

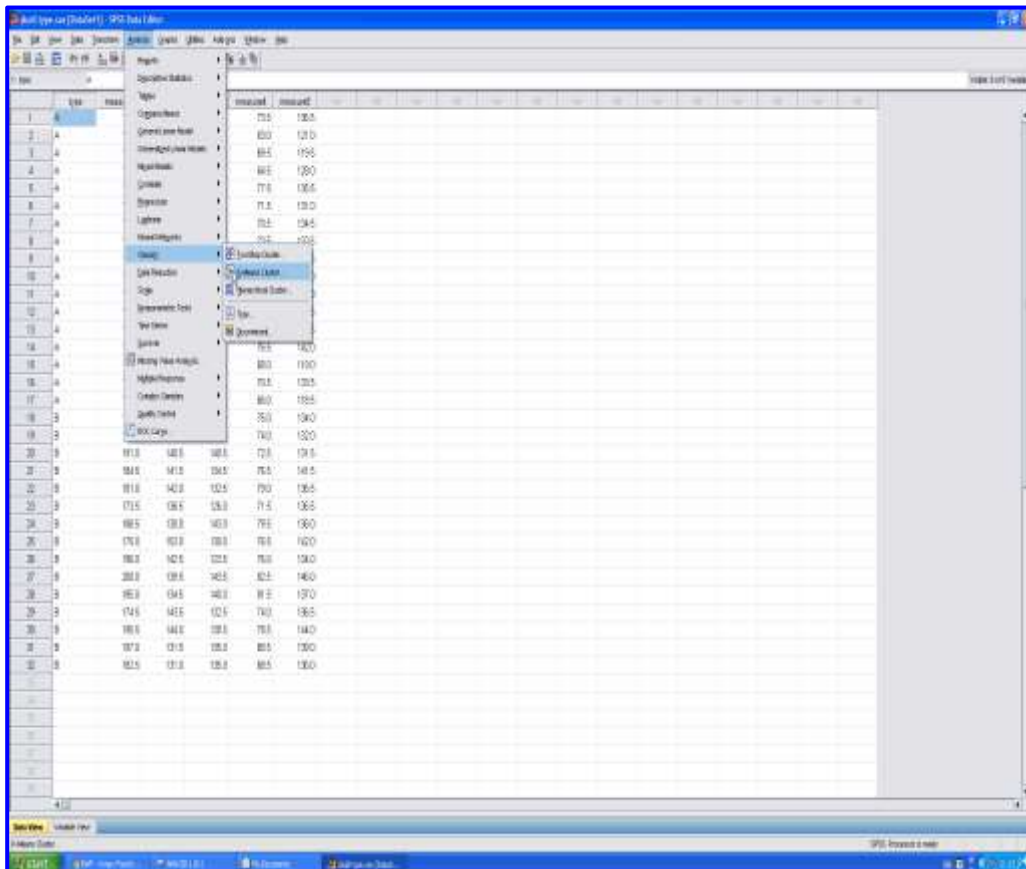


## Cluster Analysis Workshop#2-K-Means Clustering Method

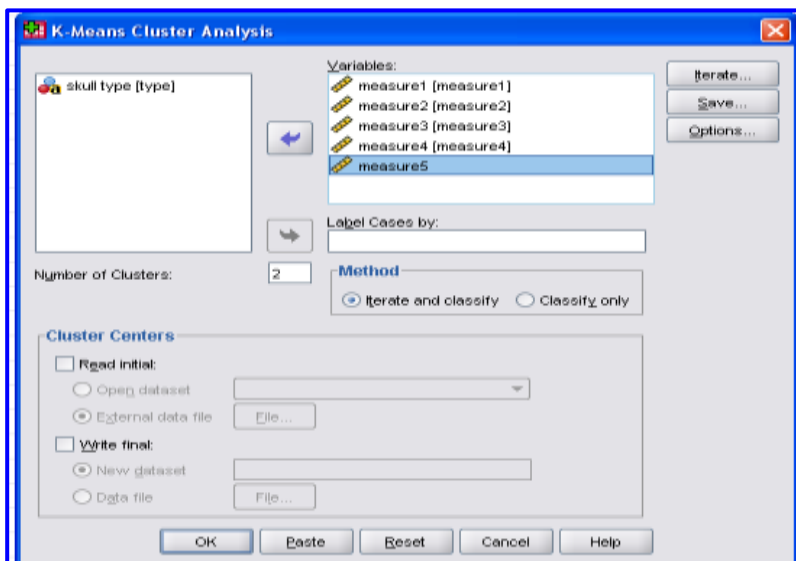
ภาพด้านล่างแสดงให้เห็นข้อมูลเกี่ยวกับกะโหลกศีรษะสมัยโบราณซึ่งเก็บได้จากสถานที่สองแห่ง และนักวิจัยได้กำหนดมาตรวัด 5 อย่างเพื่อใช้บ่งบอกลักษณะของกะโหลกศีรษะแต่ละชิ้น

	type	measure1	measure2	measure3	measure4	measure5
1	A	190.5	152.5	145.0	73.5	136.5
2	A	172.5	132.0	125.5	63.0	121.0
3	A	167.0	130.0	125.5	69.5	119.5
4	A	169.5	150.5	133.5	64.5	128.0
5	A	175.0	138.5	126.0	77.5	135.5
6	A	177.5	142.5	142.5	71.5	131.0
7	A	179.5	142.5	127.5	70.5	134.5
8	A	179.5	138.0	133.5	73.5	132.5
9	A	173.5	135.5	130.5	70.0	133.5
10	A	162.5	139.0	131.0	62.0	126.0
11	A	178.5	135.0	136.0	71.0	124.0
12	A	171.5	148.5	132.5	65.0	146.5
13	A	180.5	139.0	132.0	74.5	134.5
14	A	183.0	149.0	121.5	76.5	142.0
15	A	169.5	130.0	131.0	68.0	119.0
16	A	172.0	140.0	136.0	70.5	133.5
17	A	170.0	126.5	134.5	66.0	118.5
18	B	182.5	136.0	138.5	76.0	134.0
19	B	179.5	135.0	128.5	74.0	132.0
20	B	191.0	140.5	140.5	72.5	131.5
21	B	184.5	141.5	134.5	76.5	141.5
22	B	181.0	142.0	132.5	79.0	136.5
23	B	173.5	136.5	126.0	71.5	136.5
24	B	188.5	130.0	143.0	79.5	136.0
25	B	175.0	153.0	130.0	76.5	142.0
26	B	196.0	142.5	123.5	76.0	134.0
27	B	200.0	139.5	143.5	82.5	146.0
28	B	185.0	134.5	140.0	81.5	137.0
29	B	174.5	143.5	132.5	74.0	136.5
30	B	195.5	144.0	138.5	78.5	144.0
31	B	197.0	131.5	135.0	80.5	139.0
32	B	182.5	131.0	135.0	68.5	136.0

