

การใช้โปรแกรม STATA พื้นฐาน

กฤษฎา ดนุเดช วงศ์เวชวิวัฒน์ ภ.บ., วท.ม.
คณะเภสัชศาสตร์ มหาวิทยาลัยสยาม

STATA เป็นโปรแกรมสำเร็จรูปทางสถิติที่มีการใช้งานอย่างแพร่หลายในกลุ่มนักวิจัยสาขาต่าง ๆ โดยเฉพาะการวิจัยในสาขาด้านวิทยาศาสตร์สุขภาพ เนื่องจากโปรแกรมมีศักยภาพ ครอบคลุม ทั้งด้านการวิเคราะห์ข้อมูล ด้านการจัดการข้อมูล ด้านการนำเสนอข้อมูลและด้านการจัดการผลลัพธ์ โปรแกรมมีระบบการทำงานที่ง่ายต่อการเรียนรู้และใช้งาน มีระบบการสนับสนุนช่วยเหลือผ่านเว็บไซต์ของโปรแกรม

เนื้อหาของพื้นฐานการใช้โปรแกรม STATA เบื้องต้นประกอบด้วยส่วนต่าง ๆ ได้แก่

1. หน้าต่างบนโปรแกรม STATA
2. แถบเครื่องมือและคำสั่ง
3. สัญลักษณ์ที่ใช้ในการใช้งาน
4. การกำหนดชื่อและของแฟ้ม
5. การใช้งานคำสั่ง

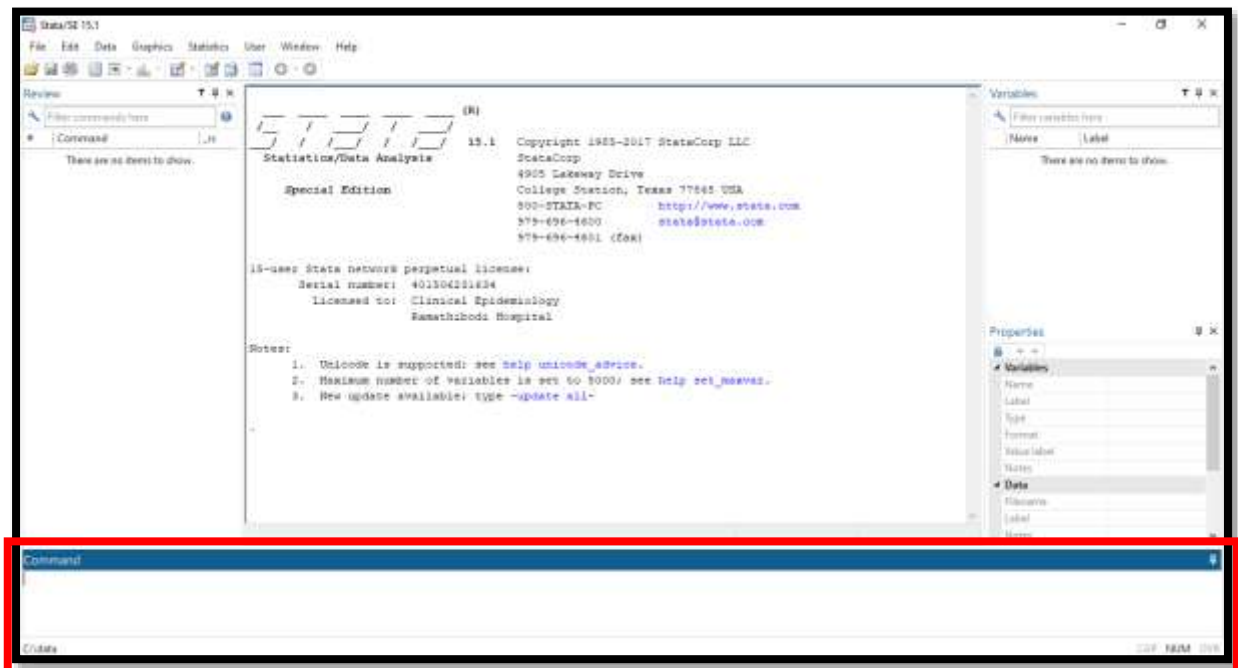
หน้าต่างบนโปรแกรม STATA

ลักษณะของหน้าต่างของโปรแกรม STATA ผู้ใช้งานสามารถกำหนดลักษณะและรูปแบบต่าง ๆ รวมถึงคุณลักษณะในการใช้งาน การกำหนดรูปแบบของการแสดงผลลัพธ์และการกำหนดคุณสมบัติของหน้าต่างโปรแกรมได้เพื่อความสะดวกในการใช้งาน



