Basic Stata and biostatistics for DNCD

THANANAN JIVARAMONAIKUL, MD.

Outline

1. Import file, open file, save file, save command, ปรับแต่งหน้าตา
2. Type of data and data management
3. Study design : cross-sectional, case-control, cohort
4. Type of statistic :
   - Descriptive: means, S.D., median, quarter, IQR
   - Analytic : comparison, association
5. Measure in epidemiology
   - Frequency : prevalence, incidence
   - Comparison : Chi2 test, t-test, Wilcoxan test
   - Association : RR, OR, Regression (linear, logistic, Poisson)
6. Confounding & effect modification

Stata/MP 16.0

File Edit Data Graphics Statistics User Window Help

History

Filter commands here

#	Command	_rc

There are no items to show.

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   16.0   Copyright 1985-2019 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     MP - Parallel Edition            College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

30-student 2-core Stata lab perpetual license:
       Serial number:  501606204774
         Licensed to:  Clinical Epidemiology and Clinical Statistic Center
                       Faculty of Medicine, Chiang Mai University

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  More than 2 billion observations are allowed; see help obs_advice.
      3.  Maximum number of variables is set to 5000; see help set_maxvar.

.
```

Command

Variables

Filter variables here

Name	Label

There are no items to show.

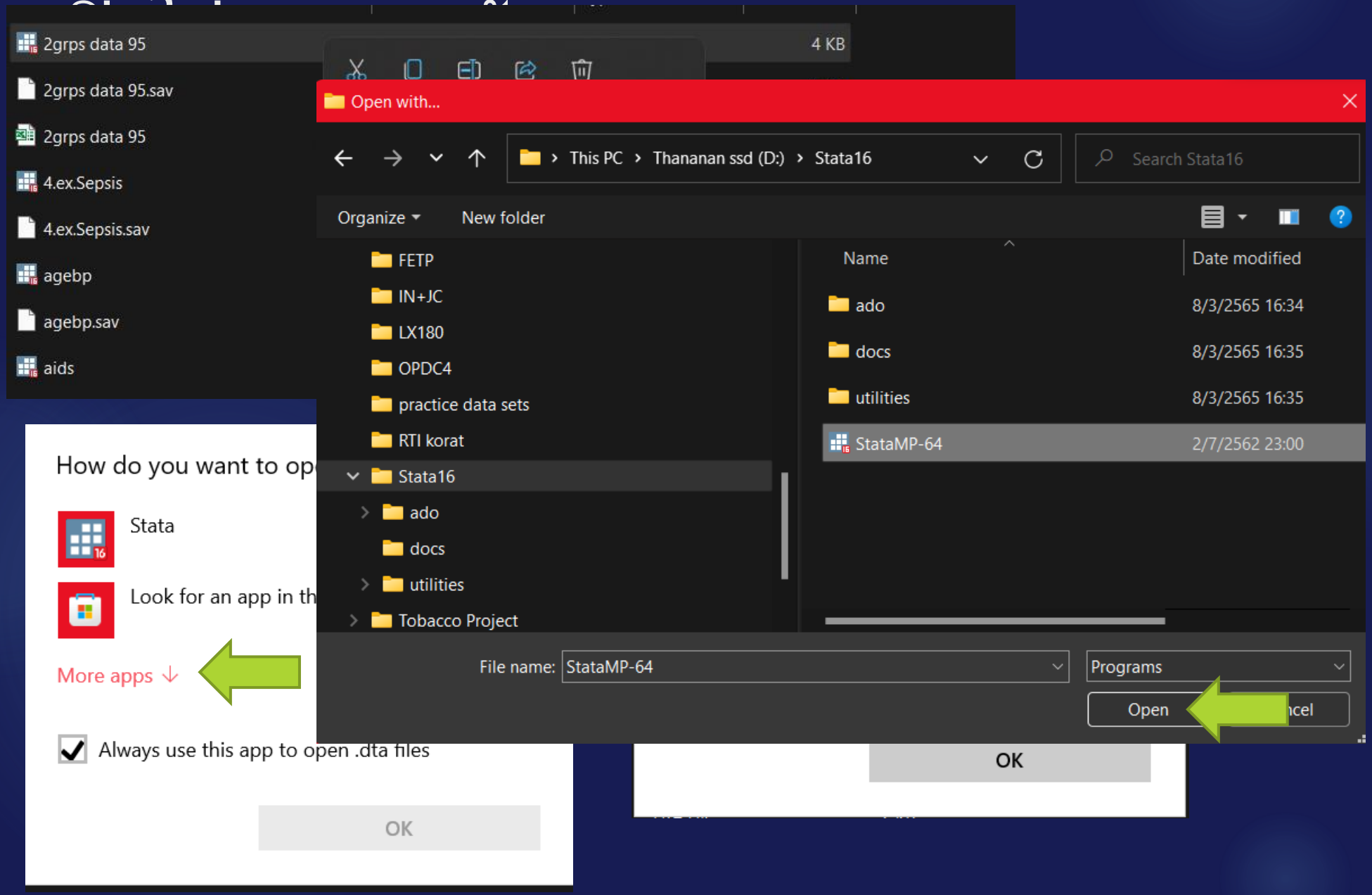
Properties

Variables	
Name	
Label	
Type	
Format	
Value label	
Notes	
Data	
Frame	default
Filename	
Label	
Notes	
Variables	0
Observations	0

C:\Program Files\Stata16

CAP NUM OVR

30°C มีแดดส่วนใหญ่ ENG 8:51 2/3/2565


2grps data 95
4 KB
2grps data 95.sav
2grps data 95
4.ex.Sepsis
4.ex.Sepsis.sav
agebp
agebp.sav
aids
Open with...
This PC › Thananan ssd (D:) › Stata16
Search Stata16
Organize
New folder
FETP
IN+JC
LX180
OPDC4
practice data sets
RTI korat
Stata16
ado
docs
utilities
Tobacco Project
Name
Date modified
ado
8/3/2565 16:34
docs
8/3/2565 16:35
utilities
8/3/2565 16:35
StataMP-64
2/7/2562 23:00
File name:
StataMP-64
Programs
Open
How do you want to op
Stata
Look for an app in th
More apps
Always use this app to open .dta files
OK
OK

นำเข้าไฟล์ xls/cvs

Stata/MP 16.0

File Edit Data Graphics Statistics User Window Help

- Open... Ctrl+O
- Save Ctrl+S
- Save as... Ctrl+Shift+S
- View...
- Do...
- Filename...
- Change working directory...
- Log
- Import
  - Excel spreadsheet (*.xls;*.xlsx)
  - Text data (delimited, *.csv, ...)
  - SPSS data (*.sav)
  - SAS data (*.sas7bdat)
  - Text data in fixed format
  - Text data in fixed format with a dictionary
  - Unformatted text data
  - SAS XPORT Version 8 (*.v8xpt)
  - SAS XPORT Version 5 (*.xpt)
  - Federal Reserve Economic Data (FRED)
  - Haver Analytics database
  - ODBC data source
  - dBase (*.dbf)
- Export
- Print
- Example datasets...
- Recent files
- Exit

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   16.0   Copyright 1985-2019 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     MP - Parallel Edition            College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

                                      ense:

                                      logy and Clinical Statistic Center
                                      e, Chiang Mai University

                                      unicode advice.
                                      s are allowed; see help obs advice.
                                      set to 5000; see help set maxvar.
```

Variables

Filter variables here

Name	Label

There are no items to show.

Properties

Variables	
Name	
Label	
Type	
Format	
Value label	
Notes	
Data	
Frame	default
Filename	
Label	
Notes	
Variables	0
Observations	0

C:\Program Files\Stata16

CAP NUM OVR

30°C มีแดดส่วนใหญ่ ENG 8:55 2/3/2565

Stata/MP 16.0

File Edit Data Graphics Statistics User Window Help

- Open... Ctrl+O
- Save Ctrl+S
- Save as... Ctrl+Shift+S
- View...
- Do...
- Filename...
- Change working directory...
- Log
- Import
  - Excel spreadsheet (*.xls;*.xlsx)
  - Text data (delimited, *.csv, ...)
  - SPSS data (*.sav)
  - SAS data (*.sas7bdat)
  - Text data in fixed format
  - Text data in fixed format with a dictionary
  - Unformatted text data
  - SAS XPORT Version 8 (*.v8xpt)
  - SAS XPORT Version 5 (*.xpt)
  - Federal Reserve Economic Data (FRED)
  - Haver Analytics database
  - ODBC data source
  - dBase (*.dbf)
- Export
- Print
- Example datasets...
- Recent files
- Exit

```
  /__    /   ____/   /   ____/
  ___/   /   /___/   /   /___/   16.0   Copyright 1985-2019 StataCorp LLC
    Statistics/Data Analysis            StataCorp
                                        4905 Lakeway Drive
       MP - Parallel Edition            College Station, Texas 77845 USA
                                        800-STATA-PC        http://www.stata.com
                                        979-696-4600        stata@stata.com
                                        979-696-4601 (fax)

30-student 2-core Stata lab perpetual license:

                                   logy and Clinical Statistic Center
                                   e, Chiang Mai University

                                   unicode advice.
                                   s are allowed; see help obs_advice.
                                   set to 5000; see help set_maxvar.

                                   ds\HBS.xlsx", sheet("HBS") firstrow
```

Variables

Filter variables here

Name	Label

There are no items to show.

Properties

Variables	
Name	
Label	
Type	
Format	
Value label	
Notes	
Data	
Frame	default
Filename	
Label	
Notes	
Variables	0
Observations	0

C:\Program Files\Stata16

CAP NUM OVR

30°C มีแดดส่วนใหญ่ ENG 9:08 2/3/2565

Stata/MP 16.0

File Edit Data Graphics Statistics User Window Help

History

Filter commands here

#	Command	_rc
1	import excel "C:\Users\Tha...	
2	clear	

```
  /__    /   ____/
 ___/   /   /___/
  Statistics/Data A

    MP - Parallel

30-student 2-core S
      Serial numbe
        Licensed t

Notes:
     1.  Unicode i
     2.  More than
     3.  Maximum n

. import excel "C:\
(215 vars, 207,191

. clear

.
```

Command

import delimited - Import delimited text data

File to import:

C:\Users\Thananan\Desktop\Tobacco Project\Health Sit\Microdata HBS 2564\Microdata HBS 2564.csv

First row as variable names: Automatic

Set ranges...

Variable-name case: Lower

Floating point precision: Use default

Text encoding: UTF-8

Quote binding: Loose

Quote stripping: Automatic

Delimiter: Automatic

Treat sequential delimiters as one

Numeric parsing rules

Use locale based parsing for numbers

Locale: English (United States)

Override decimal separator:

Override grouping separator:

Preview:

#	reg	cwt	area	month_	yr	hh_no	type	member	listing	enum	hl1	hl3	hl4	hl5	hl6
2	1	10	1	2	64	5	1	1	1	11	1	1	2	51	1
3	1	10	1	2	64	6	1	1	2	11	1	1	1	19	1
4	1	10	1	2	64	7	1	2	2	11	1	1	1	49	1
5	1	10	1	2	64	7	1	2	2	11	2	9	2	82	1
6	1	10	1	2	64	8	1		5	24					
7	1	10	1	2	64	9	1	2	1	11	1	1	2	66	1
8	1	10	1	2	64	9	1	2	1	11	2	9	2	87	1
9	1	10	1	2	64	10	1	3	1	11	1	1	1	40	1
10	1	10	1	2	64	10	1	3	1	11	2	9	2	75	1
11	1	10	1	2	64	10	1	3	1	11	3	14	2	11	1
12	1	10	1	2	64	11	1	1	1	11	1	1	1	59	1
13	1	10	1	2	64	12	1	1	1	11	1	1	1	53	2
14	1	10	1	2	64	13	1	3	1	11	1	1	1	67	1
15	1	10	1	2	64	13	1	3	1	11	2	2	2	64	1
16	1	10	1	2	64	13	1	3	1	11	3	14	2	68	1
17	1	10	1	2	64	14	1		2	12					
18	1	10	1	2	64	15	1	3	3	11	1	1	1	25	2
19	1	10	1	2	64	15	1	3	3	11	2	17	2	35	2

Variable	Type
reg	numeric
cwt	numeric
area	numeric
month_	numeric
yr	numeric
hh_no	numeric
type	numeric
member	str
listing	numeric
enum	numeric
hl1	str
hl3	str

To change the data type for a column, right-click on the selected column and choose the appropriate type.
Note: importing string data as numeric can result in loss of data.

OK Cancel Submit

Variables

Filter variables here

Name	Label

There are no items to show.

Properties

Variables

Name
Label
Type
Format
Value label
Notes

Data

Frame	default
Filename	
Label	
Notes	
Variables	0
Observations	0

C:\Program Files\Stata16

CAP NUM OVR

30°C มีแดดส่วนใหญ่ ENG 9:09 2/3/2565

Stata/MP 16.0

File Edit Data Graphics Statistics User Window Help

History

Filter commands here

#	Command	_rc
1	import excel "C:\Users\Tha...	
2	clear	
3	import delimited "C:\Users\...	

```
                                  4905 Lakeway Drive
     MP - Parallel Edition        College Station, Texas 77845 USA
                                  800-STATA-PC        http://www.stata.com
                                  979-696-4600        stata@stata.com
                                  979-696-4601 (fax)

30-student 2-core Stata lab perpetual license:
       Serial number:  501606204774
         Licensed to:  Clinical Epidemiology and Clinical Statistic Center
                       Faculty of Medicine, Chiang Mai University

Notes:
      1.  Unicode is supported; see help unicode advice.
      2.  More than 2 billion observations are allowed; see help obs advice.
      3.  Maximum number of variables is set to 5000; see help set maxvar.

. import excel "C:\Users\Thananan\Downloads\HBS.xlsx", sheet("HBS") firstrow
(215 vars, 207,191 obs)

. clear

. import delimited "C:\Users\Thananan\Desktop\Tobacco Project\Health Sit\Microdata HBS 2564\Microdata HBS 2564.csv", encoding(UTF-8)
(215 vars, 207,191 obs)

.
```

Command

Variables

Filter variables here

Name	Label
reg	REG
cwt	CWT
area	AREA
month_	MONTH_
yr	YR
hh_no	HH_NO
type	TYPE
member	MEMBER
listing	LISTING
enum	ENUM
hl1	HL1
hl3	HL3
hl4	HL4

Properties

Variables

Name	
Label	
Type	
Format	
Value label	
Notes	

Data

Frame	default
Filename	
Label	
Notes	
Variables	215
Observations	207,191

C:\Program Files\Stata16

CAP NUM OVR

30°C มีแดดส่วนใหญ่ ENG 9:09 2/3/2565

ปรับแต่งหน้าตา Stata เพิ่มสีสันก่อนเริ่มใช้งาน

Stata/MP 16.0

File Edit Data Graphics Statistics User Window Help

History

Filter commands here

#	Command	_rc

There are no items to show.

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   16.0   Copyright 1985-2019 StataCorp LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     MP - Parallel Edition            College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

30-student 2-core Stata lab perpetual license:
       Serial number:  501606204774
         Licensed to:  Clinical Epidemiology and Clinical Statistic Center
                       Faculty of Medicine, Chiang Mai University

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  More than 2 billion observations are allowed; see help obs_advice.
      3.  Maximum number of variables is set to 5000; see help set_maxvar.

.
```

Select all Ctrl+A
Clear results
Preferences...
Font...
Print...

http://www.stata.com

Command

Variables

Filter variables here

Name	Label

There are no items to show.

Properties

Variables	
Name	
Label	
Type	
Format	
Value label	
Notes	
Data	
Frame	default
Filename	
Label	
Notes	
Variables	0
Observations	0

C:\Program Files\Stata16

CAP NUM OVR

30°C มีแดดส่วนใหญ่

ไทย 8:52 2/3/2565

Stata/MP 16.0

File Edit Data Graphics Statistics User Window Help

History

Filter commands here

Command _rc

There are no items to show.

```
  ___  ____  ____  ____  ____  (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/
  Statistics/Data Analysis

    MP - Parallel Edition


30-student 2-core Stata lab p
      Serial number:  501600
        Licensed to:  Clinic
                      Facult

Notes:
      1.  Unicode is supporte
      2.  More than 2 billion
      3.  Maximum number of v

.
```

Command

General preferences

- General
- Results
- Viewer
- Data Editor
- Internet

Colors

Color scheme: Custom 2

Results	☐ Bold	
Standard	☐ Bold	
Error	☐ Bold	
Input	☑ Bold	
Link	☐ Bold	☑ Underline
Hilite	☑ Bold	☐ Underline
Background		

Behavior

1 Line gap

200 Scrollback buffer size in kilobytes (takes effect after restart)

OK Cancel Apply

Variables

Filter variables here

Name Label

There are no items to show.

Properties

Variables

Name
Label
Type
Format
Value label
Notes

Data

Frame default
Filename
Label
Notes
Variables 0
Observations 0

C:\Program Files\Stata16

CAP NUM OVR

30°C มีแดดส่วนใหญ่

ไทย 8:53 2/3/2565

Data management

Type of data (1)

NOMINAL DATA

- values that the data may have do not have specific order
- values act as labels with no real meaning
- categories, states
  - Binomial: two possible values
  - Multinomial: more than two possible values

e.g. Health status healthy =1 sick=2

e.g. Treatment new regimen = 1 standard regimen = 2

e.g. hair color brown =1 blond =2 black =100

ORDINAL DATA

- values with some kind of ordering
- data that has been measured or counted

e.g. social class: upper=1 middle = 2 working = 3

e.g. glioblastoma tumor grade: 1 2 3 4 5

e.g. position in a race: 1st 2nd 3rd

Type of data (3)

DISCRETE

- distinct or separate parts, with no finite detail

e.g. children in family

CONTINUOUS

- between any two values, there would be a third

e.g. between meters there are centimeters

INTERVAL

- equal intervals between values and an arbitrary zero on the scale

  e.g. temperature gradient

RATIO

- equal intervals between values and an absolute zero

  e.g. weight, body mass index

Examples of coding

Site Number - Subject Number - Chk Initials Date of Exam dd mm yy

Physical Exam

1. Height ___ cm
2. Weight ___.___ kg
3. Temperature ___.___ °C
4. Pulse ___ /min
5. Respiration ___ /min
6. Blood Pressure ___ / ___ mmHg (systolic / diastolic)

Cont.

Instructions: Physical exam directed by history, systems review and symptoms only.

Cat. 1 2 99

Body System	Normal	Abnormal	Not Done	Comment on all abnormalities
7. HEENT	☐	☐	☐	
8. Lymph Nodes	☐	☐	☐	
9. Cardiovascular	☐	☐	☐	

see pp. 4 & 7

Index P56

Page 50

Genentech, Inc.
Genentech, Inc.
Genentech, Inc.
Genentech, Inc.
Genentech, Inc.

Study #: V0633g
Drug: MN Recombinant Glycoprotein 120/HIV-1 Vaccine
Study: Phase I/II Recovering IVDU in Thailand

Subject Identification
Subject Number ___ - ___
Subject Initials ___ ___
Specimen Date: ___ Mo ___ Day ___ Yr

LABORATORY DATA (MONTH 12.5)

CHEMISTRY	Results	ND
Creatinine		☐
ALT/SGPT		☐

HEPATITIS	N/A	Pos	Neg	ND
HBs Ag	☐	☐	☐	☐
Anti HBc	☐	☐	☐	☐
HCV Ab	☐	☐	☐	☐

HEMATOLOGY	Results	ND
WBC		☐
RBC		☐
Hemoglobin		☐
Hematocrit		☐
Platelet Count		☐
Segmented Neutrophils		☐
Banded Neutrophils		☐
Lymphocytes		☐
Monocytes		☐
Eosinphils		☐
Basophils		☐
Other, Specify: ___		☐
[redacted]		☐
[redacted]		☐

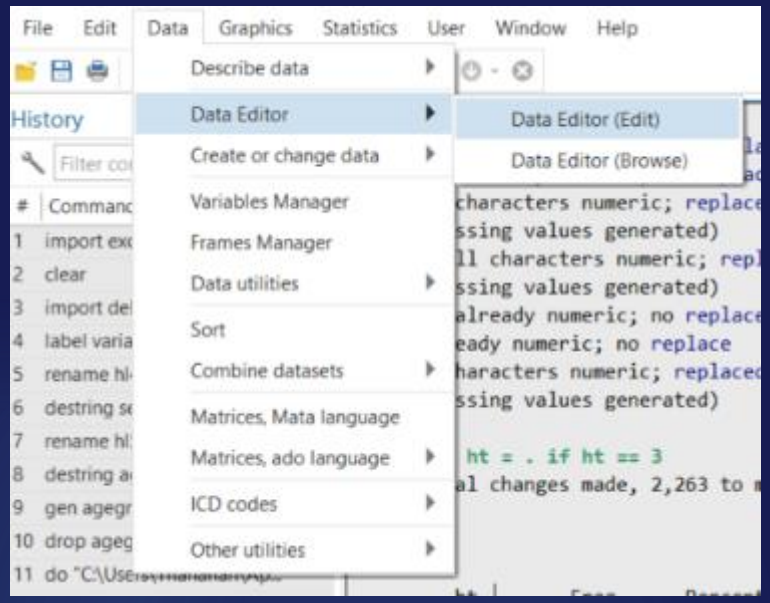
Cont. 99

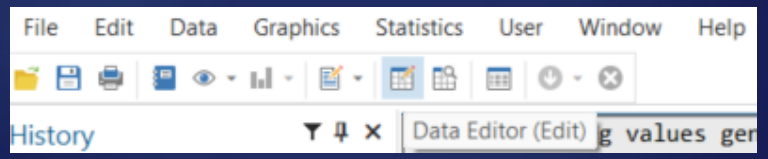
URINALYSIS	Results	ND
Urine Albumin		☐
Urine Nitrite		☐
Urine Esterase		☐
Urine Glucose		☐

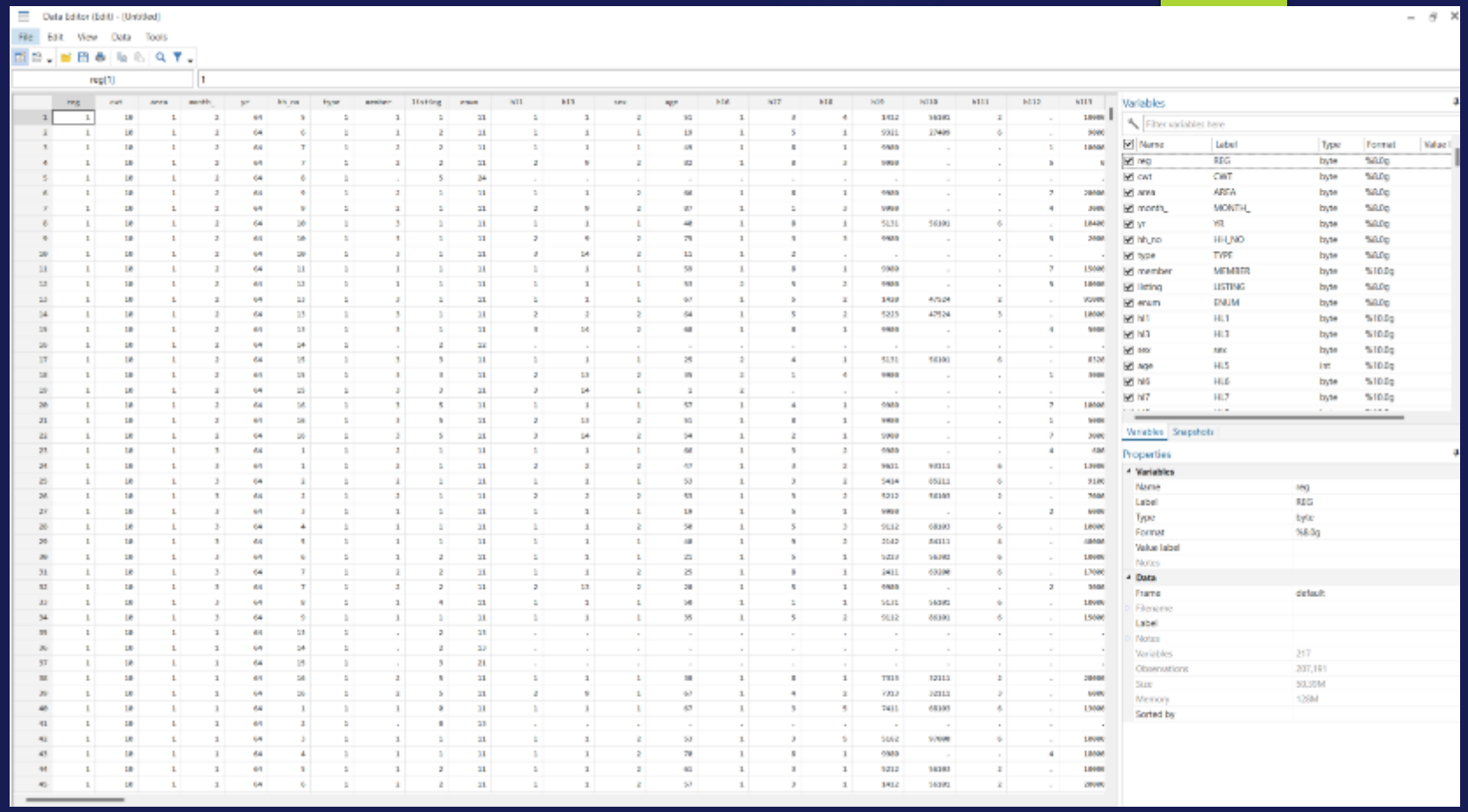
	N/A	Pos	Neg	ND
Urine Pregnancy	☐	☐	☐	☐
Urine Opiates	☐	☐	☐	☐

If "Not Done", please check ND boxes to indicate this.
N/A means "Not Applicable" (for example, Pregnancy - males).

Type of data

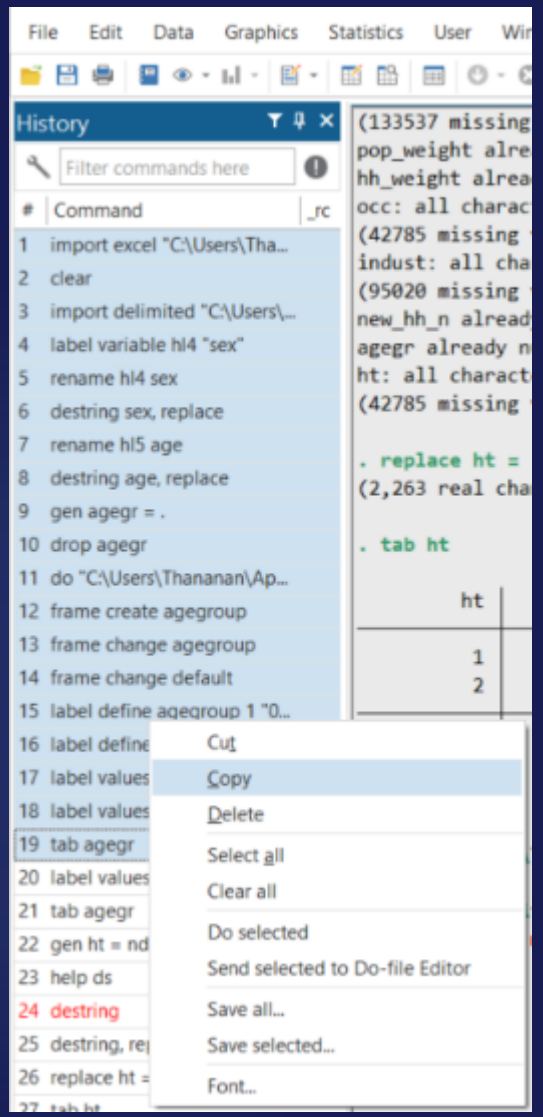


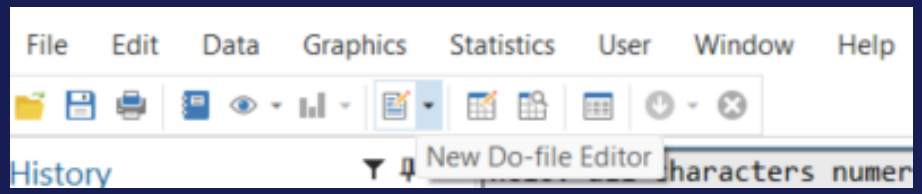


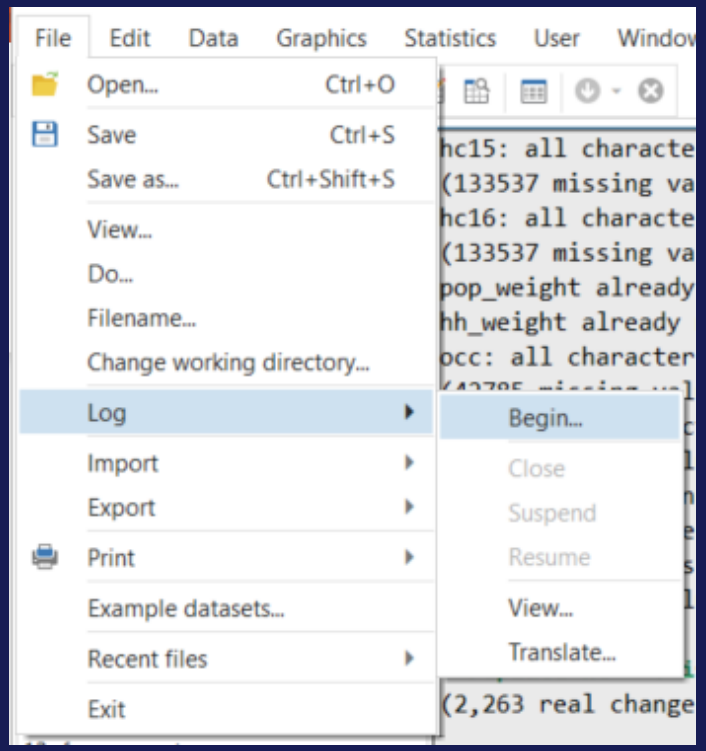


- Str2 = red / byte = black

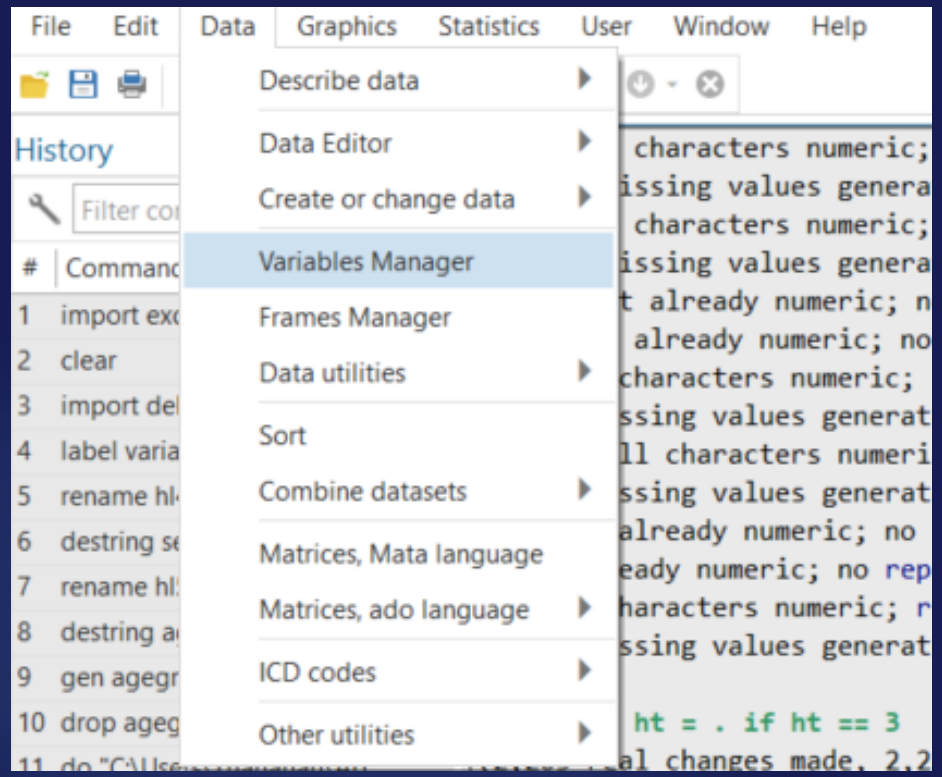
Save command in Stata

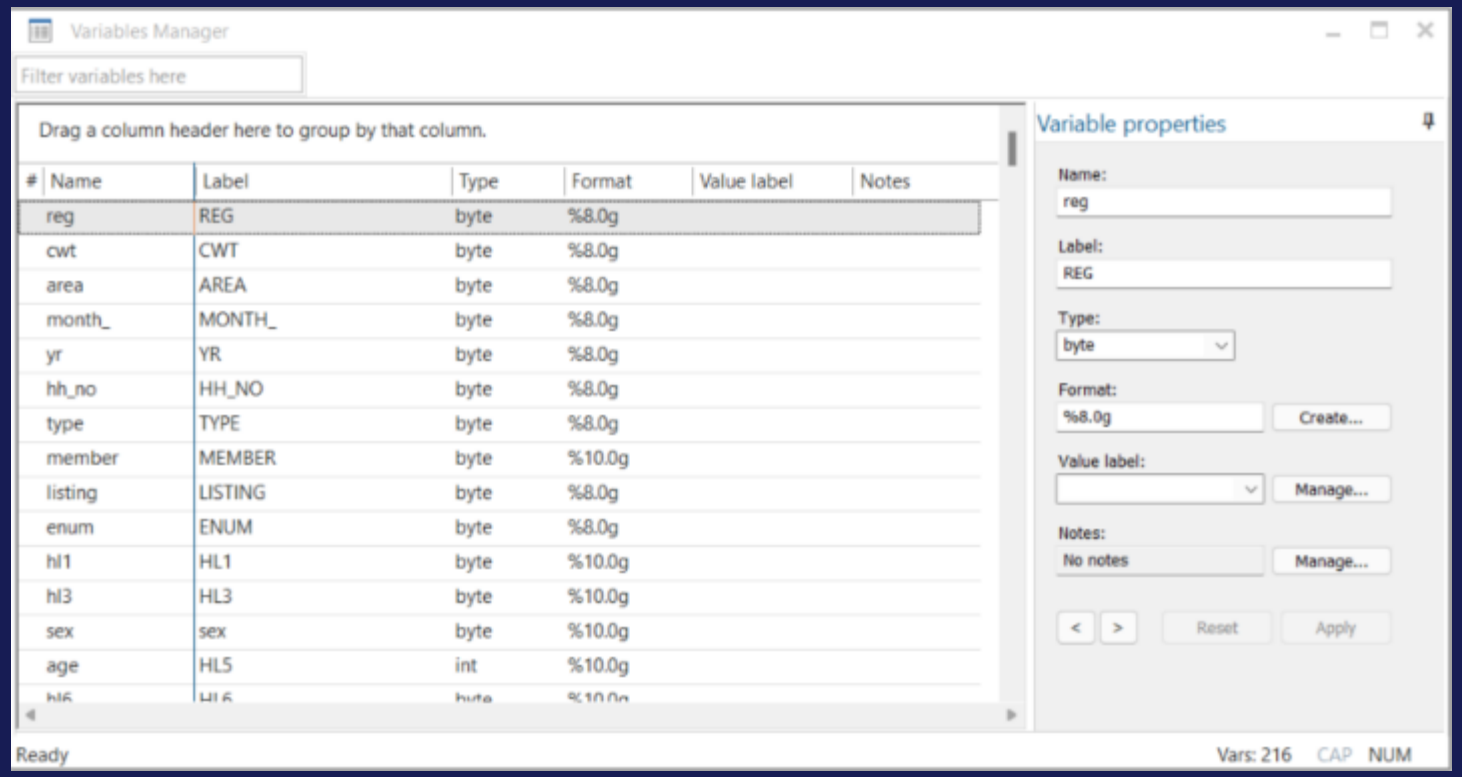






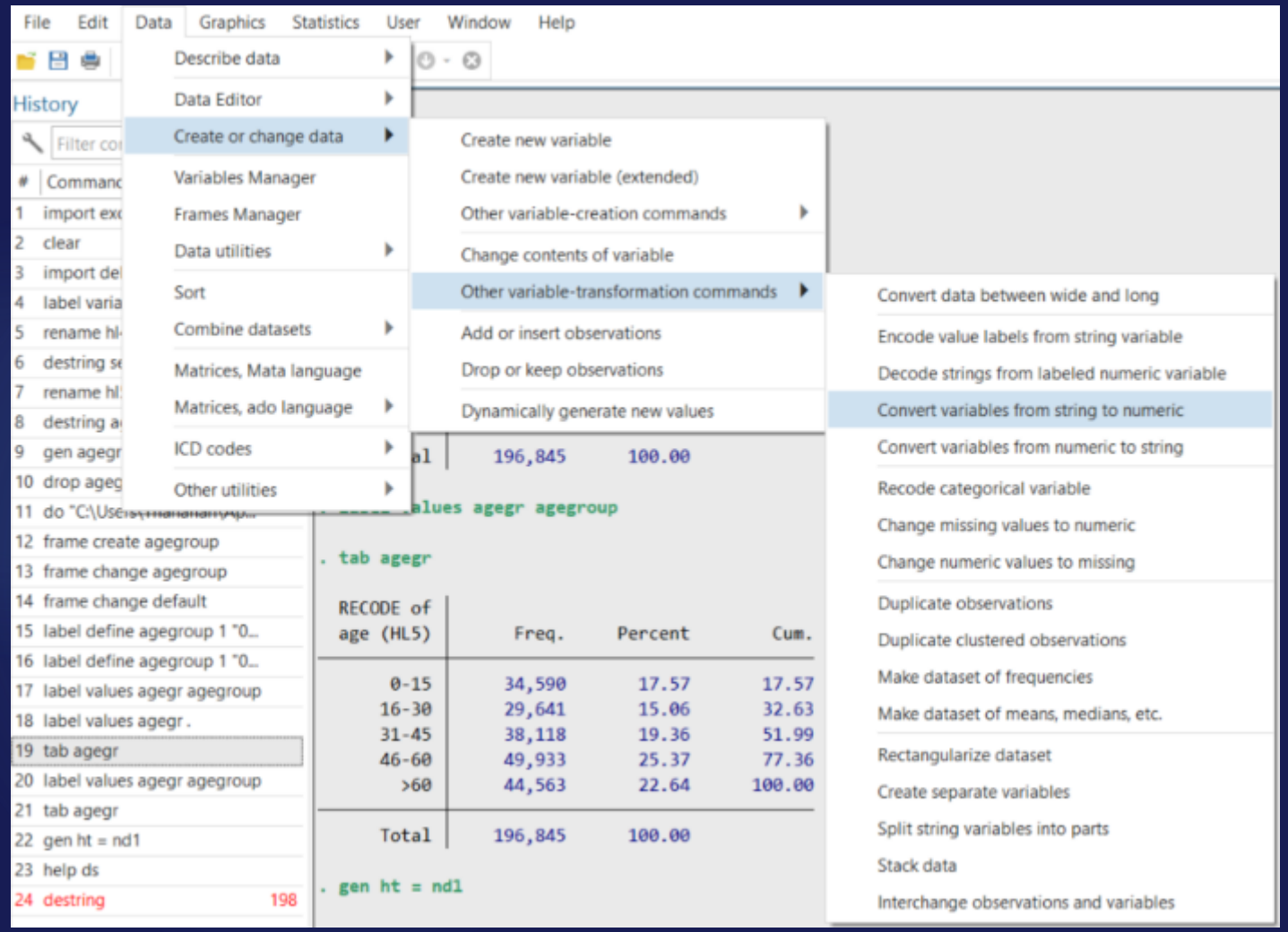
Rename variable / destring

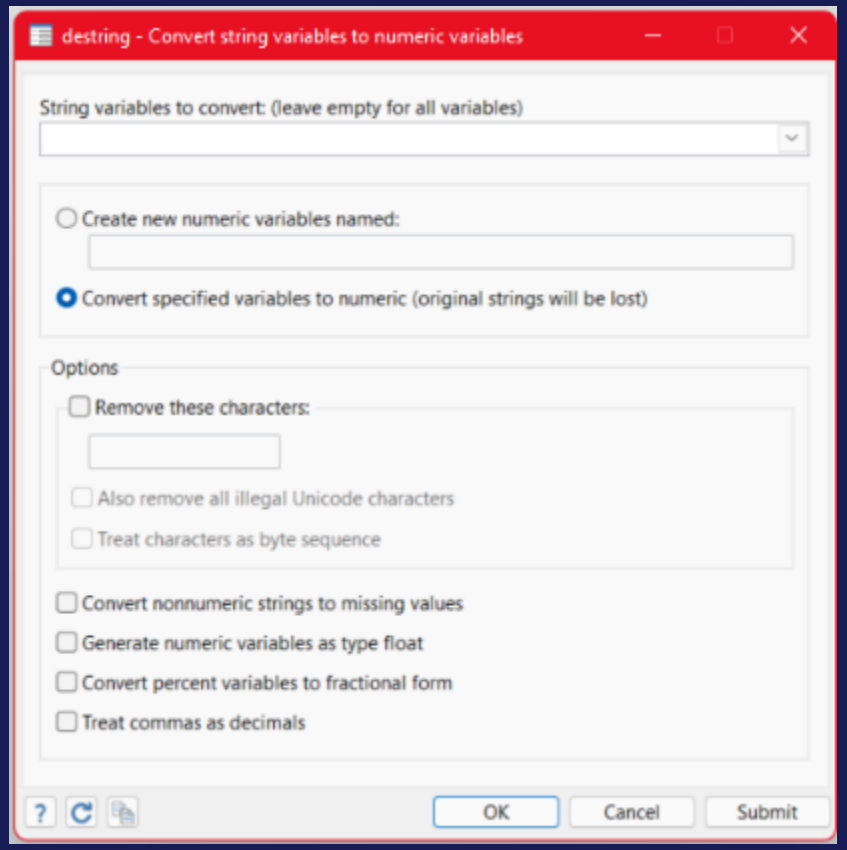





