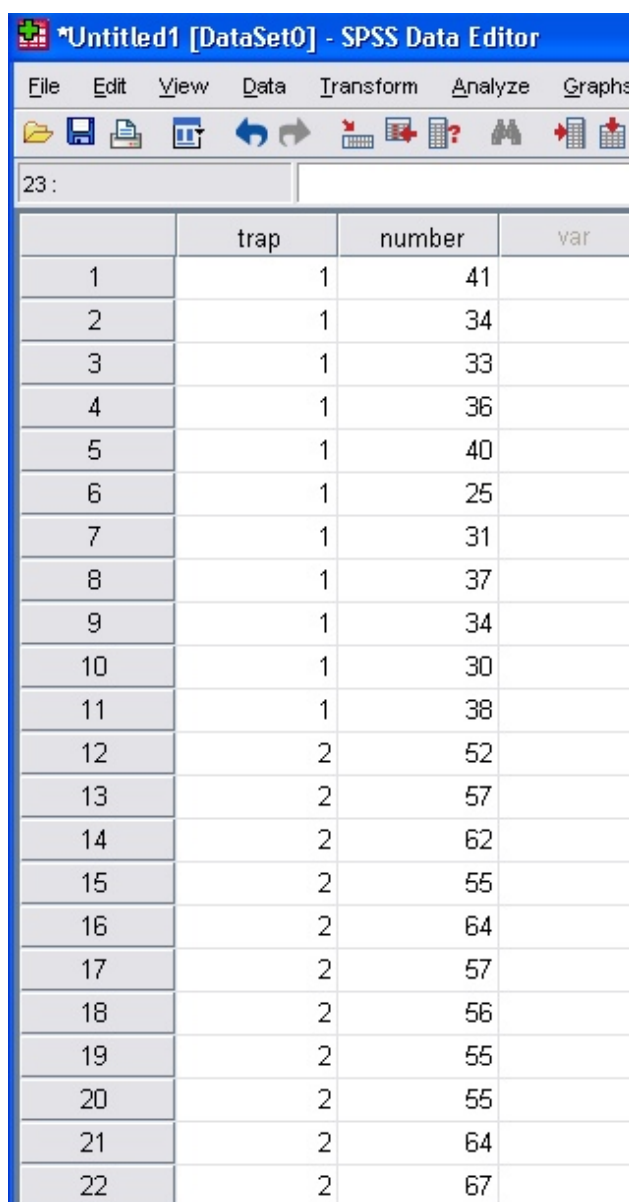
$$H_0: \mu_1 = \mu_2$$

$$H_a: \mu_1 \neq \mu_2$$

$$\alpha = 0.05$$

วิธีทำ เตรียมข้อมูล ป้อนสิ่งที่ต้องการศึกษา (case) ในแถวนอน (row) และข้อมูลค่าสังเกต (variable) ในแนวตั้ง (column) ใช้โปรแกรม SPSS for Windows

1. ในตารางใส่ข้อมูล trap หมายถึง ชนิดของกับดัก มี 2 ชนิด ใส่ 1 แทนกับดักชนิดที่ 1 จำนวน 11 อัน และ 2 แทนกับดักชนิดที่ 2 จำนวน 11 อัน number หมายถึง จำนวนผีเสื้อกลางคืนที่จับได้ในแต่ละกับดักแต่ละชนิด



	trap	number	var
1	1	41	
2	1	34	
3	1	33	
4	1	36	
5	1	40	
6	1	25	
7	1	31	
8	1	37	
9	1	34	
10	1	30	
11	1	38	
12	2	52	
13	2	57	
14	2	62	
15	2	55	
16	2	64	
17	2	57	
18	2	56	
19	2	55	
20	2	55	
21	2	64	
22	2	67	

2. ทดสอบว่าข้อมูลของแต่ละกับดักมีการกระจายแบบปกติหรือไม่ โดยอาจจะต้องแยก column แต่ละกับดัก number1 คือ จำนวนผีเสื้อกลางคืนที่จับได้จากกับดักแบบที่ 1 number 2 คือ จำนวนผีเสื้อกลางคืนที่จับได้จากกับดักแบบที่ 2 การทดสอบทำได้โดยไปที่เมนู Analyze เลือกคำสั่ง Descriptive Statistic แล้วเลือก Descriptives แล้วพิจารณาค่า mean และ variance ดังผลลัพธ์ดังนี้

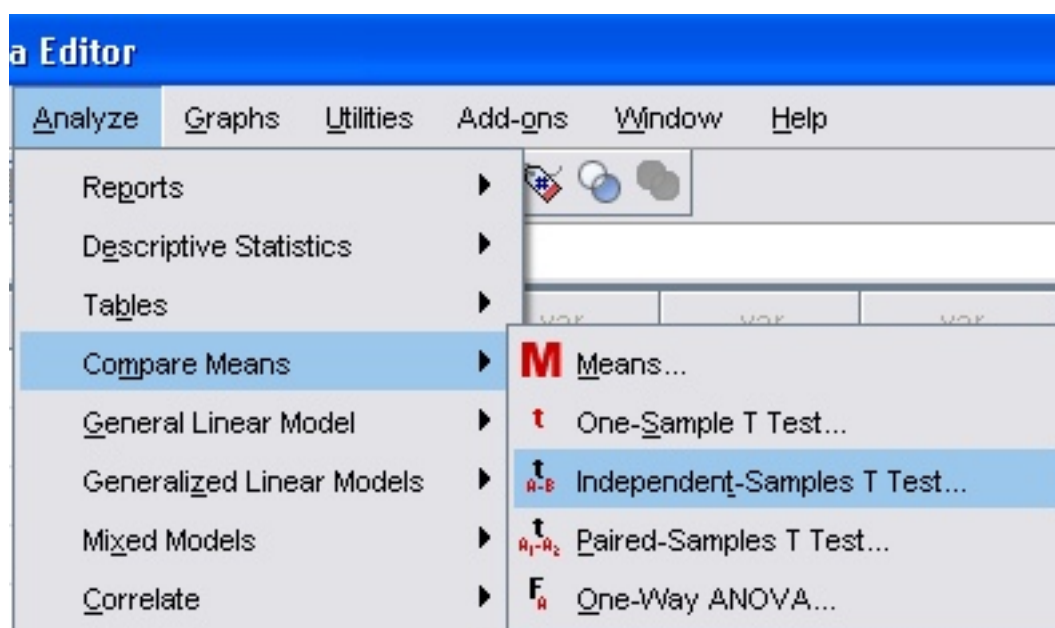
Descriptive Statistics

	N	Mean	Variance
number1	11	34.45	21.873
number2	11	58.55	23.473
Valid N (listwise)	11		

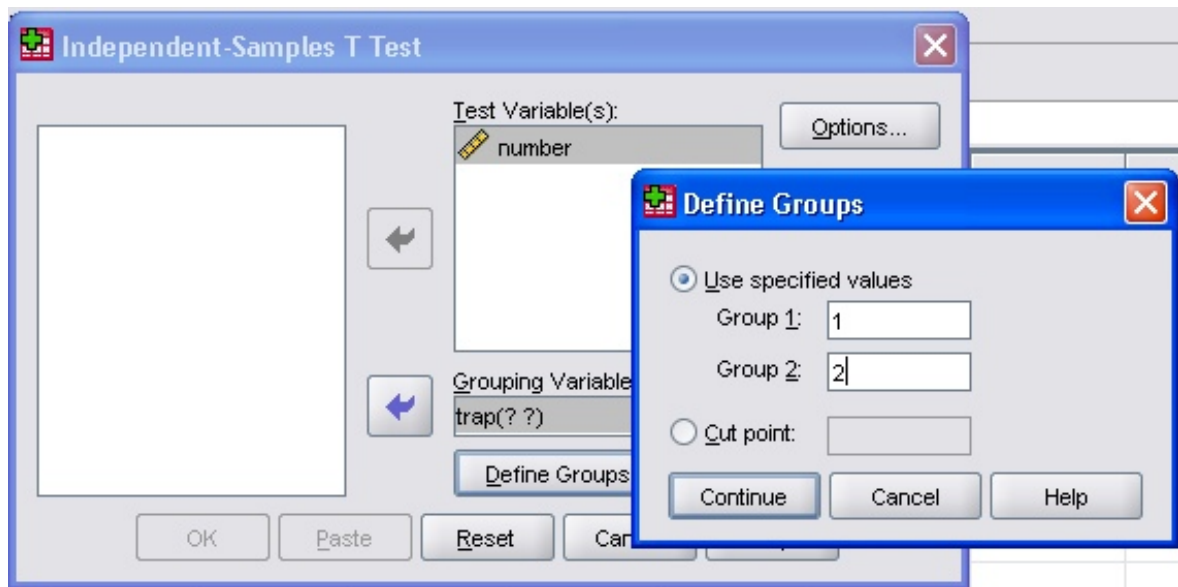
จากตาราง mean ทั้งสองค่าพบว่าค่า mean มีค่าสูงกว่า variance แสดงว่า ข้อมูลมีการกระจายแบบปกติ สามารถนำมาใช้คำนวณได้เลย

3. เปรียบเทียบค่าเฉลี่ยของข้อมูล 2 ชุด

จากเมนู Analyze เลือกคำสั่ง Compare Means แล้วใช้คำสั่งย่อย Independent-Sample T Test ดังภาพ



หน้าจอจะขึ้นหน้าต่าง Independent-Samples T Test เลือก number มาใส่ในช่อง Test Variable(s) และเลือก trap มาใส่ในช่อง Grouping Variable ซึ่งปรากฏข้อความ trap(??) ข้างใต้มีข้อความ Define Groups ให้คลิก box นี้ใส่ค่า Use Specified Values คือ Group 1: 1 และ Group 2: 2 เนื่องจากเราใช้เครื่องหมาย 1 แทนกับดักแบบที่ 1 และเครื่องหมาย 2 แทนกับดักแบบที่ 2 แล้วกด Continue และ OK ตามลำดับ



จะได้ผลลัพธ์ดังนี้

T-Test

Group Statistics

trap		N	Mean	Std. Deviation	Std. Error Mean
number	1	11	34.45	4.677	1.410
	2	11	58.55	4.845	1.461