ส่วนประกอบของหน้าต่างของโปรแกรมประกอบด้วย

1. บริเวณสำหรับป้อนคำสั่งผ่านแป้นพิมพ์



2. บริเวณแสดงผลลัพธ์ของคำสั่ง



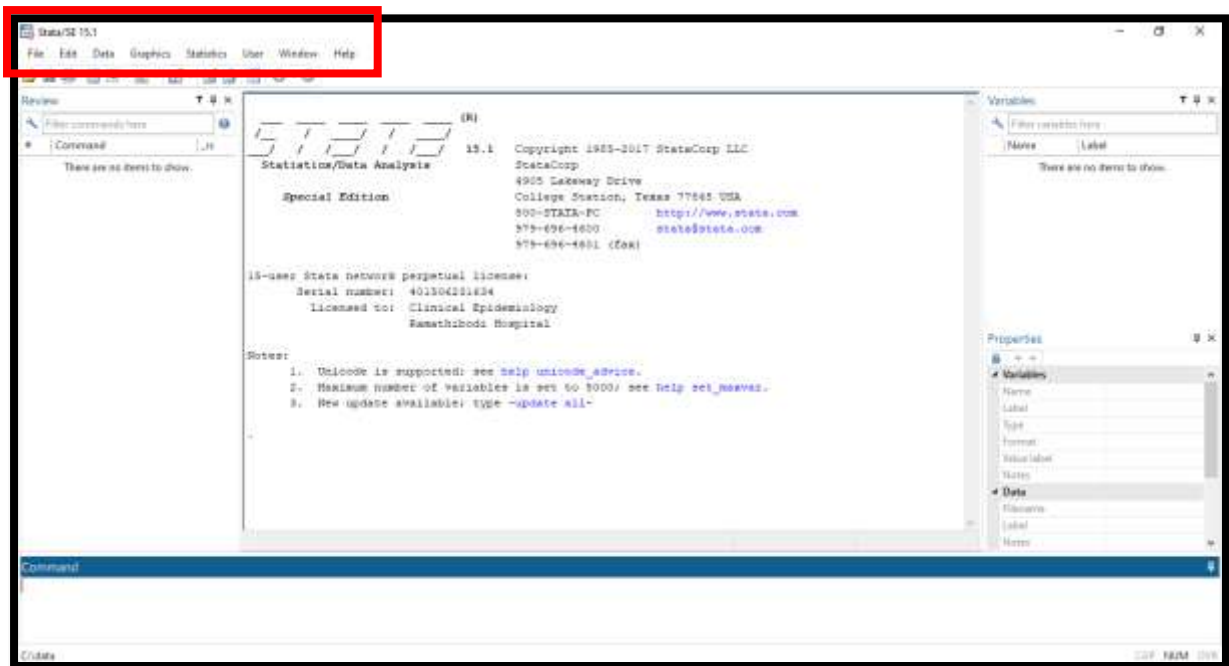
3. บริเวณแสดงคำสั่งที่เคยป้อนและทำการวิเคราะห์คำสั่งแล้ว