```
. label variable hl4 "sex"

. rename hl4 sex
```

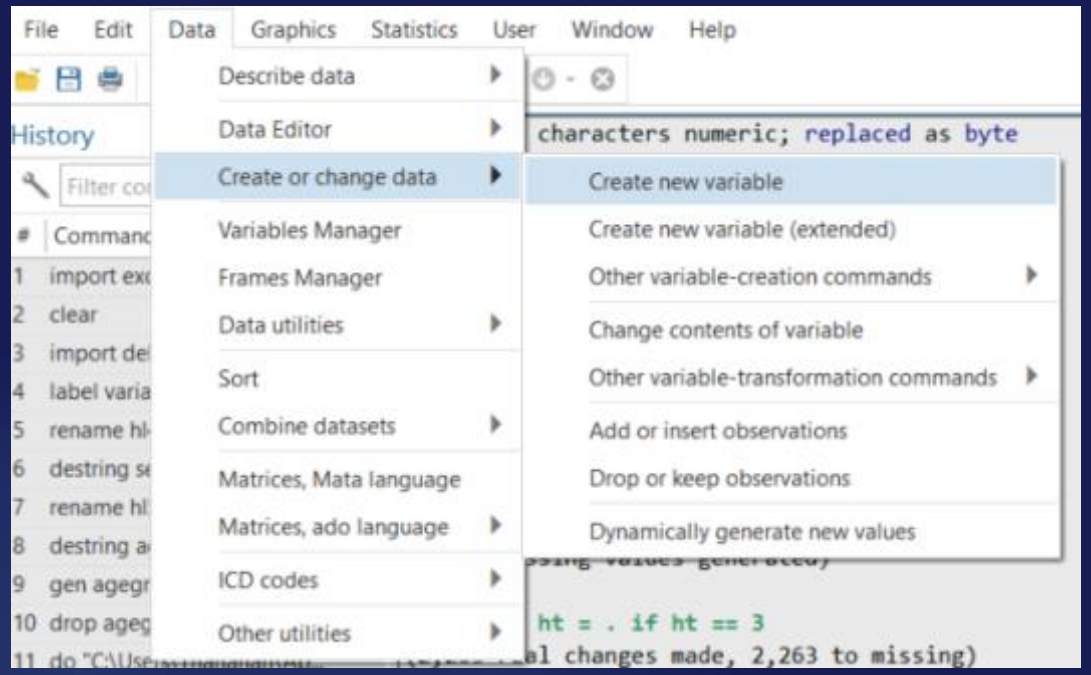
Destring data from string to numeric

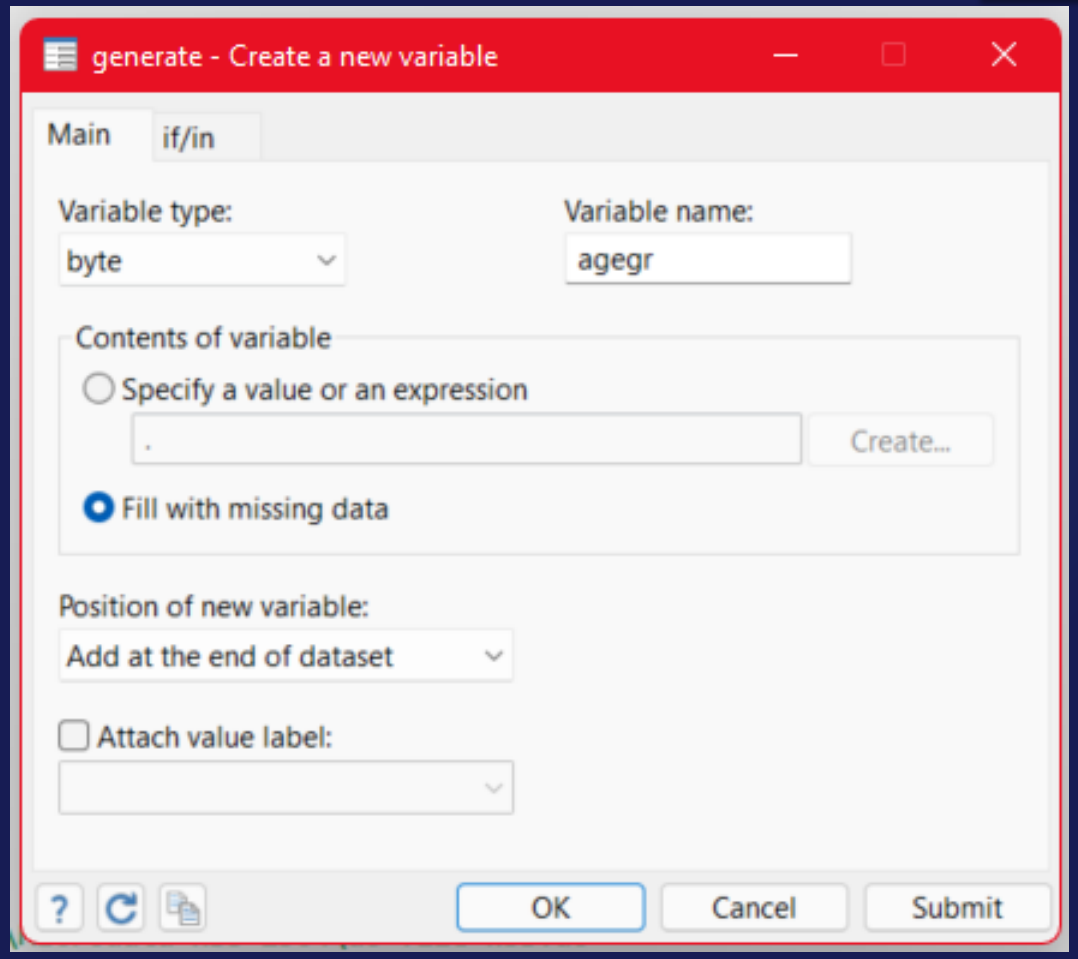





```
. destring sex, replace
sex: all characters numeric; replaced as byte
(10346 missing values generated)
```

Generate new variable

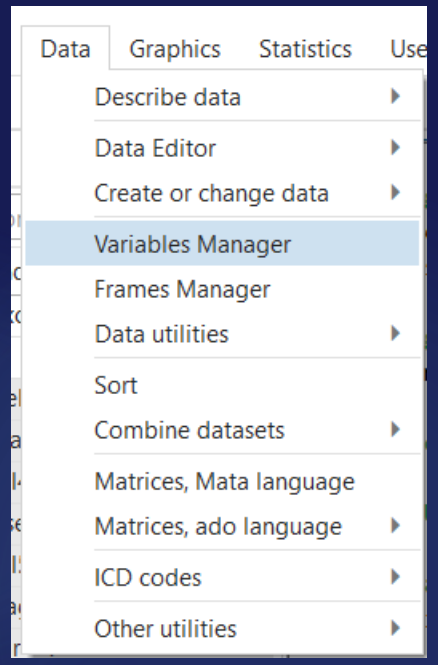


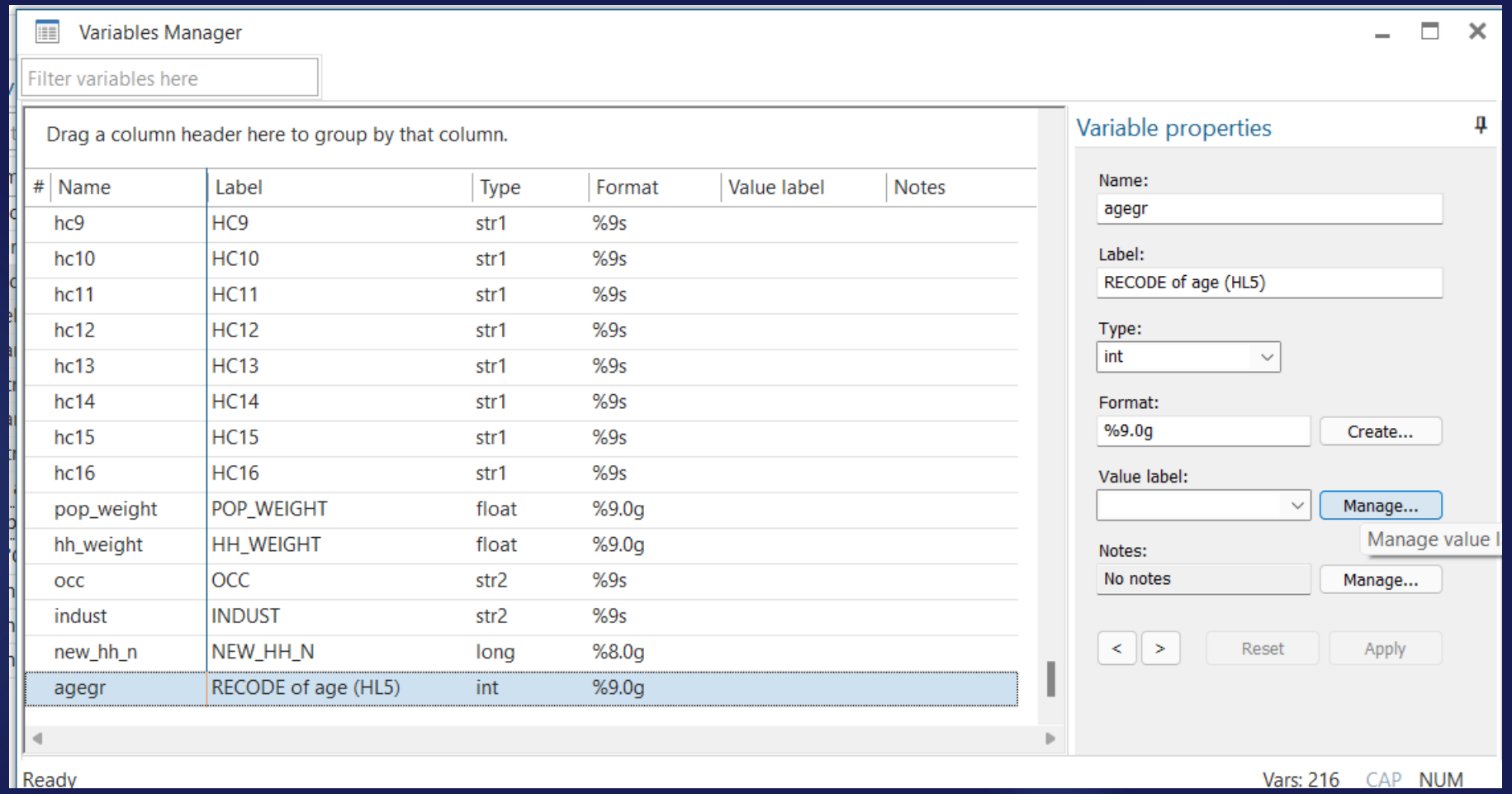



```
. recode age min/15=1 16/30=2 31/45=3 46/60=4 61/max=5, gen (agegr)
(195350 differences between age and agegr)
```

```
. gen agegr = .
```

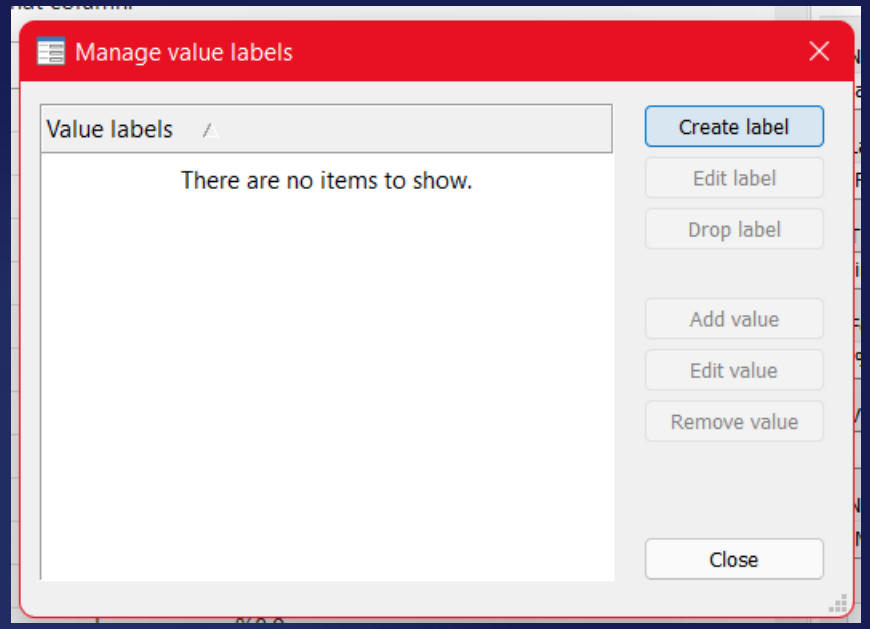
Label data in variable (1)

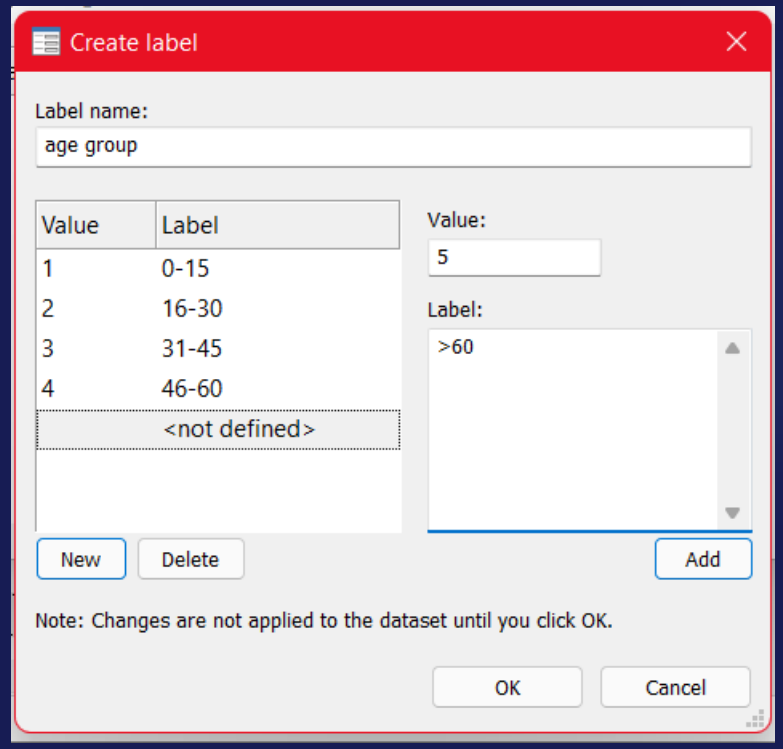




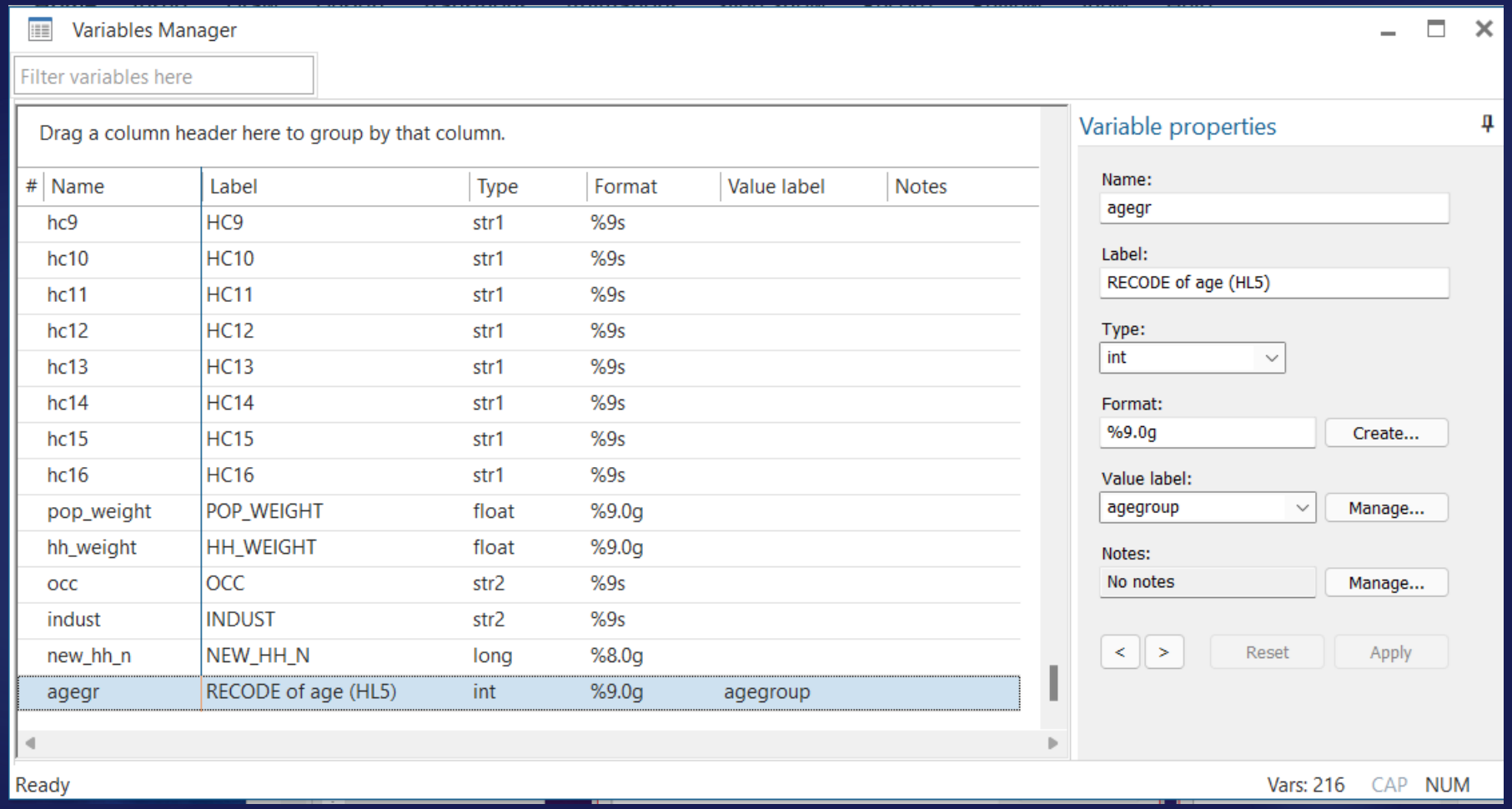
#	Name	Label	Type	Format	Value label	Notes
	hc9	HC9	str1	%9s		
	hc10	HC10	str1	%9s		
	hc11	HC11	str1	%9s		
	hc12	HC12	str1	%9s		
	hc13	HC13	str1	%9s		
	hc14	HC14	str1	%9s		
	hc15	HC15	str1	%9s		
	hc16	HC16	str1	%9s		
	pop_weight	POP_WEIGHT	float	%9.0g		
	hh_weight	HH_WEIGHT	float	%9.0g		
	occ	OCC	str2	%9s		
	indust	INDUST	str2	%9s		
	new_hh_n	NEW_HH_N	long	%8.0g		
	agegr	RECODE of age (HL5)	int	%9.0g		

Label data in variable (2)





Label data in variable (3)



#	Name	Label	Type	Format	Value label	Notes
	hc9	HC9	str1	%9s		
	hc10	HC10	str1	%9s		
	hc11	HC11	str1	%9s		
	hc12	HC12	str1	%9s		
	hc13	HC13	str1	%9s		
	hc14	HC14	str1	%9s		
	hc15	HC15	str1	%9s		
	hc16	HC16	str1	%9s		
	pop_weight	POP_WEIGHT	float	%9.0g		
	hh_weight	HH_WEIGHT	float	%9.0g		
	occ	OCC	str2	%9s		
	indust	INDUST	str2	%9s		
	new_hh_n	NEW_HH_N	long	%8.0g		
	agegr	RECODE of age (HL5)	int	%9.0g	agegroup	

Label data in variable (3)


```
. tab agegr

RECODE of
age (HL5)  |      Freq.     Percent        Cum.
-----------+-----------------------------------
         1 |     34,590       17.57       17.57
         2 |     29,641       15.06       32.63
         3 |     38,118       19.36       51.99
         4 |     49,933       25.37       77.36
         5 |     44,563       22.64      100.00
-----------+-----------------------------------
     Total |    196,845      100.00
```

```
. label values agegr agegroup

. tab agegr

RECODE of
age (HL5)  |      Freq.     Percent        Cum.
-----------+-----------------------------------
      0-15 |     34,590       17.57       17.57
     16-30 |     29,641       15.06       32.63
     31-45 |     38,118       19.36       51.99
     46-60 |     49,933       25.37       77.36
       >60 |     44,563       22.64      100.00
-----------+-----------------------------------
     Total |    196,845      100.00
```

Label data in variable (4)


```
. tab sex

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |     92,869       47.18       47.18
          2 |    103,976       52.82      100.00
------------+-----------------------------------
      Total |    196,845      100.00
```