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
number	Equal variances assumed	.283	.600	-11.865	20	.000	-24.091	2.030	-28.326	-19.856
	Equal variances not assumed			-11.865	19.975	.000	-24.091	2.030	-28.326	-19.855

พิจารณาที่ Sig. (คือความน่าจะเป็น P) ที่ t-test for Equality of Means ได้ Sig. (2-tailed) มีค่าเท่ากับ .000 ซึ่งมีค่าน้อยกว่า 0.05 แสดงว่าจำนวนแมลงที่จับได้ด้วยกับดักสองแบบแตกต่างกันอย่างมีนัยสำคัญ คือ กับดักแบบที่ 2 ดักจับแมลงได้มากกว่าหรือมีประสิทธิภาพในการดักจับผีเสื้อกลางคืนดีกว่ากับดักแบบที่ 1 ดังนั้นไม่ยอมรับ H_0

การรายงานผล: ผลการวิเคราะห์พบว่ากับดักแมลงแบบที่ 2 สามารถดักจับผีเสื้อกลางคืนได้จำนวนมากกว่ากับดักที่ 1 อย่างมีนัยสำคัญ ($t_{20} = 311.865$, $P = 0.000$)

5. การเปรียบเทียบหลายตัวแทน (Multi-samples Test)

ถ้าต้องการทดสอบตัวแทนที่มากกว่า 2 ชุดขึ้นไป สามารถทดสอบได้โดยวิธีการวิเคราะห์ค่าความแปรปรวน (Analysis of Variance, ANOVA) หลักการของ Analysis of Variance คือ การแบ่งแยกผลกำลังสอง (Sum of square) ของ variance ทั้งหมดออกเป็นส่วนๆ ตามสาเหตุของความแตกต่างนั้น ค่าที่คำนวณได้เมื่อนำมาเปรียบเทียบกับค่าความคลาดเคลื่อนซึ่งเป็นตัวมาตรฐานใช้สำหรับทดสอบจะทำให้ทราบว่ามีความแตกต่างหรือไม่และความแตกต่างนั้นมาจากสาเหตุอะไร วิธีนี้ต้องการ assumption ดังนี้

- i) ตัวแทนมาจากประชากรที่มีการกระจายแบบปกติ
- ii) ประชากรมี variance เท่ากัน ดังนั้น mean และ variance เป็นอิสระต่อกัน
- iii) อิทธิพลจาก treatment และสิ่งแวดล้อมอื่นๆ รวมกันโดยวิธีบวก

5.1 แผนการทดลองแบบสุ่มตลอด (Completely Random Design, CRD) และการวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียว (One-Way ANOVA; F-test)

การวางแผนการทดลองแบบสุ่มตลอดเป็นแผนการทดลองแบบหนึ่งที่ทำการศึกษาทดลองนิยมใช้กันมาก เพราะมีความง่าย และสะดวกต่อการปฏิบัติ แต่แผนการทดลองนี้จะมีคามแม่นยำก็ต่อเมื่อสิ่งที่นำมา ทดลอง หรือ ข้ำของการทดลองทุกๆ ข้ำ จะต้องมีความสม่ำเสมอคล้ายคลึงกันมากที่สุด เช่น อายุ น้ำหนัก ขนาด สายพันธุ์ เป็นต้น

การวิเคราะห์ความแปรปรวนแบบจำแนกทางเดียวเป็นการจำแนกข้อมูลด้วยตัวแปรหรือปัจจัยเพียงปัจจัยเดียว เป็นการเปรียบเทียบค่าเฉลี่ยตัวแปรเชิงกลุ่มที่สนใจเปรียบเทียบตั้งแต่ 2 กลุ่มขึ้นไป

Source of variation	Sum of square	Degrees of freedom	Mean square (Variance)
Between Groups	S_2	$i-1$	$S_2/2 = S_2/i-1$
Within Groups (residual variance)	S_1	$N-i$	$S_1/2 = S_1/N-1$
Total	S_T	$N-1$	

ทดสอบด้วย F-test $F = S_2/2/S_1/2$

ในการทดสอบใช้ error mean square เป็นมาตรฐานในการทดสอบว่า ความแตกต่างระหว่าง between groups นั้นมีมากเกินกว่าระดับที่กำหนดให้หรือไม่ ถ้ามากกว่าสรุปผลว่า ความแตกต่างนั้นๆ เนื่องมาจากผลของ within groups จริง

ตัวอย่าง จากการศึกษาปริมาณออกซิเจนละลายน้ำ (DO) และปริมาณไนเตรท (nitrate) จากการสุ่มตัวอย่างน้ำ 4 บ่อ โดยมีจำนวนซ้ำไม่เท่ากัน ได้ผลดังตารางที่ 5 อยากทราบว่า DO และ nitrate ของน้ำแต่ละบ่อแตกต่างกันหรือไม่

ตารางที่ 5 ค่า DO ที่วัดได้ในแต่ละครั้งของตัวอย่างน้ำแต่ละบ่อ

บ่อที่ 1	บ่อที่ 2	บ่อที่ 3	บ่อที่ 4
6.90	6.90	4.50	4.40
7.10	6.80	4.70	4.60
7.00		4.60	

สมมติฐาน $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1 : \mu_i \neq \mu_j$ อย่างน้อย 1 คู่ ; $i \neq j$

หรือ $H_0 : \text{ค่าเฉลี่ย DO แต่ละ station เท่ากัน}$

$H_1 : \text{ค่าเฉลี่ย DO อย่างน้อย 2 station แตกต่างกัน}$

การสรุปผลการทดสอบ ถ้ากำหนดระดับความเชื่อมั่นที่ 95%

➢ $P > 0.05$ ยอมรับ H_0 แสดงว่าค่าเฉลี่ย DO แต่ละ station เท่ากัน

➢ $P \leq 0.05$ ปฏิเสธ H_0 แสดงว่า อย่างน้อย 2 station มีค่าเฉลี่ย DO แตกต่างกัน

(เมื่อปฏิเสธ H_0 จะไม่ทราบว่าค่าเฉลี่ยตัวอย่างใดบ้างที่ไม่เท่ากัน จะต้องทำการทดสอบ

ต่อไปว่าค่าเฉลี่ยใดบ้างไม่เท่ากัน โดยการเปรียบเทียบเชิงซ้อน (multiple comparison) หรือ การทดสอบความแตกต่างของค่าเฉลี่ยแต่ละคู่)

การเตรียมข้อมูล ให้กรอกข้อมูลจำนวนซ้ำ (replication) ของแต่ละ station ให้อยู่ในแนวตั้ง ดังภาพ (กรณีที่ข้อมูลมีจำนวนมาก อาจเตรียมข้อมูลใน Excel แล้ว Copy มา Paste ในหน้าต่าง Data View ของ SPSS ได้)

พิจารณาว่าข้อมูลมีการกระจายแบบปกติหรือไม่ โดยพิจารณาค่าของ mean และ variance ของค่า DO แต่ละบ่อ

Descriptive Statistics

	N	Mean	Variance
station1	3	7.0000	.010
station2	2	6.8500	.005
station3	3	4.6000	.010
station4	2	4.5000	.020
Valid N (listwise)	2		

จากตาราง ค่า mean ของ DO ของทุกบ่อมีค่ามากกว่าค่า variance แสดงว่าข้อมูลมีการกระจายแบบปกติ

data_pond_ecology1.sav [DataSet0] - SPSS Data Editor

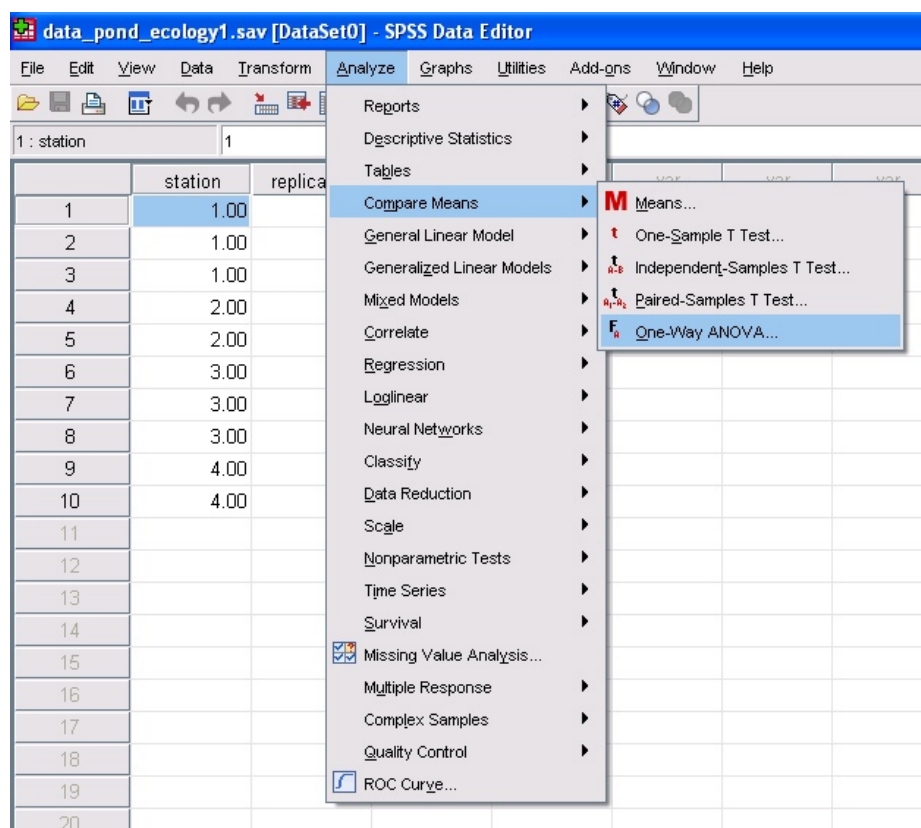
File Edit View Data Transform Analyze Graphs Utilities Add-ons Window