4. บริเวณแสดงชื่อตัวแปร



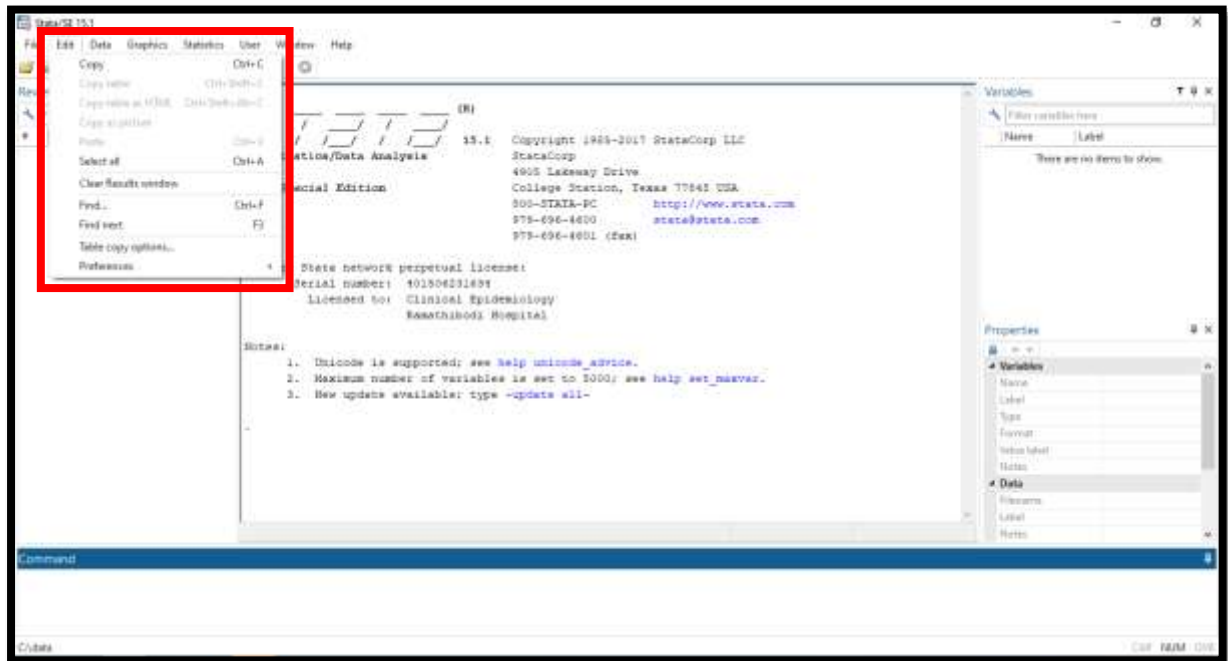
แถบเครื่องมือและคำสั่งสำเร็จรูป



1. ไฟล์ เป็นรายการคำสั่งที่มีลักษณะการทำงานเกี่ยวกับ การเปิดแฟ้มข้อมูล การเปิดแฟ้มข้อมูลรูปภาพ การเปิดแฟ้มข้อมูลที่มีการใช้ล่าสุด การบันทึกแฟ้มข้อมูล การเปิดแฟ้มบันทึกผลลัพธ์ หรือคำสั่ง การนำเข้า และนำออกแฟ้มข้อมูล การเรียกใช้ชุดข้อมูลตัวอย่างที่มากับโปรแกรม การพิมพ์งานและการออกจากโปรแกรม เป็นต้น



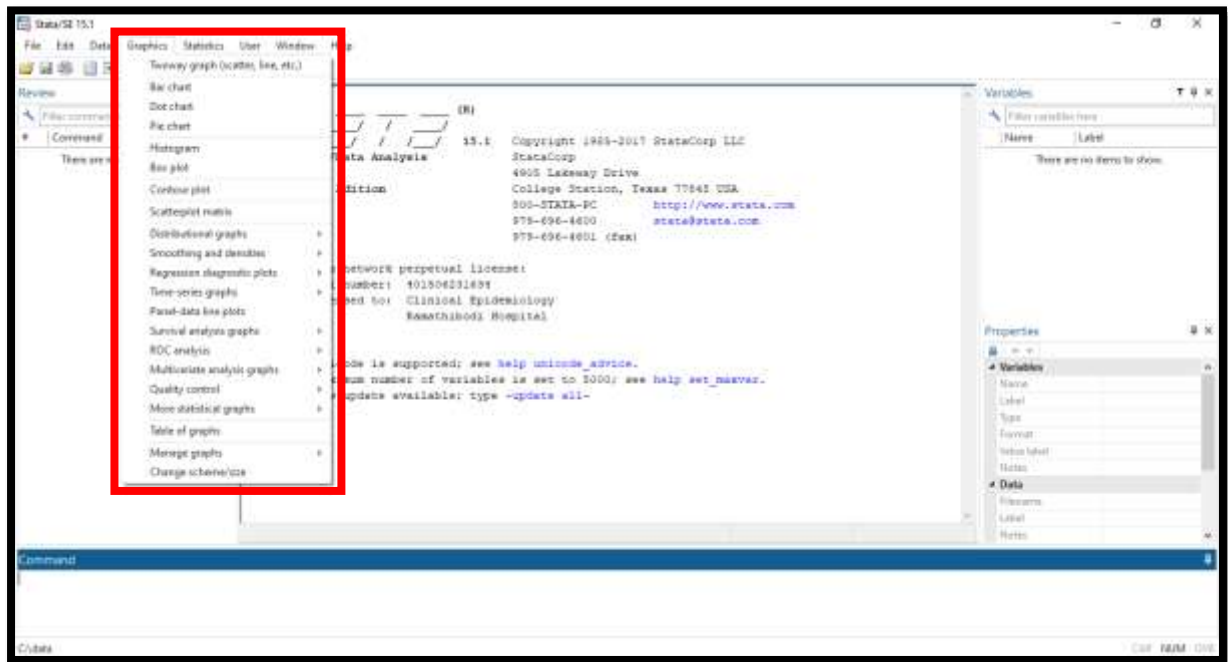
2. การแก้ไข เป็นส่วนของคำสั่งที่ใช้ในการปรับลักษณะของคำสั่งและหน้าต่างคำสั่ง เช่น การคัดลอก การเลือกคำสั่ง การค้นหาคำสั่งหรือผลลัพธ์ เป็นต้น



