ตัวอย่างคำสั่งที่ใช้บ่อยในโปรแกรม Stata

การวิเคราะห์ข้อมูลพื้นฐาน

```
. describe

Contains data from C:\Documents and Settings\Owner\My Documents\For_study_cm.511.dta
  obs:           616
 vars:            13                          10 Jul 2008 22:38
 size:        14,168 (91.2% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
class           byte   %8.0g                  Class : nominal scale
age             byte   %8.0g                  Age in year: ratio scale
gender          byte   %8.0g                  Gender: nominal scale;
                                                1=male,2=female
bmi1            float  %9.0g                  BMI1: ratio scale
bmi_code1       byte   %8.0g                  BMI_code1: ordinal scale;
                                                1=underweight,2=normal,3=risk
                                                to overweight,4=obesity1
dep_score1      byte   %8.0g                  Dep_Score1: ratio scale
dep_code1       byte   %8.0g                  Dep_Code1: ordinal scale;
                                                0=no,1=early depress,2=major
                                                depress
bmi2            float  %9.0g                  BMI2: ratio scale
bmi_code2       byte   %8.0g                  BMI_code2: ordinal scale;
                                                1=underweight,2=normal,3=risk
                                                to overweight,4=obesity1
dep_score2      byte   %8.0g                  Dep_Score2: ratio scale
dep_code2       byte   %8.0g                  Dep_Code2: ordinal scale;
                                                0=no,1=early depress,2=major
                                                depress
ess_score       byte   %8.0g                  Epworth sleepiness score: ratio
                                                scale
ess_code        byte   %8.0g                  ESS_Code: ordinal scale; 0=no,
                                                1=has a problem
-------------------------------------------------------------------------------
Sorted by:  class
```

. summarize age : แสดงผลรวมของ 1 ตัวแปร

```
    Variable |     Obs        Mean   Std. Dev.       Min        Max
-------------+-----------------------------------------------------
         age |     599    20.72788   1.852705         15         27
```

. ci age : แสดงค่า 95% Confidence interval

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

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

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

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

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

. tab class : แสดงตาราง 1 ตัวแปร

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

. tab gender bmi_code1 : แสดงตาราง 2 ตัวแปร

```
           |                      BMI_code1
    Gender |         1          2          3          4          5 |     Total
-----------+--------------------------------------------------------+----------
         1 |        24        124         25         28          8 |       209
         2 |        79        196         21         12          2 |       310
-----------+--------------------------------------------------------+----------
     Total |       103        320         46         40         10 |       519
```

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

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

. sort class : แสดงตาราง 3 ตัวแปร

. by class: tab gender dep_code2

```
-------------------------------------------------------------------------------
-> class = 1
           |            Dep_Code2
    Gender |         0          1          2 |     Total
-----------+------------------------------------+----------
         1 |        40          1          2 |        43
         2 |        69          1          2 |        72
-----------+------------------------------------+----------
     Total |       109          2          4 |       115
-------------------------------------------------------------------------------
-> class = 2
           |            Dep_Code2
    Gender |         0          1          2 |     Total
-----------+------------------------------------+----------
         1 |        34          2          0 |        36
         2 |        57          4          1 |        62
-----------+------------------------------------+----------
     Total |        91          6          1 |        98
-------------------------------------------------------------------------------
-> class = 3
           |            Dep_Code2
    Gender |         0          1          2 |     Total
-----------+------------------------------------+----------
         1 |        31          3          1 |        35
         2 |        29          4          4 |        37
-----------+------------------------------------+----------
     Total |        60          7          5 |        72
```

```
-> class = 4
           |       Dep_Code2
    Gender |         0          2 |     Total
-----------+----------------------+----------
         1 |        29          0 |        29
         2 |        62          3 |        65
-----------+----------------------+----------
     Total |        91          3 |        94
-> class = 5
           |            Dep_Code2
    Gender |         0          1          2 |     Total
-----------+-------------------------------+----------
         1 |        24          1          2 |        27
         2 |        32          2          0 |        34
-----------+-------------------------------+----------
     Total |        56          3          2 |        61
-> class = 6
           |            Dep_Code2
    Gender |         0          1          2 |     Total
-----------+-------------------------------+----------
         1 |        24          2          0 |        26
         2 |        33          3          1 |        37
-----------+-------------------------------+----------
     Total |        57          5          1 |        63
```

<u>การวิเคราะห์ปัจจัยเสี่ยงโดยไม่คำนึงถึงผลกระทบจากปัจจัยอื่น</u> <u>(Bivariate analysis)</u>