1 : station 1

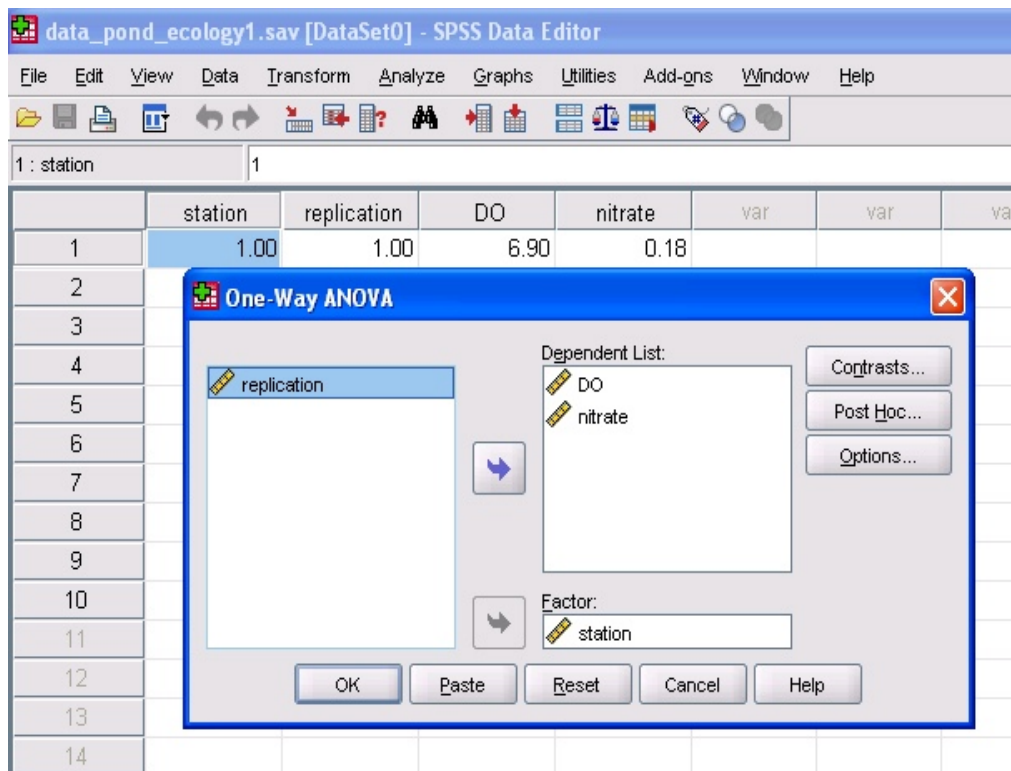
	station	replication	DO	nitrate	var
1	1.00	1.00	6.90	0.18	
2	1.00	2.00	7.10	0.17	
3	1.00	3.00	7.00	0.19	
4	2.00	1.00	6.90	0.23	
5	2.00	2.00	6.80	0.21	
6	3.00	1.00	4.50	0.35	
7	3.00	2.00	4.70	0.34	
8	3.00	3.00	4.60	0.36	
9	4.00	1.00	4.40	0.40	
10	4.00	2.00	4.60	0.45	
11					

Data View Variable View

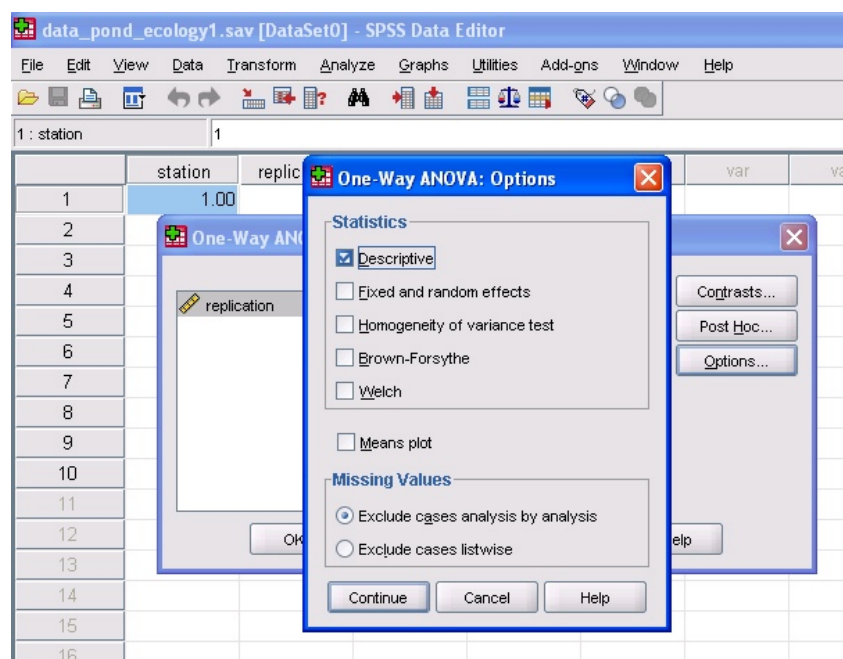
เมื่อได้ข้อมูลเรียบร้อยแล้ว ไปที่เมนู Analyze --> Compare Means --> One-Way ANOVA ดังภาพ



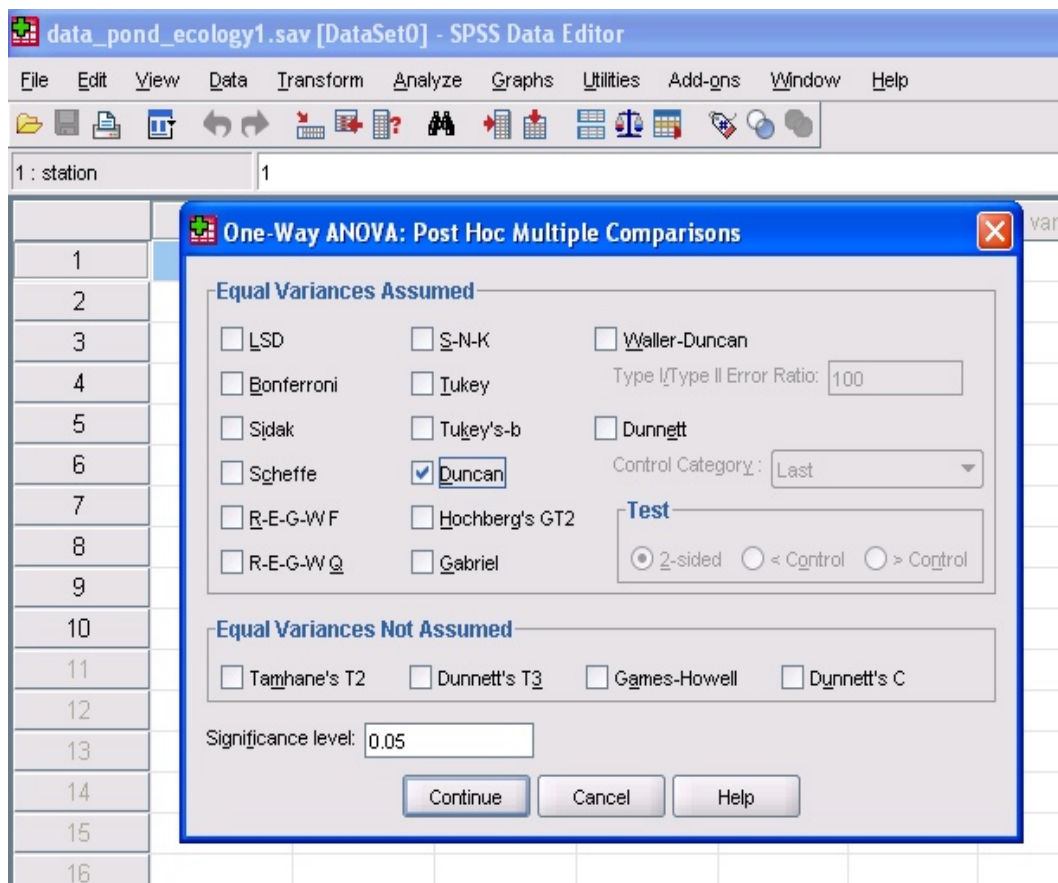
จะปรากฏหน้าต่างดังภาพ แล้วเลือกค่าตัวแปรที่ต้องการทดสอบ (ในที่นี้ คือ DO และ nitrate) ไปทางด้านขวาในช่อง Dependent List และเลือกกลุ่มที่ต้องการเปรียบเทียบ (ในที่นี้ คือ station (บ่อ)) ใส่ในช่อง Factor ดังภาพ



กด Options ถ้าต้องการแสดงค่าสถิติพื้นฐาน คลิก box ของ Descriptive แล้วกด Continue



หากผลการทดสอบทางสถิติที่ได้พบว่ามีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และถ้าต้องการเปรียบเทียบเชิงซ้อนว่าค่าเฉลี่ย station ไตแตกต่างกันบ้าง ให้ กด Post Hoc โดยเลือกวิธีการต่าง ๆ ได้ (ในที่นี้เลือก Duncan) แล้วกด Continue



เมื่อกด OK จะได้ Output ออกมาดังนี้

Oneway

Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
						Lower Bound	Upper Bound		
DO	1	3	7.0000	.10000	.05774	6.7516	7.2484	6.90	7.10
	2	2	6.8500	.07071	.05000	6.2147	7.4853	6.80	6.90
	3	3	4.6000	.10000	.05774	4.3516	4.8484	4.50	4.70
	4	2	4.5000	.14142	.10000	3.2294	5.7706	4.40	4.60
	Total	10	5.7500	1.25897	.39812	4.8494	6.6506	4.40	7.10
nitrate	1	3	.1800	.01000	.00577	.1552	.2048	.17	.19
	2	2	.2200	.01414	.01000	.0929	.3471	.21	.23
	3	3	.3500	.01000	.00577	.3252	.3748	.34	.36
	4	2	.4250	.03536	.02500	.1073	.7427	.40	.45
	Total	10	.2880	.10283	.03252	.2144	.3616	.17	.45

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
DO	Between Groups	14.200	3	4.733	436.923	.000
	Within Groups	.065	6	.011		
	Total	14.265	9			
nitrate	Between Groups	.093	3	.031	100.876	.000
	Within Groups	.002	6	.000		
	Total	.095	9			

การอ่านผลการทดสอบทางสถิติ

จากตาราง พบว่า ค่า P-value ของ DO และ nitrate เท่ากับ .000 ซึ่งมีค่าน้อยกว่าระดับความเชื่อมั่น .05 (95%) ปฏิเสธ H_0 แสดงว่าค่าเฉลี่ยของ DO และ nitrate ในแต่ละบ่อมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($F_{3,9} = 436.923$, $P = 0.000$; $F_{3,9} = 100.876$, $P = 0.000$) ตามลำดับ และมีความแตกต่างกันอย่างมีนัยสำคัญยิ่งทางสถิติ ($P < 0.01$)

Post Hoc Tests

Homogeneous Subsets

DO

Duncan

station	N	Subset for alpha = 0.05	
		1	2
4 a	2	4.5000	
3 a	3	4.6000	
2 b	2		6.8500
1 b	3		7.0000
Sig.		.333	.165

Means for groups in homogeneous subsets are displayed.

nitrate

Duncan

station	N	Subset for alpha = 0.05			
		1	2	3	4
1 a	3	.1800			
2 b	2				
3 c	3		.2200		
4 d	2				
Sig.		1.000	1.000	1.000	1.000

Means for groups in homogeneous subsets are displayed.