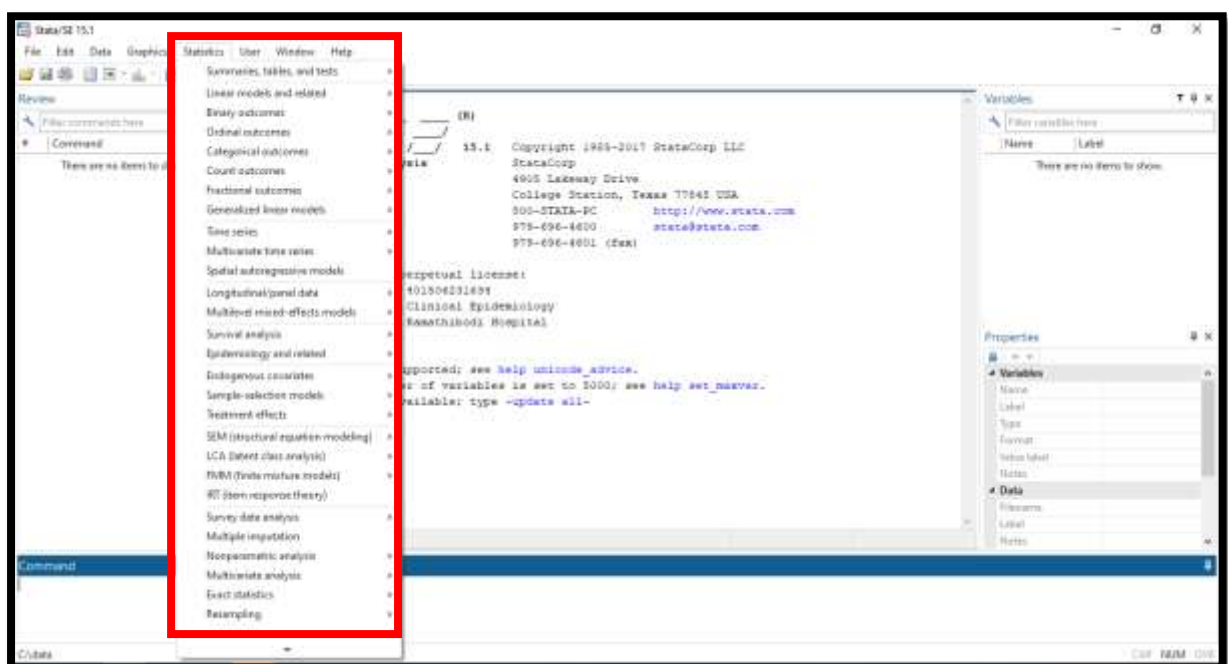
3. ข้อมูล เป็นแถบคำสั่งที่ใช้ในการแสดงผลข้อมูลที่ไม่ได้ทำการวิเคราะห์หรือการค้นหาข้อมูลหรือปรับเปลี่ยนลักษณะของรูปแบบข้อมูล การอธิบายข้อมูล การแก้ไขตัวแปร การจัดเรียงข้อมูล การรวมชุดข้อมูล การกำหนดรายละเอียดต่างๆของตัวแปร



