

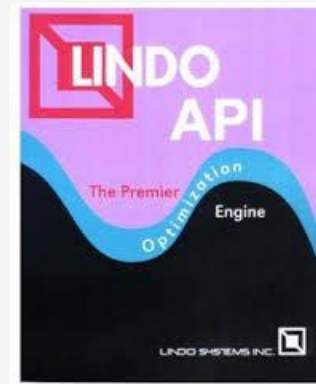
วิชา โปรแกรมสำเร็จรูปทางสถิติ (2204-2109)

บทที่ 8 โปรแกรมสำเร็จรูปที่ใช้ในการวิเคราะห์ข้อมูล

Asst. Prof. Juthawut Chantharamalee
Assistant Professor in Computer Science
(Chairperson of B.Sc. Program in Computer Science)
Office. Suan Dusit University, Phone. (+66) 2244-5691
Email. juthawut_cha@dusit.ac.th, jchantharamalee@gmail.com

8.1 การเลือกโปรแกรมสำเร็จรูปที่ใช้ในการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลทางสถิตินั้น เป็นการประมวลผลข้อมูลที่มีปริมาณมาก ดังนั้นการนำคอมพิวเตอร์มาช่วย ในการวิเคราะห์ข้อมูล จะช่วยประหยัดเวลาและค่าใช้จ่าย มีความสะดวกมากกว่าที่จะคำนวณด้วยเครื่องคิดเลข ทั่ว ๆ ไป และในการใช้คอมพิวเตอร์ส่วนใหญ่นักวิจัยนิยมที่จะใช้โปรแกรมสำเร็จรูป (Package Program) มากกว่าที่จะเขียนโปรแกรมขึ้นมาเอง



8.2 ขั้นตอนการวิเคราะห์ข้อมูลทางสถิติด้วยโปรแกรมสำเร็จรูป

- 1. การเตรียมข้อมูล** ผู้วิจัยต้องเตรียมเครื่องมือในการรวบรวมข้อมูลที่เหมาะสมกับการวิเคราะห์ด้วยคอมพิวเตอร์ (ซึ่งแตกต่างจากการวิเคราะห์ด้วยมือ) โดยในแบบสอบถามนั้นจะต้องกำหนดชื่อตัวแปร (Variable Name) และค่าของตัวแปร (Value) ให้เป็นตัวเลขเท่านั้น
- 2. การสร้างแฟ้มข้อมูล** ต้องกำหนดชื่อตัวแปร (Variable Name) ให้สอดคล้องกับที่กำหนดไว้ในแบบสอบถาม และสร้างคู่มือลงรหัส (Code book) ที่กำหนดสิ่งที่เป็นประโยชน์ต่อการวิเคราะห์ข้อมูล ซึ่งประกอบด้วย การกำหนดชื่อตัวแปร (Variable Name หรือ Name) ตำแหน่งทศนิยมของค่าของตัวแปร (Decimals) คำอธิบายชื่อตัวแปร (Label) ค่าของตัวแปร (Value) ความหมายของค่าของตัวแปร (Value Label) และระดับการวัดข้อมูล (Measure)

8.2 ขั้นตอนการวิเคราะห์ข้อมูลทางสถิติด้วยโปรแกรมสำเร็จรูป

3. การบันทึกข้อมูล ข้อมูลที่จะบันทึกต้องเป็นตัวเลขเท่านั้น โดยกำหนดค่าของตัวแปรที่เป็นตัวเลือกในแบบสอบถามให้เป็นตัวเลขเสียก่อน

4. การวิเคราะห์ข้อมูลทางสถิติ ผู้วิจัยสามารถวิเคราะห์ข้อมูลได้โดยง่าย ทั้งค่าสถิติพื้นฐานและค่าสถิติสำหรับทดสอบสมมติฐาน โดยใช้คำสั่งจาก Menu bar เลือก Analyze แล้วเลือกค่าสถิติที่ต้องการวิเคราะห์จาก Window ที่ปรากฏตามลำดับ ก็จะได้ผลการวิเคราะห์ตามต่อ