การอ่านผลการเปรียบเทียบเชิงซ้อน Post Hoc Tests

การรายงานผล: ผลการเปรียบเทียบเชิงซ้อน เรียงลำดับจากน้อยไปมาก และแบ่งออกเป็น 2 กลุ่ม ที่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติในแต่ละกลุ่ม

Subset 1: บ่อที่ 4 กับบ่อที่ 3 นั้นไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($P>0.05$)

Subset 2: บ่อที่ 2 กับบ่อที่ 1 นั้นไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติเช่นกัน ($P>0.05$)

การสรุปผลการทดลองนิยมใช้ตัวอักษรบนค่า mean $\pm$ S.D. ถ้าอักษรเหมือนกันแสดงว่าไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($P>0.05$)

5.2 การวิเคราะห์ความแปรปรวนแบบจำแนกสองทาง (Two-Way ANOVA)

เมื่อข้อมูลมีการจำแนกสองทาง การวางแผนการทดลองแบบสุ่มในบล็อก (RBD) การทดสอบใช้หลักการเดียวกับ One-Way Analysis of Variance คือใช้ error mean square เป็นเกณฑ์ในการทดสอบความแตกต่างระหว่าง among groups หรือ within groups

Source of variation	Sum of square	Degrees of freedom	Mean square (Variance)
Among Groups	S_3	$n-1$	$S_3^2 = S_3 / n-1$
Within Groups	S_2	$i-1$	$S_2^2 = S_2 / i-1$
Residual (error)	S_1	$(n-1)(i-1)$	$S_1^2 = S_1 / (n-1)(i-1)$
Total	S_T	$ni-1 = N-1$	

เปรียบเทียบระหว่างกลุ่ม $F = S_3^2 / S_1^2$

เปรียบเทียบระหว่าง sample $F = S_3^2 / S_1^2$

ตัวอย่าง จากการตรวจปริมาณ Ca ในเลือดปลานิลเพศเมียและเพศผู้ในระดับอายุแตกต่างกันได้ผลดังตารางที่ 6 อยากทราบว่าปริมาณ Ca ในเลือดปลานิลแตกต่างกันเนื่องจากเพศหรืออายุหรือไม่

ตารางที่ 6 ปริมาณ Ca ในเลือดปลานิลเพศเมียและเพศผู้ในระดับอายุแตกต่างกัน 30 วัน 60 วัน และ 75 วัน

เพศผู้ 30 วัน	เพศผู้ 60 วัน	เพศผู้ 75 วัน	เพศเมีย 30 วัน	เพศเมีย 60 วัน	เพศเมีย 75 วัน
16.87	19.07	32.45	15.86	17.20	30.54
16.18	18.77	28.71	14.92	17.64	32.41
17.12	17.63	34.65	15.63	17.86	28.97
16.83	16.99	28.29	15.24	16.78	28.46
17.19	18.04	24.46	14.80	16.92	29.65

1. พิจารณาว่าข้อมูลมีการกระจายแบบปกติหรือไม่

Descriptive Statistics

	N	Mean	Variance
Male30	5	16.8380	.159
Male60	5	18.1000	.712
Male75	5	29.7120	15.623
Female30	5	15.2900	.205
Female60	5	17.2860	.222
Female75	5	30.0060	2.414
Valid N (listwise)	5		

ผลแสดงว่าข้อมูลมีการกระจายแบบปกติ

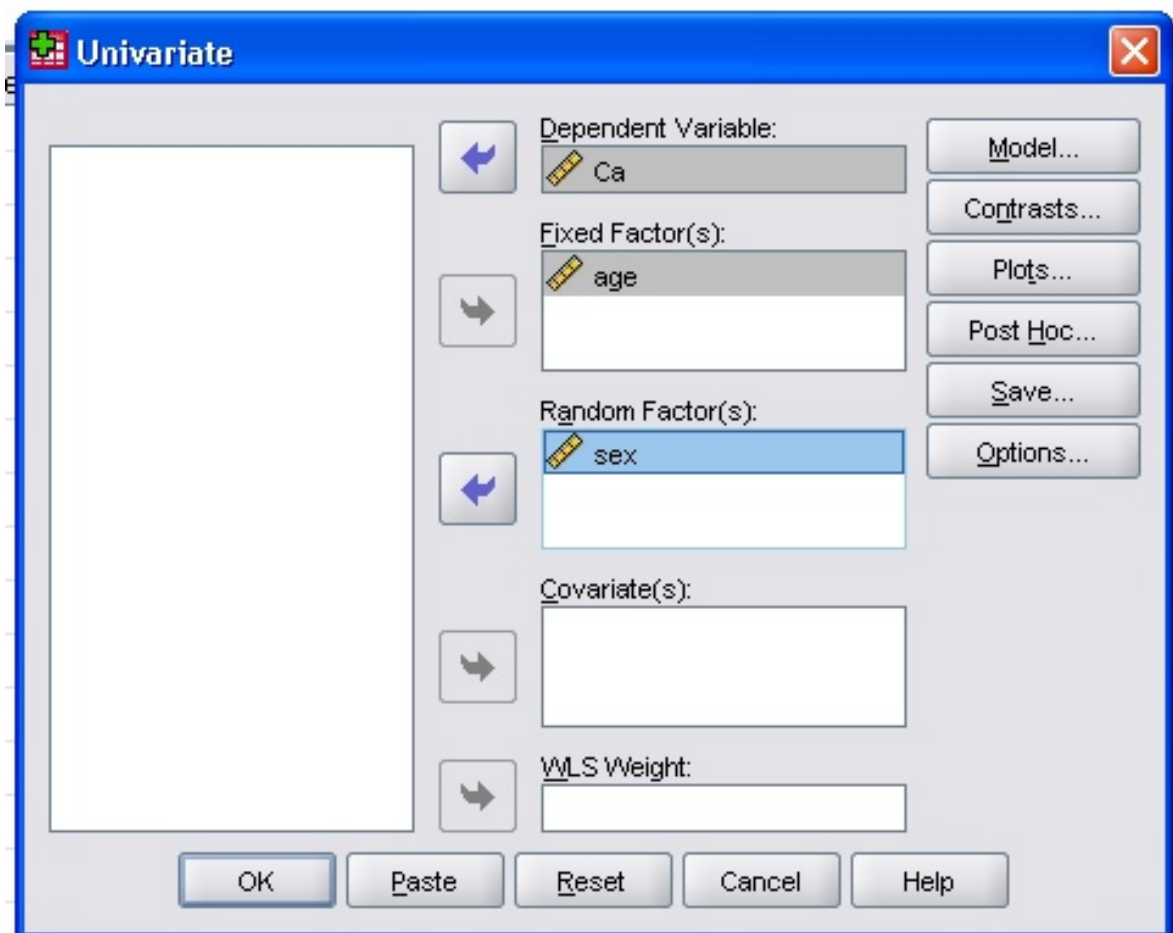
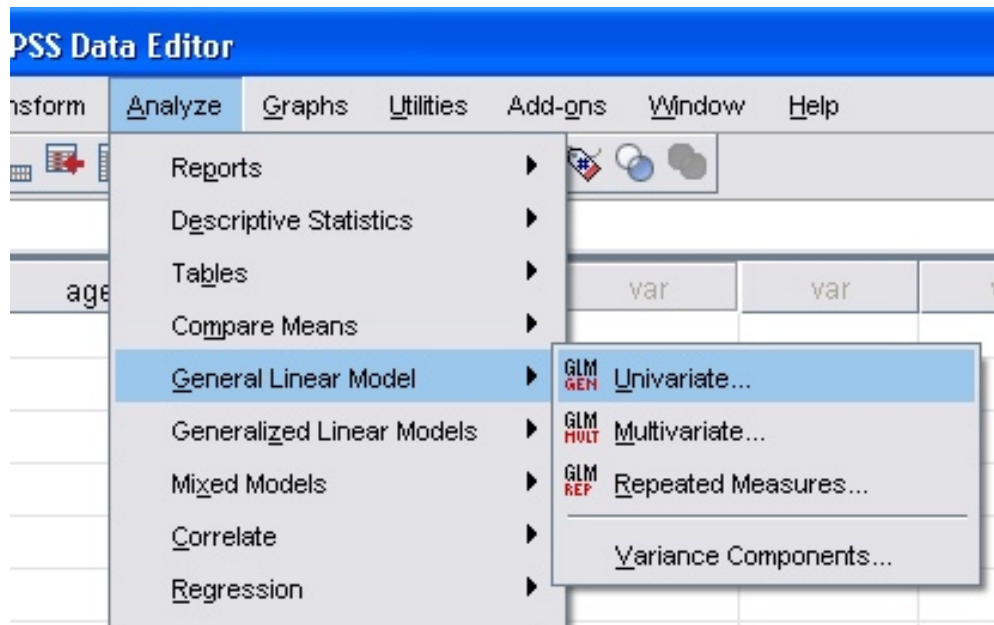
2. วิเคราะห์ Analysis of Variance โดยใช้โปรแกรม SPSS