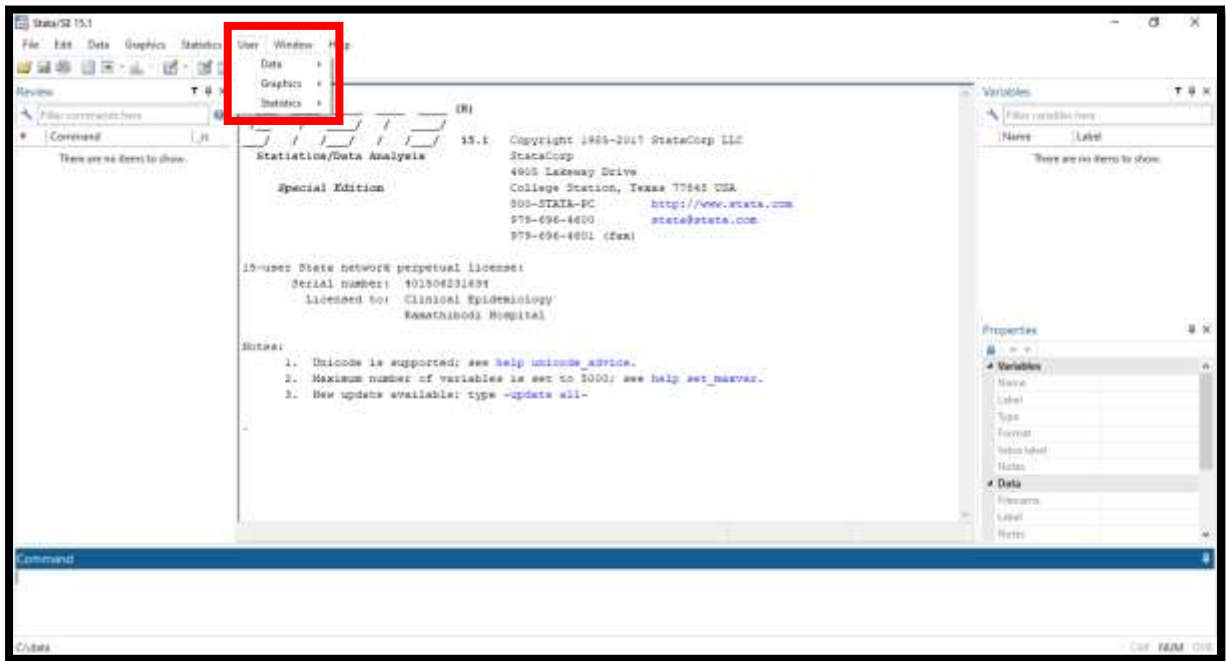
4. กราฟิก เป็นรายการคำสั่งที่ใช้ในการนำเสนอข้อมูลด้วยวิธี ต่างๆ ได้แก่ กราฟเส้น (line) แผนภูมิแท่ง (bar chart) กราฟฮิสโตแกรม (histogram) และกราฟประกอบวิธีการวิเคราะห์ทางสถิติ ได้แก่ time-series, survival analysis, ROC และ multivariate analysis เป็นต้น



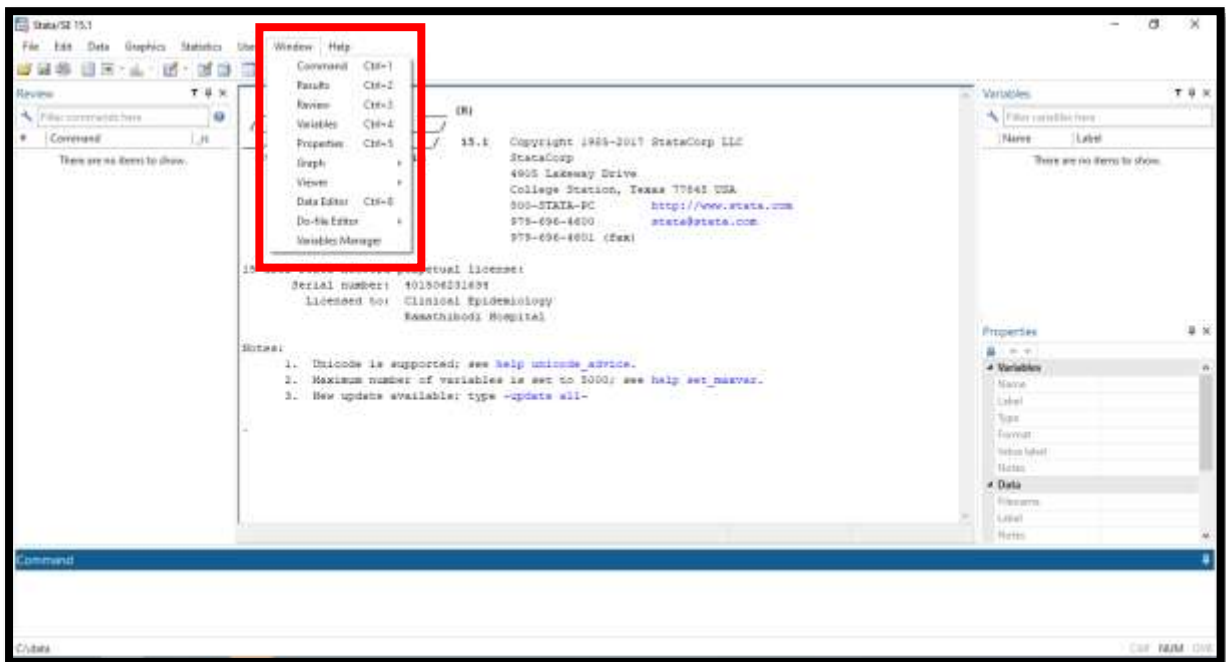
5. สถิติ เป็นรายการคำสั่งที่ใช้ในการวิเคราะห์ข้อมูลทางสถิติ ได้แก่ การแจกแจงความถี่และสถิติพรรณนา การวิเคราะห์ความสัมพันธ์และสมการถดถอยเชิงเส้น การวิเคราะห์ที่มี outcome แบบต่างๆ เช่น binary outcomes, ordinal outcomes, categorical outcomes และ count outcomes เป็นต้น การวิเคราะห์แบบ exact การวิเคราะห์แบบไม่อิงพารามิเตอร์ การวิเคราะห์ทางระบาดวิทยา และการวิเคราะห์แบบ multivariate analysis รวมถึงวิธีการคำนวณขนาดตัวอย่าง เป็นต้น