```
. label define sex 1 "male" 2 "female", replace

. label values sex sex

. tab sex

        sex |      Freq.     Percent        Cum.
------------+-----------------------------------
       male |     92,869       47.18       47.18
     female |    103,976       52.82      100.00
------------+-----------------------------------
      Total |    196,845      100.00
```

การลบตัวแปร

- คลิกขวาที่ชื่อตัวแปร แล้วกด drop
- พิมพ์โค้ด drop ตามด้วยชื่อตัวแปร

พูดจาภาษา stata (1)

- Gen มาจาก generate หมายถึง สร้างตัวแปรใหม่
- Replace หมายถึง แทนที
- Recode หมายถึงแทนค่า
- If หมายถึง ถ้า โดย = ที่มาหลัง if ต้องเป็น ==
- & หมายถึง และ
- | (อยู่ตรง ข.ไข่) หมายถึง หรือ
- . “” หมายถึง ค่าว่าง
- != หมายถึง ไม่เท่ากับ
- + - * / ^ บวก ลบ คูณ หาร ยกกำลัง
- Clear หมายถึงล้างข้อมูลทั้งหมด
- Use หมายถึงใช้ข้อมูลหรือเปิดข้อมูล
- Disp หรือ display แปลว่า แสดงผล
- Tab หรือ tabulate หมายถึงตาราง

Ex. Gen bmi = weight/(height^2)
สร้างตัวแปรชื่อ bmi เท่ากับ
ค่าในตัวแปรชื่อ weight หาร (ค่าในตัวแปรชื่อ height ยกกกำลัง 2)

Ex. Gen diff = pre-post if age !=. & (pre-post)>0
สร้างตัวแปรชื่อ diff เท่ากับ ค่าในตัวแปรชื่อ pre ลบด้วย ค่าในตัวแปรชื่อ post เมื่อ age ไม่เท่ากับค่าว่าง และ ค่าใน pre ลบด้วย post มากกว่า 0

Ex. Replace age = . If age == 999 | age > 100

แทนที่ ตัวแปรชื่อ age ให้เท่ากับค่าว่าง ถ้า ตัวแปรชื่อ age เท่ากับ 999 หรือมากกว่า 100

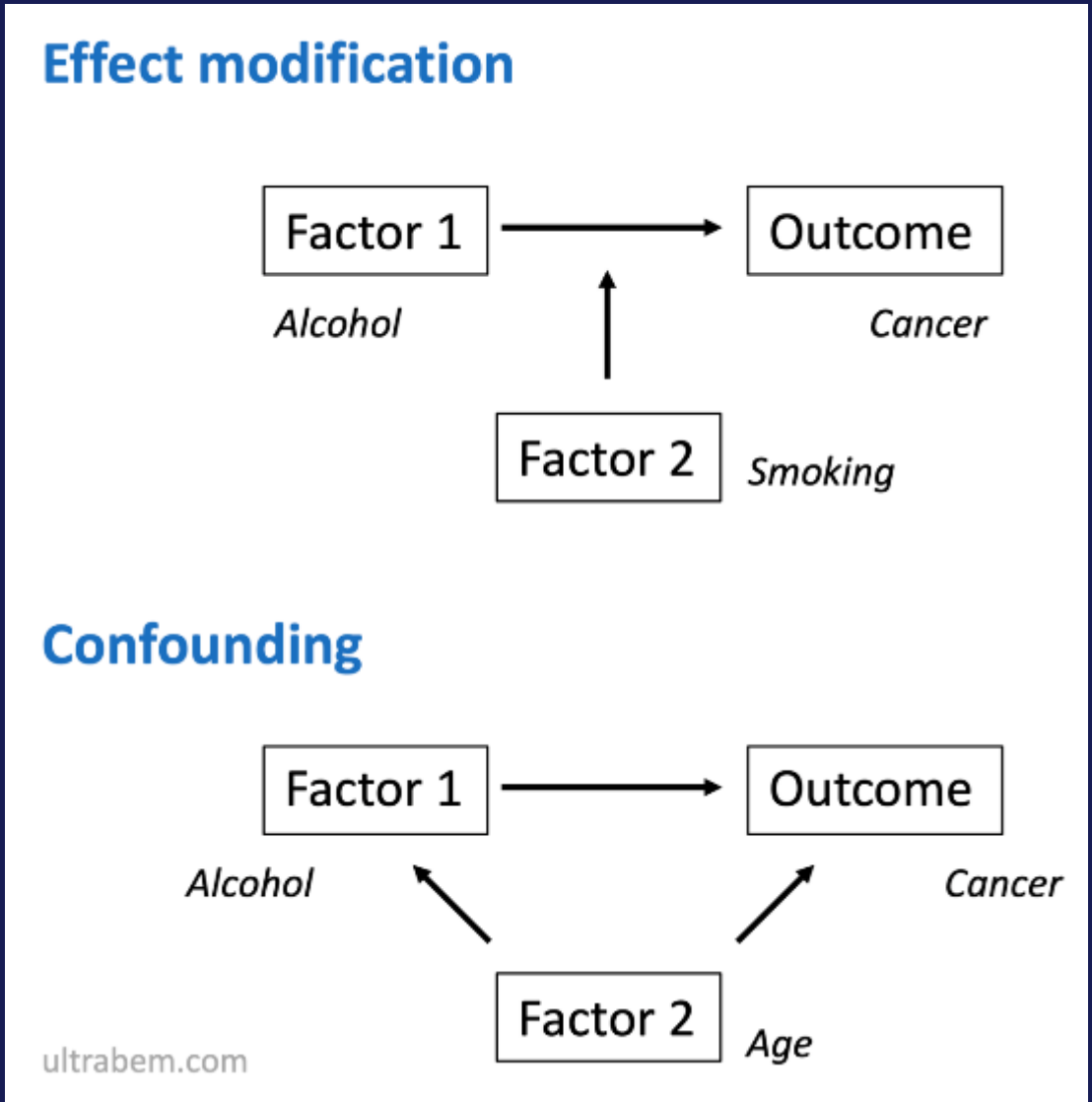
Ex. Recode age min/20=1 21/60=2 61/max=3, gen (agegr)
แปลงค่า age ค่าน้อยสุดถึง20เท่ากับ 1 ค่า21ถึง60เท่ากับ 2 ค่า61ถึงสูงสุดเท่ากับ 3
สร้างเป็นตัวแปรชื่อ agegr

Ex. Disp 52/5*100 แสดงผล 52 หารด้วย 5 คูณ 100

Ex. Tab var1 var2 สร้างตาราง var1 เป็น row และ var2 เป็น

Terminology

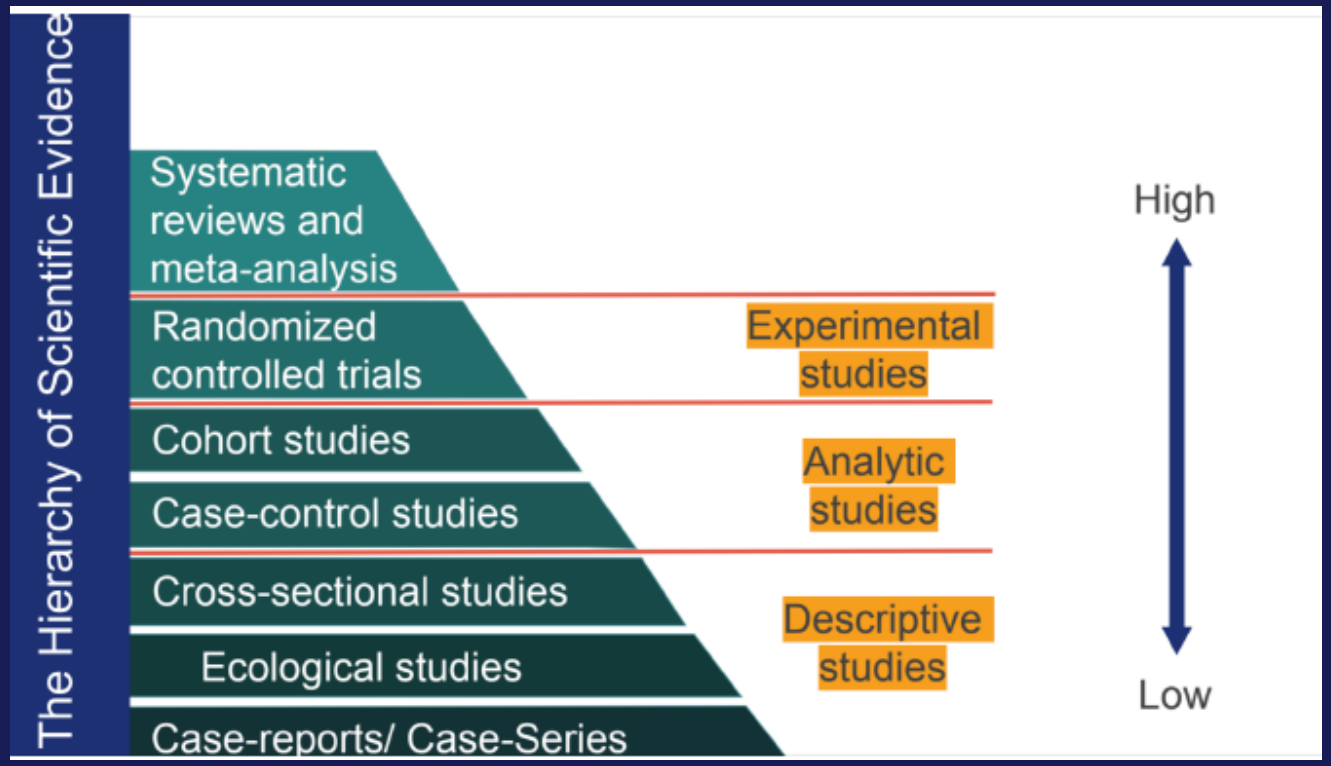
- Constants ค่าคงที่
- Variable ตัวแปร
- Independent ตัวแปรตาม
- Dependent ตัวแปรต้น
- Extraneous ตัวแปรภายนอก
- Confounding
  - มีความสัมพันธ์กับทั้ง predictive และ outcome variable โดยต้องเป็น cause of outcome แต่ไม่ได้มี predictor เป็น cause (ไม่ได้เป็น intermediate) หลังจาก adjust ต่างจาก crude
- Effect modifier/Interaction
  - ทำให้ ความสัมพันธ์ ระหว่าง predictor กับ outcome เสีย homogeneity
  - stratified specific risk ratio แตกต่างกัน



Study Design

Type of study

- **Observational**
  - Descriptive (Parameter Estimation)
    - Cross-sectional
  - Analytic (Hypothesis testing)
    - Cross-sectional
    - Case-control
    - Cohort
- **Experimental**
  - True experimental >> RCT, non-RCT
  - Quasi experimental



Observational study

	Case	Non-case	Total
Exposed	A	B	A+B
Not exposed	C	D	C+D
Total	A+C	B+D	A+B+C+D

Cohort
Exposure □ Outcome ปัจจัยไปหาผลลัพท์
บอก causality ได้ ศึกษาได้ทั้งไปข้างหน้าและย้อนหลัง
ใช้ศึกษากับ outcome ที่เป็น incidence
เริ่มศึกษาจากตอนที่ยังไม่เริ่มป่วย

Smoking (y/n)□ Cancer (y/n)
Diabetes(y/n)□ Chronic kidney disease

Case-Control
Outcome □ Exposure
บอก causality ได้
ใช้ศึกษากับ outcome ที่เป็น prevalence

Cancer (y/n)□ Smoking (y/n)
Stroke after surgery(y/n)□ surgery time

Cross-sectional
Exposure <-> Outcome
บอกได้แค่สัมพันธ์กันหรือไม่ บอกค่าเป็นเหตุผลไม่ได้

Smoking (y/n)<-> Cancer (y/n)
Depress <-> Quality of life

Cross-sectional

- Relationship of Exposure <-> Outcome
- Prevalence Odds ratio

Population at risk → Sampling → Representative sample →

	Case	Non-case	Total
Exposed	A	B	A+B
Not exposed	C	D	C+D
Total	A+C	B+D	A+B+C+D

Cross-sectional

- Single point in time (Snapshot) -> Exposure and Outcome
- measured at one point in time or over a period
- No Follow up
- ทำง่าย ใช้เวลาน้อย
- ข้อมูล Individual
- Measure of frequency : **Prevalence**
- Measure of association: **Prevalence ratio (PR),Prevalence Odds ratio(POR)**
- ใช้กับ โรคที่ไม่ทราบ onset ex. โรคเรื้อรัง
- ได้ข้อมูล POR <- Overestimate risk มากกว่า PR
- No Temporal sequence **บอกได้แค่ว่ามีAssociation ไม่ได้บอกว่าอะไรเป็นสาเหตุ**
- ใช้กับRare exposure หรือ Rare outcome ยาก
- Prevalence-incidence bias (Length Bias / Survival bias)-> เจอโรคที่มี long durationได้มากกว่า

Case-control

	Case	Non-case	Total
Exposed	A	B	A+B
Not exposed	C	D	C+D
Total	A+C	B+D	A+B+C+D

Sampling

Case population

Non-case population

- **causality**
- **Outcome □ Exposure**
- **Odds ratio**

Target population
V
Source population
V
Eligible population(Characteristics)
V
Study Participants

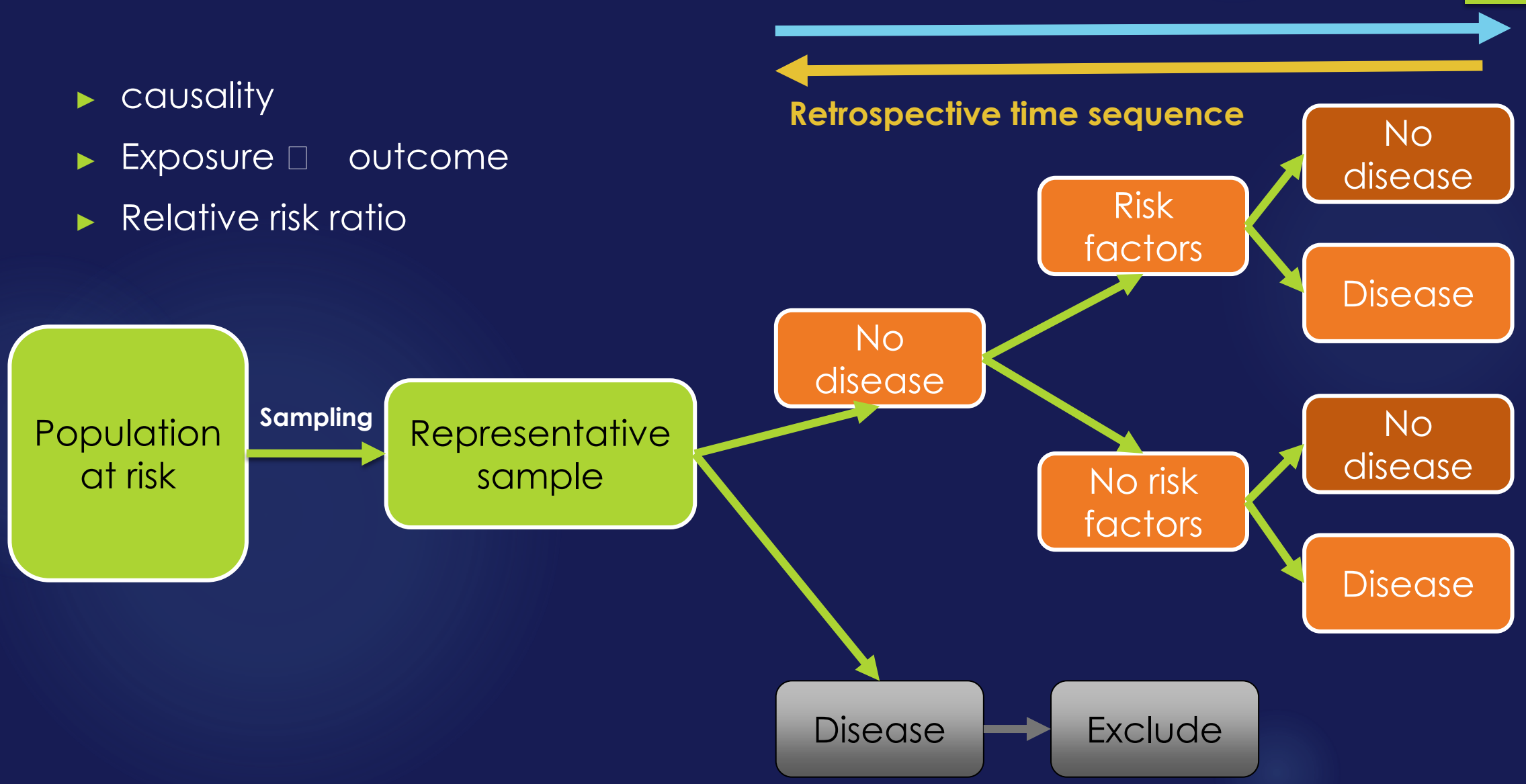
Case-control

- Pro
  - เลือกMultiple Exposureได้
  - ใช้กับRare Diseaseได้
  - ไม่มี fellow up
- Cons
  - บอก Casual Temporalityได้
  - Uncertain temporal sequence (ถ้ามี Information bias)
- Measure of frequency: **Prevalence**
- Measure of association: **Odds Ratio**

- Bias ที่เกิดขึ้นบ่อย
  - Selection bias
    - Participants not represent Population
  - Information bias
    - Memory bias -> Can't remember -> Misclassification exposure
    - Recall bias -> Case try to recall more than control ต้องมีcomparison group
      ex. แม่มีลูกเป็นโรคจะrecallมากกว่าแม่มีลูกปกติ
    - Interview bias -> ex. Review caseมากกว่าControl

Cohort

- causality
- Exposure □ outcome
- Relative risk ratio



Cohort

- Disease-free Population follow up over time
- ติดตามได้ Incident (New case) / AR
- Exposure -> Follow -> Outcome
  - Fixed Exposure (ex Blood group),
  - Time-dependent Exposure (ex Blood Sugar)
  - Hard Outcome ex. Death , Disease
  - Intermediate Outcome ex. CD4
- Comparison group ต้องมี as similar as possible
- ดีสุดต้องInternal Comparison (เปรียบเทียบในCohortเดียวกัน) แต่ถ้าไม่ได้ก็ใช้ Comparison Group (External Comparison)

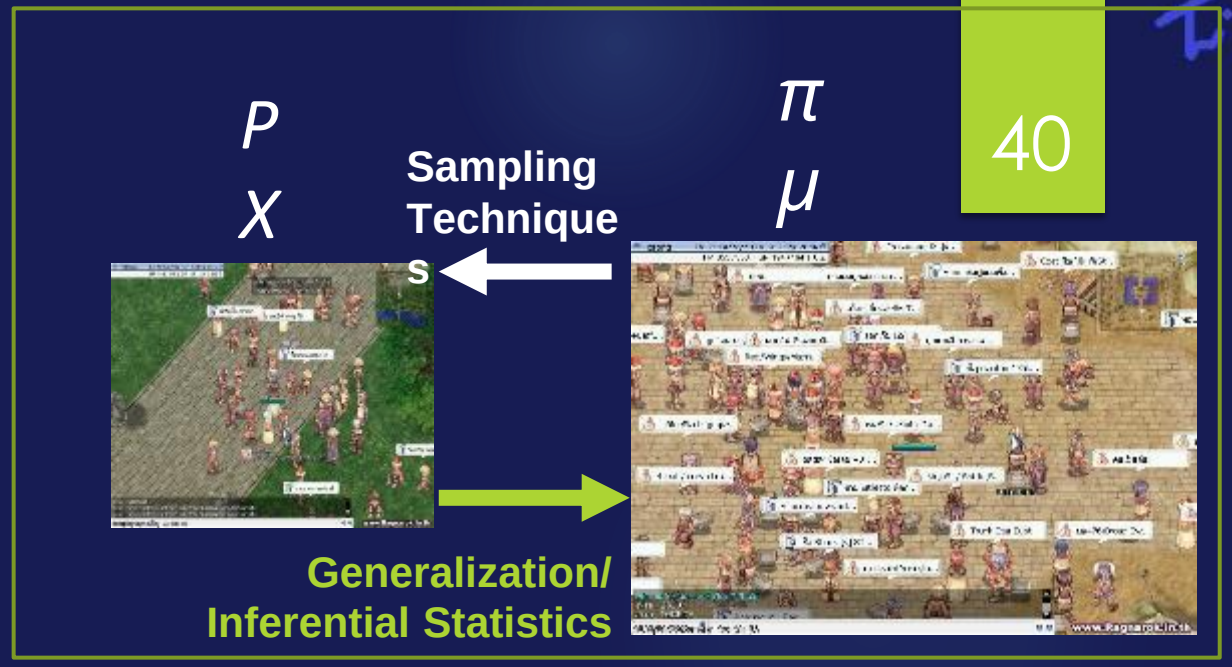
- Pro
  - ใช้กับRare Exposure, multiple outcome,
  - บอก temporal sequenceได้(Exposureเกิดก่อนOutcome)
  - บอกCause -> Effect Relation (Temporality) ได้ดีกว่า Case-Control
- Con
  - แพง ใช้เวลามาก
  - Loss Follow up
- Measure of frequency : **Incidence**
- Measurements Association : **Relative Risk**

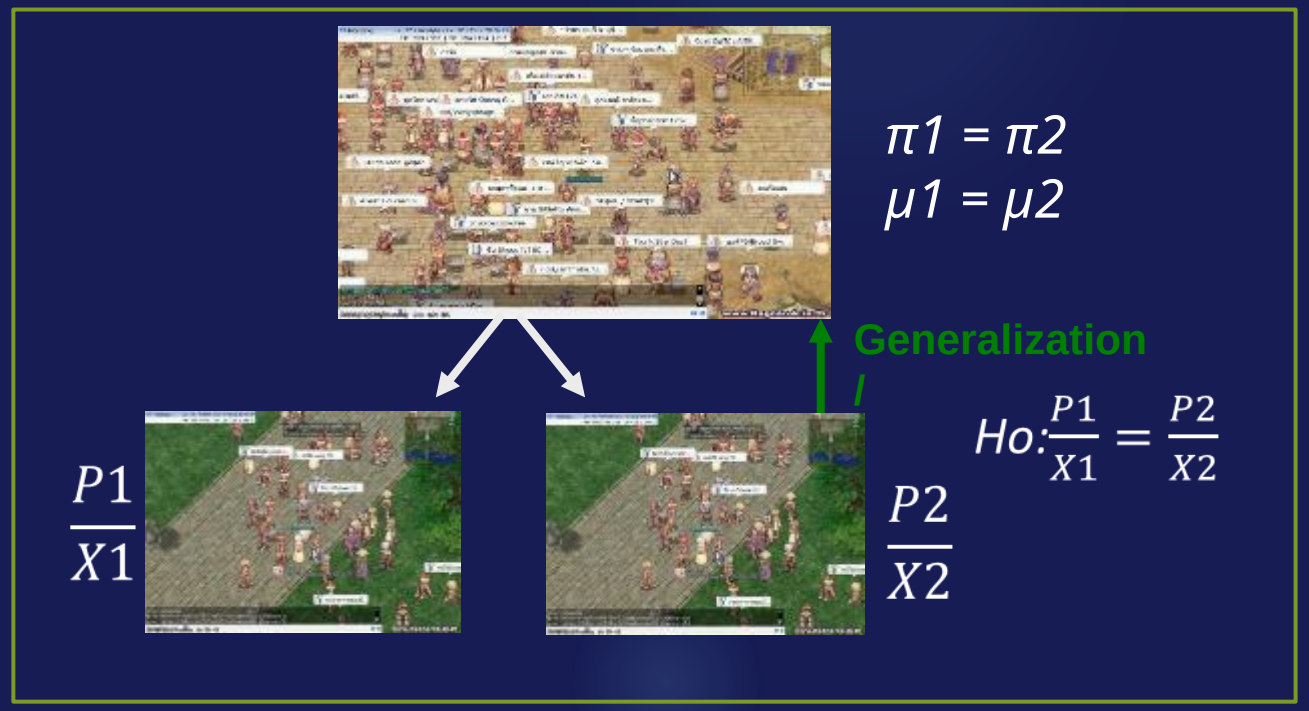
Types of statistic

Types of statistic

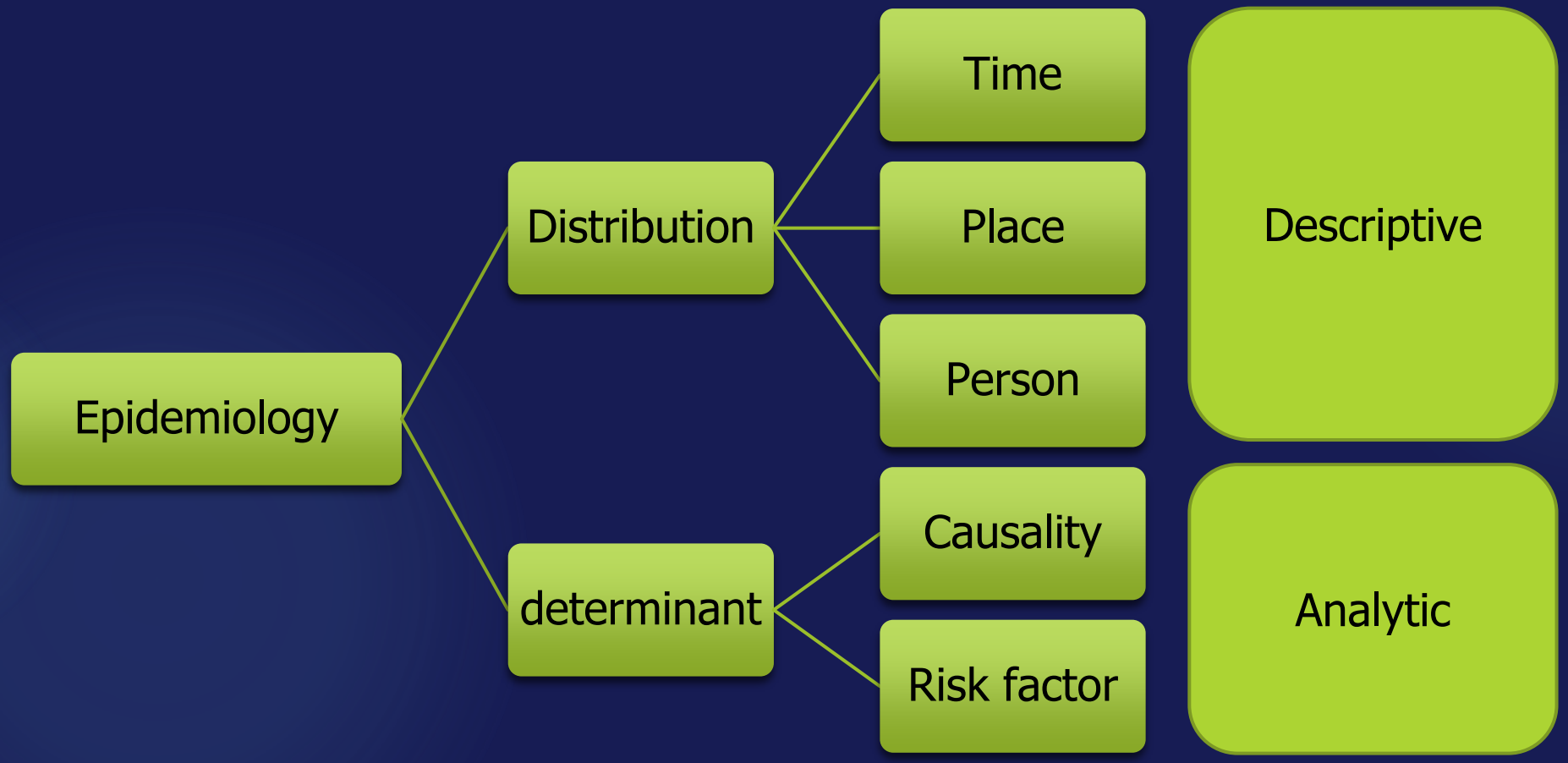
By Level of Generalization

- Descriptive Statistics
- Inferential Statistics
  - Parameter Estimation
  - Hypothesis Testing
    - Comparison
    - Association
    - Multivariable data analysis
    - Multivariate data analysis

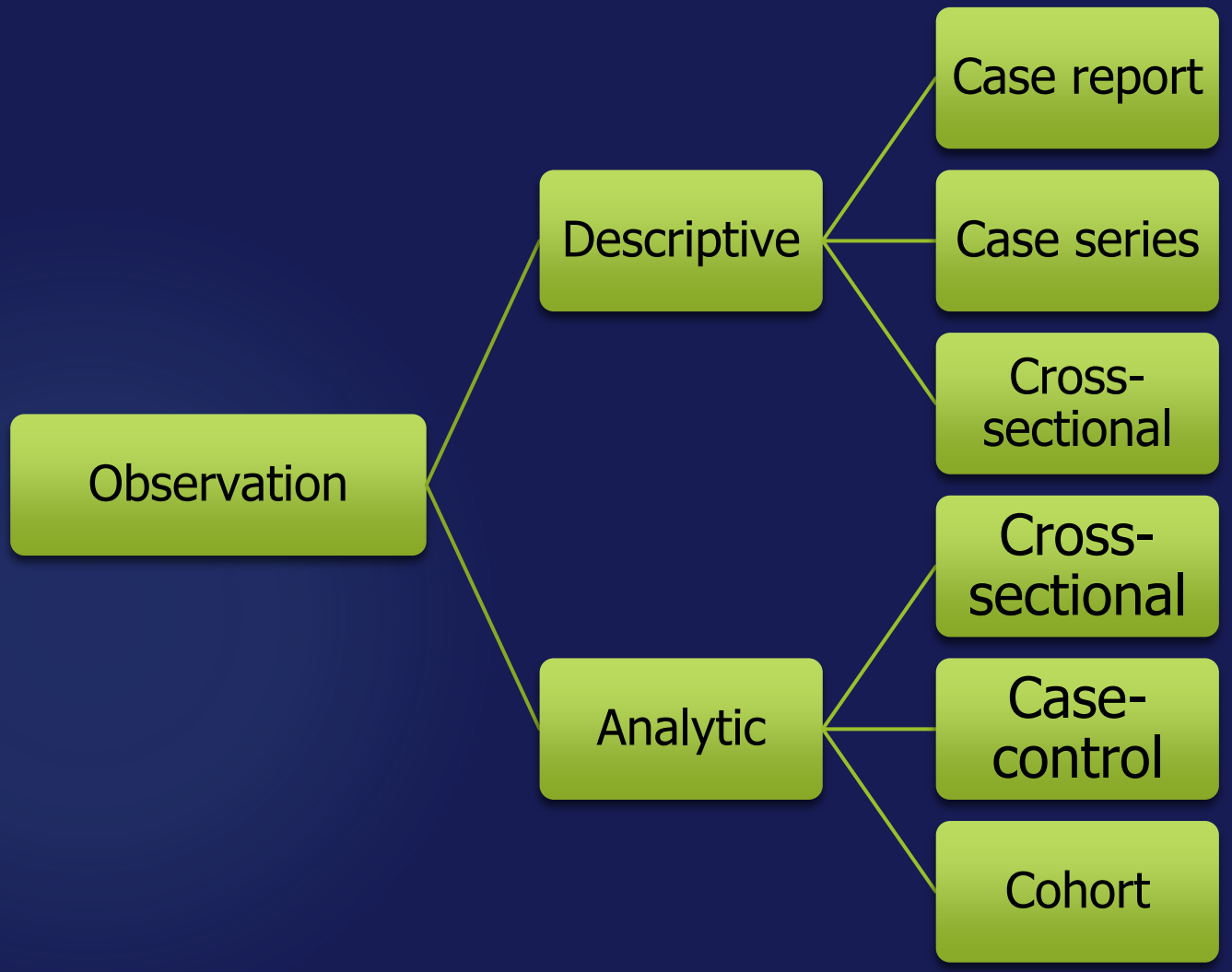




Epidemiological study design



Epidemiological study design

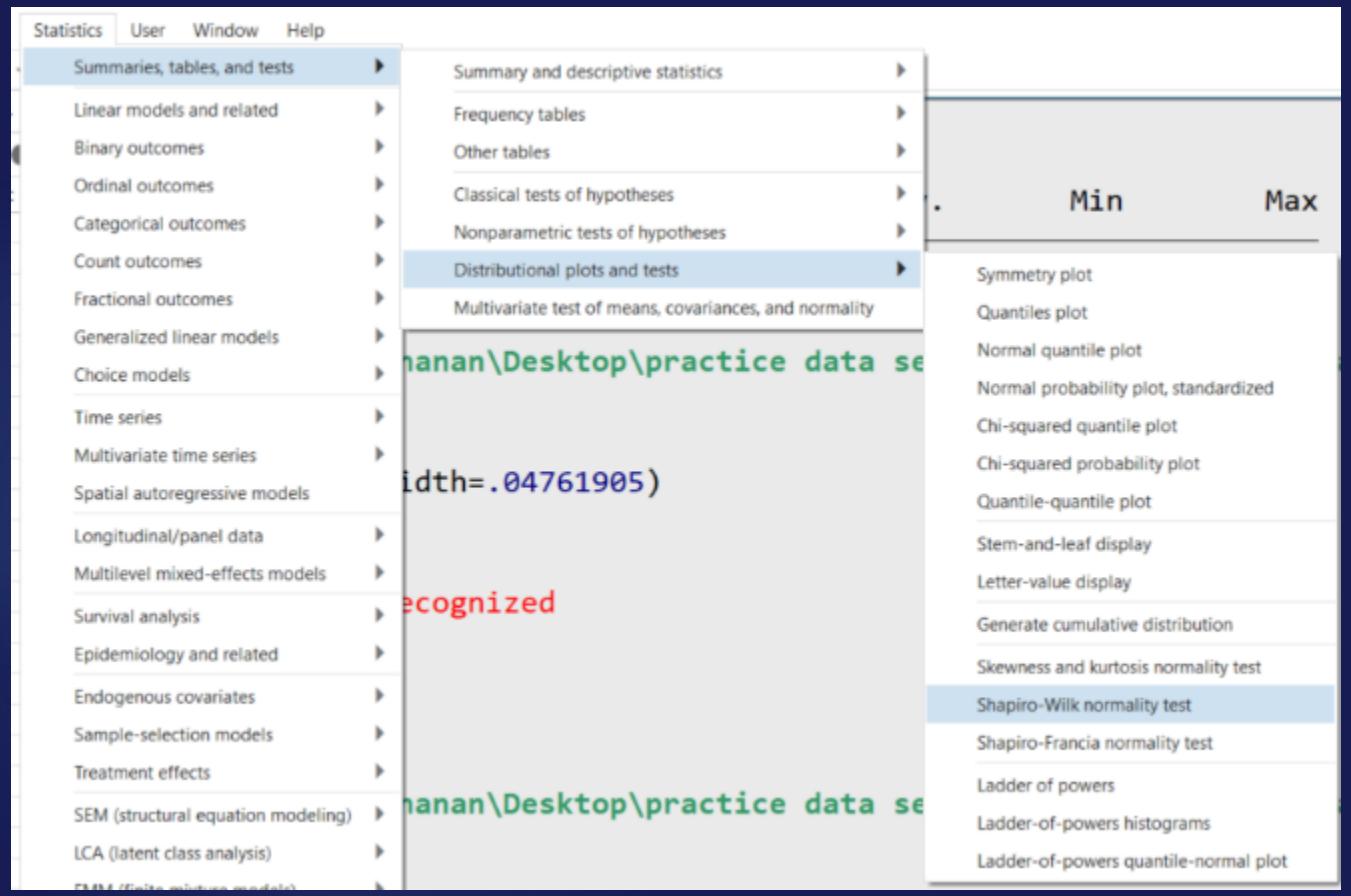


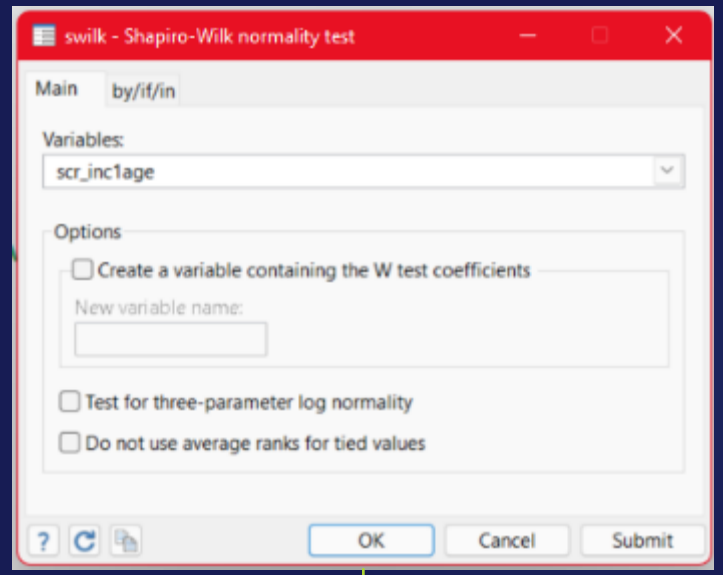
Descriptive study

Describe sample statistics

- Categorical Variables
  - Ratio, Proportion, Percent (%)
- Continuous Variable
  - Normal Distribution
    - Mean, SD
  - Not normal Distribution
    - Median, Range/IQR

Normal distribution testing : Shapiro-Wilk test



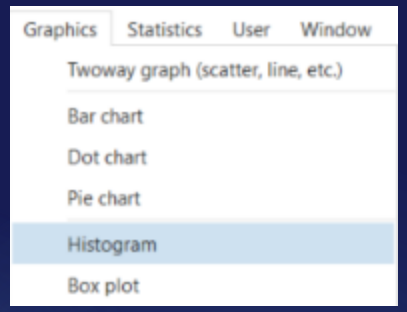


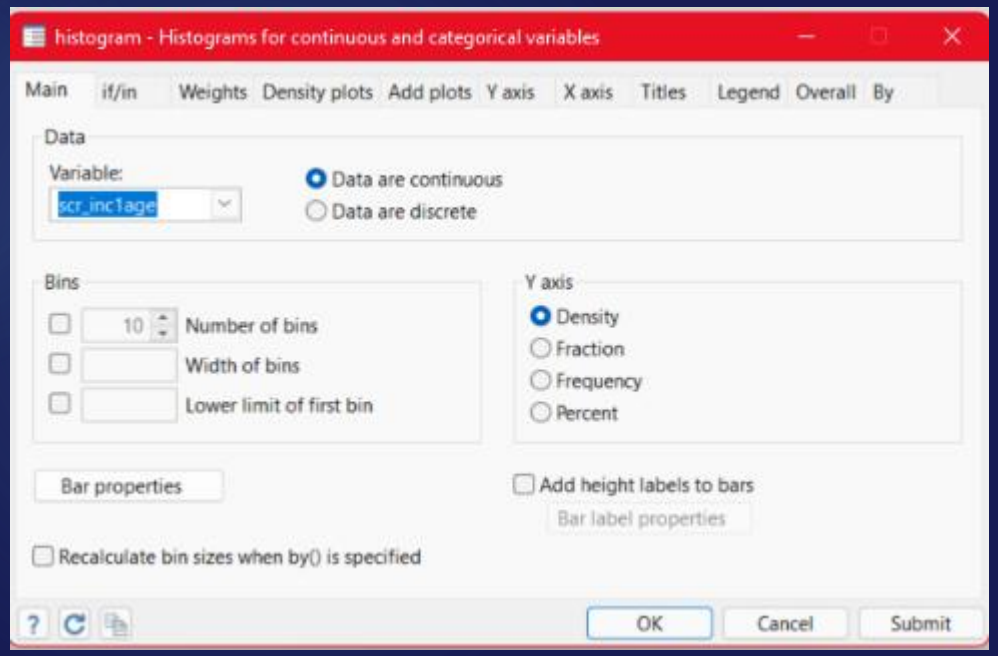

```
. swilk scr_inc1age

                    Shapiro-Wilk W test for normal data
```

Variable	Obs	W	V	z	Prob>z
scr_inc1age	480	0.98312	5.475	4.079	0.00002

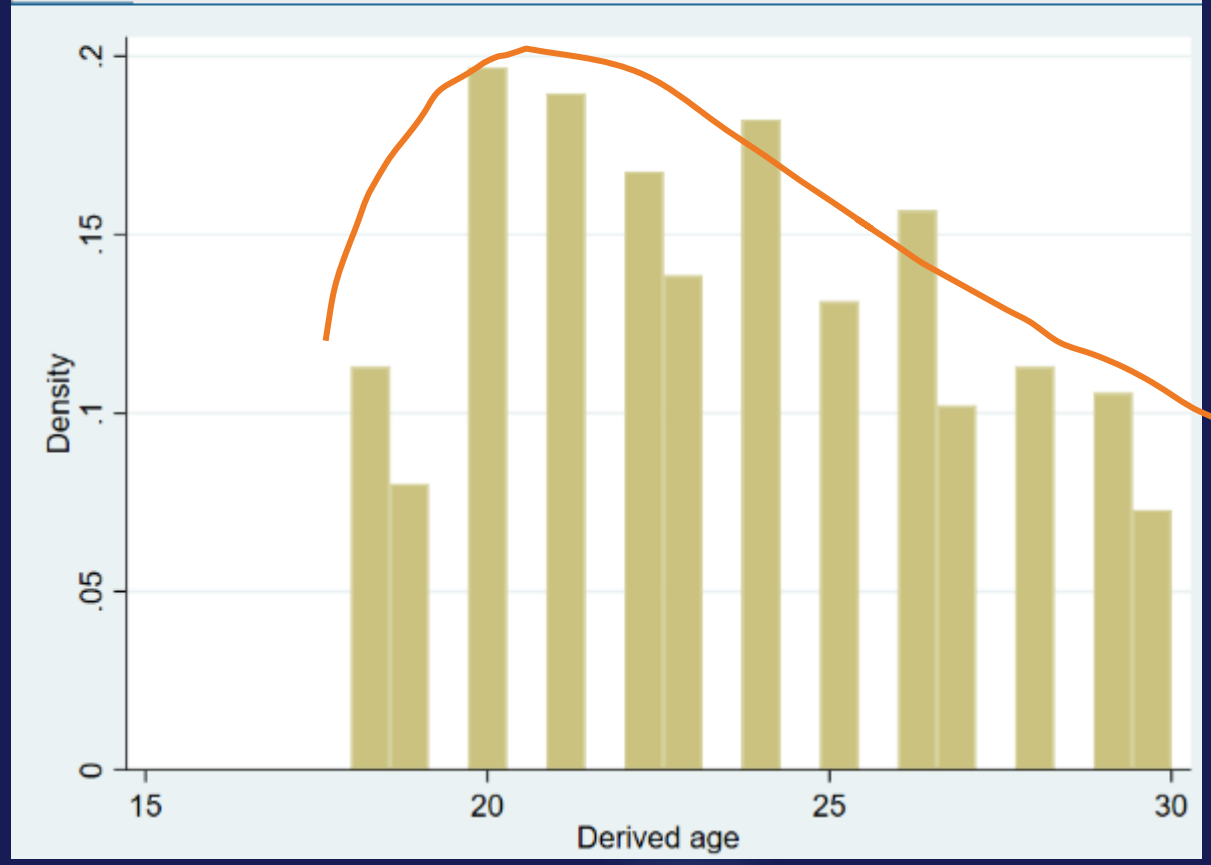
If P-value > 0.05 แปลว่า เป็น normal distribution

Normal distribution testing : Histogram



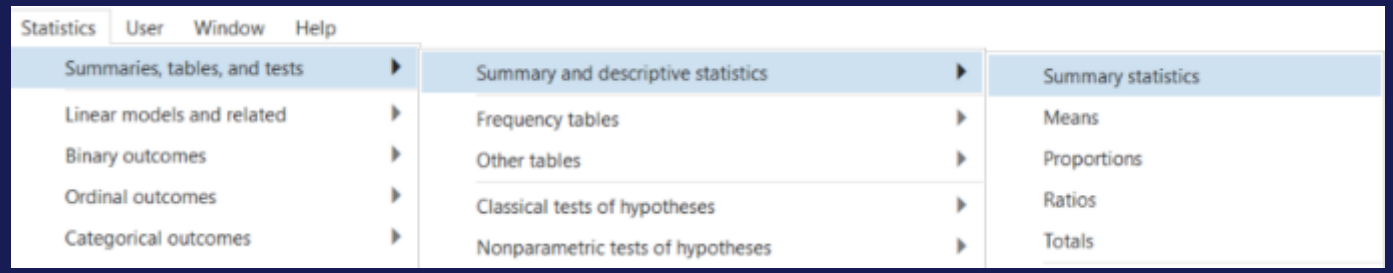


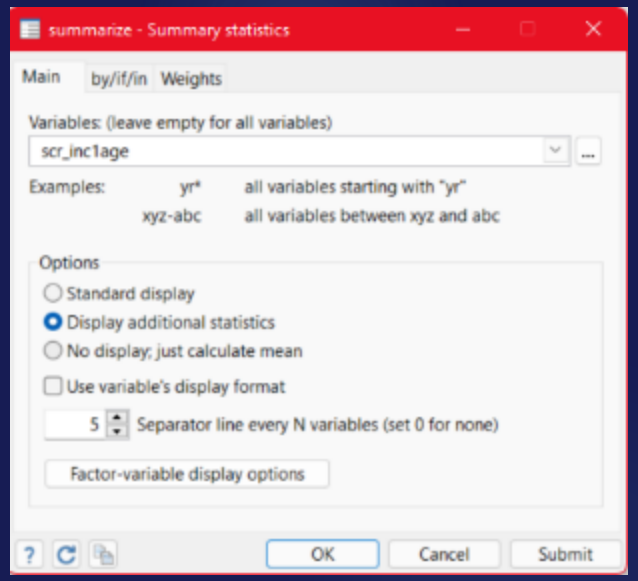

```
. clear

. use "C:\Users\Thananan\Desktop\practice data sets\hiv vaccine trial.dt

. hist scr_inc1age
(bin=21, start=18, width=.57142857)
```



ไม่ใช่โค้งปกติ เบ้ขวา

Continuous Variable






```
. sum scr_inc1age, detail

                        Derived age
-------------------------------------------------------------
      Percentiles      Smallest
 1%           18             18
 5%           18             18
10%           19             18       Obs                 480
25%           21             18       Sum of Wgt.         480