ใส่ข้อมูลลงในตารางในที่นี่มี 2 เพศ ดังนั้นมี 2 categories (แทนเพศผู้ด้วย 1 และแทนเพศเมียด้วย 2) แต่ละเพศมี 3 ช่วงอายุ (แทนอายุ 30 วันด้วย 1 อายุ 60 วันด้วย 2 และอายุ 75 วันด้วย 3) column ที่ 1 เป็นเพศ (เรียงเพศผู้ = 1 และเพศเมีย = 2) column ที่ 2 เป็นอายุ (1 = 30 วัน 2 = 60 วัน 3 = 75 วัน) column ที่ 3 เป็นปริมาณ Ca ในเลือดหน่วยเป็นมิลลิกรัมต่อลิตร

	sex	age	Ca
1	1	1	16.87
2	1	1	16.18
3	1	1	17.12
4	1	1	16.83
5	1	1	17.19
6	1	2	19.07
7	1	2	18.77
8	1	2	17.63
9	1	2	16.99
10	1	2	18.04
11	1	3	32.45
12	1	3	28.71
13	1	3	34.65
14	1	3	28.29
15	1	3	24.46
16	2	1	15.86
17	2	1	14.92
18	2	1	15.63
19	2	1	15.24
20	2	1	14.80
21	2	2	17.20
22	2	2	17.64
23	2	2	17.89
24	2	2	16.78
25	2	2	16.92
26	2	3	30.54
27	2	3	32.41
28	2	3	28.97
29	2	3	28.46
30	2	3	29.65

3. เปรียบเทียบค่าเฉลี่ยของข้อมูล

จากเมนู Analyze เลือกคำสั่ง General Linear Model แล้วใช้คำสั่งย่อย Univariate หน้าจอจะขึ้น หน้าต่าง Univariate เลือก Ca มาไว้ในช่อง Dependent Variable และเลือก age และ sex มาไว้ในช่อง Fixed Factors และ Random Factors ตามลำดับ แล้วกด OK



จะได้ผลลัพธ์ดังนี้

Univariate Analysis of Variance

Between-Subjects Factors

		N
age	1	10
	2	10
	3	10
sex	1	15
	2	15

Tests of Between-Subjects Effects

Dependent Variable:Ca

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	13489.985	1	13489.985	3.785E3	.010
	Error	3.564	1	3.564 ^a		
age	Hypothesis	1136.557	2	568.279	264.347	.004
	Error	4.299	2	2.150 ^b		
sex	Hypothesis	3.564	1	3.564	1.658	.327
	Error	4.299	2	2.150 ^b		
age * sex	Hypothesis	4.299	2	2.150	.667	.522
	Error	77.336	24	3.222 ^c		

a. MS(sex)

b. MS(age * sex)

c. MS(Error)

พิจารณาที่ Sig. พบว่า sex และ intercept sex*age ค่ามากกว่า 0.05 แสดงว่าปริมาณ Ca ในเลือดปลาทั้งสองเพศไม่แตกต่างกัน และไม่มี interactions ระหว่างเพศกับอายุของปลา แต่ปริมาณ Ca ในเลือดปลาอายุต่างกันแตกต่างกันอย่างมีนัยสำคัญ ($P < 0.05$)

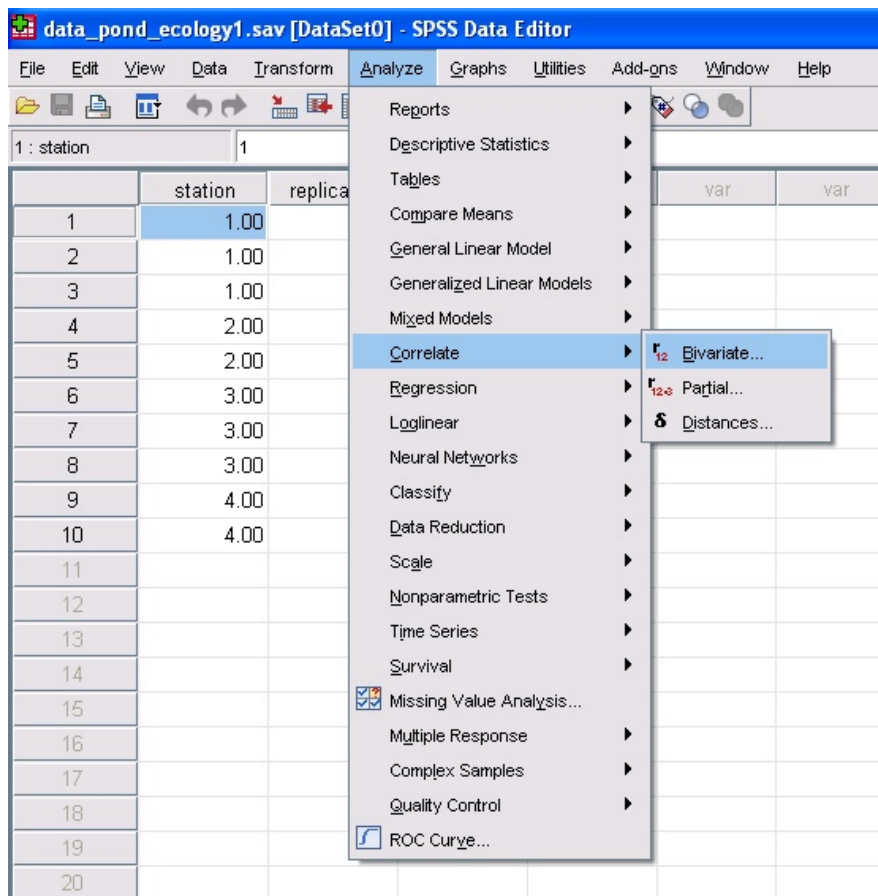
การรายงานผล: ปริมาณ Ca ในเลือดปลาทั้งสองเพศไม่แตกต่างกัน ($F_{1,29} = 1.658$, $P = 0.327$) และไม่มี interactions ระหว่างเพศกับอายุของปลา ($F_{2,29} = 0.667$, $P = 0.522$) แต่ปริมาณ Ca ในเลือดปลาอายุต่างกันแตกต่างกันอย่างมีนัยสำคัญ ($F_{2,29} = 264.347$, $P = 0.004$)

ในกรณีมีปัจจัยเดียว วางแผนการทดลองแบบ RBD (ใช้ Simple Factorial) จะไม่มี interaction เกิดขึ้น หากการทดสอบได้ผลว่ามี interaction แสดงว่าการวางแผนการทดลองไม่เหมาะสม คือ ไม่จำเป็นต้องทำบล็อก ควรเปลี่ยนแผนการทดลองเป็น CRD

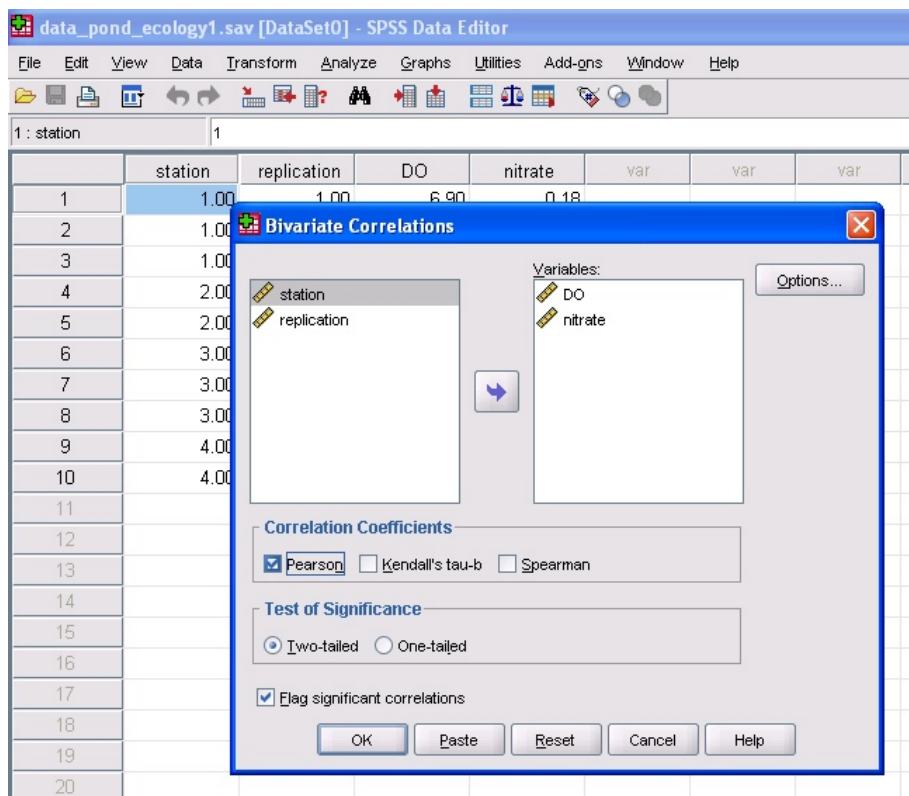
หากมีมากกว่า 1 ปัจจัย (factorial experiments) ใช้ General Factorial กำหนดปัจจัยหนึ่งเป็นปัจจัยกำหนด (fixed factor) และอีกปัจจัยหนึ่งเป็นปัจจัยสุ่ม (random factor)

6. การวิเคราะห์สหสัมพันธ์ (correlation)

การหาความสัมพันธ์ระหว่างตัวแปร ตัวอย่าง ต้องการหาความสัมพันธ์ระหว่าง DO กับ nitrate ไปที่เมนู Analyze --> Correlate --> Bivariate ดังภาพ



จะขึ้นหน้าต่าง Bivariate Correlation จากนั้นเลือกตัวแปรที่ต้องการหาความสัมพันธ์ไปทางขวา Variables (กรณีที่มีตัวแปรจำนวนมากสามารถเลือกใส่ในช่องได้หลายตัวแปร ผลการวิเคราะห์ที่ออกมาโปรแกรมจะหาความสัมพันธ์ในแต่ละตัวแปรเป็นคู่ๆ ให้) คลิก box ค่าสัมประสิทธิ์สหสัมพันธ์ (coefficient of correlation, r) ที่ต้องการ (ในที่นี้เลือก Pearson) แล้วกด OK