8.3 ความสามารถของโปรแกรมสำเร็จรูปทางสถิติทั่ว ๆ ไป

1. การจัดการกับข้อมูลและไฟล์ข้อมูล

- สามารถรับข้อมูลไปทำการวิเคราะห์ได้โดยตรง
- สามารถรับข้อมูลที่สร้างขึ้นมาจากโปรแกรมอื่น ๆ เช่น โปรแกรมพิมพ์เอกสาร
- การเปลี่ยนรูปข้อมูล จากข้อมูลเดิมโดยวิธีการทางคณิตศาสตร์
- การคัดเลือกข้อมูล ตามเงื่อนไขหรือโดยการสุ่มตัวอย่าง

2. การวิเคราะห์ข้อมูลด้วยวิธีการทางสถิติ

8.4 โปรแกรมกระดาษาทำการ (Spread Sheet Program)

ความสามารถของโปรแกรมกระดาษาทำการ

1. การวิเคราะห์ข้อมูลด้วยวิธีการทางสถิติ
2. การแจกแจงความถี่ แบบทางเดียวหรือหลายทาง
3. การคำนวณค่าสถิติเบื้องต้น เช่น ค่าผลรวม ค่าเฉลี่ย ค่าความแปรปรวน
4. การทดสอบสมมติฐาน เกี่ยวกับค่าเฉลี่ยด้วยวิธีการของ Z-test, T-test, F-test
5. การประมาณค่าเฉลี่ย ร้อยละ
6. การหาค่าความสัมพันธ์ด้วยวิธีการของ Pearson, Spearman, Kendall
7. การพยากรณ์ด้วยวิธีของ Time Series และ Regression

8.5 การเตรียมโปรแกรม Microsoft Excel สำหรับวิเคราะห์ข้อมูล

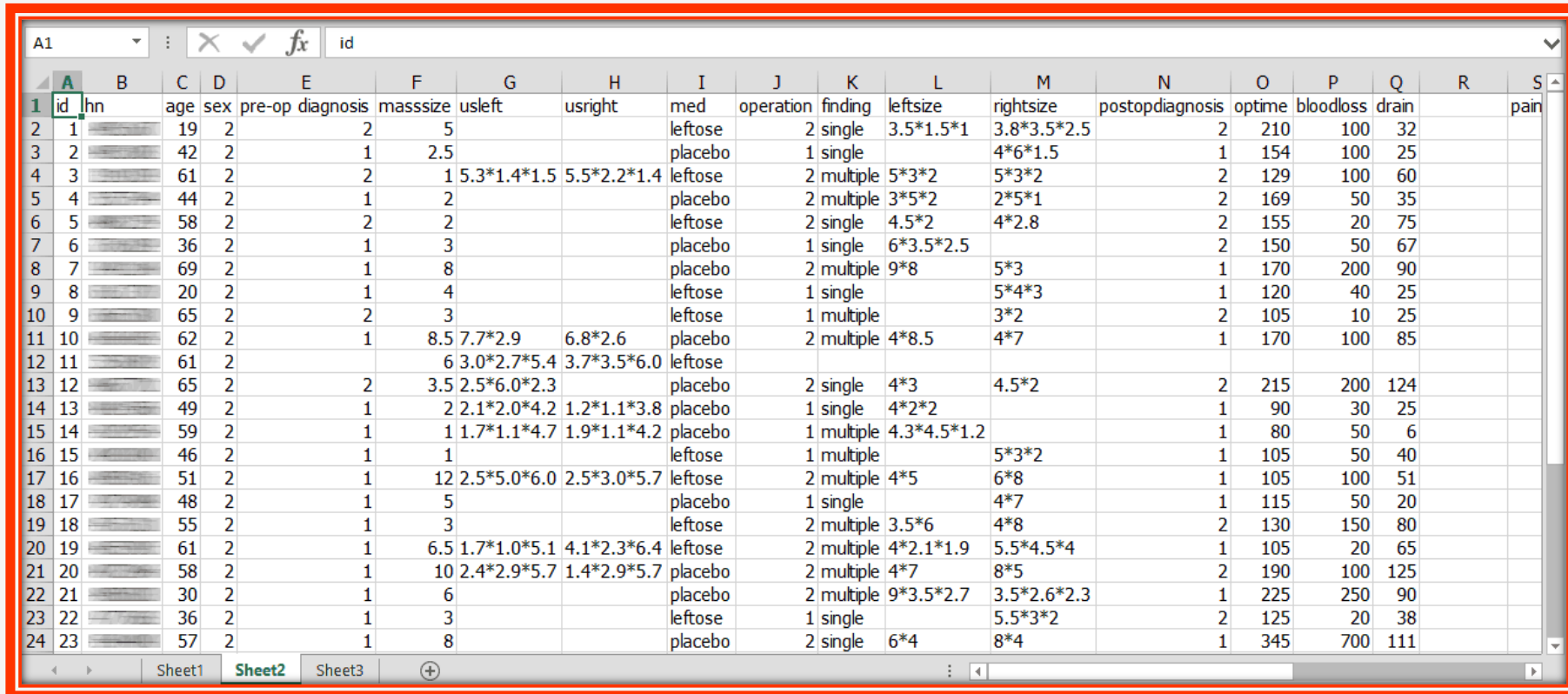
เป็นโปรแกรมสำหรับงานเกี่ยวกับการคำนวณตัวเลข การวิเคราะห์ข้อมูล การจัดเก็บ บันทึกข้อมูลของตาราง การสร้างกราฟ วิเคราะห์ข้อมูล และยังสามารถซ่อนข้อความ แทรกรูปภาพ สัญลักษณ์พิเศษต่าง ๆ ของตัวเลข มีฟังก์ชันในการคำนวณให้ผู้ใช้สามารถเลือกใช้มากมาย จึงทำให้สามารถนำมาใช้ในการวิเคราะห์คำนวณค่าตัวเลขต่าง ๆ ได้สะดวก ตลอดจนพัฒนาให้เป็นระบบงานที่มีขีดความสามารถสูง มีการวิเคราะห์ข้อมูลและพัฒนาระบบข้อมูล