50%           23                      Mean           23.57708
                        Largest       Std. Dev.      3.399358
75%           26             30
90%           29             30       Variance       11.55563
95%           29             30       Skewness       .1815262
99%           30             30       Kurtosis        1.99793
```

ไม่ใช่ normal distribution

- Median = 23
- Q1 = 21
- Q3 =26
- IQR = Q3-Q1= 5

Normal distribution

- Mean = 23.58
- S.D. = 3.40

Measures in epidemiology

การวัดทางระบาดวิทยา

Aims of **Epidemiologic** Research

1. Describe → Measure of Frequency
    a. How common of CHD among adults in Province A?
    b. What is the frequency of CHD among males and females?
2. Explain → Measure of Association
    a. Why men are more likely than women to develop CHD?
    b. Does smoking increase the risk of CHD?
3. Predict → Measure of Impact
    a. How many CHD cases would occur if we provided a specific intervention?
    b. How many new CHD cases will occur in province A next year?
4. What could be done to prevent new cases? And how?

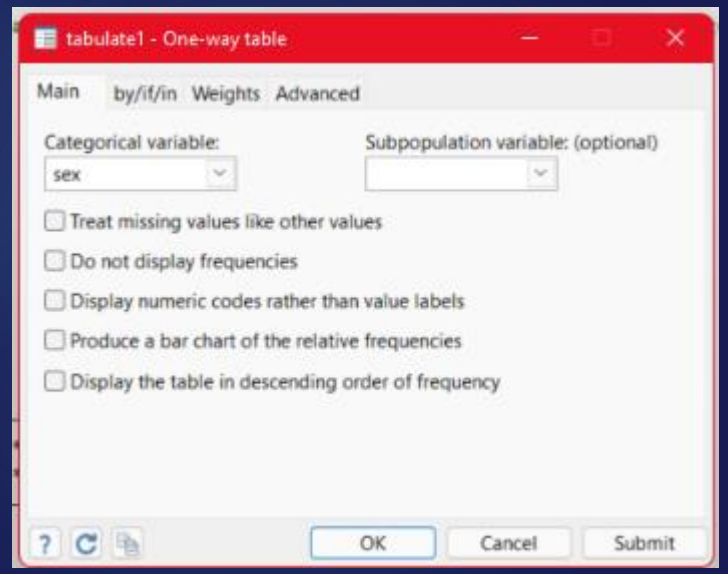
Measures of frequency

การวัดความถี่/การกระจาย

Measure of FREQUENCY

- Ratio
  - relative magnitude of two quantities or a **comparison of any two values**
  - The numerator and denominator need **not be related**
  - **=A/B**
- Proportion
  - the **comparison of a part to the whole**
  - It is a type of ratio in which the **numerator is included in the denominator**
  - **=A/(A+B)**
- Rate
  - Measure an event occurs in a defined population **over a specified period of time**
  - **= A/time**

Proportion สัดส่วน ตัวตั้ง **เป็น** ส่วนหนึ่งของ ตัวหาร




```
. tab sex
```

sex	Freq.	Percent	Cum.
male	92,869	47.18	47.18
female	103,976	52.82	100.00
Total	196,845	100.00	

Ratio อัตราส่วน ตัวตั้ง ไม่ใช่ ส่วนหนึ่งของตัวหาร

sex	Freq.	Percent	Cum.
male	92,869	47.18	47.18
female	103,976	52.82	100.00
Total	196,845	100.00	

```
. display 103976/92869
1.1195986
```

Female : Male = 1.12 : 1

Rate, Ratio or Proportion ?

Indicator	ตัวตั้ง Numerator	ตัวหาร Denominator	Rate/Ratio/Proportion
General fertility rate อัตราเจริญพันธุ์ทั่วไป	จำนวนเด็กเกิดมีชีพทั้งหมด	จำวนผู้หญิงอายุ 15-49 ปี	Ratio Ratio
Infant mortality rate	จำนวนทารกที่ตายในชวบปีแรก	จำนวนเด็กเกิดมีชีพทั้งหมดในปีนั้น	Proportion
Case-fatality rate อัตราป่วยตาย	จำนวนผู้ป่วยที่เสียชีวิต	จำนวนผู้ป่วยทั้งหมดในช่วงเวลานั้น	Proportion
Mortality rate อัตราตาย	จำนวนผู้เสียชีวิตในช่วงเวลานั้น	จำนวนประชากรทั้งหมดในช่วงเวลานั้น	Proportion
Attack rate อัตราป่วย	จำนวนผู้ป่วยใหม่ในช่วงเวลานั้น	จำนวนประชากรทั้งหมดในตอนที่เริ่มทำการศึกษา	

Prevalence

Prevalence (1) ความชุก

- **Measure existing** or presenting of "states" **at a specified time**
- **Count both new and old cases**

$$\frac{\text{จำนวนผู้ป่วยทั้งหมด ทั้งเก่าและใหม่ ณ ช่วงเวลานั้น}}{\text{จำนวนประชากรทั้งหมด ณ ช่วงเวลานั้น}}$$

1. **Point Prevalence**
2. **Period prevalence**

Prevalence (2)

-

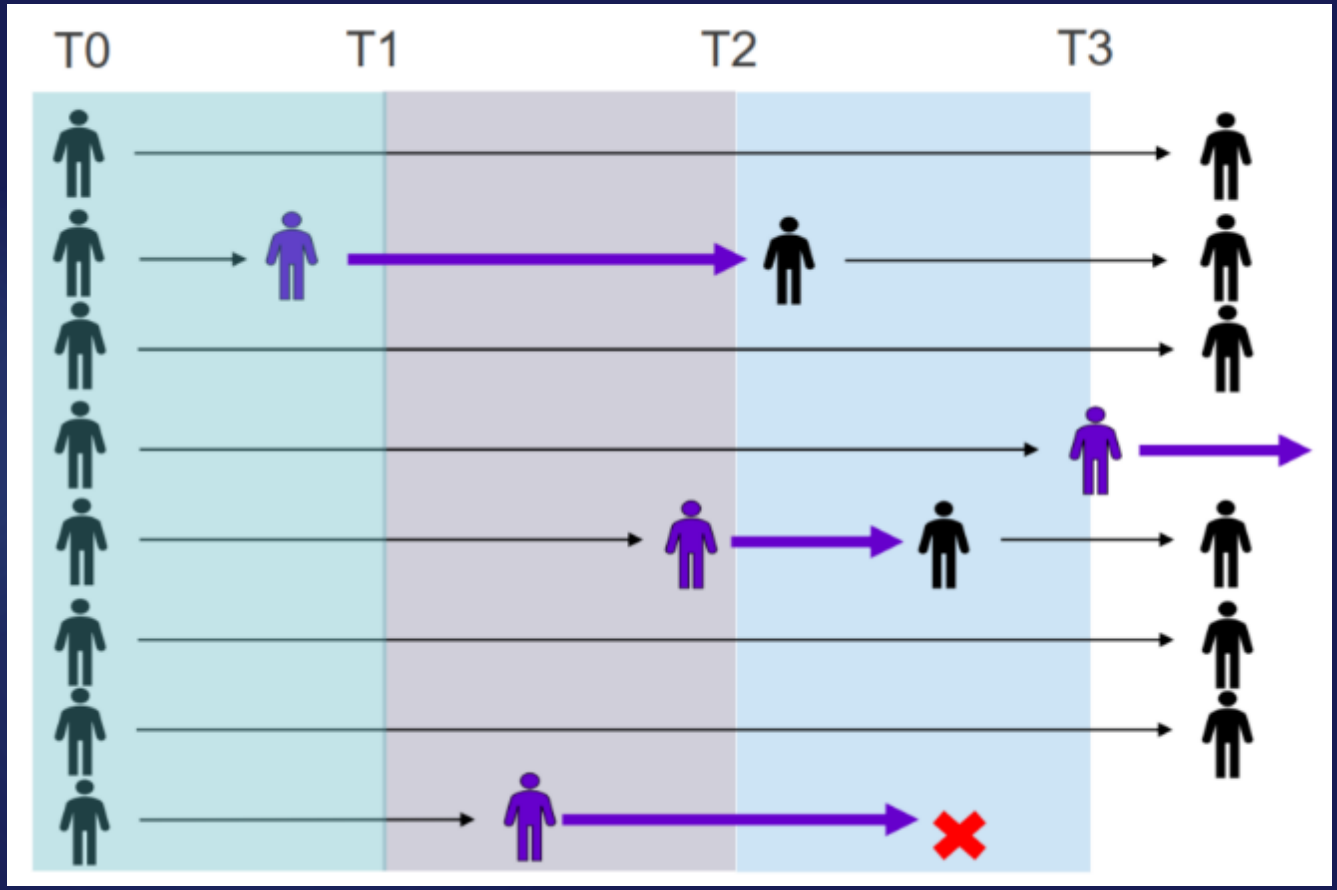
Point prevalence ; at time ณเวลาจุดใดจุดหนึ่ง

$$\frac{\text{จำนวนผู้ป่วยทั้งหมดทั้งเก่าและใหม่ในกลุ่มประชากรนั้น}}{\text{จำนวนประชากร ณ เวลาที่กำหนด}}$$

Example: วันที่ 1 มกราคม 2565 บริษัท ก มีพนักงาน 5,000 คน และมี 100 คน ป่วยเป็นเบาหวาน
จงหาความชุกของโรคเบาหวานในบริษัท ก ณวันที่ 1 มกราคม 2565

- = 100/5000 = 0.02 = 2 per 100 =2%

Point Prevalence



= 4/8 = 0.5 = 50%

Prevalence (3)

- **Period prevalence** ; during interval ในช่วงเวลาที่กำหนด

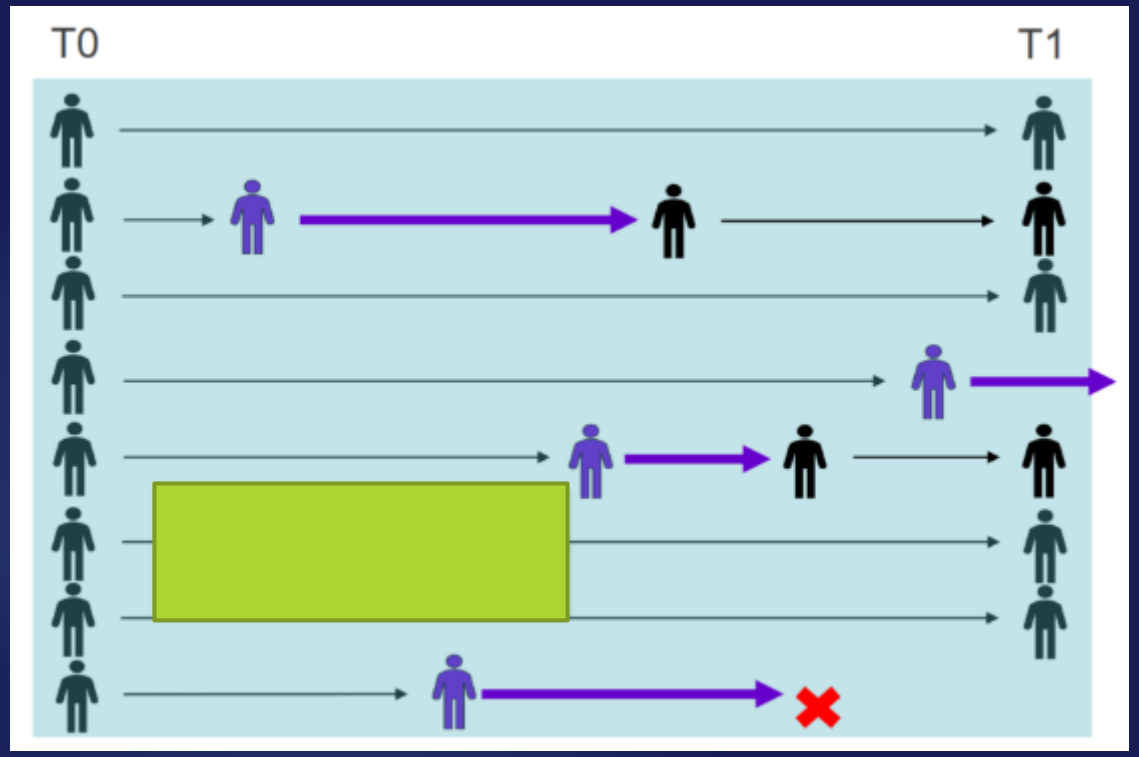
$$\frac{\text{จำนวนผู้ป่วยทั้งหมดที่ปรากฏในช่วงระยะเวลาที่กำหนด}}{\text{ค่าเฉลี่ยของจำนวนประชารในช่วงระยะเวลาที่กำหนด}}$$

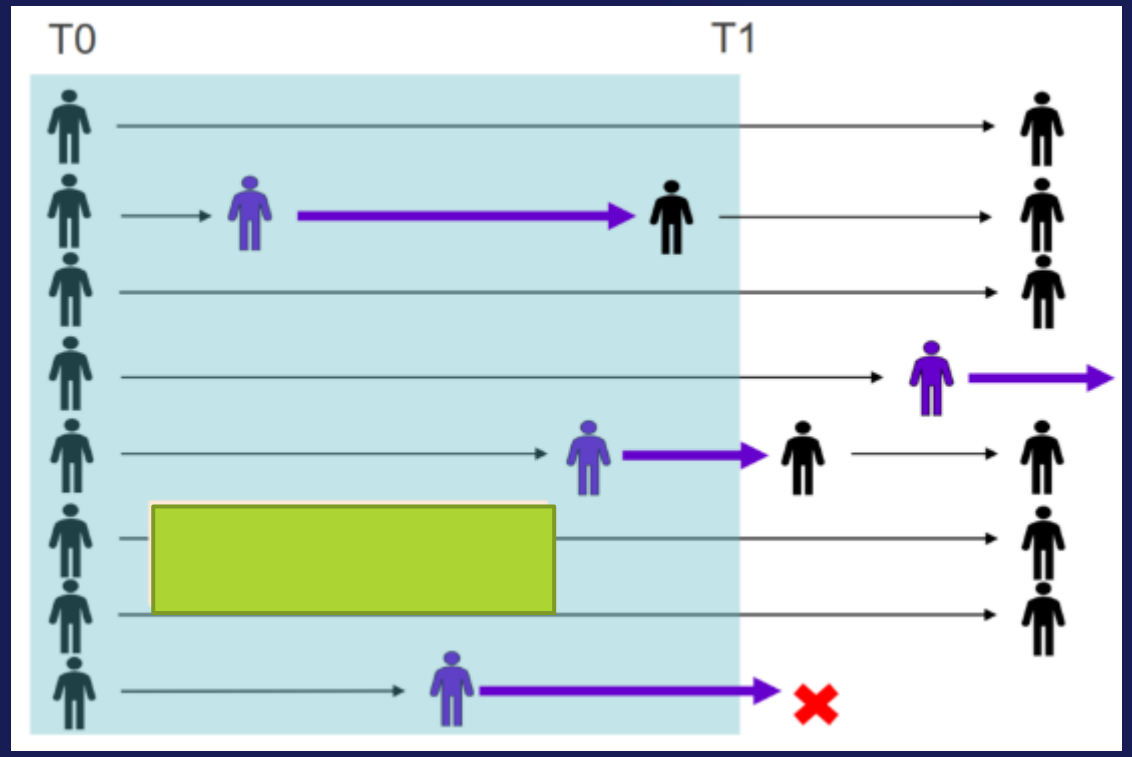
Example: ในปี 2564 บริษัท ก มีพนักงาน 5000 คน ในเดือนมกราคม และ 4900 คน ในเดือนธันวาคม มีพนักงานที่ได้รับการวินิจฉัยว่าเป็นโรคมะเร็งก่อนวันที่ 1 มกราคม 2564 จำนวน 100 คน และมีผู้ที่ได้รับการวินิจฉัยว่าเป็นโรคมะเร็งในปี 2564 อีก 5 คน ในจำนวนผู้ป่วย 105 คนนี้เสียชีวิตในปี 2564 จำนวน 10 คน

จงหาความชุกในปี 2564 (1 มกราคม – 31 ธันวาคม 2564)

- = 105/[(5000+4900)/2] = 0.021 = 2.1% = 21 per 1000

Period Prevalence





Incidence

Incidence (1) อุบัติการณ์

- Measure happening or occurrence of "events/processes" during a specified period of time
- Count **only new cases**, i.e. new events
- วัดสัดส่วนหรืออัตราของผู้ป่วยที่เกิดโรค "**รายใหม่" (new case**) ภายในช่วงเวลาใดเวลาหนึ่ง
  ที่พัฒนามาจาก**ประชากรที่มีความเสี่ยงต่อการเกิดโรค (Population at risk) แต่ไม่เคยเป็นโรคมาก่อน**

1. **Incidence proportion**
2. **Incidence rate ; เวลาเป็นตัวหาร**

Incidence (2)

1. **Incidence proportion**: จำนวนการเกิดโรคหรือเหตุการณ์ใดเหตุการณ์หนึ่งใหม่ภายในช่วงเวลาใดเวลาหนึ่ง **เปรียบเทียบกับจำนวนของประชากรที่มีโอกาสที่จะเกิด**โรคหรือเหตุการณ์นั้นๆทั้งหมด โดย**ไม่ได้คำนึงถึงจำนวนประชากรที่สูญหายไป**ระหว่างการศึกษา (Loss to follow - up)

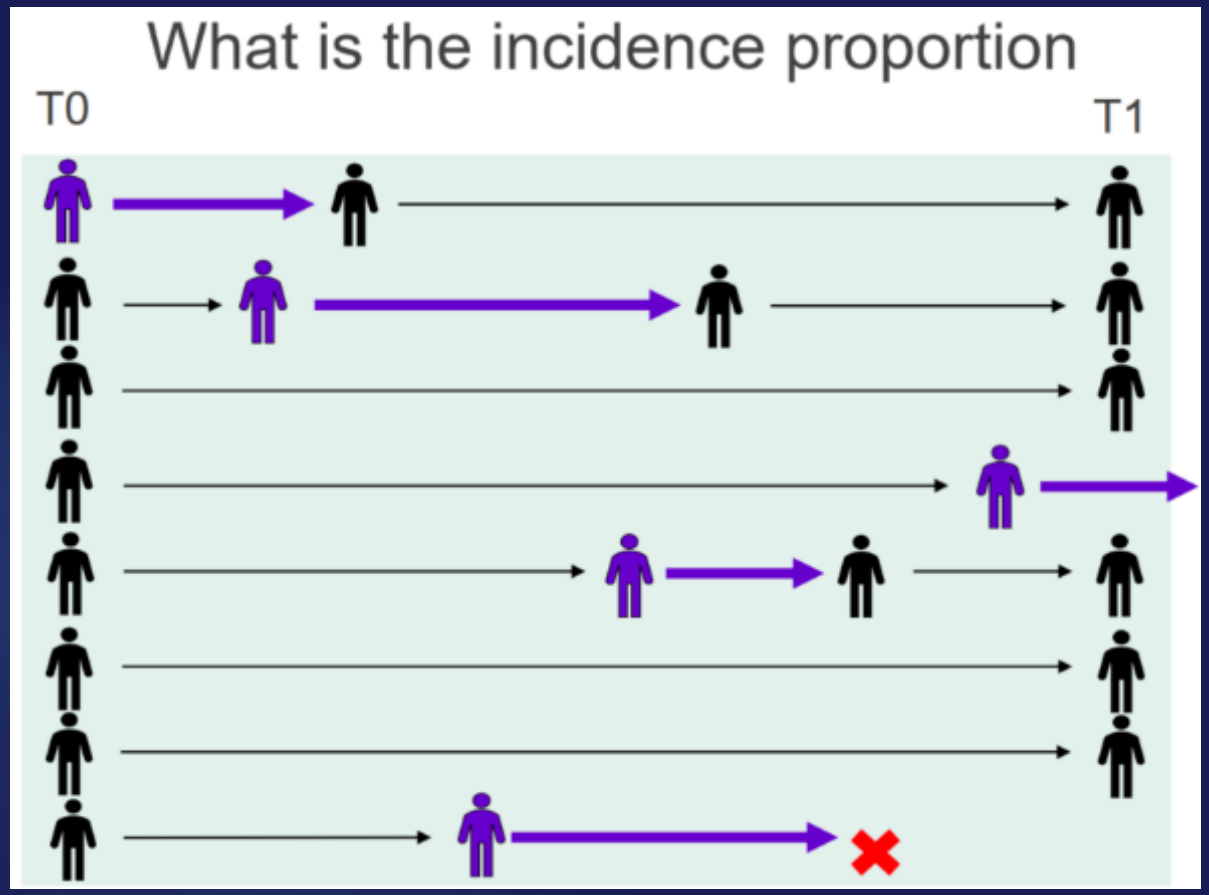
Syn: **attack rate, risk, probability of getting disease, cumulative incidence**

$$\frac{\text{จำนวนผู้ป่วย}\textbf{รายใหม่}\text{ในช่วงเวลาที่กำหนด}}{\text{จำนวน}\textbf{คนที่เสี่ยง}\text{ต่อการเกิดโรค}\textbf{ทั้งหมด}\text{ในช่วงเวลาที่กำหนด}}$$

- Example: ในระหว่างการระบาดของเชื้อไวรัสโคโรนา 2019 ผู้ป่วย 50 จาก 2000 คนเสียชีวิต จงหาความเสี่ยงต่อการเสียชีวิตในผู้ป่วยกลุ่มนี้

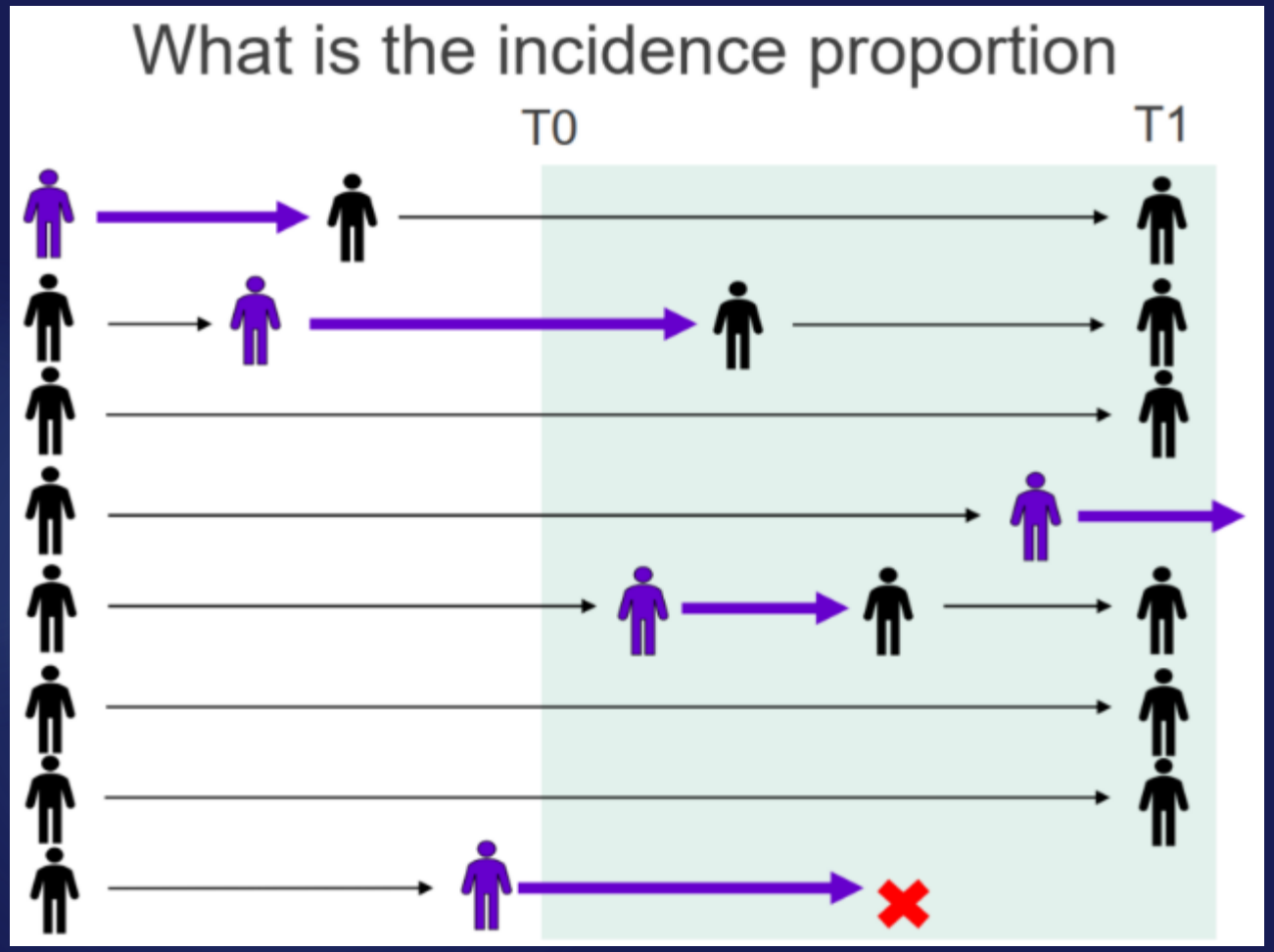
= 50/2000 = 0.025 =25 per 1000 = 2.5%

Incidence proportion



=4/(8-1) = 4/7 = 57.14%

Incidence proportion



= 2/6 = 33.33%

Incidence proportion


```
. clear

. use "C:\Users\Thananan\Desktop\practice data sets\lowbwt_dat.dta"

. tab low

     low birth weight |      Freq.     Percent        Cum.
----------------------+-----------------------------------
birth weight ge 2500g |       130       68.78       68.78
birth weight lt 2500g |        59       31.22      100.00
----------------------+-----------------------------------
                Total |       189      100.00
```

Incidence (3)

2. **Incidence rate**: การวัดอุบัติการณ์การเกิดโรคจากประชากรที่มีความเสี่ยงในการเกิดโรคที่มี**ระยะเวลาในการติดตามแต่ละรายแตกต่างกัน** การวัดอุบัติการณ์แบบ Incidence density จะวัดตัวเศษเหมือนกับ incidence proportion แต่ตัวส่วนจะเปลี่ยนจากจำนวนคนที่เสี่ยงต่อการเกิดโรค เป็น **ผลรวมของระยะเวลาที่ผู้ป่วยเสี่ยงต่อการเกิดโรค (Time at risk)** การรายงานจะอยู่ในรูป **person - time**

$$\frac{\text{จำนวนผู้ป่วยรายใหม่ในช่วงเวลาที่กำหนด}}{\text{ผลรวมของระยะเวลาเสี่ยงที่จะเกิดโรค}}$$

Example: ในระหว่างการระบาดของเชื้อไวรัสโคโรนา 2019 ผู้ป่วย 50 จาก 2000 คนเสียชีวิต
จงหาความเสี่ยงต่อการเสียชีวิตในผู้ป่วยกลุ่มนี้
= 50/2000 = 0.025 =25 per 1000 = 2.5%

Incidence (4)

2. **Incidence rate**: การวัดอุบัติการณ์การเกิดโรคจากประชากรที่มีความเสี่ยงในการเกิดโรคที่มี**ระยะเวลาในการติดตามแต่ละรายแตกต่างกัน** การวัดอุบัติการณ์แบบ Incidence density จะวัดตัวเศษเหมือนกับ incidence proportion แต่ตัวส่วนจะเปลี่ยนจากจำนวนคนที่เสี่ยงต่อการเกิดโรค เป็น **ผลรวมของระยะเวลาที่ผู้ป่วยเสี่ยงต่อการเกิดโรค (Time at risk)** การรายงานจะอยู่ในรูป **person - time**

$$\frac{\text{จำนวนผู้ป่วยรายใหม่ในช่วงเวลาที่กำหนด}}{\text{ผลรวมของระยะเวลาเสี่ยงที่จะเกิดโรค}}$$

Incidence rate

$$\frac{\text{จำนวนผู้ป่วยรายใหม่ในช่วงเวลาที่กำหนด}}{\text{ผลรวมของระยะเวลาเสี่ยงที่จะเกิดโรค}}$$

=3/(3+3+4+1+2+2+4) = 3/18 = 0.17

ปี 2560	ปี 2561	ปี 2562	ปี 2563	ปี 2654
			3	
			3	
				4
	1			
		2		
		X 2		
				4
				4

Other measure of frequency

- Attack rate = incidence proportion
- Crude mortality rate
- Case-fatality rate

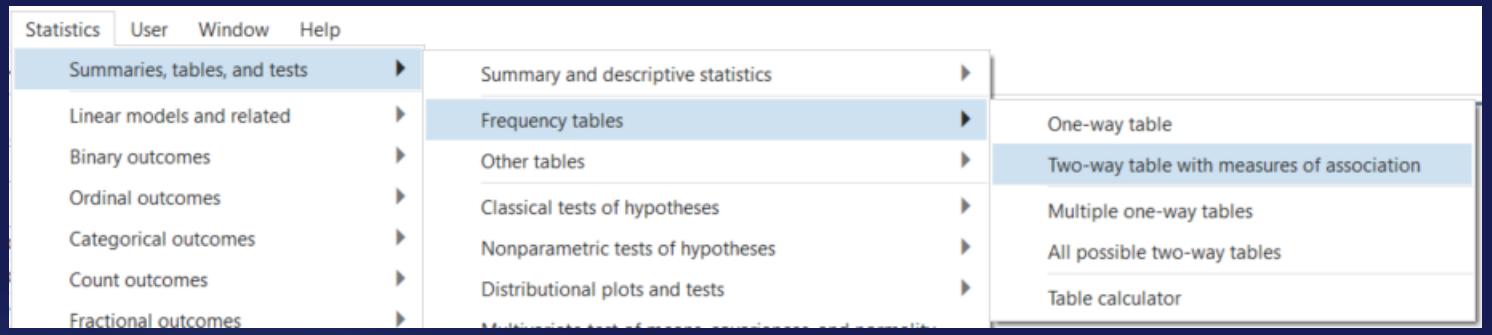
Attack rate (AR) (1)

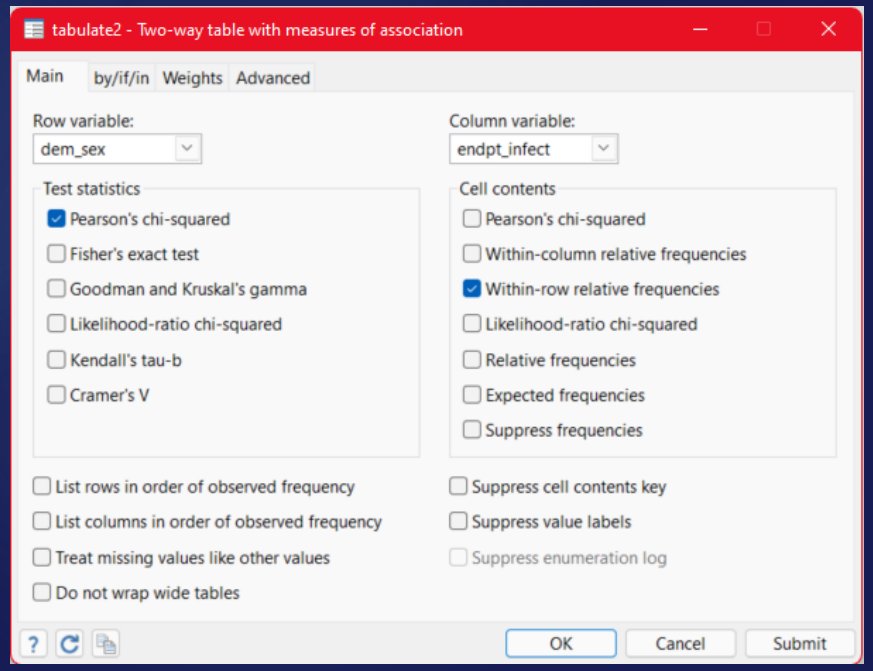
- $= \dfrac{\text{Number of incident cases during an epidemic period}}{\text{Population at risk at the beginning of the epidemic period}}$

- **Overall** attack rate: among total population
- **Specific** attack rate: among subpopulation
  - Age-specific attack rate
  - Gender-specific attack rate
  - Food-specific attack rate
- Differences among specific attack rates are important for generating hypotheses!

Attack rate (AR) (2)

```
. use "C:\Users\Thananan\Desktop\practice data sets\hiv vaccine trial.dt
```





Male AR =25.25%

Female AR = 27.32%

Overall AR=26.04%

```
. tabulate dem_sex endpt_infect, chi2 row
```

Key
frequency *row percentage*

Sex	end point - confirmed HIV + No	Yes	Total
Male	222 74.75	75 25.25	297 100.00
Female	133 72.68	50 27.32	183 100.00
Total	355 73.96	125 26.04	480 100.00

Pearson chi2(1) = 0.2519 Pr = 0.616

Pr>0.05 ไม่มีความแตกต่างกัน AR ระหว่าง male และ fe

Attack rate (AR) (3)

หา attack rate ของกลุ่มที่ได้และไม่ได้วัคซีน

```
. use "C:\Users\Thananan\Desktop\practice data sets\hiv vaccine trial.dt
```

AR ในกลุ่มที่ได้ placebo = 30.58
AR ในกลุ่มที่ได้ vaccine = 21.43
Overall AR = 26.04

Pr < 0.05 มีความแตกต่างกันของ AR ระหว่าง กลุ่ม placebo และกลุ่มที่ได้ vaccine

```
. tabulate trtmt endpt_infect, chi2 row
```

Key
frequency *row percentage*

vaccine or placebo group	end point - confirmed HIV + No	Yes	Total
Placebo	168 69.42	74 30.58	242 100.00
Vaccine	187 78.57	51 21.43	238 100.00
Total	355 73.96	125 26.04	480 100.00

Pearson chi2(1) = 5.2159 Pr = 0.022

Crude mortality rate

- It is a proportion of total deaths by all causes among total population in a given period
- $\frac{\text{จำนวนผู้เสียชีวิตทั้งหมดในปีปฏิทินนั้น}}{\text{ประชากรกลางปี}}$

Case-fatality rate

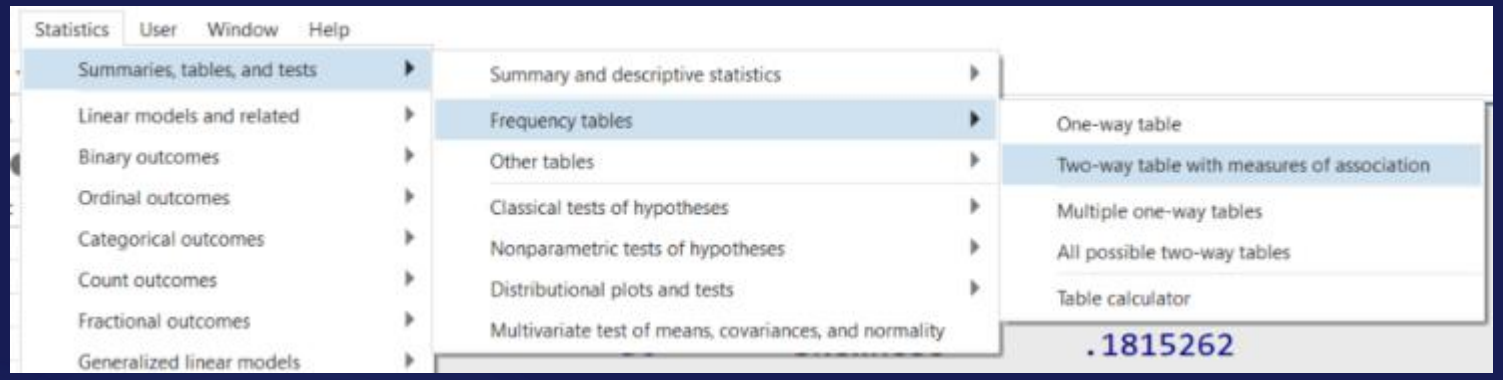
- It is a proportion of deaths by a specific disease among all deaths in a given period
- $\frac{\text{จำนวนผู้เสียชีวิตจากโรคนั้นในปีปฏิทินนั้น}}{\text{จำนวนผู้เสียชีวิตทั้งหมดในปีปฏิทินนั้น}}$

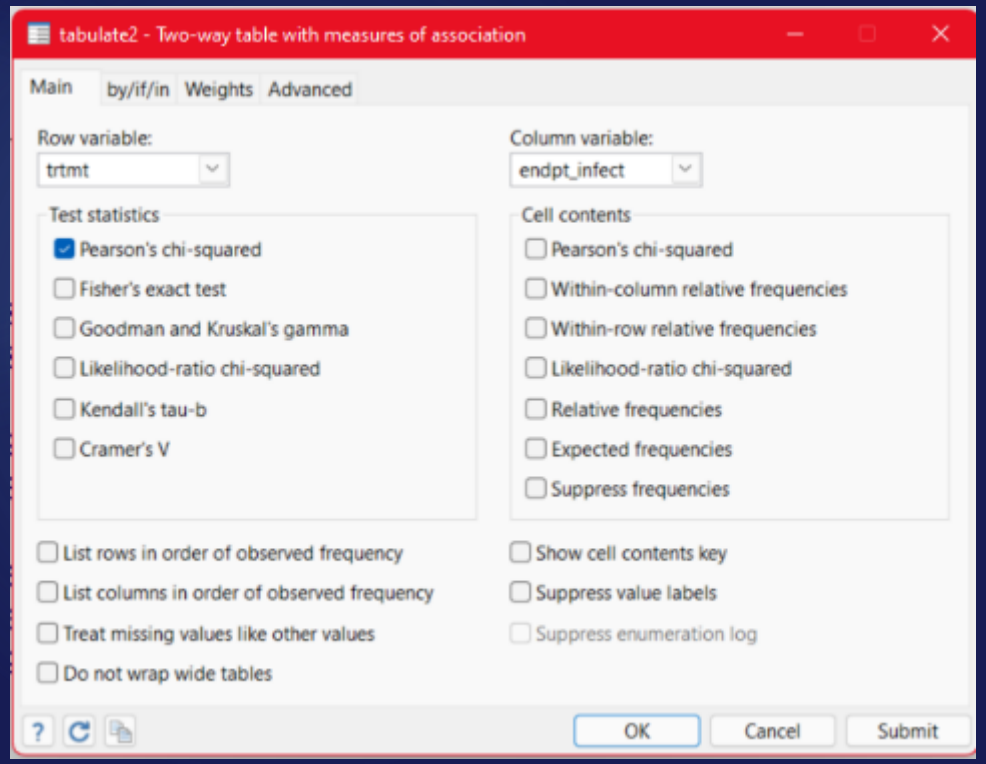
Comparison testing

Comparison between group

- Categorical outcome
  - Chi-square (χ^2) test
- Continuous Variable
  - Normal Distribution
    - t-test (2 groups) F-test/ANOVA (>= 2 groups)
  - Not normal Distribution
    - Wilcoxon test, Mann-Whitney U test

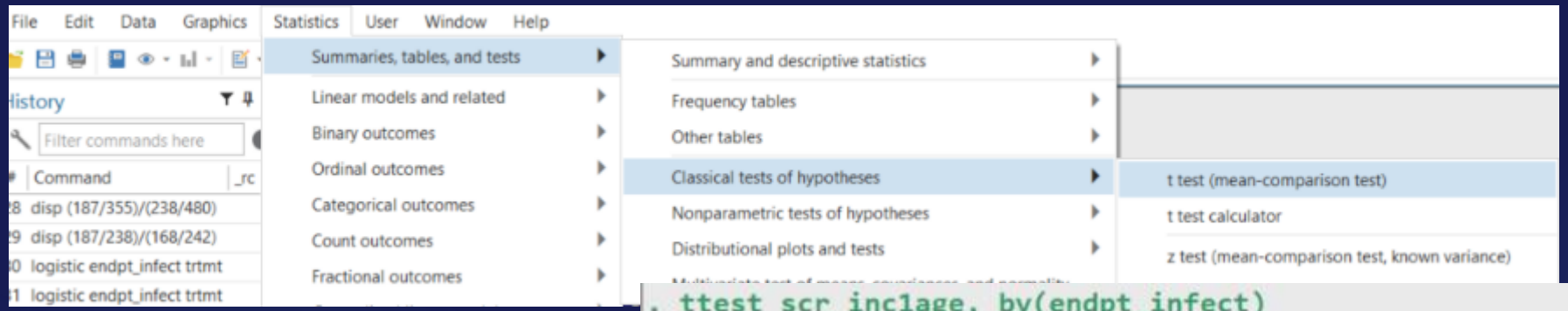
Chi-square (c2) test

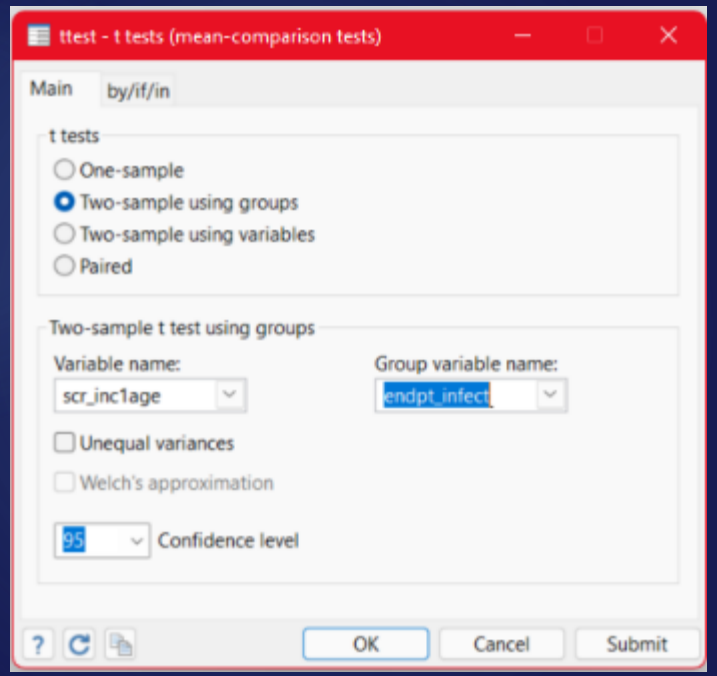