ได้ผลการวิเคราะห์ดังนี้

Correlations

Correlations

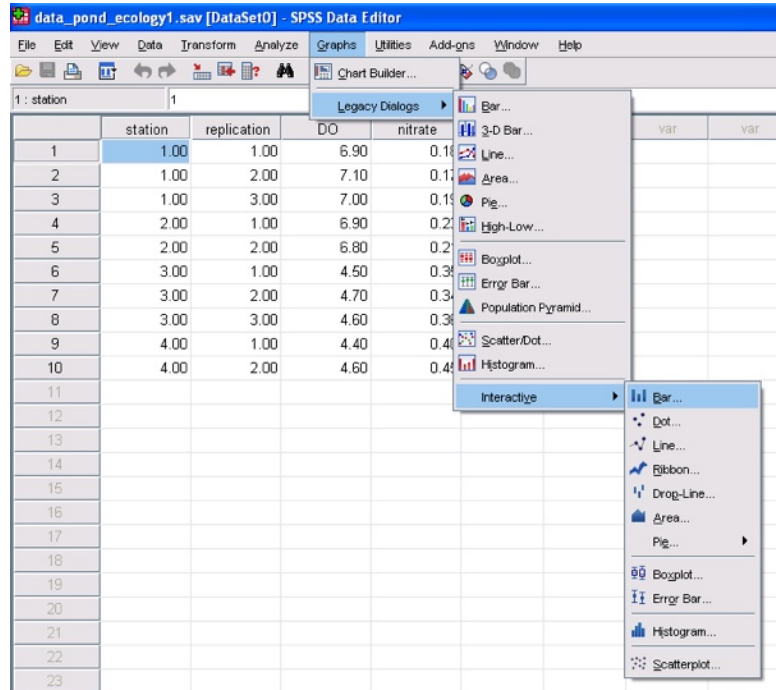
		DO	nitrate
DO	Pearson Correlation	1	-.950**
	Sig. (2-tailed)		.000
	N	10	10
nitrate	Pearson Correlation	-.950**	1
	Sig. (2-tailed)	.000	
	N	10	10

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

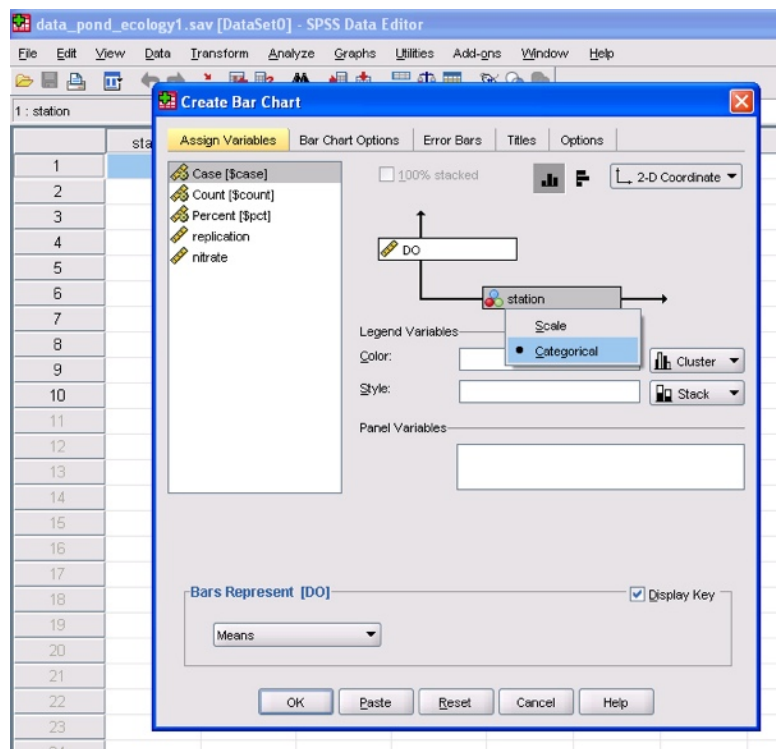
ผลการวิเคราะห์ความสัมพันธ์ พบว่า DO และ nitrate มีความสัมพันธ์กันเชิงลบอย่างมีนัยสำคัญทางสถิติที่ระดับความเชื่อมั่น .01 โดยมีค่าสัมประสิทธิ์สหสัมพันธ์ (r) = -0.950

6. การสร้างกราฟอย่างง่าย

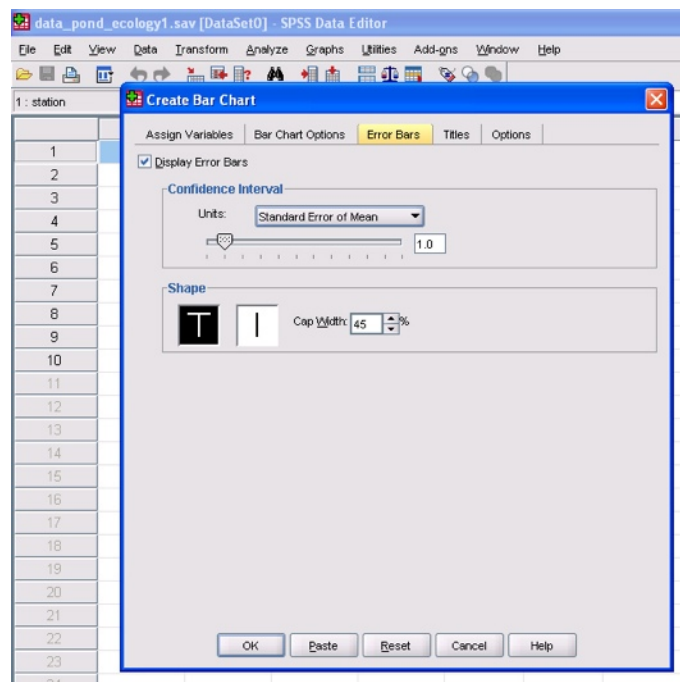
การแสดงข้อมูลในรูปภาพ (graphical method) เป็นการแสดงค่าสถิติเชิงพรรณนา อย่างหนึ่ง ทำให้เข้าใจได้ง่าย การสร้างกราฟใน SPSS อย่างง่าย ทำได้โดย ตัวอย่าง ต้องการทำกราฟแท่ง (bar) ของค่าเฉลี่ย DO แต่ละ station โดยมี error bar ด้วย ไปที่เมนู Graphs --> Legacy Dialogs --> Interactive --> Bar ดังภาพ



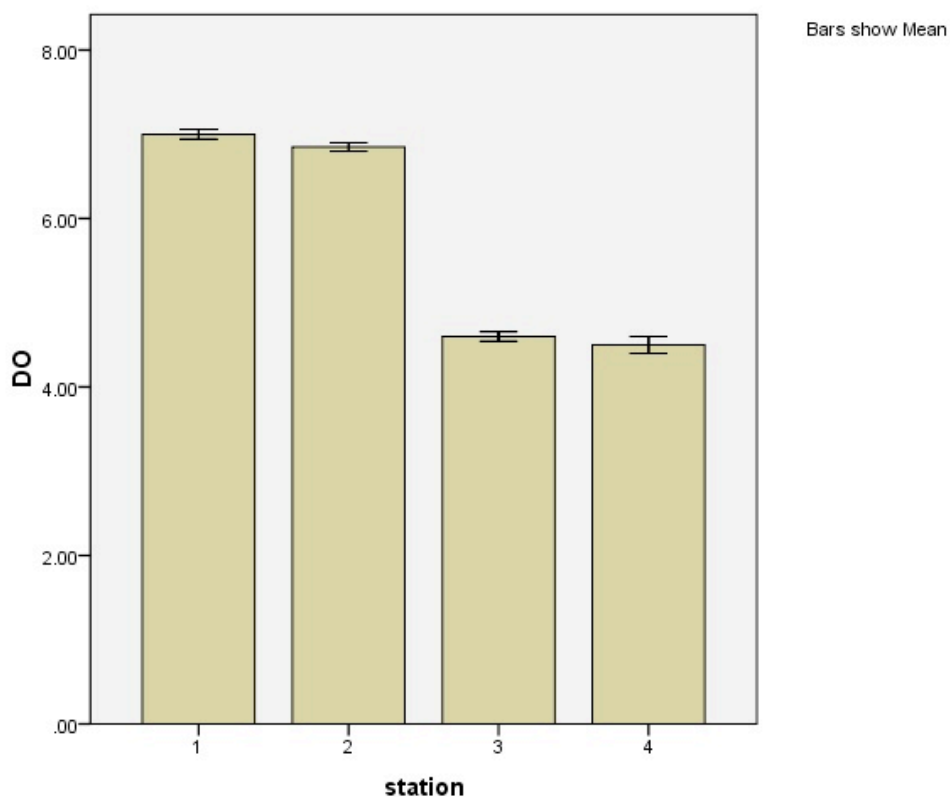
ทำการลาก DO ไปใส่ช่องในแกนตั้ง (y) และ station ในแกนนอน (x) คลิกขวาเลือกการวัดข้อมูล station เป็น Categorical



ถ้าต้องการใส่ error bar ไปที่หน้าต่าง Error Bars เลือกการกระจายของข้อมูลที่ต้องการวัด โดยค่า S.E. mean เป็นค่าที่นิยมใช้กันมาก



ได้ผลกราฟดังนี้ และสามารถแสดงการเปรียบเทียบเชิงซ้อนได้ด้วย



7. การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (simple linear regression analysis)