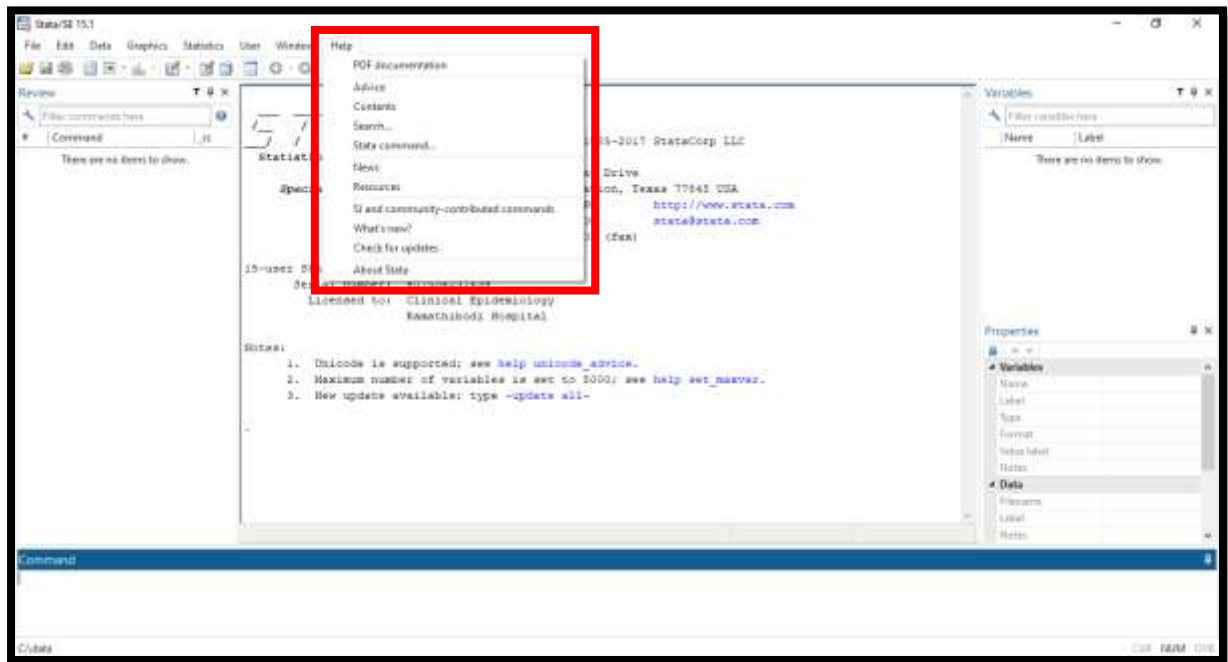
6. ผู้ใช้งาน เป็นรายการคำสั่งที่เปิดโอกาสให้ผู้ใช้งานกำหนดคำสั่งเพิ่มเติม ให้สอดคล้องกับลักษณะงานที่ต้องใช้บ่อย โดยครอบคลุมประเด็น ได้แก่ Data Graphics และ Statistics เป็นต้น



7. หน้าต่าง เป็นรายการคำสั่งที่ใช้ในการกำหนดหน้าต่างในการใช้งาน ได้แก่ หน้าต่าง Results, Graph, Viewer, Data Editor และ Do-file Editor เป็นต้น



8. ช่วยเหลือ เป็นรายการคำสั่งที่ใช้ในการเรียกความช่วยเหลือ แบบต่างๆของโปรแกรม ได้แก่ การดูเอกสารคู่มือ วิธีการใช้งานจากรายการคำสั่งย่อย Contents การสืบค้นวิธีการใช้คำสั่งจากรายการคำสั่งย่อย Search... การ update โปรแกรม และรายละเอียดเกี่ยวกับโปรแกรม STATA เป็นต้น