สามารถใช้ทั้งแผนภูมิและแผนผังลำดับงาน เพื่ออธิบายแนวคิดที่สลับซับซ้อน ได้อย่างมีประสิทธิภาพ



8.5 การเตรียมโปรแกรม Microsoft Excel สำหรับวิเคราะห์ข้อมูล

1. ทำข้อมูลในลักษณะตาราง โดยให้แถวตามแนวนอน (Row) แถวแรก เป็นชื่อของตัวแปร และในแถวถัด ๆ มาเป็นข้อมูลจริง และให้เป็น 1 แถวต่อ 1 คนเท่านั้น (โปรแกรมจะไม่อ่านหลายแถว



id	hn	age	sex	pre-op	diagnosis	mass	size	usleft	usright	med	operation	finding	leftsize	rightsize	postopdiagnosis	optime	bloodloss	drain	pain
1		19	2		2		5			leftose	2 single	3.5*1.5*1	3.8*3.5*2.5	2	210	100	32		
3		42	2		1	2.5				placebo	1 single		4*6*1.5	1	154	100	25		
4		61	2		2			5.3*1.4*1.5	5.5*2.2*1.4	leftose	2 multiple	5*3*2		2	129	100	60		
5		44	2		1	2				placebo	2 multiple	3*5*2	2*5*1	2	169	50	35		
6		58	2		2	2				leftose	2 single	4.5*2	4*2.8	2	155	20	75		
7		36	2		1	3				placebo	1 single	6*3.5*2.5		2	150	50	67		
8		69	2		1	8				placebo	2 multiple	9*8	5*3	1	170	200	90		
9		20	2		1	4				leftose	1 single		5*4*3	1	120	40	25		
10		65	2		2	3				leftose	1 multiple		3*2	2	105	10	25		
11		62	2		1	8.5	7.7*2.9		6.8*2.6	placebo	2 multiple	4*8.5	4*7	1	170	100	85		