- คลิก Analyze>Classify>K-means clustering

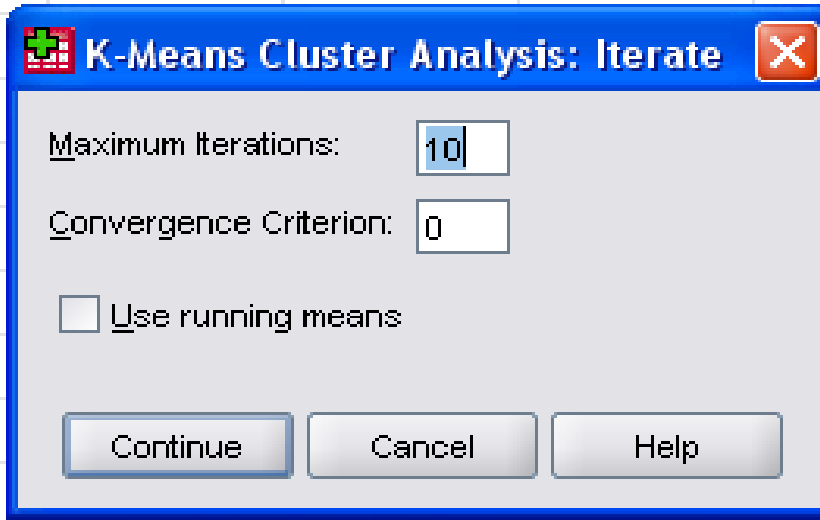


จะปรากฏ menu ดังต่อไปนี้



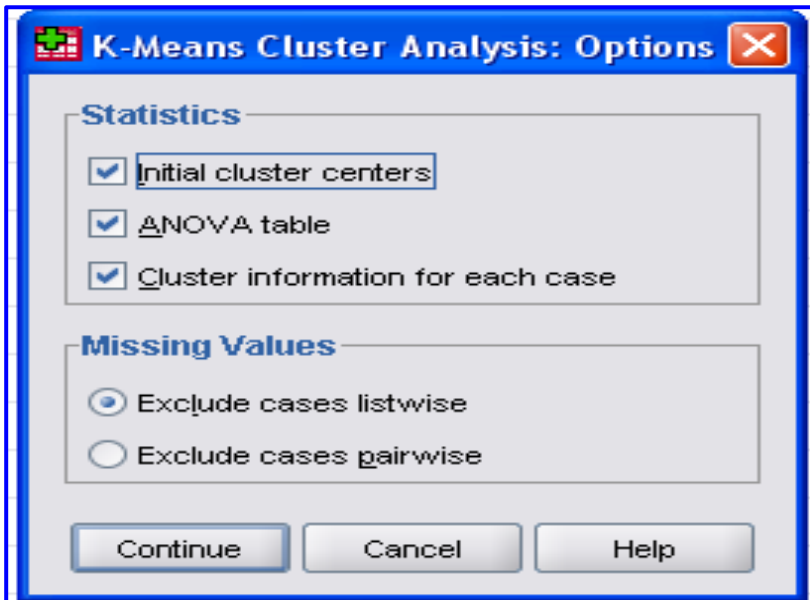
- คลิกย้ายตัวแปรทั้งหมดไปอยู่ในกล่อง Variables
- เลือกจำนวนcluster เป้าหมายที่ต้องการจัดกลุ่ม ในที่นี้ระบุ 2
- เลือกวิธีการซึ่งมีให้เลือกสองอย่างคือ
  - iterate and classify:ดำเนินการรวมกลุ่มและแยกแยะกลุ่มให้เห็น
  - classify:แยกแยะกลุ่มให้เห็น
 ในที่นี้เลือก iterate and classify

- คลิกปุ่ม Iterate จะปรากฏ menu



จำนวนรอบของiteration ที่ตั้งเป็น default=10 และ convergence criterion (เกณฑ์ตั้งค่าของการเปลี่ยนแปลงในการ iterate แต่ละครั้ง) ที่ตั้งเป็นค่า default=0 ในที่นี้เราใช้ค่า default ทั้งสอง คลิก Continue

- คลิก ปุ่ม Options จะปรากฏ menu



คลิกค่าสถิติที่ต้องการ

- Initial cluster centers หาค่าเฉลี่ยของ cluster เริ่มแรก
- ANOVA table เพื่อพิจารณาค่า Mean squared error
- Cluster information for each case เพื่อเห็นข้อมูลของแต่ละCluster ว่าประกอบด้วย case ไตบ้าง