```
. tab trtmt endpt_infect, chi2

vaccine or   | end point - confirmed
   placebo   |          HIV +
     group   |        No        Yes |     Total
-----------+----------------------+----------
   Placebo   |       168         74 |       242
   Vaccine   |       187         51 |       238
-----------+----------------------+----------
     Total   |       355        125 |       480

          Pearson chi2(1) =   5.2159   Pr = 0.022
```

t-test (2 groups)



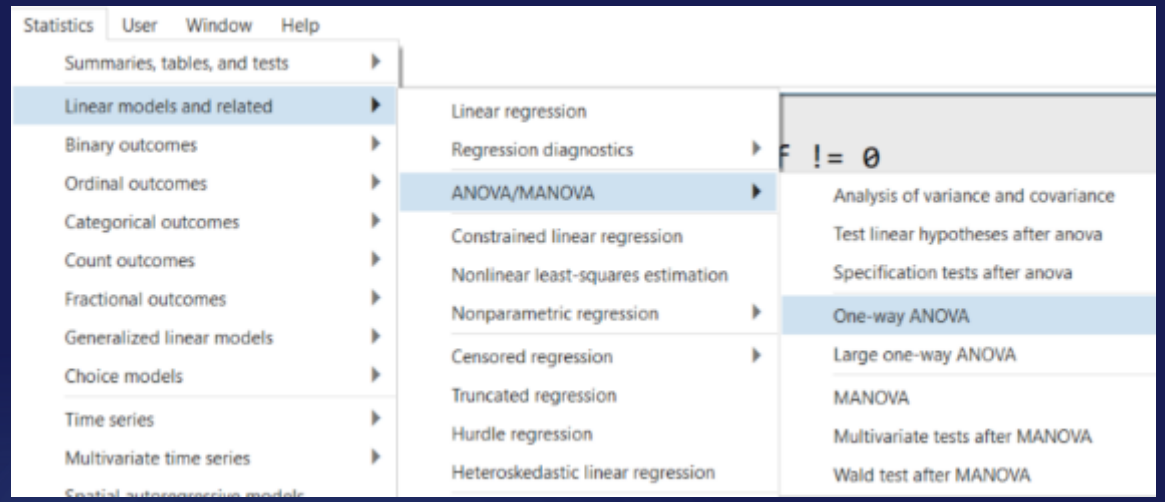


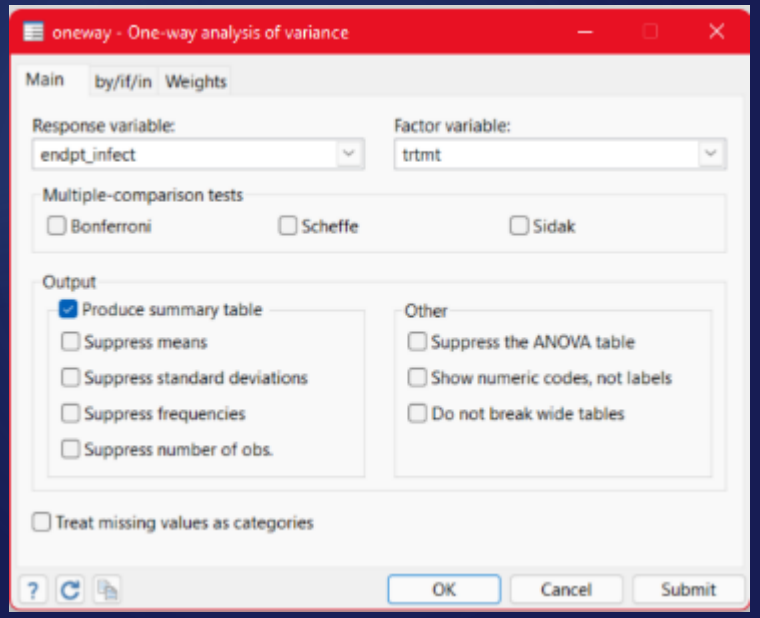

```
. ttest scr_inc1age, by(endpt_infect)

Two-sample t test with equal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
No	355	23.47324	.1844279	3.474888	23.11053	23.83595
Yes	125	23.872	.2835207	3.169858	23.31083	24.43317
combined	480	23.57708	.1551587	3.399358	23.27221	23.88196
diff		-.3987606	.3534473		-1.093263	.2957419

```
    diff = mean(No) - mean(Yes)                                   t =  -1.1282
Ho: diff = 0                                     degrees of freedom =      478

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.1299         Pr(|T| > |t|) = 0.2598          Pr(T > t) = 0.8701
```

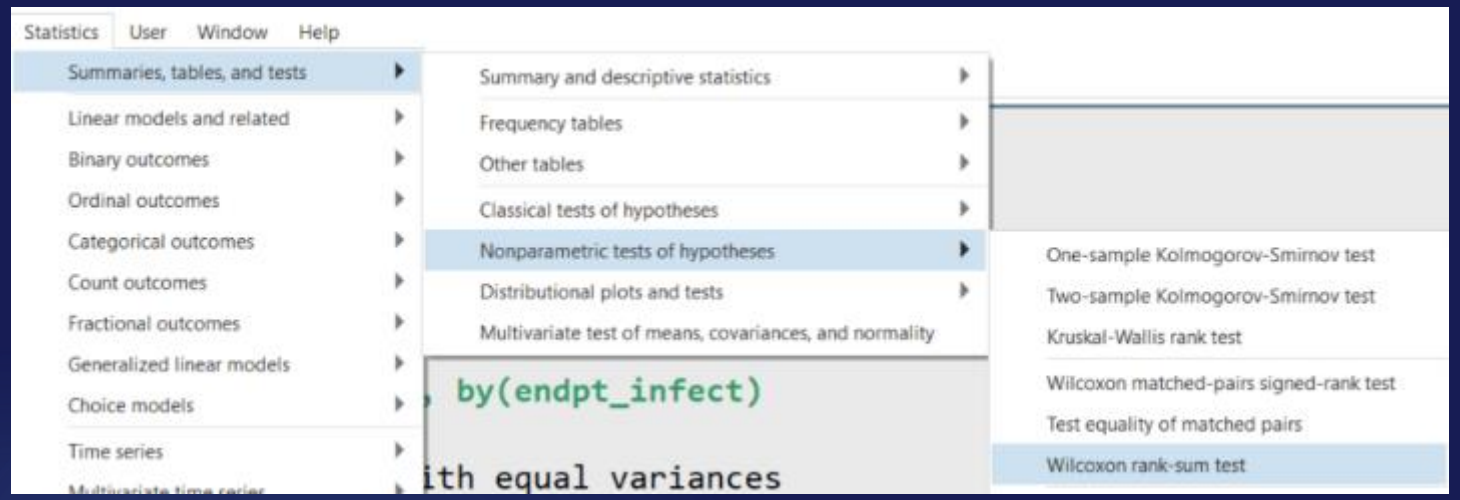
F-test/ANOVA (>= 2 groups)

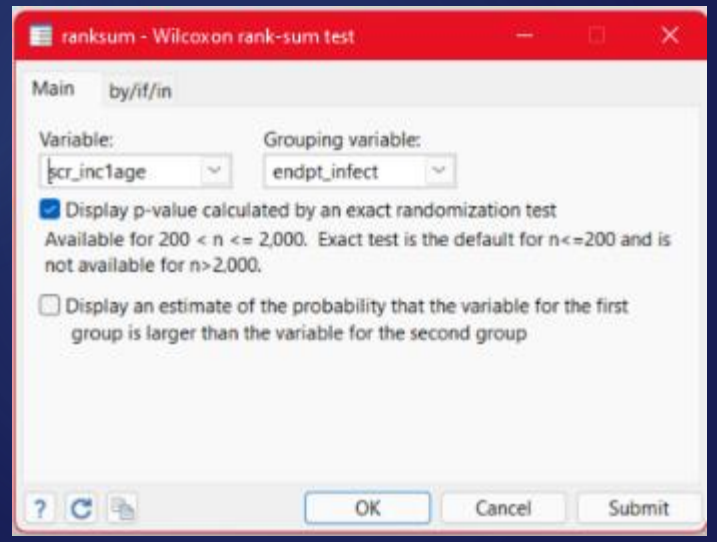





```
. oneway endpt_infect trtmt

                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      1.00458727      1   1.00458727      5.25     0.0224
 Within groups      91.4433294    478   .191304036
------------------------------------------------------------------------
    Total           92.4479167    479   .193001914

Bartlett's test for equal variances:  chi2(1) =   3.1913  Prob>chi2 = 0.074
```

Wilcoxon test,






```
. ranksum scr_inc1age, by(endpt_infect) exact

Two-sample Wilcoxon rank-sum (Mann-Whitney) test

endpt_infect |      obs    rank sum    expected
-------------+---------------------------------
          No |      355       83815     85377.5
         Yes |      125       31625     30062.5
-------------+---------------------------------
    combined |      480      115440      115440

unadjusted variance  1778697.92
adjustment for ties   -13277.47
                     ----------
adjusted variance    1765420.45

Ho: scr_in~e(endpt_~t==No) = scr_in~e(endpt_~t==Yes)
            z =  -1.176
   Prob > |z| =   0.2396
   Exact Prob =   0.2400
```

Measures of association

การวัดความสัมพันธ์

Measures of association

- การวัดความสัมพันธ์ทางสถิติระหว่างตัวแปรต้นและตัวแปรตาม
- อาจจะเป็นสาเหตุหรือไม่ใช่สาเหตุก็ได้
- มักใช้ใน observation study
- การเลือกใช้ขึ้นอยู่กับ study design
- แบ่งเป็น 2 ประเภทตามการเปรียบเทียบ
  - **Ratio** scale
    - Risk ratio: ratio of incidence proportion (risk)
    - Rate ratio: ratio of incidence rate
    - Odds ratio: ratio of odds
    - Prevalence ratio: ratio of prevalence
  - **Difference** scale
    - Risk difference
    - Rate difference
    - Prevalence difference

	Case	**Non-case**	**Total**
Exposed	A	B	A+B
Not exposed	C	D	C+D
Total	A+C	B+D	A+B+C+D

Prevalence ratio

- ใช้ใน cross-sectional study
- $\frac{Prevalence \text{ ในกลุ่ม } exposed}{Prevalence \text{ ในกลุ่ม } unexposed}$
- $\frac{Pe}{Pu} = \frac{A(A+B)}{C/(C+D)}$

	Case	Non-case	Total
Exposed	A	B	A+B
Not exposed	C	D	C+D
Total	A+C	B+D	A+B+C+D

- **Interpretation:** ความชุกของการป่วยในกลุ่ม exposed เป็น PR เท่าของกลุ่ม unexposed

prevalence of having the disease in exposed group is 6 times of that in unexposed group

Prevalence difference (PD)

- A difference of two prevalence
- Cross-sectional study
- $PE - Pu$
- Prevalence of having disease in exposed group is PD higher than that in unexposed group
- ความชุกของการป่วยในกลุ่ม exposed สูงกว่า exposed PD%

Risk Ratio

Risk ratio / Relative risk (RR) (1)

- ใช้ใน Cohort study
- $\frac{Risk \text{ ของกลุ่ม } exposed}{Risk \text{ ของกลุ่ม } unexposed}$
- $\frac{Incidence \text{ ของกลุ่ม } exposed}{Incidence \text{ ของกลุ่ม } unexposed}$
- $IE \ / \ Iu$
- $\frac{A/(A+B)}{C/(C+D)}$

	ท้องเสีย	ไม่มีท้องเสีย	Total
กินไข่ต้ม	**60**	**40**	100
ไม่ได้กินไข่ต้ม	**10**	**90**	100
Total	70	130	200

$$= \frac{60/100}{10/100} = \frac{0.6}{0.1} = 6$$

ความเสี่ยงต่อการเกิดอาการท้องเสียในกลุ่มที่กินไข่ต้มเป็น 6 เท่าของคนที่ไม่ได้กิน

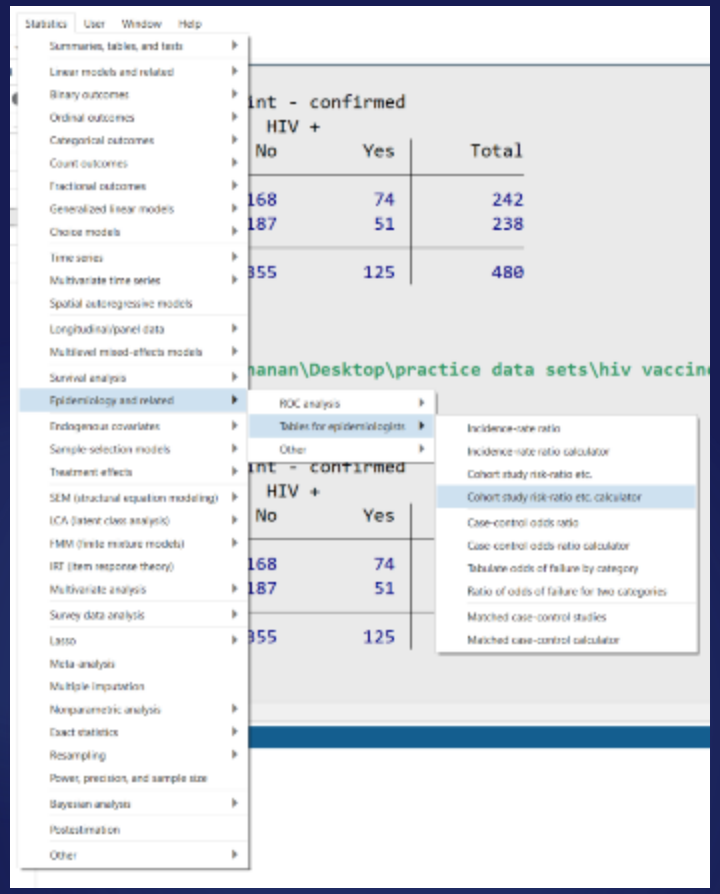
Risk ratio / Relative risk (RR) (2)


```
. clear

. use "C:\Users\Thananan\Desktop\practice data sets\hiv vaccine trial.dta"

. tab trtmt endpt_infect
```

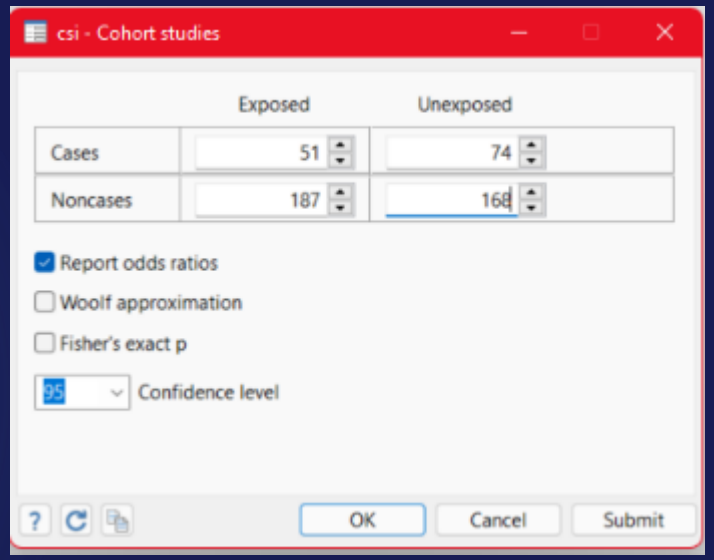
vaccine or placebo group	end point - confirmed HIV + No	Yes	Total
Placebo	168	74	242
Vaccine	187	51	238
Total	355	125	480

- $= \frac{51/238}{74/242} = 0.7$ ความเสี่ยงต่อการติดเชื้อในคนที่**ได้วัคซีน**เป็น **0.7 เท่า**ของคนที่**ไม่ได้วัคซีน**
  - $= \frac{1}{0.7} = 1.43$ ความเสี่ยงต่อการติดเชื้อในคนที่**ไม่ได้วัคซีน**เป็น **1.43 เท่า**ของค**ได้วัคซีน**

Risk ratio / Relative risk (RR) (3)



vaccine or placebo group	end point - confirmed HIV + No	Yes	Total
Placebo	168	74	242
Vaccine	187	51	238
Total	355	125	480




```
. csi 51 74 187 168, or
```

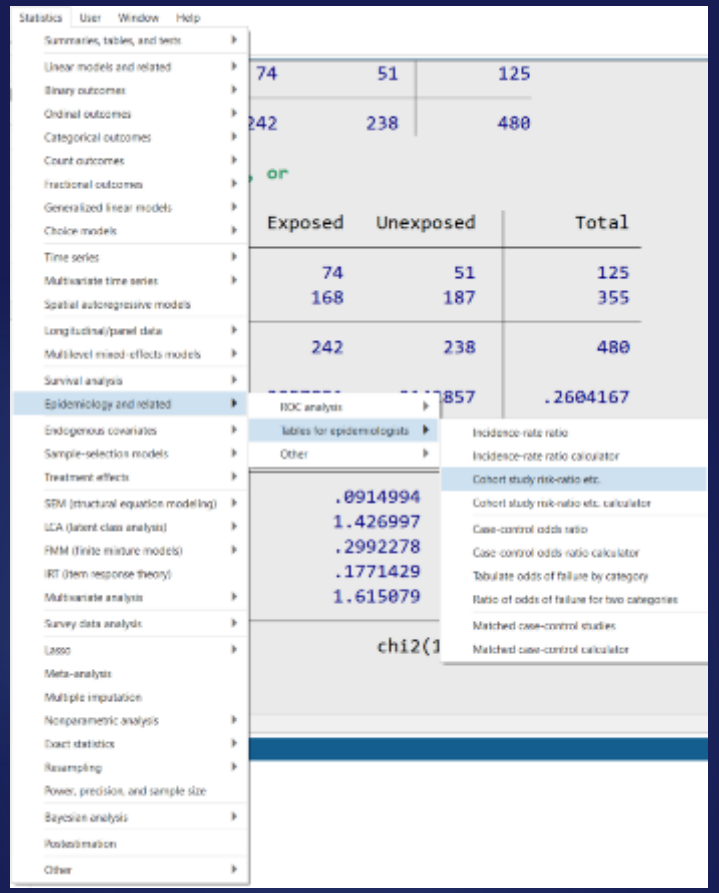
	Exposed	Unexposed	Total
Cases	51	74	125
Noncases	187	168	355
Total	238	242	480
Risk	.2142857	.3057851	.2604167

	Point estimate	[95% Conf.	Interval]
Risk difference	-.0914994	-.1695202	-.0134786
Risk ratio	.7007722	.514711	.9540921
Prev. frac. ex.	.2992278	.0459079	.485289
Prev. frac. pop	.1483671		
Odds ratio	.6191646	.4101872	.9346859

chi2(1) = 5.22 Pr>chi2 = 0.0224

ความเสี่ยงต่อการติดเชื้อในคนที่**ได้วัคซีน**เป็น **0.7 เท่า**ของคนที่**ไม่ได้วัคซีน**

Risk ratio / Relative risk (RR) (4)




```
. cs endpt_infect trtmt

                 | vaccine or placebo group
                 |   Exposed   Unexposed |     Total
-----------------+------------------------+----------
           Cases |        51         74 |       125
        Noncases |       187        168 |       355
-----------------+------------------------+----------
           Total |       238        242 |       480
                 |                        |
            Risk |  .2142857    .3057851 |  .2604167
                 |                        |
                 |   Point estimate       |  [95% Conf. Interval]
-----------------+------------------------+----------
Risk difference  |     -.0914994          |  -.1695202   -.0134786
     Risk ratio  |      .7007722          |   .514711    .9540921
Prev. frac. ex.  |      .2992278          |  .0459079     .485289
Prev. frac. pop  |      .1483671          |
-----------------+------------------------+----------
                                   chi2(1) =     5.22  Pr>chi2 = 0.0224
```

```
. tab endpt_infect trtmt

end point -      | vaccine or placebo group
confirmed HIV +  |   Placebo    Vaccine |     Total
-----------------+------------------------+----------
              No |       168        187 |       355
             Yes |        74         51 |       125
-----------------+------------------------+----------
           Total |       242        238 |       480
```

ความเสี่ยงต่อการติดเชื้อในคนที่**ได้วัคซีน**เป็น **0.7 เท่า**ของคนที่**ไม่ได้วัคซีน**

Vaccine Efficacy (VE) = 1- RR = 1-0.7 = 0.3 = 30%

วัคซีนมีประสิทธิภาพป้องกันการติดเชื้อได้ร้อยละ 30

Vaccine Efficacy(VE)

- Vaccine Efficacy = RCT effect
- Vaccine Effective = real world effect
- Vaccine efficiency = Cost / Effect

- Vaccine Efficacy(VE)

$$VE = \frac{(\text{AR nonvaccine} - \text{AR vaccine})}{\text{AR nonvaccine}}$$

$$= \frac{\text{AR nonvaccine}}{\text{AR nonvaccine}} - \frac{\text{AR vaccine}}{\text{AR nonvaccine}}$$

$$= 1 - \frac{\text{AR vaccine}}{\text{AR nonvaccine}} = 1 - RR$$

- $RR = \frac{Incidence \text{ ของกลุ่ม } exposed}{Incidence \text{ ของกลุ่ม } unexposed} = \frac{AR \text{ ของกลุ่ม } exposed}{AR \text{ ของกลุ่ม } unexposed} = \frac{AR \text{ ของกลุ่ม } vaccine}{AR \text{ ของกลุ่ม } nonvaccine}$

Risk difference (RD, attributable risk)

- difference of two incidence proportions (exposed vs unexposed group)
- Cohort study
- $IE - Iu$
- = (51/238) – (74/242)
- =21.43-30.58
- = -0.09 = 9%

```
. cs endpt_infect trtmt

                 | vaccine or placebo group
                 |   Exposed   Unexposed |     Total
-----------------+------------------------+----------
           Cases |        51         74 |       125
        Noncases |       187        168 |       355
-----------------+------------------------+----------
           Total |       238        242 |       480
                 |                        |
            Risk |  .2142857    .3057851 |  .2604167
                 |                        |
                 |     Point estimate     |  [95% Conf. Interval]
-----------------+------------------------+----------------------
Risk difference  |       -.0914994        |  -.1695202   -.0134786
     Risk ratio  |        .7007722        |   .514711    .9540921
Prev. frac. ex.  |        .2992278        |  .0459079     .485289
Prev. frac. pop  |        .1483671        |
-----------------+------------------------+----------------------
                                  chi2(1) =     5.22  Pr>chi2 = 0.0224
```

```
. tab endpt_infect trtmt, col

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

end point -  | vaccine or placebo group
confirmed    |
    HIV +    |   Placebo    Vaccine |     Total
-------------+-------------------------+----------
        No   |       168        187 |       355
             |     69.42      78.57 |     73.96
-------------+-------------------------+----------
       Yes   |        74         51 |       125
             |     30.58      21.43 |     26.04
-------------+-------------------------+----------
     Total   |       242        238 |       480
             |    100.00     100.00 |    100.00
```

- ความเสี่ยงต่อการติดเชื้อในคนที่ได้วัคซีนคือ 9% น้อยกว่าคนที่ไม่ได้วัคซีน

Rate ratio / Incidence rate ratio (IRR) (1)

- A ratio of two incidence rates
- Used in RCT or Cohort study (with person-time data)
- Rate ratio (IRR) = [A/TE] / [B/TU]

	No. of sick	**Person-time**	**Incidence rate**
Exposed	A	TE	A/TE
Unexposed	B	TU	B/TU
Total	A+B	TE+TU	(A+B)/(TE+TU)

Odds Ratio (OR)

Odds ratio

- Use in Case-control, cohort, cross-sectional study

- Odds

$$= \frac{\text{โอกาสเกิดเหตุการณ์}}{\text{โอกาสในการไม่เกิดเหตุการณ์}}$$

$$= \frac{P}{1-P}$$

	Case	**Non-case**	**Total**
Exposed	A	B	A+B
Not exposed	C	D	C+D
Total	A+C	B+D	A+B+C+D

ไม่มี Odds difference เพราะตัวหารคนละตัวกัน

Disease odds ratio (cohort study)

- $\dfrac{odds \text{ ของการเกิดโรคในกลุ่ม } exposed}{odds \text{ ของการเกิดโรคในกลุ่ม } unexposed}$
- $\dfrac{\dfrac{\text{โอกาสของการเกิดโรคในกลุ่ม } exposed}{\text{โอกาสของการไม่เกิดโรคในกลุ่ม } exposed}}{\dfrac{\text{โอกาสของการเกิดโรคในกลุ่ม } unexposed}{\text{โอกาสของการไม่เกิดโรคในกลุ่ม } unexposed}}$
- $\dfrac{\dfrac{A/(A+B)}{B/(A+B)}}{\dfrac{C/(C+D)}{D/(C+D)}} = \dfrac{\frac{A}{B}}{\frac{C}{D}} = \dfrac{A/B}{C/D} = AD/CB$

	Case	Non-case	Total
Exposed	A	B	A+B
Not exposed	C	D	C+D
Total	A+C	B+D	A+B+C+D

- **Interpretation:** การเกิดโรค ในกลุ่ม exposed เป็น OR เท่า ของกลุ่ม unexposed

Exposure odds ratio (case-control study)

- $\dfrac{odds \text{ ของการ } exposed \text{ ในกลุ่มป่วย}}{odds \text{ ของการ } exposed \text{ ในกลุ่มไม่ป่วย}}$
- $\dfrac{\dfrac{\text{โอกาสของ } exposed \text{ ในกลุ่มป่วย}}{\text{โอกาสของการ } unexposed \text{ ในกลุ่มป่วย}}}{\dfrac{\text{โอกาสของ } exposed \text{ ในกลุ่มไม่ป่วย}}{\text{โอกาสของการ } unexposed \text{ ในกลุ่มไม่ป่วย}}}$
- $\dfrac{\dfrac{A/(A+C)}{C/(A+C)}}{\dfrac{B/(B+D)}{D/(B+D)}} = \dfrac{\frac{A}{C}}{\frac{B}{D}} = \dfrac{A/C}{B/D} = AD/CB$

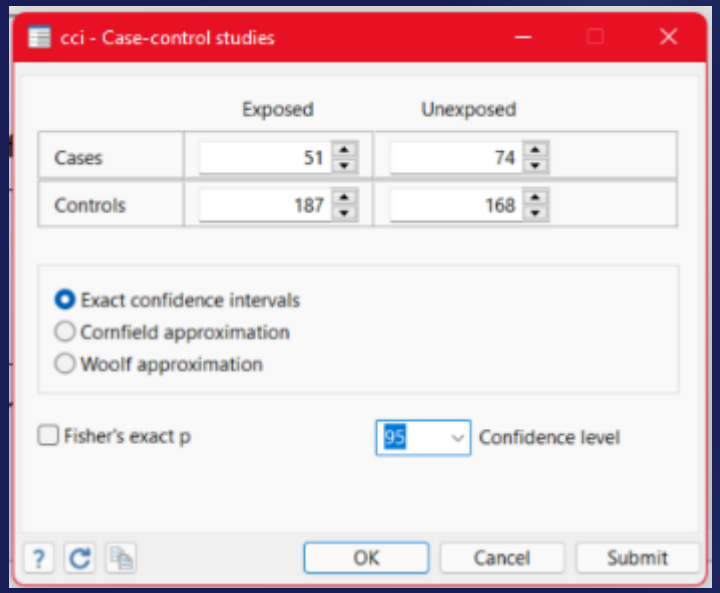
	Case	Non-case	Total
Exposed	A	B	A+B
Not exposed	C	D	C+D
Total	A+C	B+D	A+B+C+D

- **Interpretation:** การ exposed ในกลุ่มป่วย เป็น OR เท่า ของกลุ่มไม่ป่วย

Statistics User Window Help

- Summaries, tables, and tests
- Linear models and related
- Binary outcomes
- Ordinal outcomes
- Categorical outcomes
- Count outcomes
- Fractional outcomes
- Generalized linear models
- Choice models
- Time series
- Multivariate time series
- Spatial autoregressive models
- Longitudinal/panel data
- Multilevel mixed-effects models
- Survival analysis
- Epidemiology and related
  - ROC analysis
  - Tables for epidemiologists
    - Incidence-rate ratio
    - Incidence-rate ratio calculator
    - Cohort study risk-ratio etc.
    - Cohort study risk-ratio etc. calculator
    - Case-control odds ratio
    - Case-control odds-ratio calculator
    - Tabulate odds of failure by category
    - Ratio of odds of failure for two categories
    - Matched case-control studies
    - Matched case-control calculator
  - Other
- Endogenous covariates
- Sample-selection models
- Treatment effects
- SEM (structural equation modeling)
- LCA (latent class analysis)
- FMM (finite mixture models)
- IRT (item response theory)
- Multivariate analysis
- Survey data analysis
- Lasso
- Meta-analysis
- Multiple imputation
- Nonparametric analysis
- Exact statistics
- Resampling
- Power, precision, and sample size
- Bayesian analysis
- Postestimation
- Other

```
_infect, chi row

point - confirmed
    HIV +
    No          Yes        |     Total
-----------------------------------------
   133           50        |
 72.68        27.32        |
-----------------------------------------
   355          125        |
 73.96        26.04        |

hi2(1) =    0.2519
```

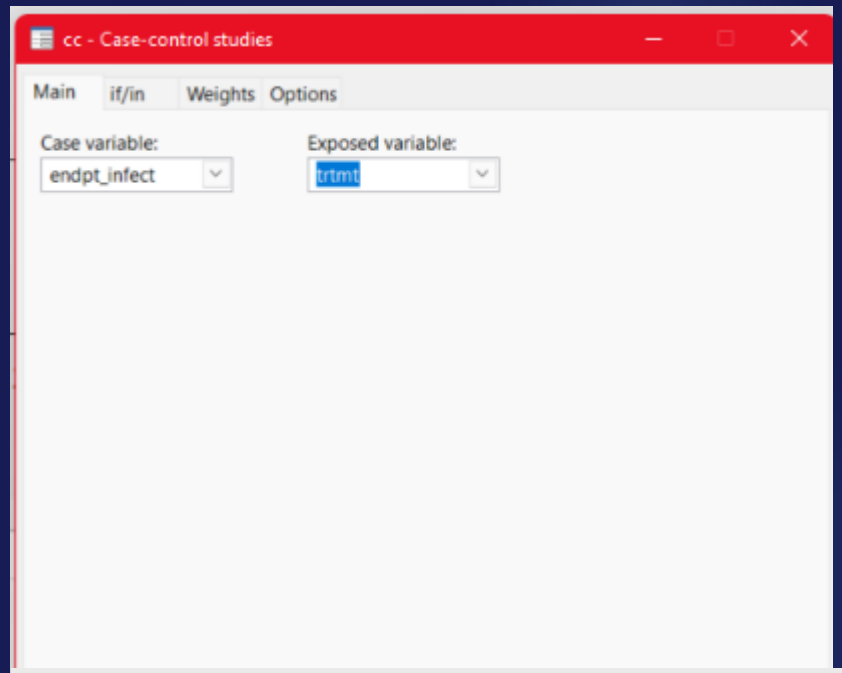



```
. cci 51 74 187 168

                                                   Proportion
                    |  Exposed   Unexposed  |     Total     Exposed
-------------------+-----------------------+-----------------------
        Cases       |       51         74   |       125      0.4080
     Controls       |      187        168   |       355      0.5268
-------------------+-----------------------+-----------------------
        Total       |      238        242   |       480      0.4958
                    |                       |
                    |   Point estimate      |  [95% Conf. Interval]
                    |                       |
      Odds ratio    |     .6191646          |   .4002905    .9546679 (exact)
 Prev. frac. ex.    |     .3808354          |   .0453321    .5997095 (exact)
 Prev. frac. pop    |     .2006091          |
                    +-----------------------+-----------------------
                                   chi2(1) =     5.22  Pr>chi2 = 0.0224
```

Statistics User Window Help

- Summaries, tables, and tests
- Linear models and related
- Binary outcomes
- Ordinal outcomes
- Categorical outcomes
- Count outcomes
- Fractional outcomes
- Generalized linear models
- Choice models
- Time series
- Multivariate time series
- Spatial autoregressive models
- Longitudinal/panel data
- Multilevel mixed-effects models
- Survival analysis
- Epidemiology and related
  - ROC analysis
  - Tables for epidemiologists
    - Incidence-rate ratio
    - Incidence-rate ratio calculator
    - Cohort study risk-ratio etc.
    - Cohort study risk-ratio etc. calculator
    - Case-control odds ratio
    - Case-control odds-ratio calculator
    - Tabulate odds of failure by category
    - Ratio of odds of failure for two categories
    - Matched case-control studies
    - Matched case-control calculator
  - Other
- Endogenous covariates
- Sample-selection models
- Treatment effects
- SEM (structural equation modeling)
- LCA (latent class analysis)
- FMM (finite mixture models)
- IRT (item response theory)
- Multivariate analysis
- Survey data analysis
- Lasso
- Meta-analysis
- Multiple imputation
- Nonparametric analysis
- Exact statistics
- Resampling
- Power, precision, and sample size
- Bayesian analysis
- Postestimation

```
133        50        183
72.68      27.32     100.00

355        125       480
73.96      26.04     100.00

hi2(1) =   0.2519   Pr = 0.616

                                  Propo
Exposed   Unexposed        Total      Ex

51          74             125
187         168            355

Point estimat

.6191646
.3808354
.2006091

chi2(1
```




```
. cc endpt_infect trtmt
```

	Exposed	Unexposed	Total	Proportion Exposed
Cases	51	74	125	0.4080
Controls	187	168	355	0.5268
Total	238	242	480	0.4958

	Point estimate	[95% Conf.	Interval]
Odds ratio	.6191646	.4002905	.9546679
Prev. frac. ex.	.3808354	.0453321	.5997095
Prev. frac. pop	.2006091		

chi2(1) = 5.22 Pr>chi2 = 0.0224

Odds ratio

- ใน Case-control study
  - ถ้า Case และ control ในการศึกษาที่สามารถเป็นตัวแทนที่ดีในประชากร
  - Exposure OR = Disease OR
- ถ้าใน rare disease OR ≈ RR
  - เราสามารถแปลผล OR แบบ RR ได้

Odds ratio in Logistic regression

- Binary logistic regression model
- Binary outcome เช่น case non-case ป่วยกับไม่ป่วย

$$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

- OR = exp β

File Edit Data Graphics Statistics User Window Help

History

Filter commands here

#	Command
1	use "C:\Users\Thananan\De...
2	tab vaccno
3	tab trtmt
4	tab endpt_infect
5	tab trtmt endpt_infect
6	clear
7	use "C:\Users\Thananan\De...
8	tab trtmt endpt_infect
9	tab endpt_infect trtmt
10	csi 74 51 168 187, or
11	cs endpt_infect trtmt
12	label define ENDPT_IN 0 "...
13	replace endpt_infect = 2 if ...
14	cs endpt_infect trtmt
15	replace endpt_infect = 0 if ...
16	recode endpt_infect 0=1 1... 1
17	recode endpt_infect 0=1 1...
18	label define ENDPT_IN 0 "Y...
19	tab endpt_infect
20	cs endpt_infect trtmt
21	label define ENDPT_IN 0 "...
22	recode endpt_infect 1=0 0...
23	cs endpt_infect trtmt
24	disp 1/0.7
25	recode endpt_infect 1=0 0... 1
26	recode endpt_infect 1=0 0...
27	cs infect2 trtmt
28	disp (187/355)/(238/480)
29	disp (187/238)/(168/242)
30	logistic endpt_infect trtmt
31	logistic endpt_infect trtmt
32	disp 1/0.7
33	csi 51 74 187 168, or
34	เพิ่ม ค่าง 1
35	tab sex 1
36	tab dem_sex
37	tab dem_sex endpt_infect
38	tab dem_sex endpt_infect, r...
39	tab dem_sex endpt_infect, ...
40	cci 51 74 187 168

- Summaries, tables, and tests
- Linear models and related
- Binary outcomes
- Ordinal outcomes
- Categorical outcomes
- Count outcomes
- Fractional outcomes
- Generalized linear models
- Choice models
- Time series
- Multivariate time series
- Spatial autoregressive models
- Longitudinal/panel data
- Multilevel mixed-effects models
- Survival analysis
- Epidemiology and related
- Endogenous covariates
- Sample-selection models
- Treatment effects
- SEM (structural equation modeling)
- LCA (latent class analysis)
- FMM (finite mixture models)
- IRT (item response theory)
- Multivariate analysis
- Survey data analysis
- Lasso
- Meta-analysis
- Multiple imputation
- Nonparametric analysis
- Exact statistics
- Resampling
- Power, precision, and sample size
- Bayesian analysis
- Postestimation
- Other

Point estimate

- Logistic regression
- Probit regression
- Complementary log-log regression
- Conditional logistic regression
- Exact logistic regression
- Skewed logistic regression
- Probit model with endogenous covariates
- Probit model with sample selection
- Heteroskedastic probit regression
- Probit with endogeneity, selection, and treatment
- GLM for the binomial family
- Bivariate probit regression
- Seemingly unrelated bivariate probit regression
- Panel regression
- Multilevel mixed-effects regression
- FMM (finite mixture models)
- Nonparametric regression
- Treatment effects
- Lasso inferential models
- Bayesian regression
- Postestimation

chi2(1) = 5.

logistic - Logistic regression, reporting odds ratios

Model | by/if/in | Weights | SE/Robust | Reporting | Maximization

Dependent variable: endpt_infect

Independent variables: trtmt

☐ Suppress constant term

Options

Offset variable:

☐ Retain perfect predictor variables

Constraints: Manage...