12		61	2			6	3.0*2.7*5.4		3.7*3.5*6.0	leftose									
13		65	2		2	3.5	2.5*6.0*2.3			placebo	2 single	4*3	4.5*2	2	215	200	124		
14		49	2		1	2	2.1*2.0*4.2		1.2*1.1*3.8	placebo	1 single	4*2*2		1	90	30	25		
15		59	2		1	1	1.7*1.1*4.7		1.9*1.1*4.2	placebo	1 multiple	4.3*4.5*1.2		1	80	50	6		
16		46	2		1	1				leftose	1 multiple		5*3*2	1	105	50	40		
17		51	2		1	12	2.5*5.0*6.0		2.5*3.0*5.7	leftose	2 multiple	4*5	6*8	1	105	100	51		
18		48	2		1	5				placebo	1 single		4*7	1	115	50	20		
19		55	2		1	3				leftose	2 multiple	3.5*6	4*8	2	130	150	80		
20		61	2		1	6.5	1.7*1.0*5.1		4.1*2.3*6.4	leftose	2 multiple	4*2.1*1.9	5.5*4.5*4	1	105	20	65		
21		58	2		1	10	2.4*2.9*5.7		1.4*2.9*5.7	placebo	2 multiple	4*7	8*5	2	190	100	125		
22		30	2		1	6				placebo	2 multiple	9*3.5*2.7	3.5*2.6*2.3	1	225	250	90		
23		36	2		1	3				leftose	1 single		5.5*3*2	2	125	20	38		
24		57	2		1	8				placebo	2 single	6*4	8*4	1	345	700	111		

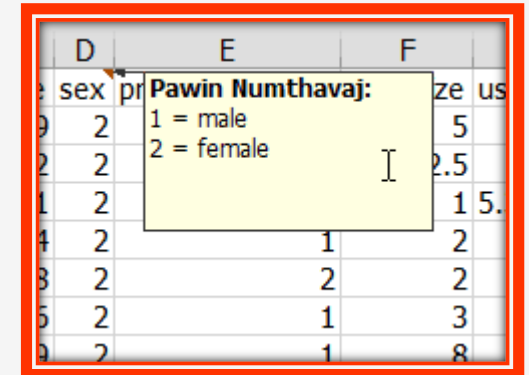
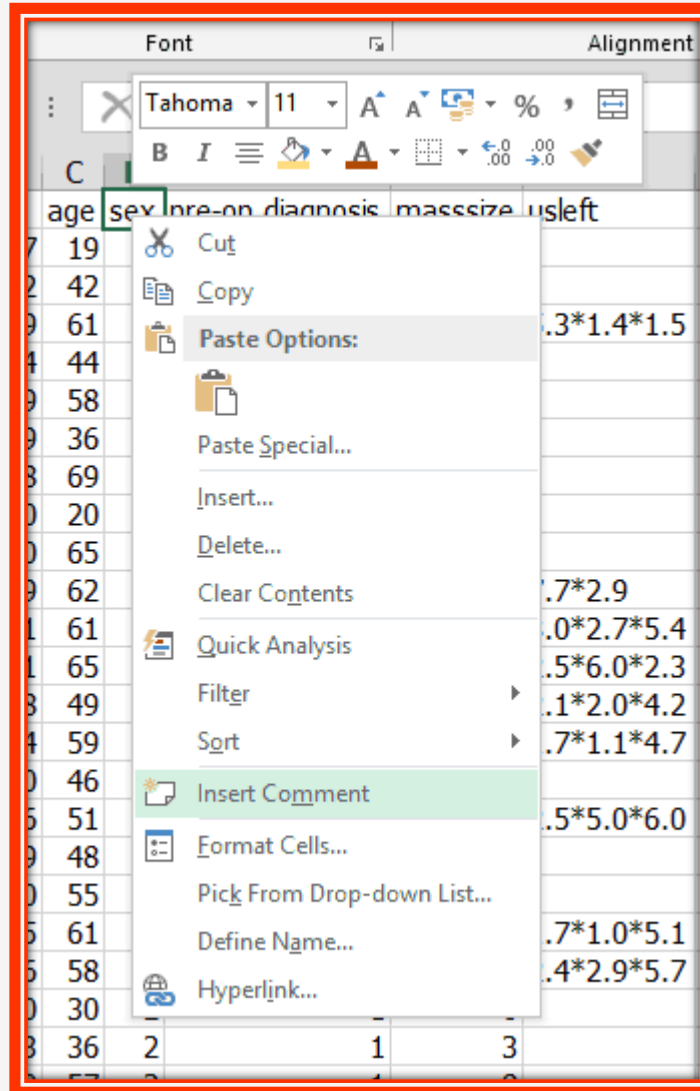
8.5 การเตรียมโปรแกรม Microsoft Excel สำหรับวิเคราะห์ข้อมูล