สัญลักษณ์ที่ใช้ในการใช้งาน

โปรแกรม STATA ได้กำหนดสัญลักษณ์ที่เกี่ยวข้องในการใช้งาน เพื่อใช้ระบุในคำสั่งหรือเงื่อนไขการทำงานให้สอดคล้องกับวัตถุประสงค์ตามที่ต้องการ โดยมีสัญลักษณ์และความหมาย ดังตาราง

ตาราง แสดงสัญลักษณ์และความหมายที่เกี่ยวข้องในการใช้งานคำสั่งและเงื่อนไข

สัญลักษณ์	ความหมาย
=	เท่ากับ***
>	มากกว่า
<	น้อยกว่า
>=	มากกว่าหรือเท่ากับ
<=	น้อยกว่าหรือเท่ากับ
~ = หรือ !=	ไม่เท่ากับ
~ หรือ !	ไม่
	หรือ
&	และ
+	บวก
-	ลบ
*	คูณ
/	หาร
^	ยกกำลัง
exp(x)	e^x
int(x)	จำนวนเต็ม
ln(x) หรือ log(x)	ลอการิทึม x ฐาน e
log10(x)	ลอการิทึม x ฐาน 10
sqrt(x)	รากที่สองของ x

*** กรณีกำหนดนิพจน์ภายใต้เงื่อนไขคำสั่ง if จะใช้เครื่องหมายเท่ากับสองตัว (= =)

การกำหนดชื่อและของแฟ้ม

การกำหนดชื่อแฟ้มข้อมูล ตัวแปร มาโคร ฟังก์ชันและคำอธิบายค่าข้อมูลในโปรแกรม STATA มีเงื่อนไขการกำหนดชื่อ ดังนี้

1. ใช้ตัวอักษร (a-z และ A-Z), ตัวเลข (0-9) หรือ เครื่องหมาย (_) เท่านั้น
2. ให้เริ่มต้นด้วยตัวอักษรและกำหนดให้กระชับ ไม่เกิน 32 ตัวอักษร
3. หลีกเลี่ยงการใช้ชื่อที่ตรงกับคำสั่งหลัก หรือชื่อฟังก์ชันที่มีในโปรแกรม
4. หากเป็นไปได้ควรกำหนดชื่อให้มีความหมายสามารถสื่อเนื้อหาเดียวกับค่าข้อมูลที่มี
5. การกำหนดควรยึดหลักให้คงที่และเป็นระบบตลอดทั้งแฟ้มข้อมูล
6. ควรกำหนดเป็นตัวพิมพ์เล็ก และใช้เครื่องหมาย (_) แทนช่องว่าง
7. กรณีตัวแปรหุ่น ควรกำหนดชื่อตามค่าที่มีของตัวแปร

STATA มีชนิดของแฟ้มข้อมูลที่ใช้งานในลักษณะที่แตกต่างกันโดยแสดงในตารางแสดงคำอธิบายชนิดของแฟ้มในโปรแกรม STATA

ชนิดของแฟ้ม	คำอธิบาย	คำสั่งที่เกี่ยวข้อง
.dta	แฟ้ม dataset	use หรือ save
.do	แฟ้ม do-file	do หรือ doedit
.ado	แฟ้ม do-file สำเร็จรูปที่มีผู้พัฒนาขึ้นและสามารถ download มาใช้งานได้จากระบบอินเทอร์เน็ต	doedit
.log	แฟ้ม log ในรูป text ใช้เก็บข้อมูลผลลัพธ์ทั้งหมด	log
.smcl	แฟ้ม log ในรูป SMCL ใช้เก็บข้อมูลคำสั่ง	cmdlog
.raw	แฟ้ม ASCII	infile,infix, insheet
.out	แฟ้มที่ถูกบันทึกด้วยคำสั่ง outsheet	outsheet
.dct	แฟ้มข้อมูล ASCII ในพจนานุกรม	infix
.gph	แฟ้มรูปภาพและแผนภูมิ	graph

การใช้งานคำสั่ง

การใช้งานคำสั่งของโปรแกรม STATA สามารถป้อนคำสั่งใช้งานได้ 3 รูปแบบคือ

1. การป้อนคำสั่งผ่านคำสั่งสำเร็จรูปบนแถบเครื่องมือ
2. การป้อนคำสั่งผ่านแป้นพิมพ์
3. การป้อนคำสั่งผ่านการใช้งานแบบ Do files