OK | Cancel | Submit

logistic - Logistic regression, reporting odds ratios

Model | by/if/in | Weights | SE/Robust | Reporting | Maximization

95 Confidence level

◉ Report odds ratios (default)
○ Report estimated coefficients

☐ Do not report constraints

◉ Report p-values, test statistics, and confidence intervals
○ Suppress confidence intervals
○ Suppress p-values and test statistics

Factor-variable reporting options | Set table formats

OK | Cancel | Submit

Odds ratio in Logistic regression

Report odds ratios

```
. logistic endpt_infect trtmt

Logistic regression                               Number of obs     =        480
                                                  LR chi2(1)        =       5.24
                                                  Prob > chi2       =     0.0221
Log likelihood = -272.65617                       Pseudo R2         =     0.0095

------------------------------------------------------------------------------
endpt_infect | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       trtmt |   .6191646   .1304974    -2.27   0.023     .4096413    .9358549
       _cons |   .4404762   .0614553    -5.88   0.000     .3350909    .5790049
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

Report coefficients

```
. logit endpt_infect trtmt

Iteration 0:   log likelihood =  -275.2763
Iteration 1:   log likelihood = -272.66478
Iteration 2:   log likelihood = -272.65617
Iteration 3:   log likelihood = -272.65617

Logistic regression                               Number of obs     =        480
                                                  LR chi2(1)        =       5.24
                                                  Prob > chi2       =     0.0221
Log likelihood = -272.65617                       Pseudo R2         =     0.0095

------------------------------------------------------------------------------
endpt_infect |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       trtmt |  -.4793841   .2107637    -2.27   0.023    -.8924733   -.0662949
       _cons |  -.8198989   .1395202    -5.88   0.000    -1.093354   -.5464443
------------------------------------------------------------------------------
```

OR = exp β = exp 0.479 = e^0.479

Coef. = ln (OR) = ln 0.619 = 0.479

Regression model

Regression model

- Prediction of outcome from exposures
- Hypothesis testing with adjusted (controlled) for confounding other variables
- Simple Y □ X ; Outcome □ Factors
  - BP □ Drug (A/B) ; BP = a + b(Drug)
  - BP □ Age ; BP = a + b(Age)
- Multiple Y □ X1 X2 X3
- BP □ Drug(A/B) Sex (M/F) Age

  BP = a + b1(Drug) + b2 (Sex) + b3 (Age)

  BP = 100 + 3 Drug + 5 Sex - 7 Age

  A=0 / B=1 M=0 / F=1

Multi-variables analysis

- Linear Regression
  - Y = continuous + normal distribution
  - Injury severity score

$$y_i = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

- Logistic Regression
  - Y = Categorical
  - Dead/ severity of injury/ bone fracture

$$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

OR = exp β

- Poisson Regression
  - Y = Incidence/count
  - Dead

$$\log[count] = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

$$\log[count \ \ or \ cases / time] = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

$$\log[rate = \lambda] = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

IRR = exp β

- Cox's Proportional Hazard Regression
  - Y = Time to event
  - Time from injury to dead

$$\log(Hz_t) = \log(Hz_0) + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

$$\log(\lambda_t) = \log(\lambda_0) + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

HR = exp β

X= continuous / categorical
; sex, age, alcohol drinking

Note: If $P_0(T)$ is small, when comparing two groups

$$OR(t) \approx RR(t) \approx IRR(t) \approx HR(t) \approx K$$

Principles for studying association

- Start with graphical display: **scatterplots**
  - Display the relationship between two quantitative variables.
  - The values of one variable appear on the horizontal axis (the x axis) and the values of the other variable on the vertical axis (the *y* axis).
  - Each individual is the point in the plot fixed by the values of both variables for that individual.
  - In regression, usually call the explanatory variable x and the response variable *y*.
- Look for **overall patterns** and for striking deviations from the pattern : interpreting scatterplots
  - Overall pattern: the relationship has ...
    - form (linear relationships, curved relationships, clusters)
    - direction (positive/negative association)
    - strength (how close the points follow a clear form?)
    - Outliers
  - For a categorical x and quantitative *y*, show the distributions of y for each category of x.
- When the overall pattern is quite regular, use a compact mathematical **model** to describe it.

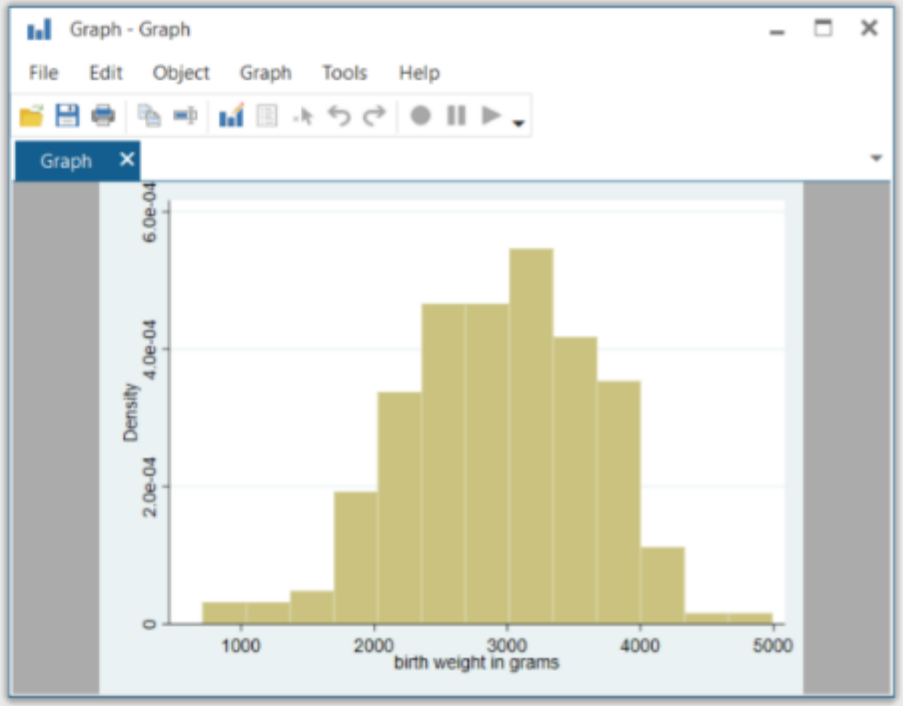
Linear Regression

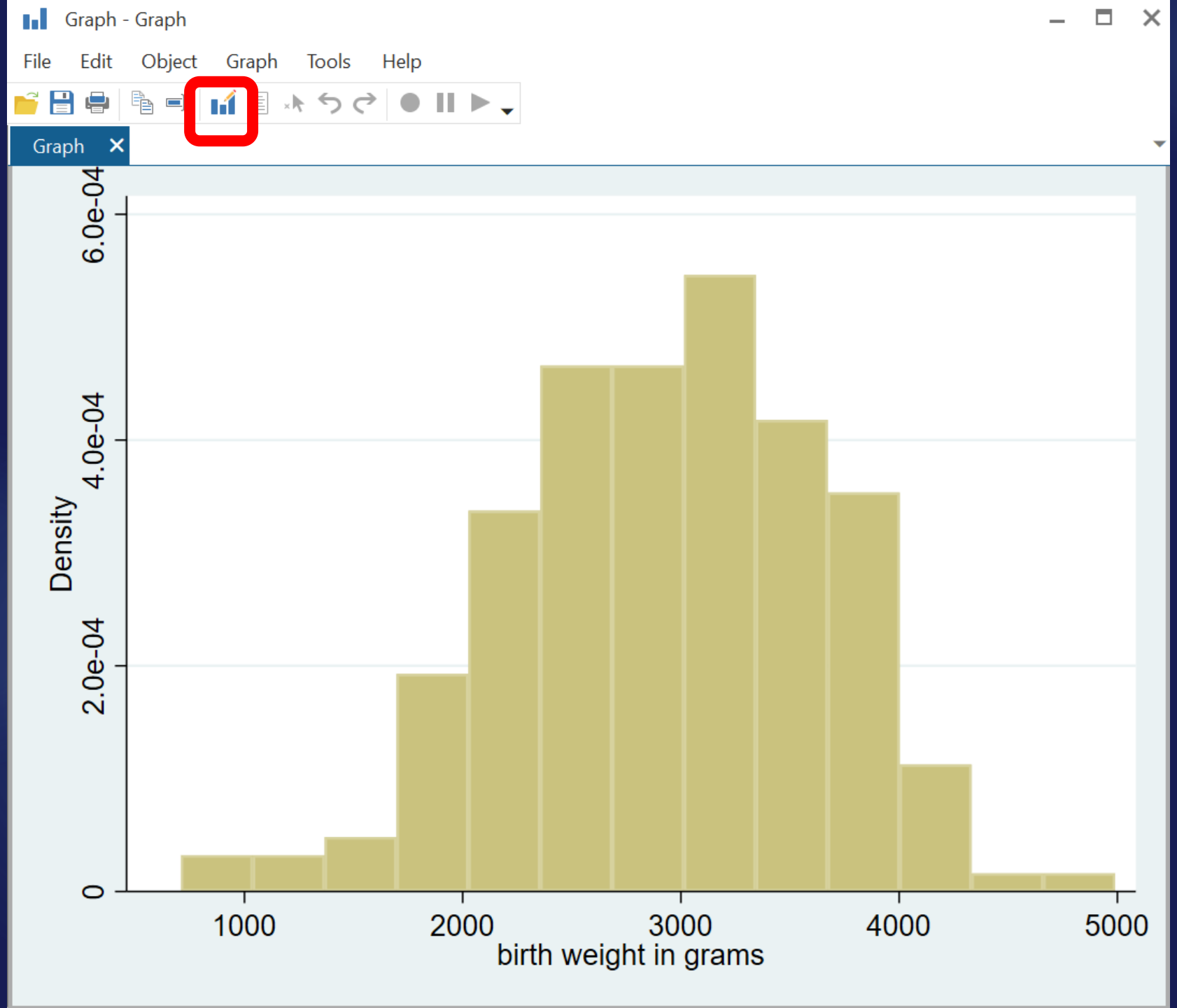
Y = CONTINUOUS + NORMAL DISTRIBUTION

$$y_i = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

Normal distribution of Y


```
.
. use "C:\Users\Thananan\Desktop\practice data sets\lowbwt_dat.dta"

.
. hist bwt
(bin=13, start=709, width=329.30769)

.
. swilk bwt

                   Shapiro-Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+-------------------------------------------------------
         bwt |        189    0.99258      1.053     0.119    0.45247
```

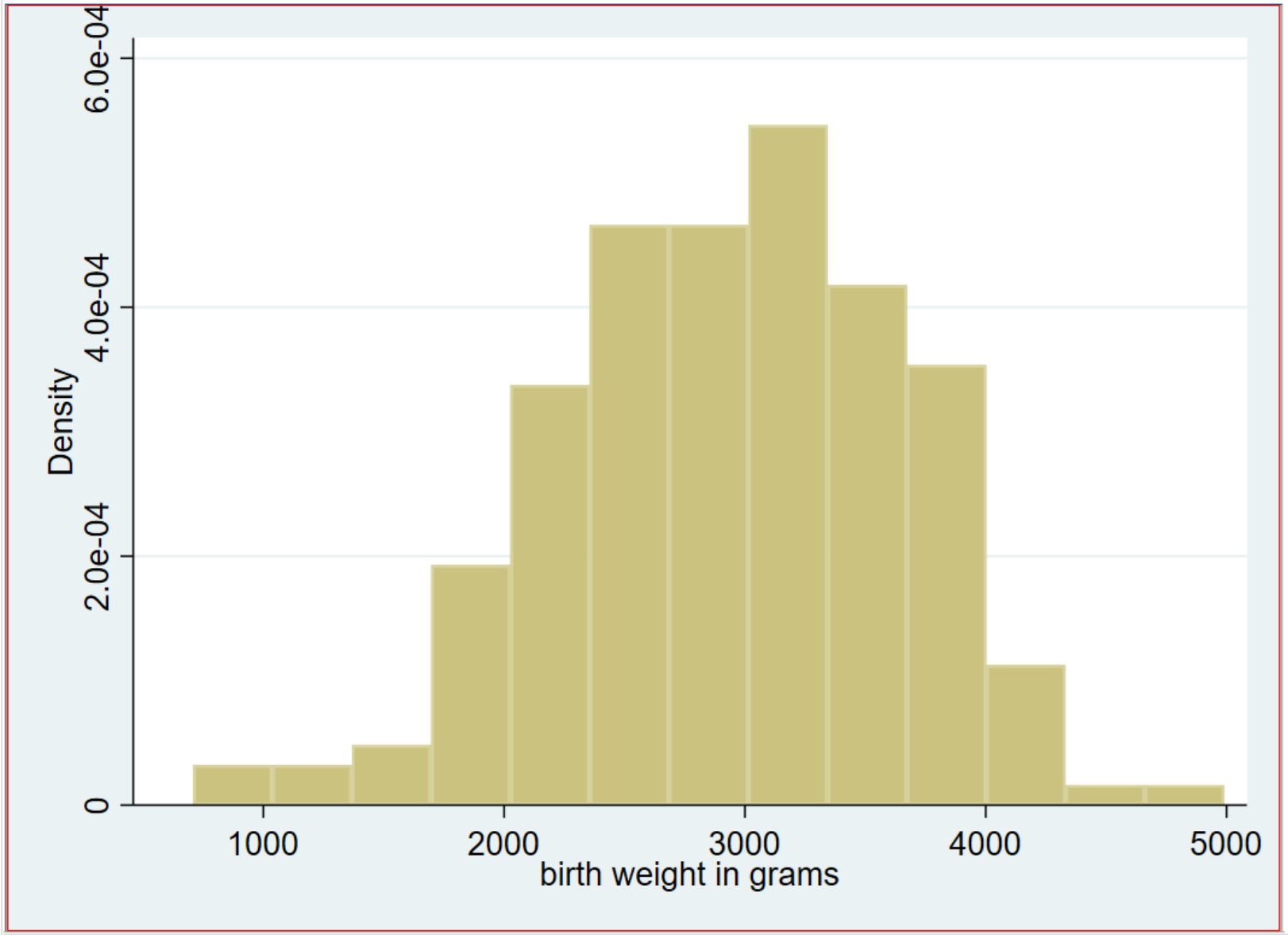



Graph - Graph
File
Edit
Object
Graph
Tools
Help
Graph
Density
0
2.0e-04
4.0e-04
6.0e-04
1000
2000
3000
4000
5000
birth weight in grams

Graph - Graph

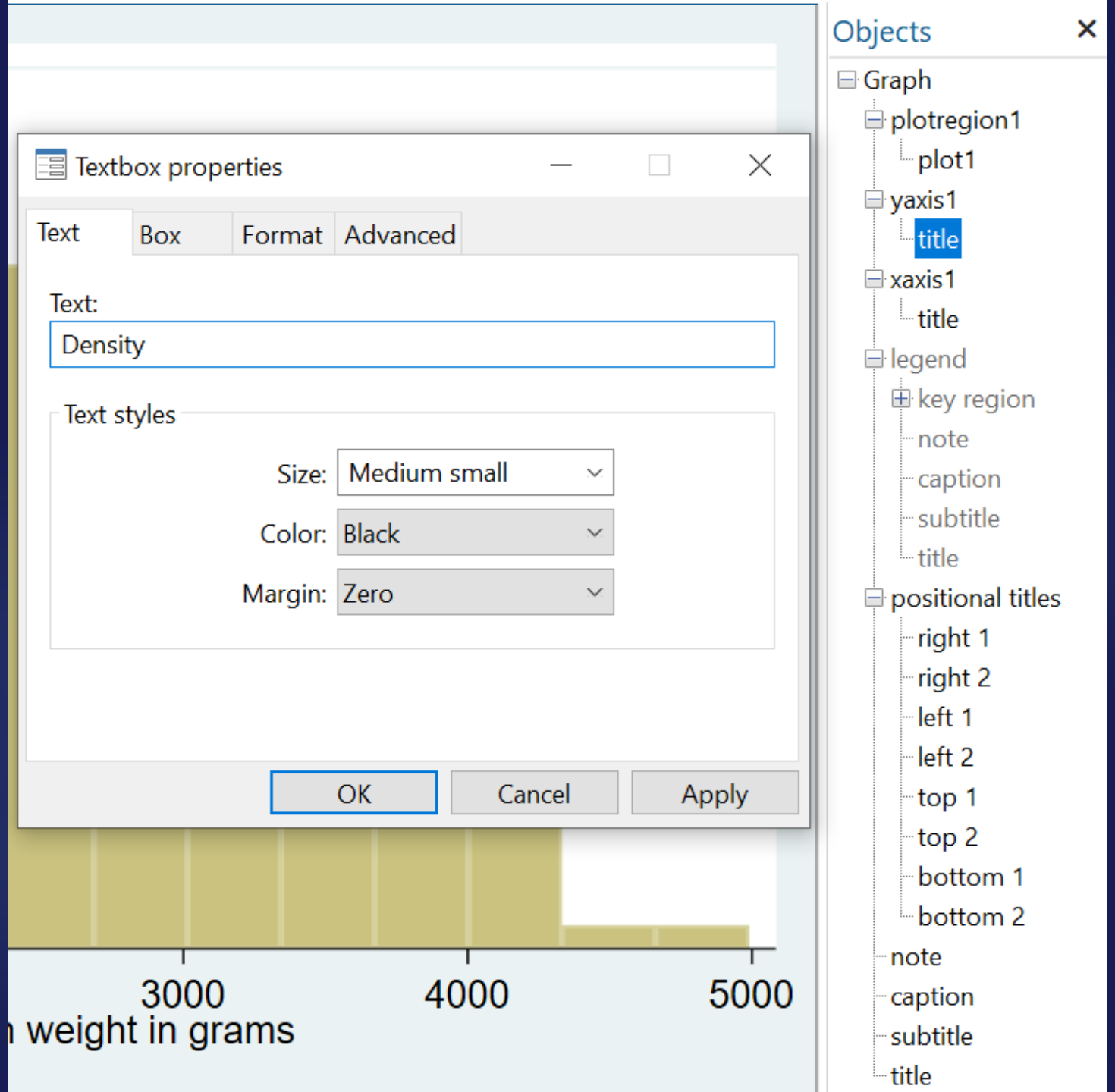
File Edit Object Graph Tools Help

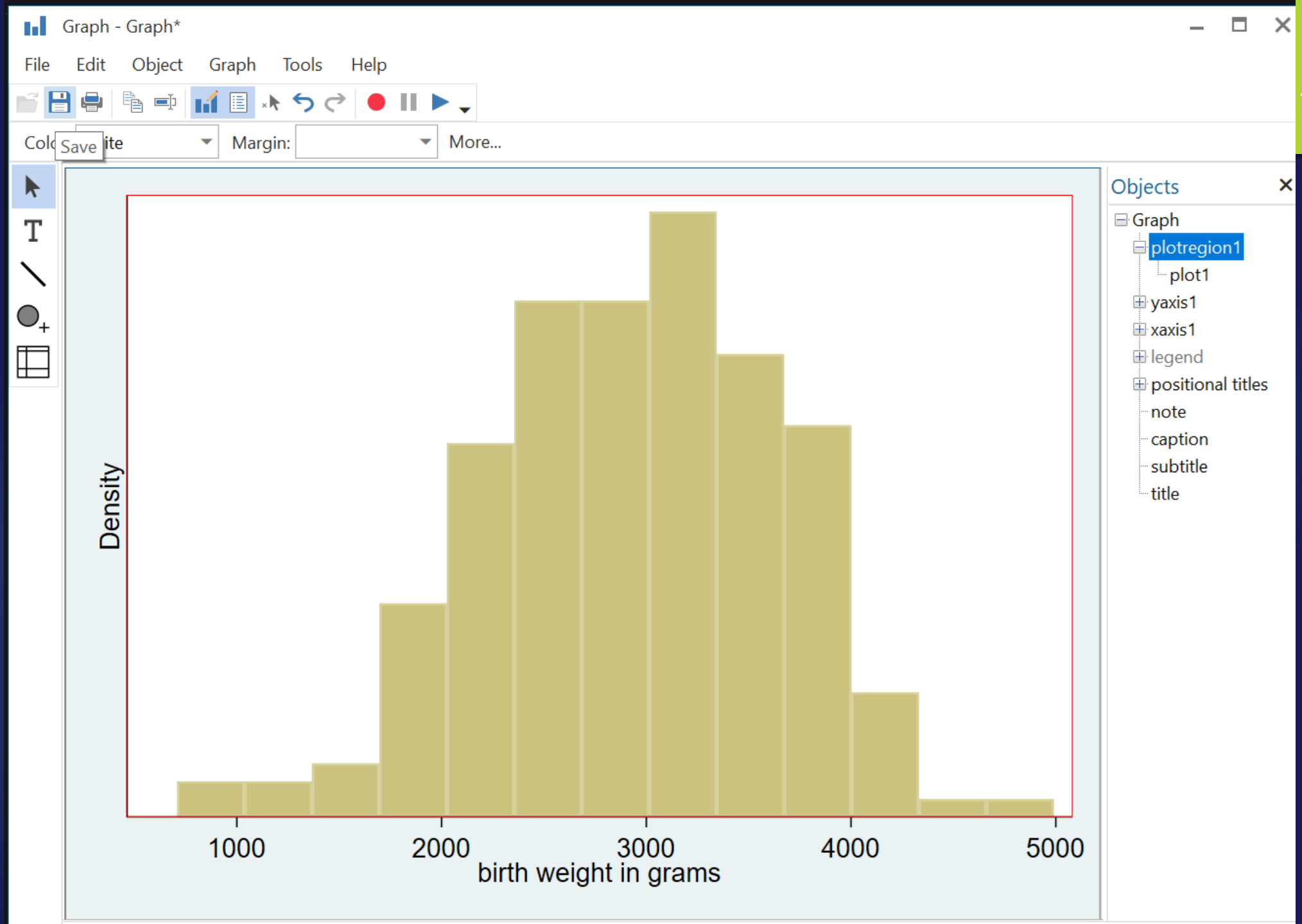
Color: Light bluish-gra Scale: 1 Aspect ratio: 0 Title: More...



Objects

- Graph
  - plotregion1
    - plot1
  - yaxis1
    - title
  - xaxis1
    - title
  - legend
    - key region
    - note
    - caption
    - subtitle
    - title
  - positional titles
    - right 1
    - right 2
    - left 1
    - left 2
    - top 1
    - top 2
    - bottom 1
    - bottom 2
  - note
  - caption
  - subtitle
  - title


Textbox properties
Text
Box
Format
Advanced
Text:
Density
Text styles
Size:
Medium small
Color:
Black
Margin:
Zero
OK
Cancel
Apply
3000
4000
5000
weight in grams
Objects
Graph
plotregion1
plot1
yaxis1
title
xaxis1
title
legend
key region
note
caption
subtitle
title
positional titles
right 1
right 2
left 1
left 2
top 1
top 2
bottom 1
bottom 2
note
caption
subtitle
title

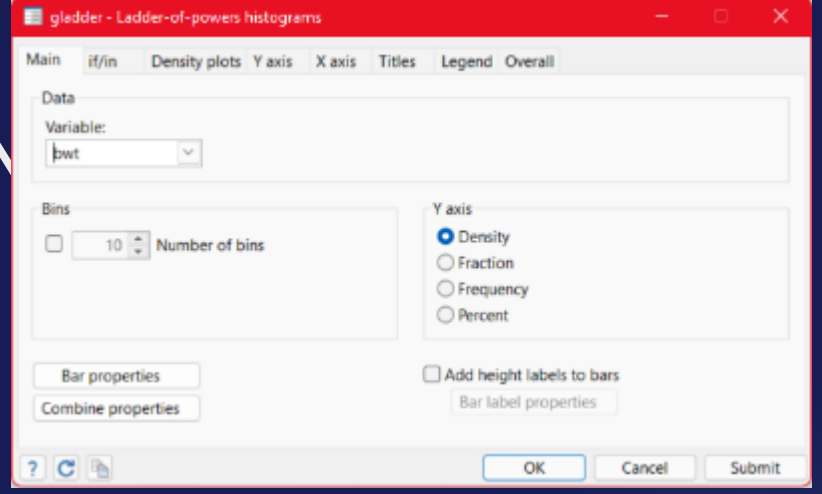

Graph - Graph*
File
Edit
Object
Graph
Tools
Help
Save
Margin:
More...
Objects
Graph
plotregion1
plot1
yaxis1
xaxis1
legend
positional titles
note
caption
subtitle
title
Density
1000
2000
3000
4000
5000
birth weight in grams

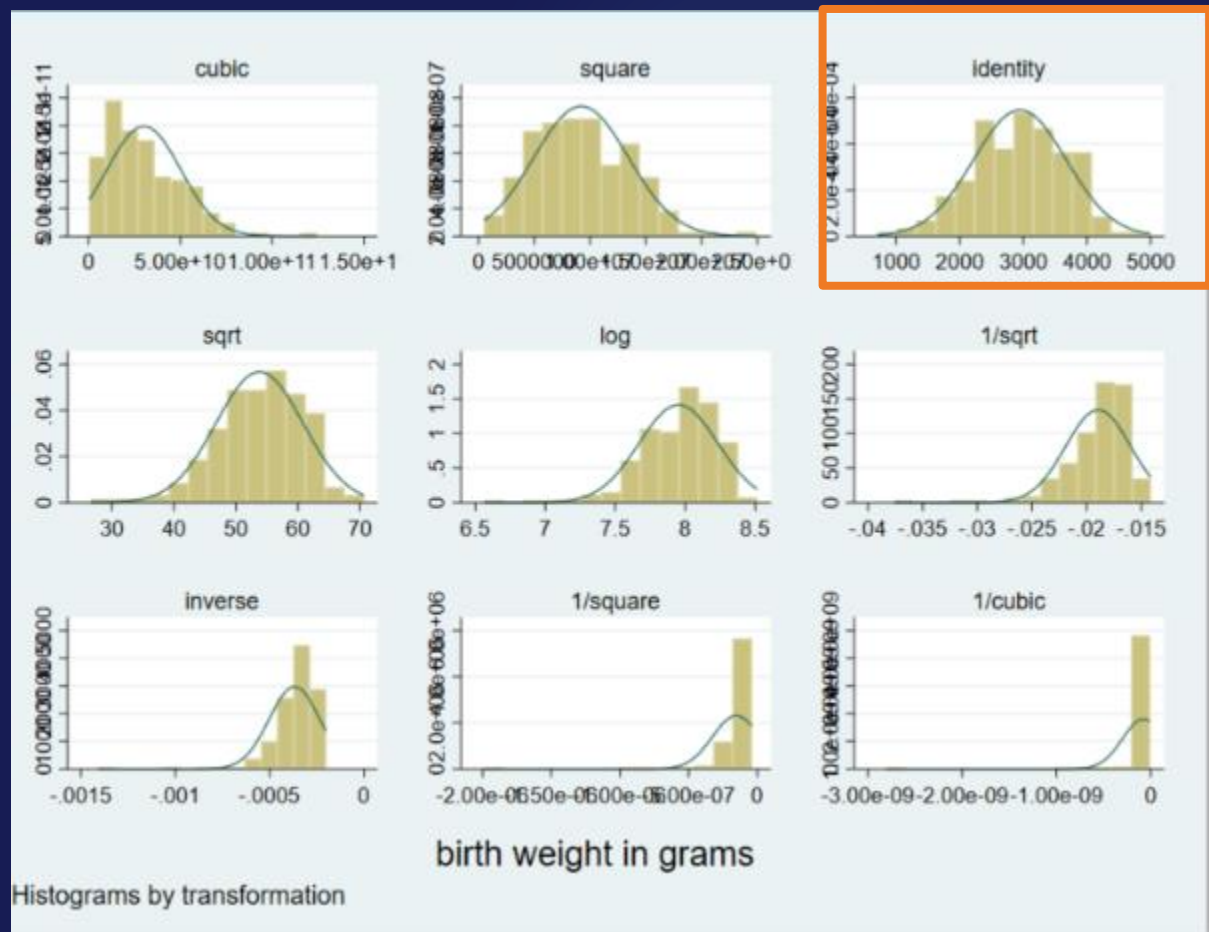
Statistics User Window Help

- Summaries, tables, and tests
  - Summary and descriptive statistics
  - Frequency tables
  - Other tables
  - Classical tests of hypotheses
  - Nonparametric tests of hypotheses
  - Distributional plots and tests
    - Symmetry plot
    - Quantiles plot
    - Normal quantile plot
    - Normal probability plot, standardized
    - Chi-squared quantile plot
    - Chi-squared probability plot
    - Quantile-quantile plot
    - Stem-and-leaf display
    - Letter-value display
    - Generate cumulative distribution
    - Skewness and kurtosis normality test
    - Shapiro-Wilk normality test
    - Shapiro-Francia normality test
    - Ladder of powers
    - Ladder-of-powers histograms
    - Ladder-of-powers quantile-normal plot
  - Multivariate test of means, covariances, and normality
- Linear models and related
- Binary outcomes
- Ordinal outcomes
- Categorical outcomes
- Count outcomes
- Fractional outcomes
- Generalized linear models
- Choice models
- Time series
- Multivariate time series
- Spatial autoregressive models
- Longitudinal/panel data
- Multilevel mixed-effects models
- Survival analysis
- Epidemiology and related
- Endogenous covariates
- Sample-selection models
- Treatment effects
- SEM (structural equation modeling)
- LCA (latent class analysis)
- FMM (finite mixture models)
- IRT (item response theory)
- Multivariate analysis
- Survey data analysis
- Lasso
- Meta-analysis
- Multiple imputation
- Nonparametric analysis
- Exact statistics
- Resampling
- Power, precision, and sample size
- Bayesian analysis
- Postestimation
- Other

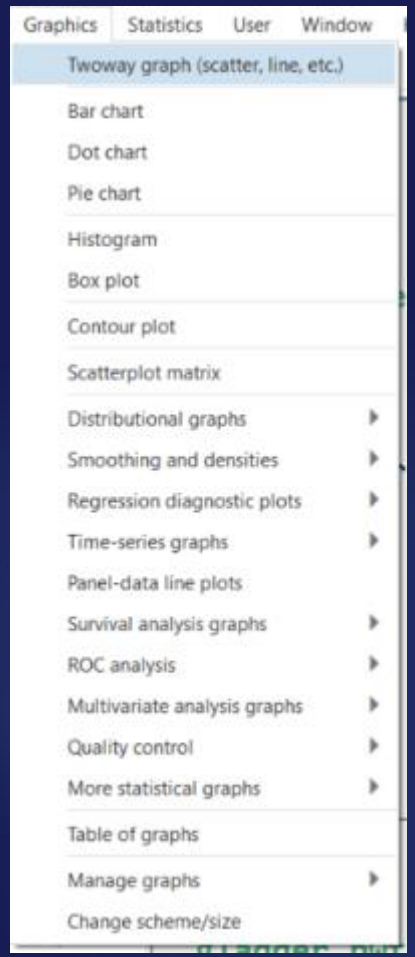
```
nanan\Desktop\practice data se

width=329.30769)