- คลิกปุ่ม Save ซึ่งจะปรากฏ menu



- คลิกเพื่อให้เห็นข้อมูลเกี่ยวกับ
- การเป็นสมาชิกในแต่ละ cluster (cluster membership)
  - ระยะห่างของ case จาก cluster center (Distance from cluster center)

ผลจากcomputer output แสดงได้ดังนี้

Initial Cluster Centers

	Cluster		
	1	2	3
measure1	200.0	171.5	170.0
measure2	139.5	148.5	126.5
measure3	143.5	132.5	134.5
measure4	82.5	65.0	66.0
measure5	146.0	146.5	118.5

ตารางด้านบนแสดงให้เห็นจุดศูนย์กลางของ cluster เริ่มแรก จะสังเกตได้ว่า ตัวแปรที่ 4 (measure4) มีค่าอยู่ใกล้ cluster center มากกว่าตัวแปรอื่น ๆ

ตารางด้านล่างแสดงผลจากการ iterate ซึ่งแสดงให้เห็นว่าหลังจาก iteration สองรอบ cluster ที่หนึ่งไม่มีการเปลี่ยนแปลงใน cluster center ส่วน Cluster 2 และ 3 การเปลี่ยนแปลงใน cluster center สิ้นสุดลงหลัง iteration รอบที่สาม

**Iteration History<sup>a</sup>**

Iteration	Change in Cluster Centers		
	1	2	3
1	13.596	14.884	10.607
2	.967	1.498	3.505
3	.000	.508	1.936
4	.000	.000	.000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is .000. The current iteration is 4. The minimum distance between initial centers is 35.711.

**Cluster Membership**

Case Number	Cluster	Distance
1	1	15.551
2	3	6.712
3	3	7.145
4	2	16.052
5	2	8.156
6	2	12.277
7	2	5.243
8	2	6.082
9	2	7.613
10	3	12.125
11	3	11.693
12	2	15.741
13	2	5.243
14	2	15.715
15	3	3.495
16	2	7.447
17	3	7.408
18	1	10.112
19	2	8.643
20	1	8.689
21	1	8.700
22	2	8.065
23	2	8.058
24	1	11.052
25	2	13.873
26	1	16.421
27	1	14.003
28	1	8.817
29	2	3.694
30	1	8.927
31	1	10.707
32	2	13.188

ตารางด้านขวาแสดงให้เห็นว่าแต่ละ case สังกัดอยู่ที่ cluster ไต และมีระยะห่างจากศูนย์กลาง ของ cluster มากน้อยเท่าใด

### Final Cluster Centers

	Cluster		
	1	2	3
measure1	191.0	176.7	170.0
measure2	139.2	141.6	132.1
measure3	138.2	131.2	130.6
measure4	77.7	72.3	66.6
measure5	138.0	135.7	121.3

ตาราง Final Cluster Centers แสดงให้เห็นค่าของตัวแปรทั้งห้าจากศูนย์กลางของ cluster ที่เป็น Solution ในขั้นสุดท้าย

### Distances between Final Cluster Centers

Cluster	1	2	3
1		17.113	30.857
2	17.113		19.356
3	30.857	19.356	

Distances between Final Cluster Centers แสดงให้เห็นระยะห่างระหว่าง cluster

### ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
measure1	996.583	2	25.024	29	39.824	.000
measure2	196.302	2	36.492	29	5.379	.010
measure3	175.811	2	27.424	29	6.411	.005
measure4	237.354	2	14.695	29	16.152	.000
measure5	580.717	2	19.190	29	30.262	.000

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

ข้อมูลใน ตาราง ANOVA ชี้ให้เห็นว่าตัวแปรทั้งห้ามีประสิทธิภาพในการแยกแยะกลุ่มได้พอสมควร โดยตัวแปรที่ 4 และ 5 สามารถแยกแยะกลุ่มได้ดีกว่าตัวแปรอื่น ๆ ที่เหลืออีก 3 ตัวแปร