2. ชื่อของตัวแปร ควรใช้ตัวอักษรภาษาอังกฤษตัวเล็ก ไม่ควรยาวเกินไป และไม่ควรมีเครื่องหมายแปลก ๆ เช่น เครื่องหมายวงเล็บ เว้นวรรค เครื่องหมาย % & + ! เนื่องจากเครื่องหมายเหล่านี้ มักจะมีความหมายในโปรแกรมวิเคราะห์ทางสถิติ ทำให้การนำข้อมูลเข้าโปรแกรม มีความผิดพลาดได้ หากจำเป็นต้องใช้เครื่องหมายแยก ให้ใช้เครื่องหมายขีดล่าง (Underscore _)

เพศ	sex, gender
อายุ	age
น้ำหนัก	weight
ส่วนสูง	height
วันเกิด	dateofbirth
วันผ่าตัด	dateofoperation
คะแนนความเจ็บปวดที่หนึ่งชั่วโมง	painscore_1hr

8.5 การเตรียมโปรแกรม Microsoft Excel สำหรับวิเคราะห์ข้อมูล

3. ไม่ควรนำสิ่งที่ไม่ใช่ข้อมูล เข้าไปไว้ในส่วนที่เป็นข้อมูลผู้ใช้งานใหญ่เมื่อใช้โปรแกรม Spreadsheet ในการลงข้อมูล มักจะพยายามใส่ข้อมูล “เกิน” กว่าที่ตั้งใจเอาไว้ในการเก็บข้อมูล เสมือนเป็นการ “ทด” ลงในตารางข้อมูล เช่น ลงข้อมูลคะแนนความเจ็บปวด ที่เป็นตัวเลขว่า “ผู้ป่วยไม่เจ็บ” หรือทดไว้ว่า “2 - จากการโทรสอบถาม” ซึ่งจะทำให้เกิดปัญหามากในการนำข้อมูลลงไปโปรแกรมวิเคราะห์ทางสถิติ



8.5 การเตรียมโปรแกรม Microsoft Excel สำหรับวิเคราะห์ข้อมูล

4. ไม่พยายามทด หรือแยกความแตกต่างในแต่ละข้อมูล ด้วย “สี” ของเซลล์

โปรแกรมทางสถิตินั้นจะมองเฉพาะข้อมูลที่เป็นข้อมูลจริง ๆ เท่านั้น จะไม่มองถึงรูปแบบของสี หรือ ฟอนต์ ในการวิเคราะห์ข้อมูล การพยายามแยกสี เช่น สีแดงสำหรับคนไข้ที่ไม่ได้รับยา สีเขียวสำหรับคนไข้ที่ได้รับยา ซึ่งเสมือนกับว่าเป็นข้อมูลของคนไข้ แต่เก็บไว้กับสีนี้จะไม่ถูกส่งต่อไปยังโปรแกรมวิเคราะห์ข้อมูล

8.5 การเตรียมโปรแกรม Microsoft Excel สำหรับวิเคราะห์ข้อมูล

5. พยายามใส่ข้อมูลที่เป็น category ให้เป็นตัวเลข โปรแกรมวิเคราะห์ทางสถิติส่วนใหญ่ จะไม่สามารถแยกความแตกต่างของตัวอักษร และสร้างออกมาเป็น category ได้ เช่น หากผู้ใช้บันทึกข้อมูลในตัวแปร sex ว่าเป็นข้อความ “male”, “female” โปรแกรมสถิติจะมองเห็นเป็นข้อความ ไม่ใช่ category ทำให้ต้องมานั่งจัด category กันใหม่ในโปรแกรมทางสถิติ รวมถึงบางทีก็พิมพ์ผิดเองจาก male เป็น mal หรือ MALE ทำให้โปรแกรมจัด category ไม่ถูกต้อง