Shapiro-Wilk W test for normal
```

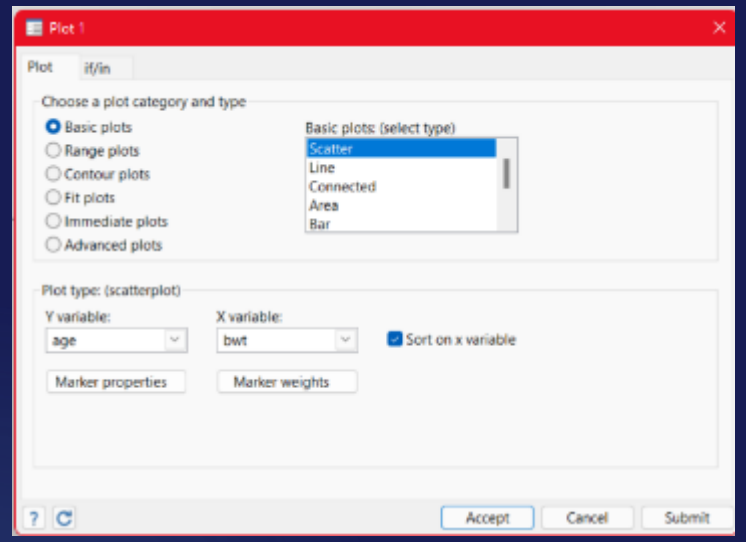
Obs	W	V	z	Prob>z
189	0.99258	1.053	0.119	0.45247

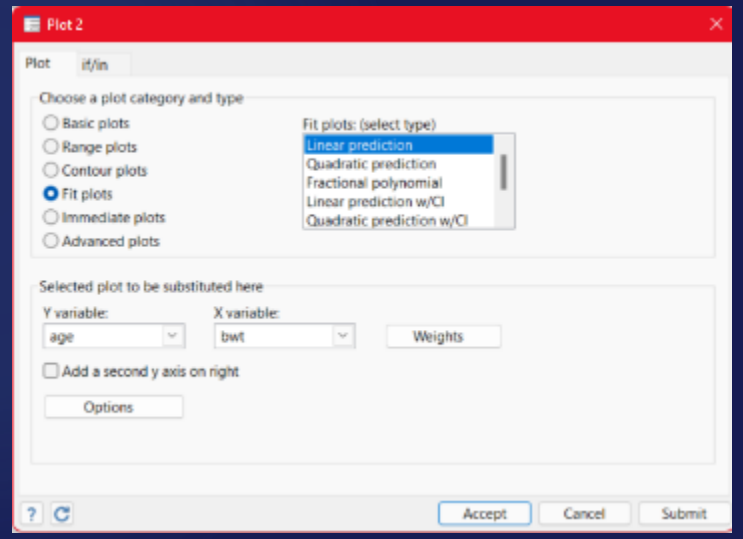


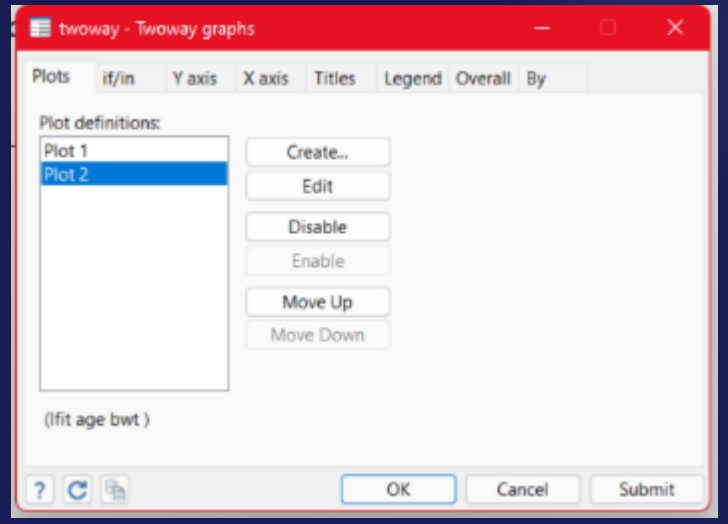


scatterplots

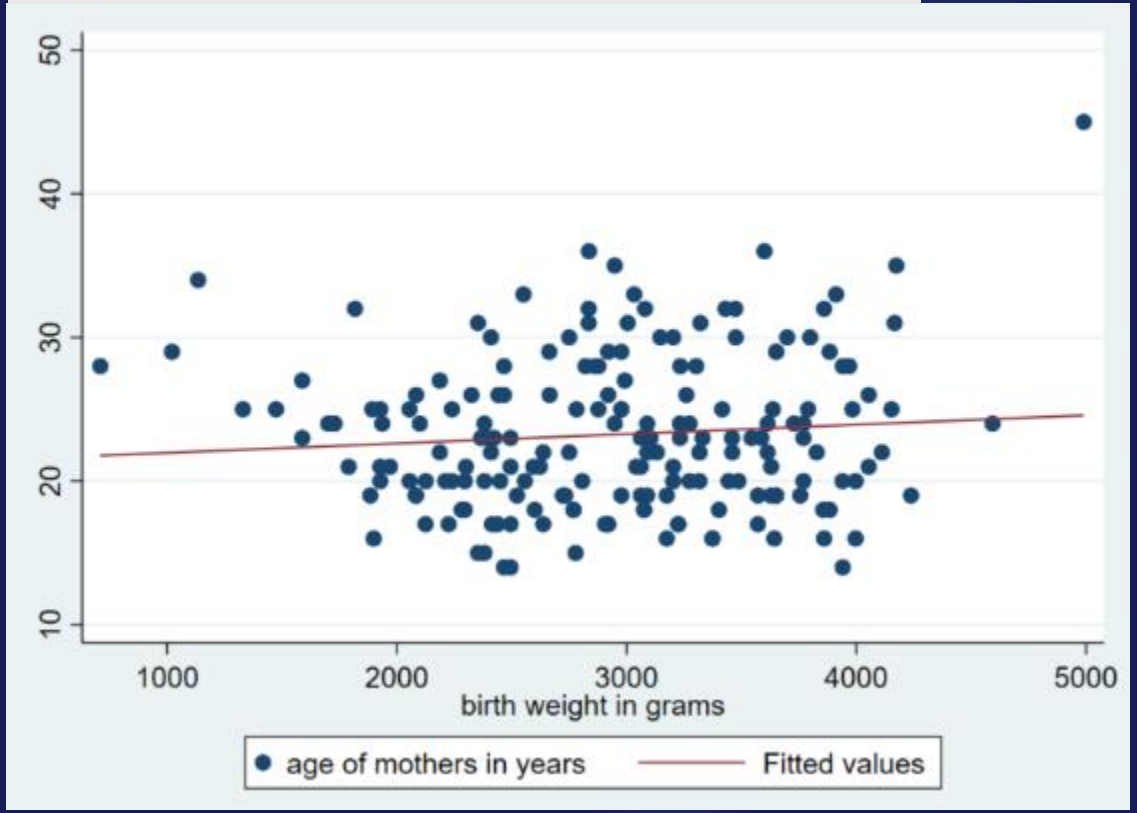




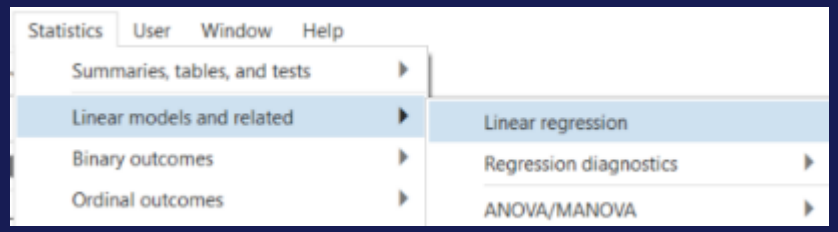


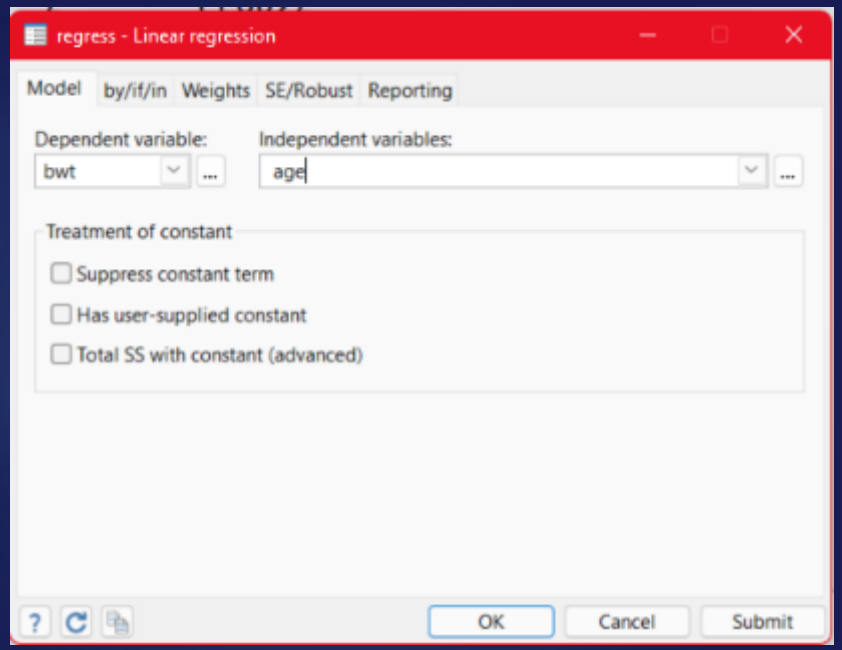



```
twoway (scatter age bwt, sort) (lfit age bwt)
```



Linear regression






```
. regress bwt age

      Source |       SS           df       MS      Number of obs   =       189
-------------+----------------------------------   F(1, 187)       =      1.52
       Model |  806926.913         1  806926.913   Prob > F        =    0.2188
    Residual |  99110125.7       187  530000.672   R-squared       =    0.0081
-------------+----------------------------------   Adj R-squared   =    0.0028
       Total |  99917052.6       188  531473.684   Root MSE        =    728.01

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   12.36433   10.02055     1.23   0.219    -7.403527    32.13219
       _cons |   2657.333    238.804    11.13   0.000     2186.236    3128.429
------------------------------------------------------------------------------
```

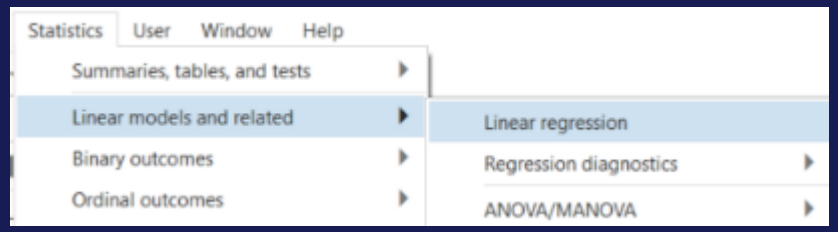
Y = a + bx

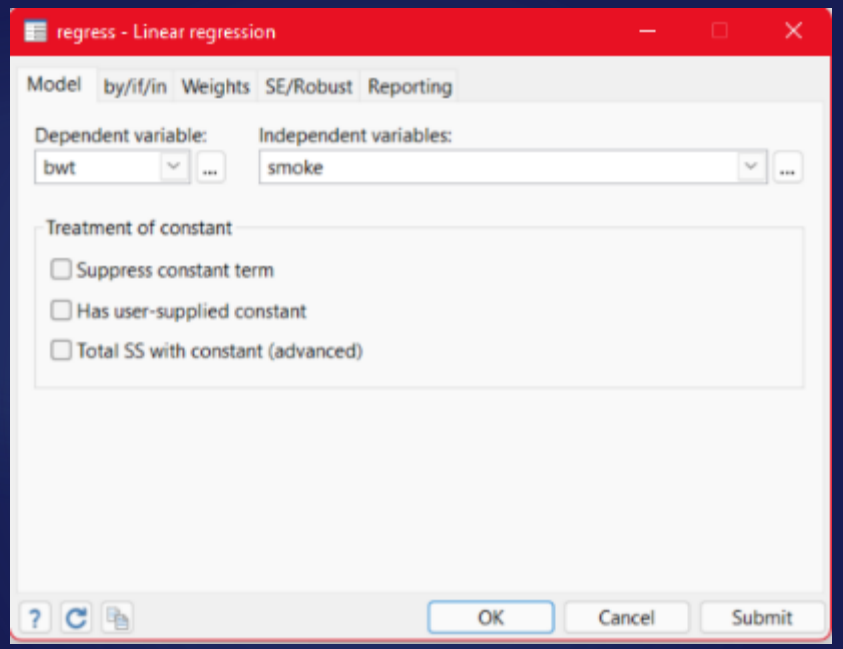
ทุก 1 หน่วย x ที่เพิ่มขึ้น จะเพิ่ม y ขึ้น b หน่วย

Bwt = 2657.33 + 12.36 (age)

ถ้าแม่มีอายุเพิ่มขึ้น 1 ปี ลูกจะมีน้ำหนักเพิ่มขึ้น 12.36 กรัม

Linear regression






```
. regress bwt smoke

      Source |       SS           df       MS      Number of obs   =       189
-------------+----------------------------------   F(1, 187)       =      6.94
       Model |  3573406.24         1  3573406.24   Prob > F        =    0.0092
    Residual |  96343646.4       187  515206.665   R-squared       =    0.0358
-------------+----------------------------------   Adj R-squared   =    0.0306
       Total |  99917052.6       188  531473.684   Root MSE        =    717.78

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |  -281.7133   106.9687    -2.63   0.009    -492.7338   -70.69274
       _cons |   3054.957   66.93324    45.64   0.000     2922.915    3186.998
------------------------------------------------------------------------------
```

Y = a + bx

ทุก 1 หน่วย x ที่เพิ่มขึ้น จะเพิ่ม y ขึ้น b หน่วย

Bwt = 3054.957 – 281.7133 (age)

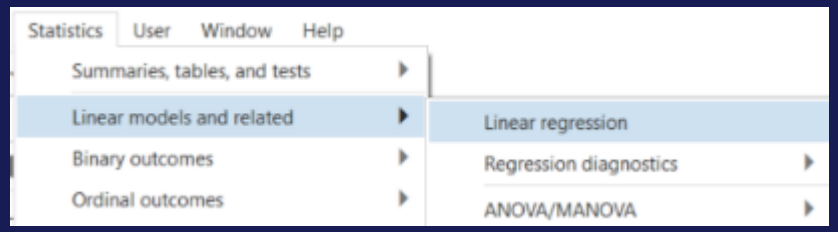
ถ้าแม่สูบบุหรี่ ลูกจะมีน้ำหนักลดลง 281.7133 กรัม

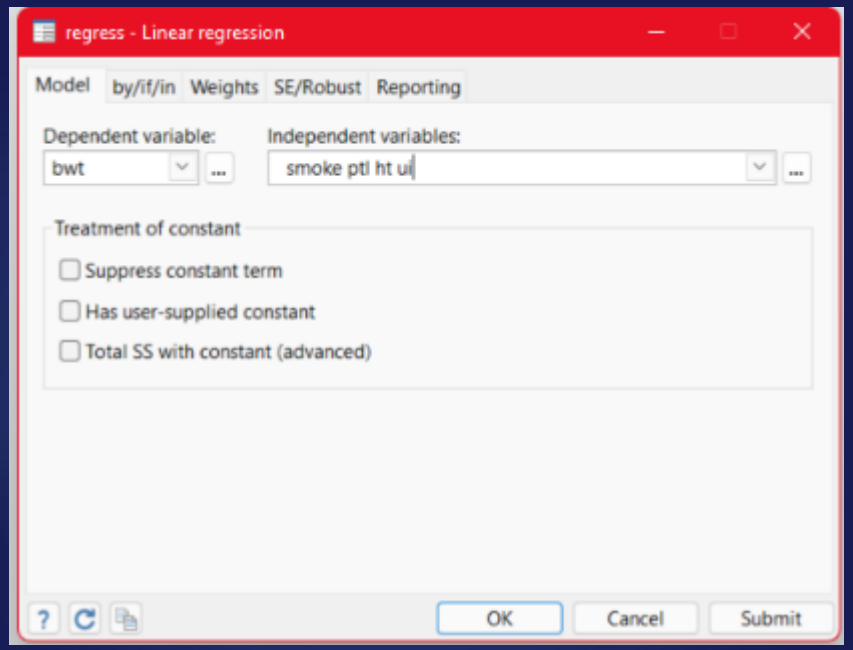
Univariate


```
. regress bwt smoke

      Source |       SS           df       MS      Number of obs   =       189
-------------+----------------------------------   F(1, 187)       =      6.94
       Model |  3573406.24         1  3573406.24   Prob > F        =    0.0092
    Residual |  96343646.4       187  515206.665   R-squared       =    0.0358
-------------+----------------------------------   Adj R-squared   =    0.0306
       Total |  99917052.6       188  531473.684   Root MSE        =    717.78

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |  -281.7133   106.9687    -2.63   0.009    -492.7338   -70.69274
       _cons |   3054.957   66.93324    45.64   0.000     2922.915    3186.998
------------------------------------------------------------------------------

. regress bwt ht

      Source |       SS           df       MS      Number of obs   =       189
-------------+----------------------------------   F(1, 187)       =      4.08
       Model |  2132014.49         1  2132014.49   Prob > F        =    0.0449
    Residual |  97785038.2       187  522914.643   R-squared       =    0.0213
-------------+----------------------------------   Adj R-squared   =    0.0161
       Total |  99917052.6       188  531473.684   Root MSE        =    723.13

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ht |  -435.5607   215.7094    -2.02   0.045    -861.0973   -10.02413
       _cons |   2972.311   54.35366    54.68   0.000     2865.086    3079.536
------------------------------------------------------------------------------
```

```
. regress bwt ui

      Source |       SS           df       MS      Number of obs   =       189
-------------+----------------------------------   F(1, 187)       =     16.34
       Model |  8028747.44         1  8028747.44   Prob > F        =    0.0001
    Residual |  91888305.2       187  491381.311   R-squared       =    0.0804
-------------+----------------------------------   Adj R-squared   =    0.0754
       Total |  99917052.6       188  531473.684   Root MSE        =    700.99

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ui |  -580.1801   143.5318    -4.04   0.000    -863.3298   -297.0304
       _cons |   3030.609   55.24543    54.86   0.000     2921.624    3139.593
------------------------------------------------------------------------------

. regress bwt ptl

      Source |       SS           df       MS      Number of obs   =       189
-------------+----------------------------------   F(1, 187)       =      4.59
       Model |  2392205.02         1  2392205.02   Prob > F        =    0.0335
    Residual |  97524847.6       187  521523.249   R-squared       =    0.0239
-------------+----------------------------------   Adj R-squared   =    0.0187
       Total |  99917052.6       188  531473.684   Root MSE        =    722.17

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ptl |  -228.6506   106.7603    -2.14   0.034      -439.26   -18.04114
       _cons |   2989.418   56.53494    52.88   0.000      2877.89    3100.947
------------------------------------------------------------------------------
```

Multivariate






```
. regress bwt smoke ptl ht ui

      Source |       SS           df       MS      Number of obs   =       189
-------------+----------------------------------   F(4, 184)       =      7.75
       Model |  14402866.8         4  3600716.7   Prob > F        =    0.0000
    Residual |  85514185.8       184  464751.01   R-squared       =    0.1441
-------------+----------------------------------   Adj R-squared   =    0.1255
       Total |  99917052.6       188  531473.684   Root MSE        =    681.73

------------------------------------------------------------------------------
         bwt |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       smoke |  -234.4121   103.4714    -2.27   0.025     -438.555   -30.26921
         ptl |  -95.53186   105.1879    -0.91   0.365    -303.0614    111.9977
          ht |  -522.2454   204.6148    -2.55   0.012    -925.9383   -118.5526
          ui |  -568.9723   144.2281    -3.94   0.000    -853.5258   -284.4187
       _cons |   3172.589   68.24453    46.49   0.000     3037.947    3307.232
------------------------------------------------------------------------------
```

Bwt = 3172.59 – 234.41(smoke) – 95.53(ptl) – 522.2454(ht) – 568.97(ui)

Baby birth weight will decrease 234.41g in smoking mother after adjusted for history of premature labor, hypertension, and uterine irritation with statistic significant.

Logistic Regression

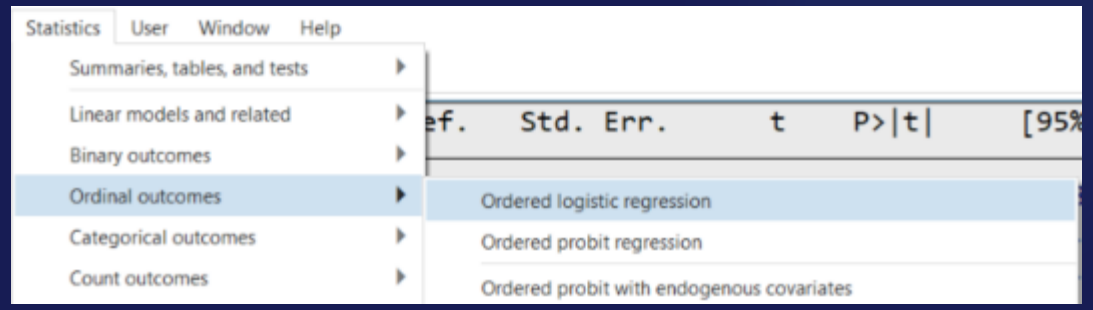
Y = CATEGORICAL

$$\log \frac{p_i}{1-p_i} = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

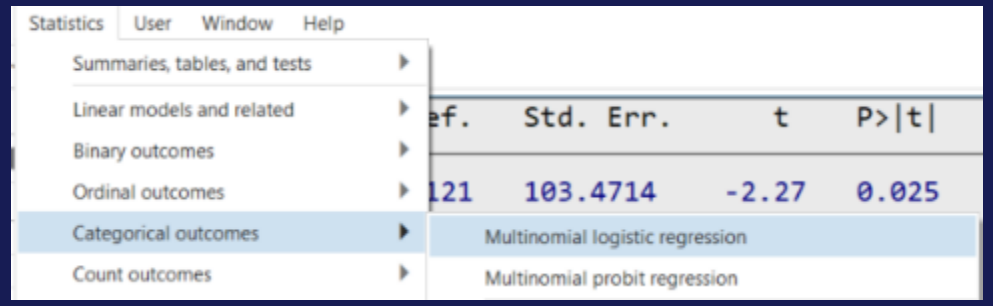
Logistic Regression

- Binary outcome
  - Outcome = yes/no

- Ordinal outcome
  - Outcome = level
  - Ex. Tumor grade



- Multinomial outcome
  - Outcome = category



Binary Logistic regression

Report odds ratios

```
. logistic endpt_infect trtmt

Logistic regression                               Number of obs     =        480
                                                  LR chi2(1)        =       5.24
                                                  Prob > chi2       =     0.0221
Log likelihood = -272.65617                       Pseudo R2         =     0.0095

------------------------------------------------------------------------------
endpt_infect | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       trtmt |   .6191646   .1304974    -2.27   0.023     .4096413    .9358549
       _cons |   .4404762   .0614553    -5.88   0.000     .3350909    .5790049
------------------------------------------------------------------------------
Note: _cons estimates baseline odds.
```

Report coefficients

```
. logit endpt_infect trtmt

Iteration 0:   log likelihood =  -275.2763
Iteration 1:   log likelihood = -272.66478
Iteration 2:   log likelihood = -272.65617
Iteration 3:   log likelihood = -272.65617

Logistic regression                               Number of obs     =        480
                                                  LR chi2(1)        =       5.24
                                                  Prob > chi2       =     0.0221
Log likelihood = -272.65617                       Pseudo R2         =     0.0095

------------------------------------------------------------------------------
endpt_infect |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       trtmt |  -.4793841   .2107637    -2.27   0.023    -.8924733   -.0662949
       _cons |  -.8198989   .1395202    -5.88   0.000    -1.093354   -.5464443
------------------------------------------------------------------------------
```

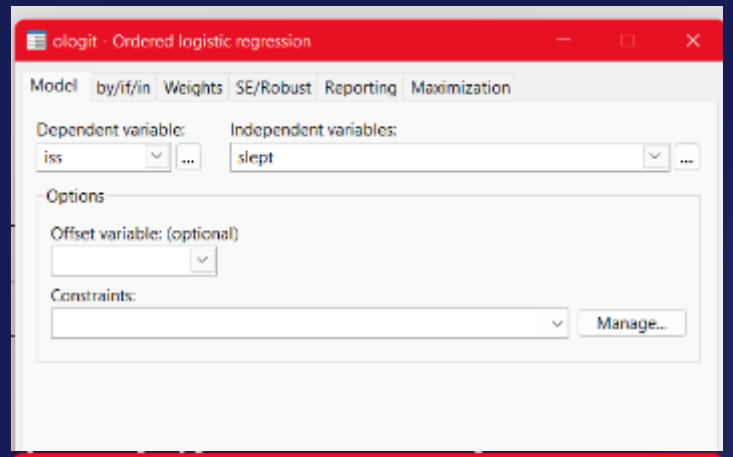
Interpretation: การได้รับวัคซีนในกลุ่มที่ติดเชื้อ เป็น 0.62 เท่า ของกลุ่มที่ไม่ติดเชื้อ

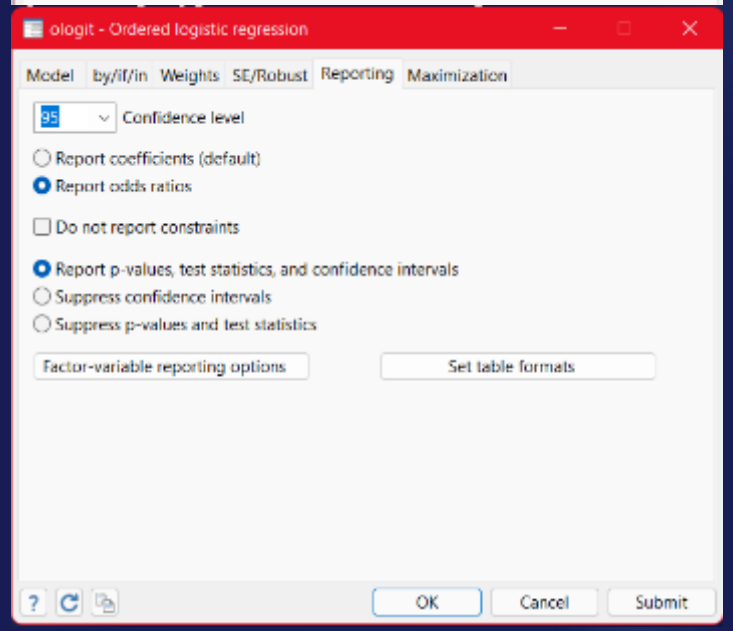
แปลแบบ disease OR = การติดเชื้อในกลุ่มได้รับวัคซีนเป็น 0.62 เท่าของกลุ่มที่ไม่ได้รับวัคซีน

แปลแบบ RR = คนที่ได้รับวัคซีนมีโอกาสติดเชื้อเป็น 0.62 เท่าของคนที่ไม่ได้รับ

Ordinal Logistic regression

- Outcome = level
- Ex. 1 2 3 4 5
- เป็นการเปรียบเทียบกับระดดับที่สูงกว่าและต่ำกว่า
- 1 เทียบกับ 2, 2 เทียบกับ 1 และ 3, 3 เทียบกับ 2 และ 4, 4 เทียบกับ 3 และ 5, 5 เทียบกับ 4





Ordinal Logistic regression


```
. use "C:\Users\Thananan\Downloads\rti korat.dta"

. tab iss
```

level of severity	Freq.	Percent	Cum.
1	38	64.41	64.41
2	11	18.64	83.05
3	1	1.69	84.75
4	4	6.78	91.53
5	1	1.69	93.22
6	4	6.78	100.00
Total	59	100.00	

```
. ologit iss slept,or

Iteration 0:   log likelihood = -50.928458
Iteration 1:   log likelihood = -50.567058
Iteration 2:   log likelihood =  -50.56545
Iteration 3:   log likelihood =  -50.56545

Ordered logistic regression                     Number of obs     =         44
                                                LR chi2(1)        =       0.73
                                                Prob > chi2       =     0.3942
Log likelihood =  -50.56545                     Pseudo R2         =     0.0071
```

iss	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
slept	1.874469	1.419475	0.83	0.407	.4249042	8.269241
/cut1	.8599954	.6808475			-.4744412	2.194432
/cut2	1.867402	.7317753			.4331486	3.301655
/cut3	2.571326	.7959231			1.011345	4.131306
/cut4	2.820017	.8267848			1.199549	4.440486

Note: Estimates are transformed only in the first equation.

Ordinal Logistic regression

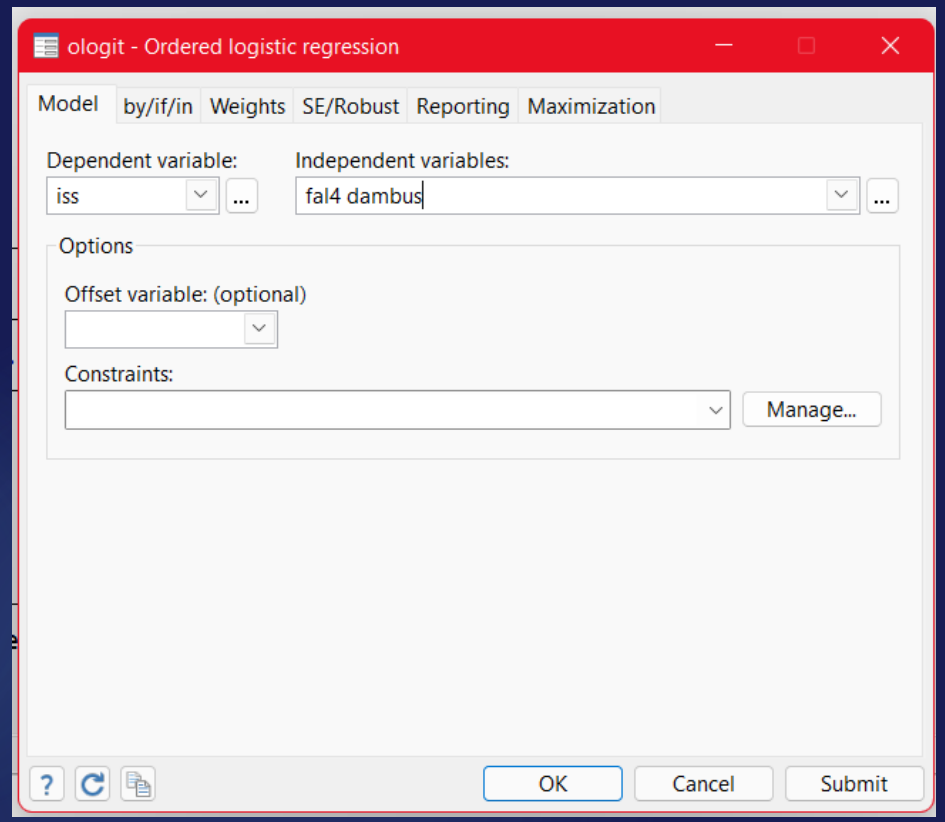
```
. ologit iss fal4,or

Iteration 0:   log likelihood = -64.879232
Iteration 1:   log likelihood = -61.732332
Iteration 2:   log likelihood = -61.662334
Iteration 3:   log likelihood = -61.662308
Iteration 4:   log likelihood = -61.662308

Ordered logistic regression                     Number of obs     =         59
                                                LR chi2(1)        =       6.43
                                                Prob > chi2       =     0.0112
Log likelihood = -61.662308                     Pseudo R2         =     0.0496

------------------------------------------------------------------------------
         iss | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        fal4 |   4.495933    2.86405     2.36   0.018     1.289965    15.66974
-------------+----------------------------------------------------------------
       /cut1 |   1.585876   .5470079                      .5137605    2.657992
       /cut2 |   2.662984   .6141425                      1.459287    3.866682
       /cut3 |   2.799218   .6259402                      1.572398    4.026038
       /cut4 |   3.501766   .7021733                      2.125531       4.878
       /cut5 |   3.751654    .738761                      2.303709    5.199599
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
```

```
. ologit iss dambus,or

Iteration 0:   log likelihood = -64.879232
Iteration 1:   log likelihood = -56.912306
Iteration 2:   log likelihood = -56.244907
Iteration 3:   log likelihood = -56.241332
Iteration 4:   log likelihood = -56.241331

Ordered logistic regression                     Number of obs     =         59
                                                LR chi2(1)        =      17.28
                                                Prob > chi2       =     0.0000
Log likelihood = -56.241331                     Pseudo R2         =     0.1331

------------------------------------------------------------------------------
         iss | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dambus |   13.53989   13.971538     3.93   0.000      3.69495    49.61599
-------------+----------------------------------------------------------------
       /cut1 |   1.256891   .3573746                      .5564498    1.957333
       /cut2 |   2.604729   .5040012                      1.616905    3.592554
       /cut3 |   2.787684   .5289695                      1.750923    3.824445
       /cut4 |   3.663075   .6500212                      2.389057    4.937094
       /cut5 |   3.939048   .6936303                      2.579557    5.298538
------------------------------------------------------------------------------
Note: Estimates are transformed only in the first equation.
```

การตกจากเก้าอี้จะทำให้การบาดเจ็บรุนแรงขึ้น 4.5 เท่า

การนั่งบริเวณที่มีการยุบตัวของตัวถังรถจะทำให้การบาดเจ็บรุนแรงขึ้น 13.54 เท่า

Ordinal Logistic regression




```
. ologit iss fal4 dambus, or

Iteration 0:   log likelihood = -64.879232
Iteration 1:   log likelihood = -54.674626
Iteration 2:   log likelihood = -53.501901
Iteration 3:   log likelihood = -53.492808
Iteration 4:   log likelihood = -53.492806

Ordered logistic regression                     Number of obs     =         59
                                                LR chi2(2)        =      22.77
                                                Prob > chi2       =     0.0000
Log likelihood = -53.492806                     Pseudo R2         =     0.1755

------------------------------------------------------------------------------
         iss | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        fal4 |    4.69239   3.313033     2.19   0.029     1.176003    18.72318
      dambus |   13.41914   9.103078     3.83   0.000     3.550581    50.71661
-------------+----------------------------------------------------------------
       /cut1 |   2.290707   .6576021                      1.001831    3.579584
       /cut2 |   3.748644   .7987795                      2.183065    5.314223
       /cut3 |   3.964235   .8323803                        2.3328    5.595671
       /cut4 |   4.940901   .9605194                      3.058317    6.823484
       /cut5 |   5.231272   .9959223                        3.2793    7.183244
------------------------------------------------------------------------------
```

การตกจากเก้าอี้จะทำให้การบาดเจ็บรุนแรงขึ้น 4.69 เท่า
เมื่อคำนึงถึงตัวแปรรบกวนจากการนั่งบริเวณที่มีการ

Multinomial Logistic regression


```
. use "C:\Users\Thananan\Desktop\Tobacco Project\Health Sit\Microdata HBS 2564\HBS64.dta"

. tab quitint
```

intention to quit	Freq.	Percent	Cum.
no	15,170	37.19	37.19
yes but never try	10,567	25.90	63.09
yes and try	15,057	36.91	100.00
Total	40,794	100.00	

```
. mlogit quitint smkhev, rrr

Iteration 0:   log likelihood =  -44287.48
Iteration 1:   log likelihood = -44111.885
Iteration 2:   log likelihood =  -44111.75
Iteration 3:   log likelihood =  -44111.75