เป็นการศึกษาถึงความสัมพันธ์ระหว่างตัวแปรเชิงปริมาณ 2 ตัว โดยที่ต้องการทราบค่าของตัวแปรตัวหนึ่งหรือต้องกำหนดค่าตัวแปรตัวหนึ่งไว้ล่วงหน้า ซึ่งตัวแปรทั้งสองมีความสัมพันธ์อยู่ในรูปเชิงเส้น ซึ่งสามารถแสดงความสัมพันธ์ในรูปสมการเชิงเส้นดังนี้

$$Y_i = \beta_0 + \beta_1 X_i + e_i ; i = 1, 2, \dots, N$$

หรือ $Y_i = a + bX_i$ โดยที่ $\beta_0 = a, \beta_1 = b$

เมื่อ Y = ตัวแปรตาม (dependent variable) เนื่องจากค่า Y ขึ้นอยู่กับค่า X
 X = ตัวแปรอิสระ (independent variable) หรือ ตัวแปรต้นเหตุ (predictor variable)

β_0 = ส่วนตัดแกน Y หรือ คือค่าของ Y เมื่อ X มีค่าเป็นศูนย์

β_1 = ความชัน (slope)

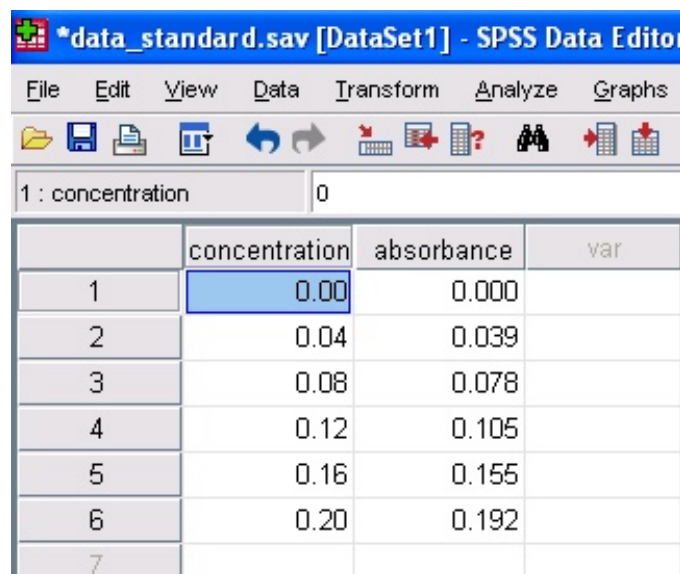
e = ความคลาดเคลื่อนอย่างสุ่ม (random error)

สมมติฐาน $H_0 : \beta_1 = 0$ หรือ $H_0 : Y$ ไม่มีความสัมพันธ์กับ X ในรูปเชิงเส้น

$H_1 : \beta_1 \neq 0$ $H_1 : Y$ มีความสัมพันธ์กับ X ในรูปเชิงเส้น

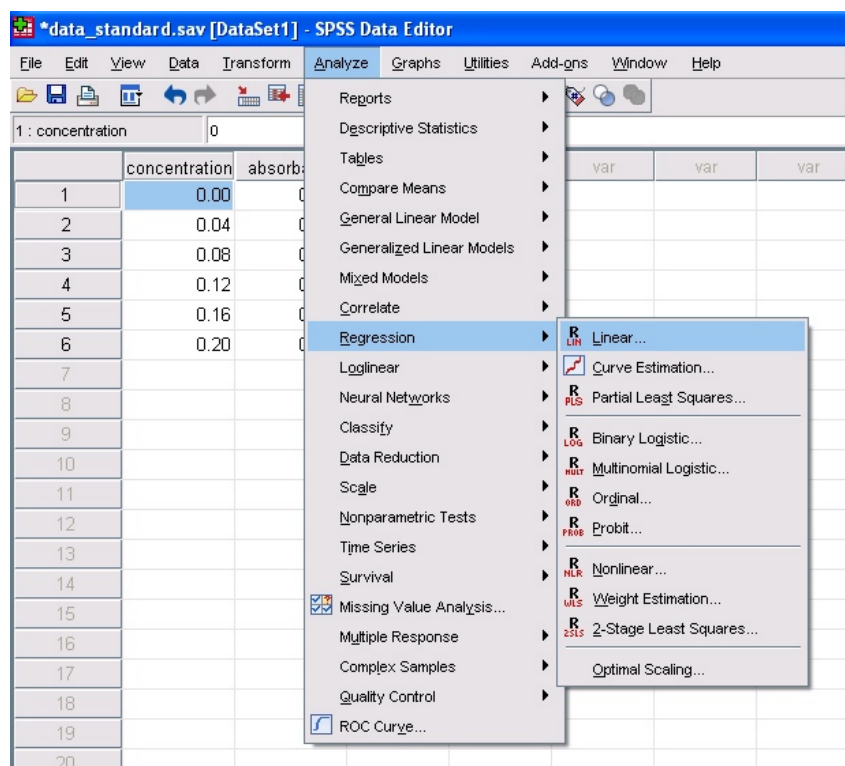
ค่าสัมประสิทธิ์การกำหนด (coefficient of determination, R square, r^2) สัมประสิทธิ์การกำหนด หมายถึง สัดส่วนที่ตัวแปร X สามารถอธิบายการเปลี่ยนแปลงของตัวแปร Y ได้ ดังนั้นถ้า r^2 มากแสดงว่า Y และ X มีความสัมพันธ์กันมาก

ตัวอย่าง การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย ระหว่างค่า concentration กับ absorbance ของแอมโมเนีย เตรียมข้อมูลดังนี้

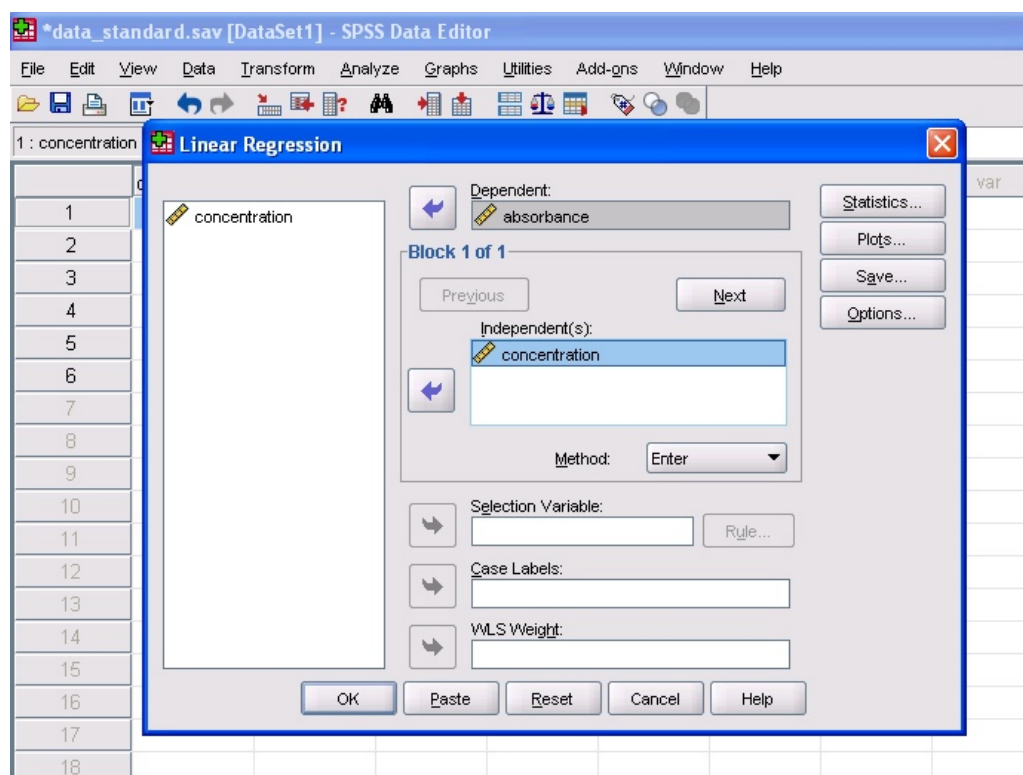


	concentration	absorbance	var
1	0.00	0.000	
2	0.04	0.039	
3	0.08	0.078	
4	0.12	0.105	
5	0.16	0.155	
6	0.20	0.192	
7			

วิเคราะห์การถดถอยเชิงเส้น ไปที่เมนู Analyze --> Regression --> Linear ดังภาพ



จะขึ้นหน้าต่าง Linear Regression จากนั้นเลือกตัวแปรอิสระ (X) ไปช่อง Independent(s) และตัวแปรตาม (Y) ไปช่อง Dependent แล้วกด OK



ผลการวิเคราะห์ปรากฏดังนี้

Regression

Variables Entered/Removed^b

Model	Variables Entered	Variables Removed	Method
1	concentration ^a		Enter

a. All requested variables entered.

b. Dependent Variable: absorbance

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.998 ^a	.996	.995	.004962

a. Predictors: (Constant), concentration

ANOVA^b

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.025	1	.025	1.034E3	.000 ^a
	Residual	.000	4	.000		
	Total	.026	5			

a. Predictors: (Constant), concentration

b. Dependent Variable: absorbance

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	.004		-.146	.891
	concentration	.954	.030	.998	32.159	.000