### Number of Cases in each

Cluster	
Cluster 1	10.000
2	16.000
3	6.000
Valid	32.000
Missing	.000

ข้อมูลจากตารางด้านซ้ายมือชี้ให้เห็นจำนวนของ case ที่มีอยู่ในแต่ละ cluster โดย cluster ที่สองจะประกอบด้วย case มากที่สุด ส่วน cluster ที่สาม จะมี case น้อยที่สุด

ข้อสังเกต:- ใน k-means clustering หากผู้วิเคราะห์กำหนด initial clusters ที่แตกต่างออกไป หรือเปลี่ยนแปลงจำนวน cluster ที่ต้องการ หรือมีการสลับ case ในไฟล์ข้อมูล ผลของการจัดกลุ่มจะแตกต่างกันออกไป

Contribution this issue: ดร . ดนัย ปัตตพงศ์

งานวิจัยที่อ้างอิงบทความนี้

สุชาวดี ปลั่งศรี “ การจัดกลุ่มขนาดผลิตภัณฑ์โดยใช้ K-means Clustering เพื่อลดต้นทุนบรรจุภัณฑ์ “. วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ สาขาบริหารธุรกิจ ภาควิชา วิศวกรรมอุตสาหการและการจัดการ มหาวิทยาลัยศิลปากร .2562

อยากเรียนรู้การนำสถิติข้างต้นนี้ไปใช้ในการวิจัยระดับสารนิพนธ์ (independent study) วิทยานิพนธ์ (thesis) ดุษฎีนิพนธ์(dissertation) ปรึกษาได้ที่ [dpattaphongse@gmail.com](mailto:dpattaphongse@gmail.com)

- \* ผู้แต่ง MBA's Made Easy (160+ issues) เอกสารวิชาการด้านศาสตร์การบริหารธุรกิจที่ช่วยให้ธุรกิจสามารถยืนหยัดและอยู่รอดได้ในภาวะที่โลกเปลี่ยนแปลงอยู่ตลอดเวลา
- \* ผู้พัฒนา FINALYSIS... a dedicated software สำหรับให้บริการนักธุรกิจที่ต้องการวิเคราะห์ความเป็นไปได้ทางการเงินของโครงการพัฒนาอสังหาริมทรัพย์ (บ้านจัดสรร/จัดสรรที่ดินเพื่อการอุตสาหกรรม/อาคารชุด/อาคารสำนักงานให้เช่า) โรงแรม โรงพยาบาลเอกชน ห้างสรรพสินค้า โรงงานน้ำตาล โรงงานกระดาษ โรงไฟฟ้าชีวมวล ฯลฯ ได้เห็นตัวเลขก่อนโครงการเกิด หลีกเลี่ยงความผิดพลาดเป็นร้อยเป็นพันล้านหากเกิดการลงทุนจริง(กำหนด DEBUT 1 เมษายน 2569)
- \* ผู้แต่งหนังสือ”การวิเคราะห์ความเป็นไปได้ทางการเงินและการจัดวงเงินเครดิตของโครงการลงทุน”ประกอบด้วยตัวอย่างของธุรกิจจริงที่ไม่เปิดเผยชื่อนับ 100 บริษัท ครอบคลุมอุตสาหกรรม 24 อุตสาหกรรม
- \* Co-developer ซอฟต์แวร์ en@gex@cel<sup>®</sup> สำหรับใช้ทดสอบ/เรียนรู้ศัพท์(ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 90 บทครอบคลุมศัพท์ระดับ SAT/IELTS/TOEFL กว่า 12,000 คำ) และไวยากรณ์อังกฤษ (ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 160 บทหรือกว่า 10,000 ข้อครอบคลุมเนื้อหาระดับอุดมศึกษาและTOEFL) มาพร้อมกับไฟล์เสียง/ไฟล์ข้อมูล/เฉลย อีกมาก(กำหนด DEBUT 1 เมษายน 2569)