Multinomial logistic regression                 Number of obs     =     40,794
                                                LR chi2(2)        =     351.46
                                                Prob > chi2       =     0.0000
Log likelihood =  -44111.75                     Pseudo R2         =     0.0040
```

quitint	RRR	Std. Err.	z	P>|z|	[95% Conf.	Interval]
no	(base outcome)					
yes_but_never_try						
smkhev	.7471556	.019123	-11.39	0.000	.7105997	.7855921
_cons	.8216297	.0157442	-10.25	0.000	.7913438	.8530746
yes_and_try						
smkhev	.6520336	.015163	-18.39	0.000	.6229818	.6824403
_cons	1.255548	.0216569	13.19	0.000	1.213811	1.298721

Note: _cons estimates baseline relative risk for each outcome.

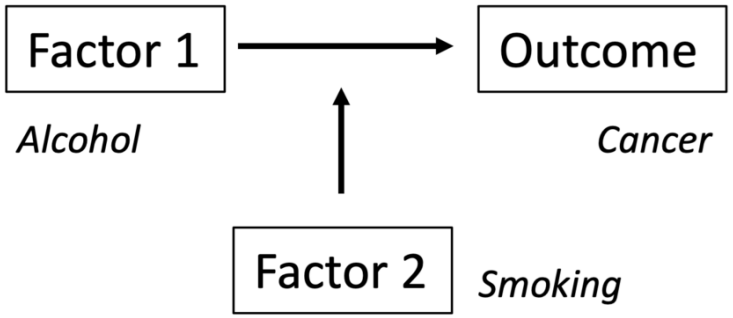
คนที่สูบบุหรี่หนักจะมีความอยากเลิกแต่ไม่เคยพยายามเลิกเป็น 0.75 เท่าเมื่อเทียบคนที่ไม่อยากเลิก

คนที่อยากเลิกแต่ไม่เคยพยายามเลิกจะเป็นคนสูบบุหรี่หนัก
เป็น 0.75 เท่าเมื่อเทียบกับคนที่ไม่อยากเลิก

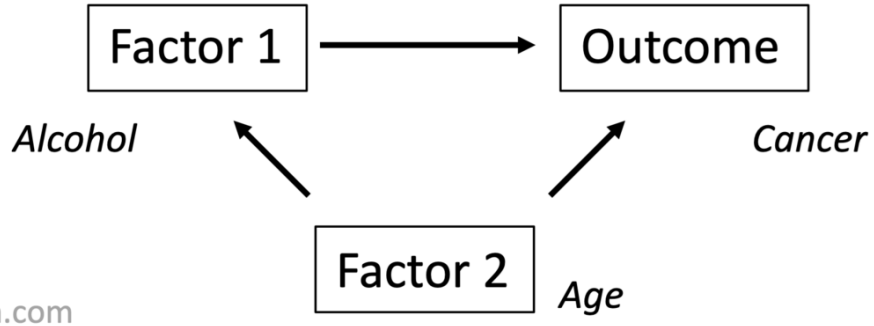
คนที่สูบบุหรี่หนักจะมีความอยากเลิกและพยายามเลิกบุหรี่เป็น 0.65 เท่าของคนที่ไม่อยากเลิก

Confounding vs effect modification

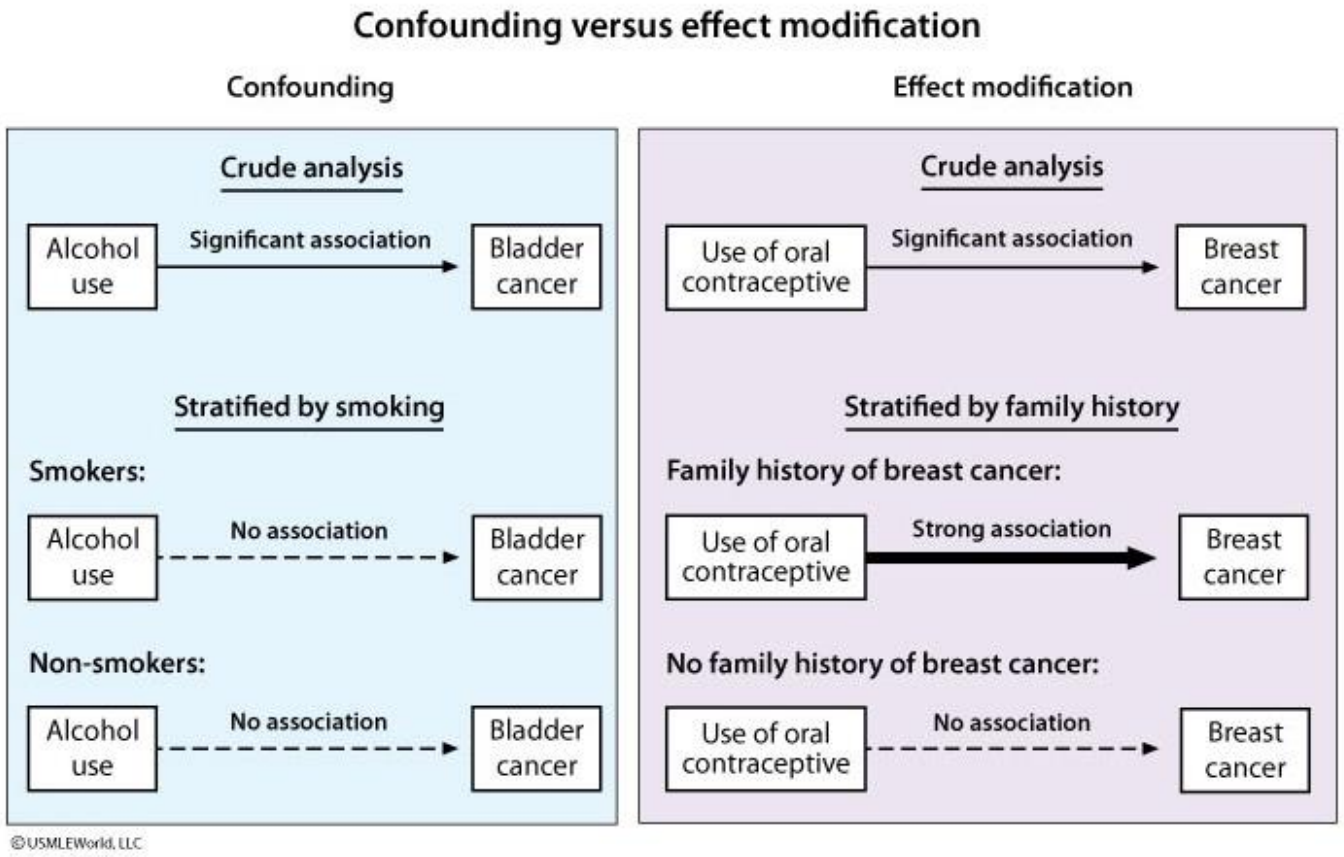
Effect modification



Confounding

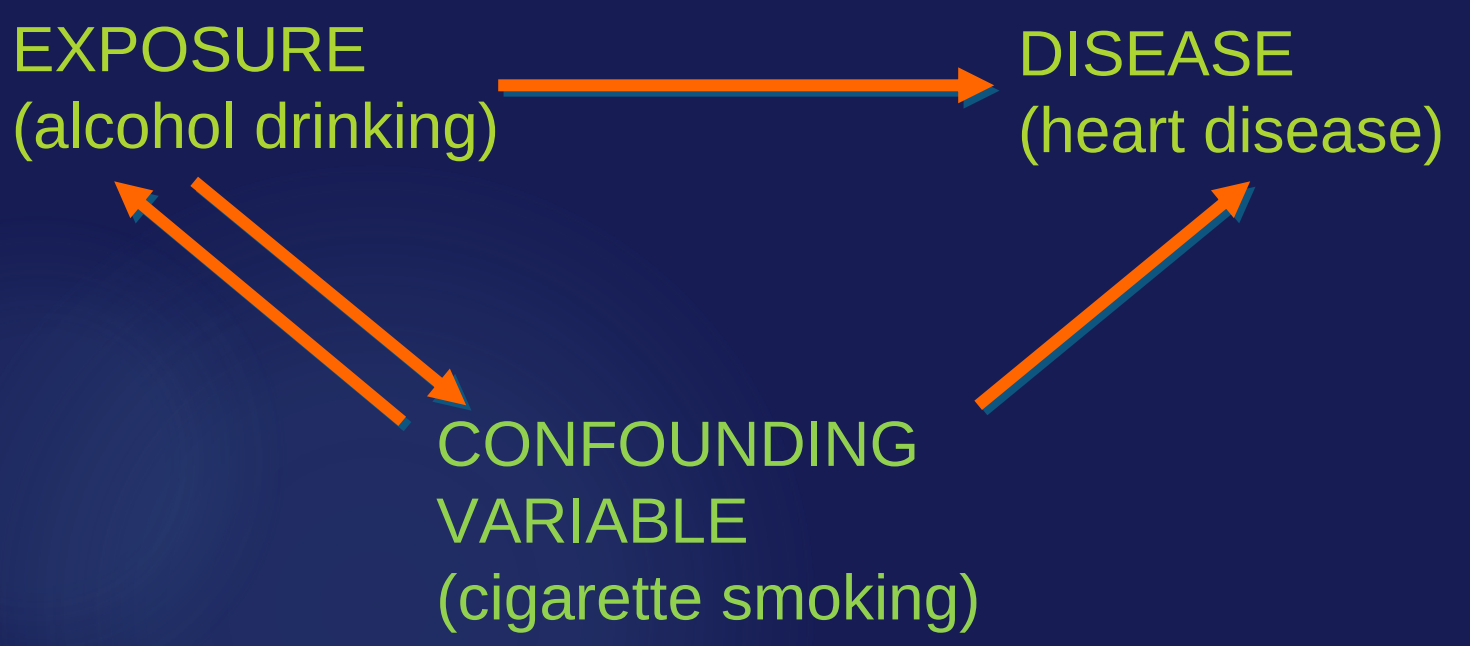


ultrabem.com



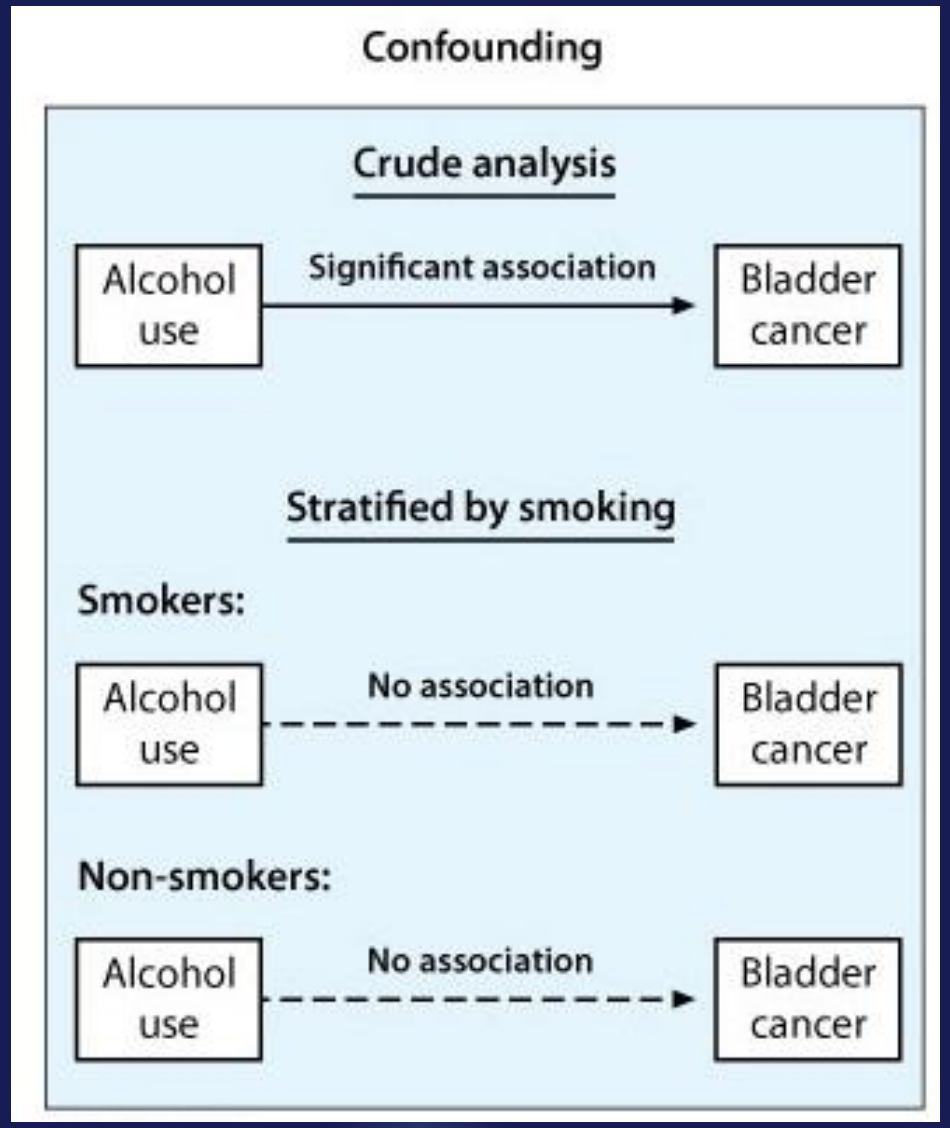
©USMLEWorld, LLC

Confounding



1. เป็น risk factor ของ outcome
2. มีความสัมพันธ์กับ exposure
3. ไม่เป็น intermediate

รายงานผลเป็น Adjusted measure



Control of confounding

Study design

- Restriction
- Matching

Analytic

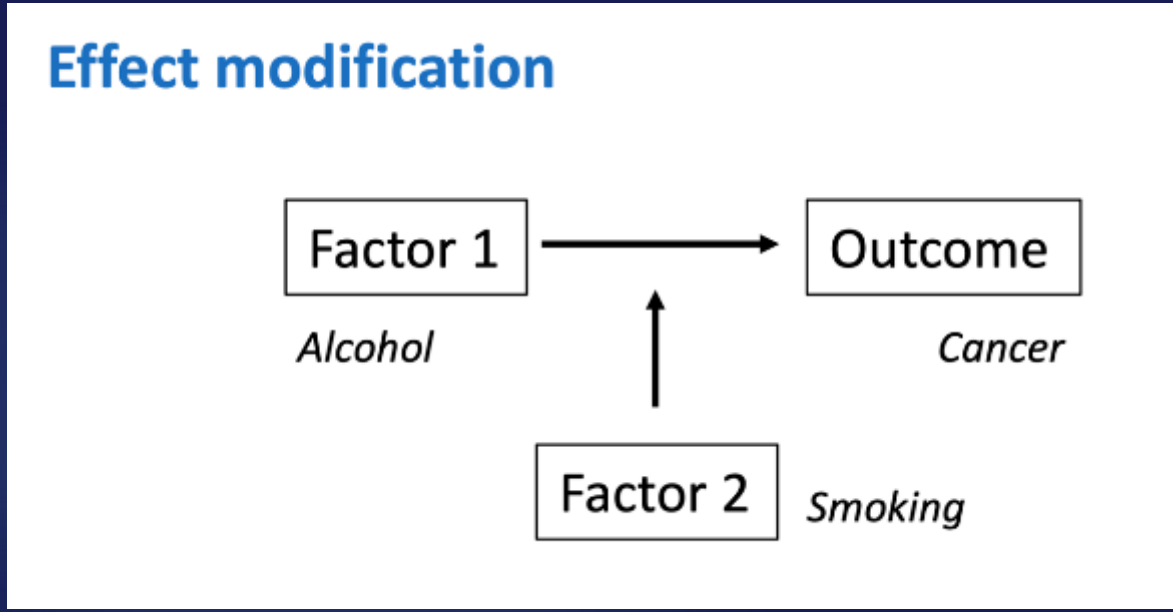
- Stratification
- Multivariate analysis
  - Crude measure (OR/RR)
  - Adjusted RR (OR/RR)

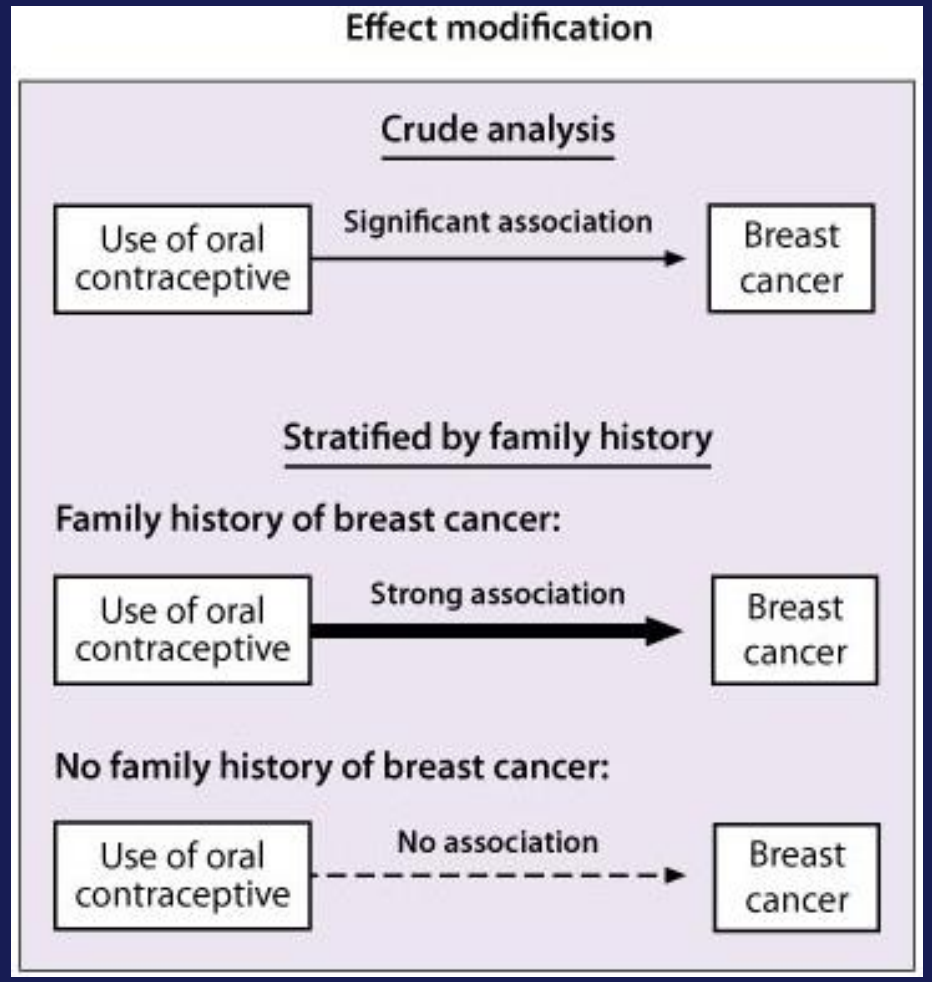
Degree of Confounding

$$= \frac{Crude\ measure}{Adjusted\ measure}$$

- >1 over estimation
- <1 under estimation

Effect modification/ Interaction





รายงานผลเป็น Crude measure จากการ stratified

Poisson Regression

Y = INCIDENCE / CONTINUOUS

$$\log[count] = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

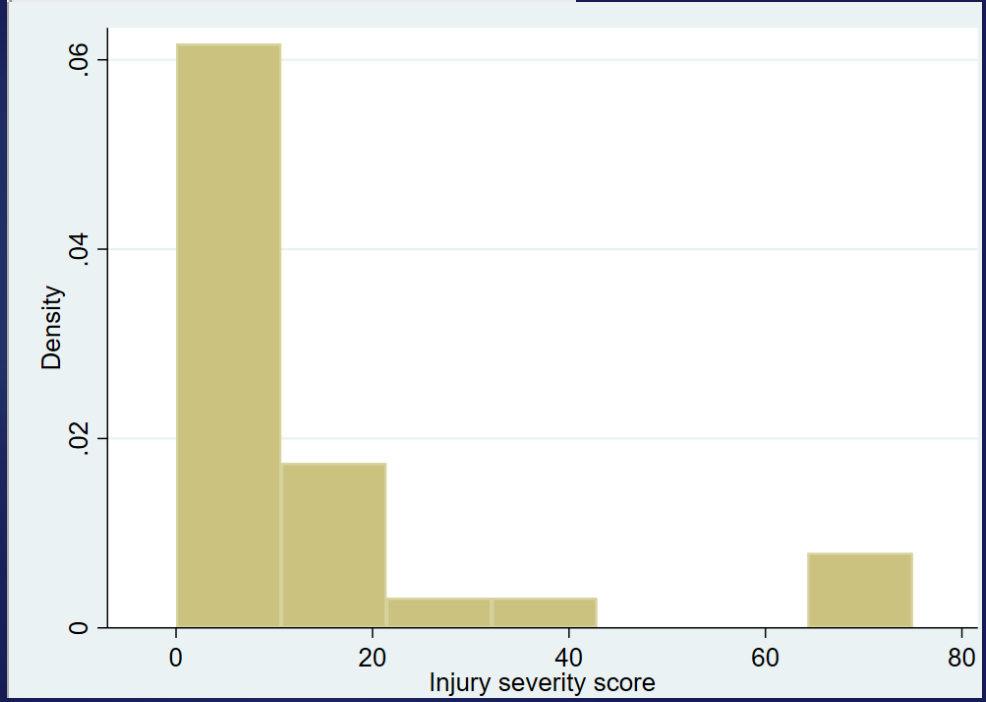
$$\log[count \;\; or\ cases / time] = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

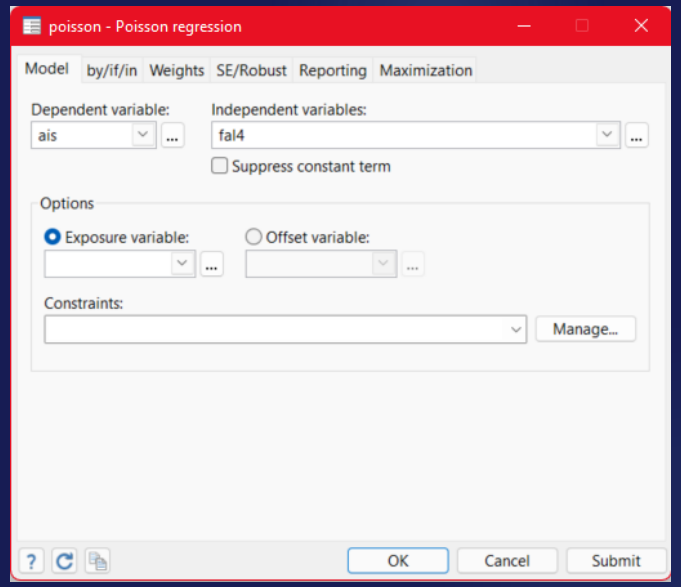
$$\log[rate = \lambda] = \beta_0 + \beta_1 x_{1i} + \square + \beta_K x_{Ki}$$

```
. use "D:\practice data sets\rti korat.dta"
```

```
. swilk ais

                   Shapiro-Wilk W test for normal data

    Variable |        Obs       W           V         z       Prob>z
-------------+-------------------------------------------------------
         ais |         59    0.59608      21.662     6.623    0.00000
```

```
. hist ais
(bin=7, start=0, width=10.714286)
```






```
. poisson ais fal4, irr

Iteration 0:   log likelihood = -591.04958
Iteration 1:   log likelihood = -590.86852
Iteration 2:   log likelihood = -590.86834
Iteration 3:   log likelihood = -590.86834

Poisson regression                              Number of obs     =         59
                                                LR chi2(1)        =     267.53
                                                Prob > chi2       =     0.0000
Log likelihood = -590.86834                     Pseudo R2         =     0.1846

------------------------------------------------------------------------------
         ais |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        fal4 |   4.777992   .5528086    13.52   0.000     3.808573    5.994163
       _cons |   3.818182   .4165978    12.28   0.000     3.083067    4.728575
------------------------------------------------------------------------------
Note: _cons estimates baseline incidence rate.
```

Univariate analysis

- poisson ais fal4, irr

```
Poisson regression                              Number of obs     =         44
                                                LR chi2(1)        =      15.03
                                                Prob > chi2       =     0.0001
Log likelihood = -593.09701                     Pseudo R2         =     0.0125

------------------------------------------------------------------------------
         ais |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       slept |   1.460446   .1486272     3.72   0.000     1.196355    1.782834
       _cons |       11.6   1.077033    26.40   0.000     9.669989    13.91522
------------------------------------------------------------------------------
```

- poisson ais fal4, irr

```
Poisson regression                              Number of obs     =         59
                                                LR chi2(1)        =     467.21
                                                Prob > chi2       =     0.0000
Log likelihood = -491.02805                     Pseudo R2         =     0.3224

------------------------------------------------------------------------------
         ais |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dambus |   4.933804   .3703446    21.26   0.000     4.258811    5.715778
       _cons |   6.431819   .3823319    31.31   0.000     5.724468    7.226575
------------------------------------------------------------------------------
```

- poisson ais slept, irr

```
Poisson regression                              Number of obs     =         59
                                                LR chi2(1)        =     467.21
                                                Prob > chi2       =     0.0000
Log likelihood = -491.02805                     Pseudo R2         =     0.3224

------------------------------------------------------------------------------
         ais |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      dambus |   4.933804   .3703446    21.26   0.000     4.258811    5.715778
       _cons |   6.431819   .3823319    31.31   0.000     5.724468    7.226575
------------------------------------------------------------------------------
```

- poisson ais hit, irr

```
Poisson regression                              Number of obs     =         59
                                                LR chi2(1)        =      18.52
                                                Prob > chi2       =     0.0000
Log likelihood = -715.37453                     Pseudo R2         =     0.0128

------------------------------------------------------------------------------
         ais |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hit |   1.039761   .0095076     4.26   0.000     1.021293    1.058564
       _cons |    10.3936   .6666055    36.50   0.000     9.165859    11.78579
------------------------------------------------------------------------------
```

Multivariate analysis; IRR


```
. poisson ais fal4 slept dambus hit, irr

Iteration 0:   log likelihood = -330.60085
Iteration 1:   log likelihood = -330.50745
Iteration 2:   log likelihood = -330.50742

Poisson regression                              Number of obs     =         44
                                                LR chi2(4)        =     540.21
                                                Prob > chi2       =     0.0000
Log likelihood = -330.50742                     Pseudo R2         =     0.4497

------------------------------------------------------------------------------
         ais |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        fal4 |   4.178341   .5660355    10.56   0.000        3.204    5.448981
       slept |   1.150043   .1176251     1.37   0.172     .9411402    1.405316
      dambus |   3.856233   .3175042    16.39   0.000     3.281552    4.531555
         hit |   1.076447   .0107098     7.40   0.000      1.05566    1.097644
       _cons |   1.477969   .2629338     2.20   0.028     1.042881    2.094575
------------------------------------------------------------------------------
Note: _cons estimates baseline incidence rate.
```

- ผู้ที่ตกจากเก้าอี้จะมีความเสี่ยงที่จะมีค่ารุนแรงการบาดเจ็บเพิ่มขึ้นเป็น 4.18 เท่าของคนที่ไม่ตก เมื่อตัดตัวรบกวนจากการหลับขณะเกิดเหตุ การนั่งบริเวณที่มีการยุบตัวของตัวถังรถและการโดนชิ้นส่วนรถกระแทกแล้ว

- คนที่นั่งบริเวณที่มีการยุบตัวขอรถมีคามเสี่ยงที่จะมีคะแนนความรุนแรงการบาดเจ็บเพิ่มขึ้น 3.86 เท่า เมื่อตัดตัวรบกวนจากการหลับขณะเกิดเหตุ การตกเก้าอี้และการโดนชิ้นส่วนรถกระแทกแล้ว

Multivariate analysis; Coef.


```
. poisson ais fal4 slept dambus hit

Iteration 0:   log likelihood = -330.60085
Iteration 1:   log likelihood = -330.50745
Iteration 2:   log likelihood = -330.50742

Poisson regression                              Number of obs     =         44
                                                LR chi2(4)        =     540.21
                                                Prob > chi2       =     0.0000
Log likelihood = -330.50742                     Pseudo R2         =     0.4497
```

ais	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
fal4	1.429914	.135469	10.56	0.000	1.1644	1.695429
slept	.1397997	.1022788	1.37	0.172	-.0606631	.3402624
dambus	1.349691	.0823353	16.39	0.000	1.188317	1.511065
hit	.0736658	.0099492	7.40	0.000	.0541657	.093166
_cons	.3906691	.1779021	2.20	0.028	.0419875	.7393508

- ผู้ที่ตกจากเก้าอี้จะมีความเสี่ยงที่จะมีค่ารุนแรงการบาดเจ็บเพิ่มขึ้นเป็น 1.43 คะแนน เมื่อตัดตัวรบกวนจากการหลับขณะเกิดเหตุ การนั่งบริเวณที่มีการยุบตัวของตัวถังรถและการโดนชิ้นส่วนรถกระแทกแล้ว

- คนที่นั่งบริเวณที่มีการยุบตัวขอรถมีคามเสี่ยงที่จะมีคะแนนความรุนแรงการบาดเจ็บเพิ่มขึ้น 1.35 เมื่อตัดตัวรบกวนจากการหลับขณะเกิดเหตุ การตกเก้าอี้และการโดนชิ้นส่วนรถกระแทกแล้ว

Data management Poisson Regression


```
. use "D:\practice data sets\depress2.dta"
```

	subj	dep1	dep2	dep3	dep4	dep5	dep6	group	pre
1	1	17	18	15	17	14	15	placebo	18
2	2	26	23	18	17	12	10	placebo	27
3	3	17	14	.	.	.	.	placebo	16
4	4	14	23	17	13	12	12	placebo	17
5	5	12	10	8	4	5	5	placebo	15
6	6	19	11.54	9	8	6.82	5.05	placebo	20
7	7	13	13	9	7	8	7	placebo	16
8	8	26	27	.	.	.	.	placebo	28
9	9	26	24	19	13.94	11	9	placebo	28
10	10	9	12	15	12	13	20	placebo	25
11	11	14	.	.	.	.	.	placebo	24
12	12	19	13	14	23	15	11	placebo	16
13	13	13	22	.	.	.	.	placebo	26

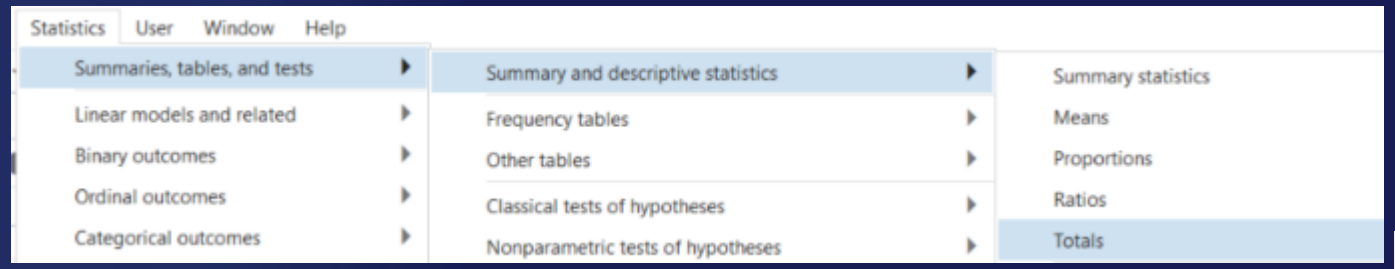
	subj	dep1	dep2	dep3	dep4	dep5	dep6	group	pre	post	diff	diff2	better	ptime
1	1	17	18	15	17	14	15	placebo	18	15	3	3	0	6
2	2	26	23	18	17	12	10	placebo	27	10	17	17	1	6
3	3	17	14	.	.	.	.	placebo	16	14	2	2	0	2
4	4	14	23	17	13	12	12	placebo	17	12	5	5	0	6
5	5	12	10	8	4	5	5	placebo	15	5	10	10	1	6
6	6	19	11.54	9	8	6.82	5.05	placebo	20	5.05	14.95	14.95	1	6
7	7	13	13	9	7	8	7	placebo	16	7	9	9	0	6
8	8	26	27	.	.	.	.	placebo	28	27	1	1	0	2
9	9	26	24	19	13.94	11	9	placebo	28	9	19	19	1	6
10	10	9	12	15	12	13	20	placebo	25	20	5	5	0	6
11	11	14	.	.	.	.	.	placebo	24	14	10	10	1	1
12	12	19	13	14	23	15	11	placebo	16	11	5	5	0	6
13	13	13	22	.	.	.	.	placebo	26	22	4	4	0	2
14	14	7	13	.	.	.	.	placebo	21	13	8	8	0	2
15	15	18	.	.	.	.	.	placebo	21	18	3	3	0	1
16	16	18	.	.	.	.	.	placebo	22	18	4	4	0	1
17	17	19	13	22	12	18	13	placebo	26	13	13	13	1	6
18	18	19	7	8	2	5	6	placebo	19	6	13	13	1	6
19	19	20	15	20	17	15	13.73	placebo	22	13.73	8.27	8.27	0	6

Rate ratio / Incidence rate ratio (IRR) (1)

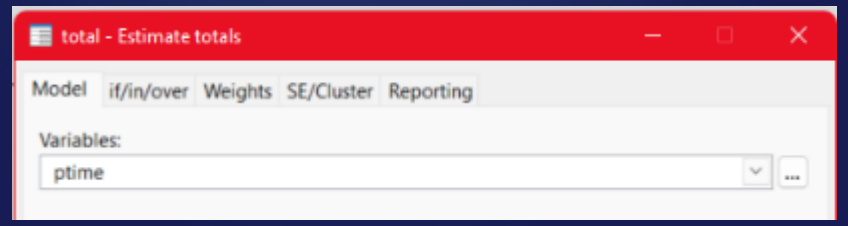
- A ratio of two incidence rates
- Used in RCT or Cohort study (with person-time data)
- Rate ratio (IRR) = [A/TE] / [B/TU]
- IRR = (25/178)/(9/117) = 1.82

	No. of sick	Person-time	Incidence rate
Exposed	A	TE	A/TE
Unexposed	B	TU	B/TU
Total	A+B	TE+TU	(A+B)/(TE+TU)

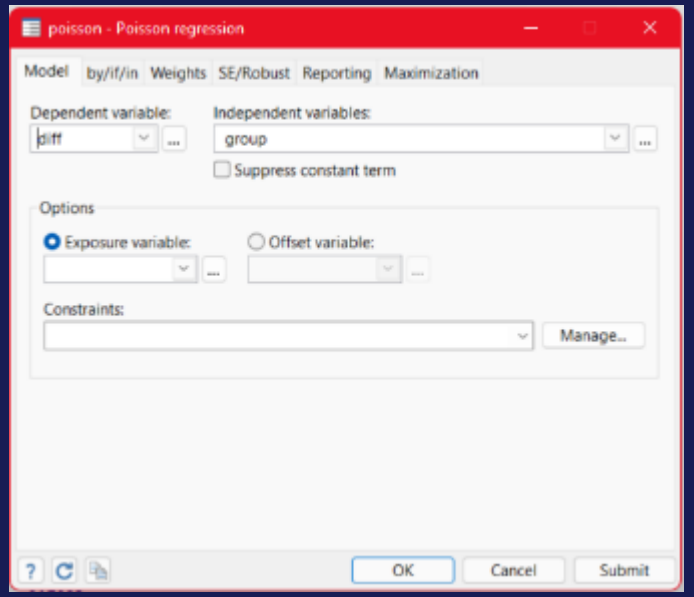
	No. of better	Person-time	Incidence rate
estrogen	25	178	0.1404= 14.04%
Placebo	9	117	0.0769= 76.69%
Total	A+B	TE+TU	(A+B)/(TE+TU)

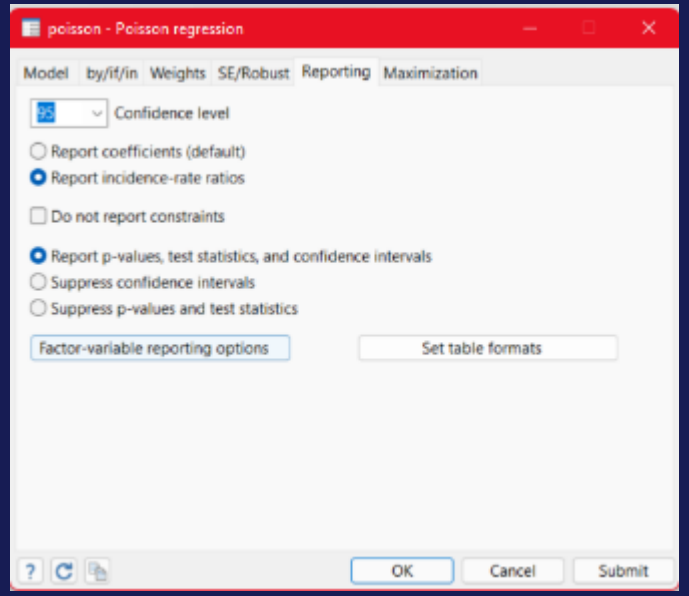



```
. tab group better

                   better
      group         0         1     Total
----------+----------------------------------
    placebo        18         9        27
   estrogen         9        25        34
----------+----------------------------------
      Total        27        34        61
```



```
. total ptime, over(group)

Total estimation                  Number of obs   =         61

--------------------------------------------------------------
             |     Total   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
c.ptime@group|
     placebo |       117   11.61895      93.75864    140.2414
    estrogen |       178   9.951397      158.0942    197.9058
--------------------------------------------------------------
```






```
. poisson diff group, irr
note: you are responsible for interpretation of noncount dep. variable

Iteration 0:   log likelihood = -218.04324
Iteration 1:   log likelihood = -218.04323

Poisson regression                              Number of obs     =         61
                                                LR chi2(1)        =      51.85
                                                Prob > chi2       =     0.0000
Log likelihood = -218.04323                     Pseudo R2         =     0.1063

------------------------------------------------------------------------------
        diff |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       group |   1.809487   .1539536     6.97   0.000     1.53156    2.137848
       _cons |   7.361852   .5221697    28.15   0.000    6.406371    8.459838
------------------------------------------------------------------------------
```

Conclusion

Type of study design

ดูว่างานเราเป็นแบบไหน

- Qualitative / quantitative
- Descriptive
- Analytic ??
- Cohort/Case control/Cross-sectional

Type of data

1. ข้อมูลเราเป็นแบบไหน
   Category / ordinal/ continue
   Normal distribution ??

1. ตัวไหนเป็น outcome / exposure

Descriptive study

- Categorical Variables
  - Ratio, Proportion, Percent (%)
- Continuous Variable
  - Normal Distribution
    - Mean, SD
  - Not normal Distribution
    - Median, Range/IQR

Analytic

- Comparison
  - Categorical outcome
    - Chi-square (χ^2) test
  - Continuous Variable
    - Normal Distribution
      - t-test (2 groups) F-test/ANOVA (>= 2 groups)
    - Not normal Distribution
      - Wilcoxon test, Mann-Whitney U test
- Association
  - Study design

Association by study design

Study design	Measure of frequency	Measure of association
Cross-sectional	Prevalence	• Prevalence ratio • Prevalence odds ratio
Case-control	Prevalence	• Odds ratio
Cohort	Incidence	• Risk ratio • Odds ratio

Regression model (Predict/ Odds ratio)

- Linear Regression
  - Y = continuous + normal distribution
- Logistic Regression
  - Y = Categorical
- Poisson Regression
  - Y = Incidence/count
- Cox's Proportional Hazard Regression
  - Y = Time to event

Y = outcome, x= exposure
X= continuous / categorical