คำแนะนำ พยายามทำ category ให้เป็นตัวเลข ซึ่งอาจทำให้สอดคล้องไปกันกับแบบเก็บข้อมูล

ตัวอย่าง เช่น ในแบบเก็บข้อมูลทำการเก็บข้อมูล “ความรุนแรงของโรค”

ความรุนแรงของโรค: 1.Mild 2. Moderate 3. Severe
ควรเก็บเป็นดังนี้ 1, 2, 3 (อาจเพิ่ม 9 ในกรณีที่ไม่ทราบข้อมูล)

8.5 การเตรียมโปรแกรม Microsoft Excel สำหรับวิเคราะห์ข้อมูล

6. ในกรณีที่ไม่มีข้อมูล
อาจเว้นว่าง หรือทำ
โค้ดตัวเลข (เช่นเลข 9)
เอาไว้ ไม่ควรใส่เป็น
ข้อความว่า “N/A”
หรือ “Unknown”
เพราะตอนนำข้อมูลเข้า
โปรแกรมจะมองเป็น
ตัวอักษรแทน

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	id	date	inclus1	inclus2	inclus3	inclus4	dateofbirth	sex	height	weight	underlying	medication	nsaid	ppi	othermed	diagnosis	d
2	1	8/9/2014	1	1	1	1	5/1/1947	1	158	61	1	1	1	2	2	2	0
3	2	8/9/2014	1	1	1	1	12/6/1989	2	158	46	2	2	2	2	2	2	0
4	3	9/9/2014	1	1	1	1	1/1/1944	1	170	74	1	1	2	2	2	1	0
5	4	18/9/2014	1	1	1	1	30/6/1976	2	158	62	2	2	2	1	2	1	0
6	5	25/9/2014	1	1	1	1	6/12/1970	1	161	83	2	2	2	2	2	2	0
7	6	27/9/2014	1	1	1	1	7/6/1951	1	163	53	2	2	2	2	2	2	0
8	7	29/9/2014	1	1	1	1	22/9/1995	2	154	72	2	2	2	2	2	2	0
9	8	13/10/2014	1	1	1	1	9/7/1959	2	157	68	1	1	2	2	2	2	0
10	9	30/10/2014	1	1	1	1	2/4/1966	2	155	55	2	2	2	2	2	2	0
11	10	1/11/2014	1	1	1	1	6/2/1987	1	170	60	2	2	2	2	2	2	1
12	11	4/11/2014	1	1	1	1	29/4/1958	1	169	70	1	1	2	2	2	2	0
13	12	24/11/2014	1	1	1	1	13/3/1952	1	166	74	1	1	2	2	2	2	0
14	13	9/12/2014	1	1	1	1	17/4/1958	2	163	45	2	2	2	2	2	2	0
15	14	20/12/2014	1	1	1	1	31/3/1947	1	172	84	1	1	2	1	1	1	0
16	15	15/12/2014	1	1	1	1	30/7/1954	1	151	85	1	1	1	1	1	1	0
17	16	28/12/2014	1	1	1	1	15/9/1956	1	156	55	2	2	2	2	2	2	0
18	17	29/12/2014	1	1	1	1	29/11/1972	2	156	44	2	2	2	2	2	1	0
19	18	31/12/2014	1	1	1	1	22/6/1962	2	154	69	1	1	2	2	2	2	0
20	19	5/1/2015	1	1	1	1	1/1/1949	2	143	61	1	1	2	2	2	2	0
21	20	7/1/2015	1	1	1	1	1/1/1952	2	152	68	1	1	2	2	2	1	0
22	21	12/1/2015	1	1	1	1	13/1/1974	1	162	58	2	2	2	1	1	1	0

Thank You

จบการนำเสนอ



Any Question