ตัวอย่างคำสั่งพื้นฐานที่ใช้บ่อยในโปรแกรม STATA

คำสั่งที่ใช้บ่อยในโปรแกรม Stata

สำหรับการวิเคราะห์ข้อมูลพื้นฐาน

. describe

```
Contains data from C:\Documents and Settings\Owner\My Documents\For_study_cm.511.dta
  obs:      616
  vars:      13                      10 Jul 2008 22:38
  size:     14,168 (91.2% of memory free)
```

variable name	storage type	display format	value label	variable label
class	byte	%8.0g		Class : nominal scale
age	byte	%8.0g		Age in year: ratio scale
gender	byte	%8.0g		Gender: nominal scale; 1=male,2=female
bmi1	float	%9.0g		BMI1: ratio scale
bmi_code1	byte	%8.0g		BMI_code1: ordinal scale; 1=underweight,2=normal,3=risk to overweight,4=obesity1
dep_score1	byte	%8.0g		Dep_Score1: ratio scale
dep_code1	byte	%8.0g		Dep_Code1: ordinal scale; 0=no,1=early depress,2=major depress
bmi2	float	%9.0g		BMI2: ratio scale
bmi_code2	byte	%8.0g		BMI_code2: ordinal scale; 1=underweight,2=normal,3=risk to overweight,4=obesity1
dep_score2	byte	%8.0g		Dep_Score2: ratio scale
dep_code2	byte	%8.0g		Dep_Code2: ordinal scale; 0=no,1=early depress,2=major depress
ess_score	byte	%8.0g		Epworth sleepiness score: ratio scale
ess_code	byte	%8.0g		ESS_Code: ordinal scale; 0=no, 1=has a problem

Sorted by: class

. summarize age : แสดงผลรวมของ 1 ตัวแปร

Variable	Obs	Mean	Std. Dev.	Min	Max
age	599	20.72788	1.852705	15	27

. ci age : แสดงค่า 95% Confidence interval

Variable	Obs	Mean	Std. Err.	[95% Conf. Interval]	
age	599	20.72788	.0756995	20.57921	20.87655

. ci age, level(99) : แสดงค่า 99% Confidence interval

Variable	Obs	Mean	Std. Err.	[99% Conf. Interval]	
age	599	20.72788	.0756995	20.53227	20.92349

. tabstat age, by (gender) stats(mean sd p25 p50 p75) : แสดงผลการวิเคราะห์ข้อมูลทางสถิติตามต้องการภายในวงเล็บหลังคำ stats ตามเพศ

Summary for variables: age					
by categories of: gender					
gender	mean	sd	p25	p50	p75
1	20.71193	1.878362	19	21	22
2	20.66379	1.789872	19	21	22
Total	20.68359	1.825354	19	21	22

. tab class : แสดงตาราง 1 ตัวแปร

Class	Freq.	Percent	Cum.
1	119	19.32	19.32
2	116	18.83	38.15
3	97	15.75	53.90
4	127	20.62	74.51
5	86	13.96	88.47
6	71	11.53	100.00
Total	616	100.00	

. tab gender bmi_code1 : แสดงตาราง 2 ตัวแปร

Gender	1	2	3	4	5	Total
1	24	124	25	28	8	209
2	79	196	21	12	2	310
Total	103	320	46	40	10	519

. tab gender bmi_code1,col ro : แสดงจำนวนเปอร์เซ็นต์ตาม column และ row

+-----+ Key +-----+ frequency row percentage column percentage +-----+						
gender	bmi_code1					
	1	2	3	4	5	Total
1	24	124	25	28	8	209
	11.48	59.33	11.96	13.40	3.83	100.00
	23.30	38.75	54.35	70.00	80.00	40.27
2	79	196	21	12	2	310
	25.48	63.23	6.77	3.87	0.65	100.00
	76.70	61.25	45.65	30.00	20.00	59.73
Total	103	320	46	40	10	519
	19.85	61.66	8.86	7.71	1.93	100.00
	100.00	100.00	100.00	100.00	100.00	100.00

. sort class : แสดงตาราง 3 ตัวแปร

. by class: tab gender dep_code2

Gender	0	1	2	Total
1	40	1	2	43
2	69	1	2	72
Total	109	2	4	115

Gender	0	1	2	Total
1	34	2	0	36
2	57	4	1	62
Total	91	6	1	98

Gender	0	1	2	Total
1	31	3	1	35
2	29	4	4	37
Total	60	7	5	72

-> class = 4				
	Dep_Code2			
Gender	0	2	Total	
1	29	0	29	
2	62	3	65	
Total	91	3	94	

-> class = 5				
	Dep_Code2			
Gender	0	1	2	Total
1	24	1	2	27
2	32	2	0	34
Total	56	3	2	61

-> class = 6				
	Dep_Code2			
Gender	0	1	2	Total
1	24	2	0	26
2	33	3	1	37
Total	57	5	1	63