a. Dependent Variable: absorbance

Coefficients จะแสดงค่าประมาณสัมประสิทธิ์ความถดถอย ดังนี้

$$a (\beta_0) = .000$$

$$b (\beta_1) = 0.954$$

ผลการทดสอบส่วนการตัดแกน Y

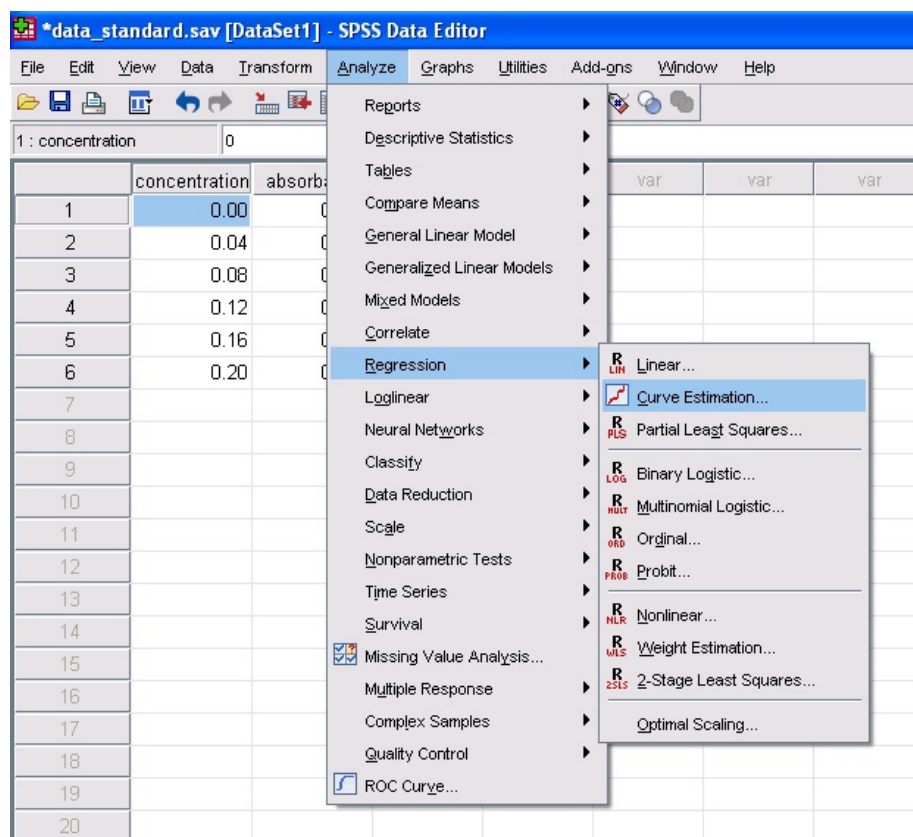
$$H_0 : \beta_0 = 0 \text{ ยอมรับ } H_0 \text{ (} t = -.146 \text{ และค่า Sig. ของ } t = .891 \text{)}$$

ผลการทดสอบความสัมพันธ์กันในรูปเชิงเส้น

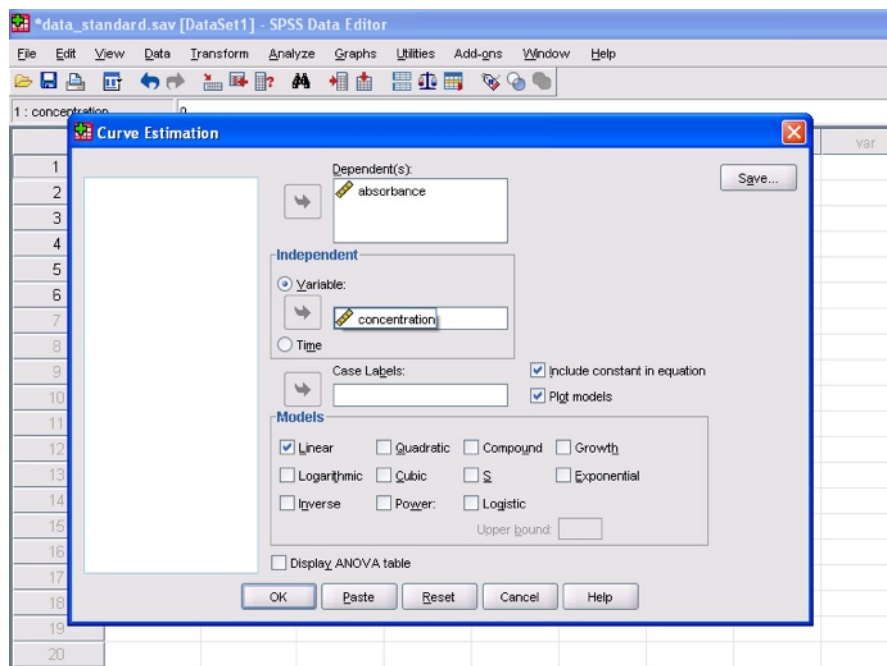
$$H_1 : \beta_1 \neq 0 \text{ ปฏิเสธ } H_0 \text{ (ค่า } t = 32.159 \text{ และค่า Sig. ของ } t = .000 \text{)}$$

แสดงว่า absorbance และ concentration มีความสัมพันธ์ในรูปเชิงเส้น

ถ้าต้องการทำกราฟ ไปที่เมนู Analyze --> Regression --> Curve Estimation ดังภาพ



จะขึ้นหน้าต่าง Curve Estimation จากนั้นเลือกตัวแปรอิสระ (x) ไปช่อง Independent(s) และตัวแปรตาม (y) ไปช่อง Dependent คลิก box Linear แล้วกด OK



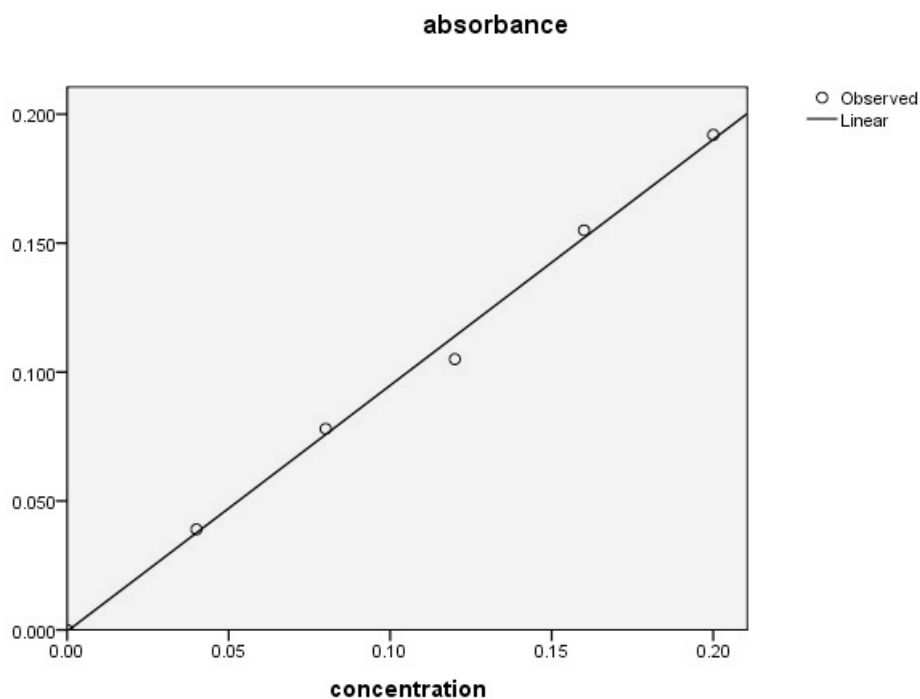
ได้ผลปรากฏดังนี้

Model Summary and Parameter Estimates

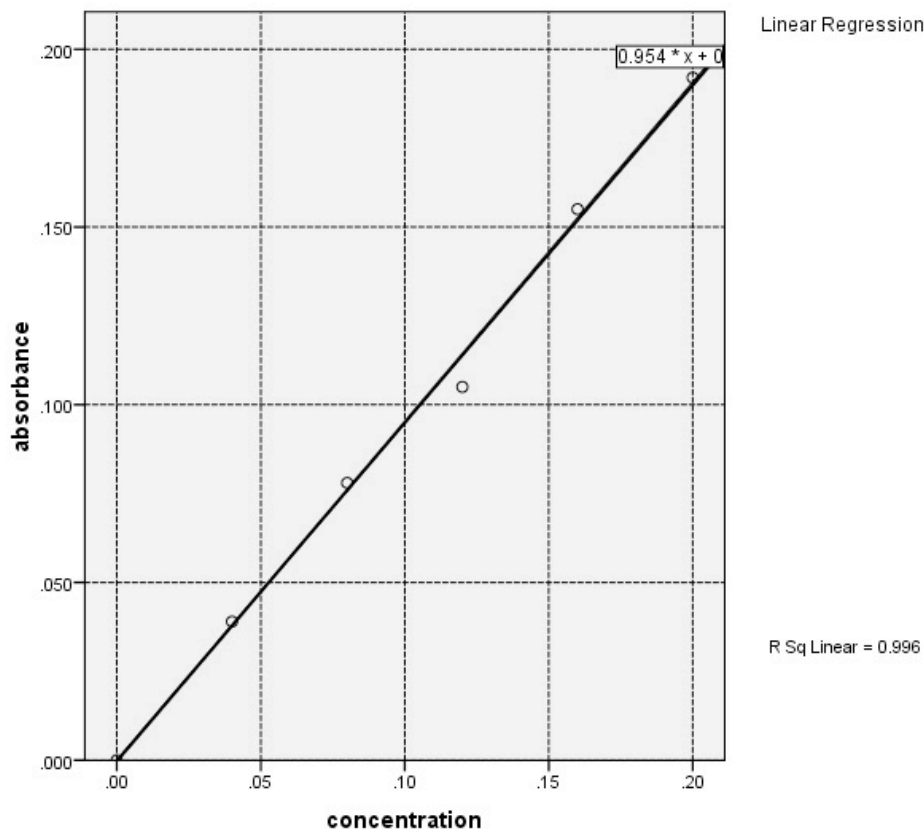
Dependent Variable:absorbance

Equation	Model Summary					Parameter Estimates	
	R Square	F	df1	df2	Sig.	Constant	b1
Linear	.996	1.034E3	1	4	.000	.000	.954

The independent variable is concentration.



กรณีทำกราฟ scatter plot แบบ linear regression ได้ผลดังนี้



บรรณานุกรม

นฤมล แสงประดับ. มมป. เอกสารประกอบการสอนวิชา 311 405 วิทยาระเบียบวิธีวิจัย. ภาควิชาชีววิทยา คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น, 29 หน้า.

บุญเสฐียร บุญสูง. 2550. เอกสารประกอบการสอนวิชา 01424381 นิเวศวิทยา เรื่อง การวิเคราะห์ข้อมูลทางนิเวศวิทยาเบื้องต้นโดยใช้โปรแกรมสำเร็จรูป SPSS for Windows. ภาควิชาสัตววิทยา คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์, 19 หน้า.

SPSS, 1933. **SPSS for Windows Base System User's Guide Release 6.0.** SPSS Inc., Chicago, USA.

Glover, T. and Mitchell, K. 2002. **An Introduction to Biostatistics.** McGraw-Hill Companies, Inc., New York, USA.