เพื่อหาขนาดของความสัมพันธ์ ได้แก่ Prevalence rate ratio ใน Cross-sectional study, Odds ratio ใน Case-Control study และ Relative risk ใน Cohort study โดยกำหนดรหัสเป็น 0 และ 1 โดยให้ 0 = ไม่เป็นโรค และไม่มีปัจจัยเสี่ยง ส่วน 1 = เป็นโรค และมีปัจจัยเสี่ยง ซึ่งต้องปรับรหัสตัวแปรให้เป็น 0 และ 1 ทั้งสองตัวแปร ในที่นี้ใช้ ess_code และ dep_code2 ซึ่ง ess_code มีค่า 0 และ 1 แต่dep_code2 มีค่า 0,1 และ 2 ปรับให้เป็น 0 และ 1 โดยสร้างตัวแปรใหม่ (generate) เป็น dep_code21 ดังนี้

```
. generate dep_code21= dep_code2
(93 missing values generated)

. replace dep_code21=1 if dep_code21==2
(18 real changes made)
```

การเปรียบเทียบข้อมูลเชิงคุณภาพ

โดยใช้ Chi-square test และ Fisher's exact test โดยให้แสดงค่า expected ดังนี้

```
. tab ess_code dep_code21 ,chi2 exact expected

+--------------------+
| Key                |
|--------------------|
|     frequency      |
| expected frequency |
+--------------------+
           |      dep_code21
  ess_code |         0          1 |     Total
-----------+----------------------+----------
         0 |       242          3 |       245
           |     225.1       19.9 |     245.0
-----------+----------------------+----------
         1 |       199         36 |       235
           |     215.9       19.1 |     235.0
-----------+----------------------+----------
     Total |       441         39 |       480
           |     441.0       39.0 |     480.0

          Pearson chi2(1) =  31.9213   Pr = 0.000
           Fisher's exact =                 0.000
   1-sided Fisher's exact =                 0.000
```

อาจใช้คำสั่ง tabi ตามด้วยกลุ่มตัวเลข (กรณีไม่ทราบข้อมูลดิบ) แทนได้ดังนี้

```
. tabi 242 3 \ 199 36, chi2 exact expected
```

<u>หมายเหตุ</u> i หมายถึง immediate

จากตารางข้างต้นจะเห็นว่าตัวแปรทั้งสองมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ จึงนำไปหาขนาดของความสัมพันธ์ต่อไป

การหาขนาดของความสัมพันธ์

ตัวแปรที่จะหาขนาดของความสัมพันธ์ต้องมีค่าเป็น 0 และ 1 ทั้งตัวแปรต้นและตัวแปรตาม ถ้าตัวแปรใดมีค่าเป็นอย่างอื่นต้องปรับให้มีค่าเป็น 0 และ 1 ซึ่งอาจจะใช้คำสั่ง generate เพื่อสร้างเป็นตัวแปรใหม่แล้วตามด้วยคำสั่ง replace เพื่อปรับค่าให้เป็น 0 และ 1

ตัวอย่างต้องการปรับตัวแปร BMI_code1 และ BMI_code2 จากเดิมมีค่าเป็น 1-5 ให้เป็นค่า 0 และ 1 โดยให้ 0 = underweight และ normal ส่วน 1 = risk to overweight, obesity1 และ obesity2 จึงสร้างตัวแปรใหม่เป็น BMI_code11 และ BMI_code21 ตามลำดับด้วยคำสั่ง generate แล้วปรับค่าตัวแปรด้วยคำสั่ง replace ดังนี้

```
. tab bmi_code1

 BMI_code1: |       Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        103       19.85       19.85
          2 |        320       61.66       81.50
          3 |         46        8.86       90.37
          4 |         40        7.71       98.07
          5 |         10        1.93      100.00
------------+-----------------------------------
      Total |        519      100.00

. generate bmi_code11= bmi_code1
(97 missing values generated)

. replace bmi_code11=0 if bmi_code11 <=2
(423 real changes made)

. replace bmi_code11=1 if bmi_code11 >=3
(193 real changes made)

. tab bmi_code11

 bmi_code11 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        423       68.67       68.67
          1 |        193       31.33      100.00
------------+-----------------------------------
      Total |        616      100.00

. tab bmi_code2

 BMI_code2: |       Freq.     Percent        Cum.
------------+-----------------------------------
          1 |         91       18.24       18.24
          2 |        314       62.93       81.16
          3 |         44        8.82       89.98
          4 |         37        7.41       97.39
          5 |         13        2.61      100.00
------------+-----------------------------------
      Total |        499      100.00

. generate bmi_code21= bmi_code2
(117 missing values generated)

. replace bmi_code21=0 if bmi_code21 <=2
(405 real changes made)
. replace bmi_code21=1 if bmi_code21 >=3
(211 real changes made)
```

```
. tab bmi_code21

 bmi_code21 |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        405       65.75       65.75
          1 |        211       34.25      100.00
------------+-----------------------------------
      Total |        616      100.00
```

ในกรณี Cross-sectional study หรือ Cohort study ใช้คำสั่ง " cs (ตัวแปร case) (ตัวแปร exposure) " จากคำสั่งดังกล่าว cs ตามด้วย Case (0 = No Case, 1 = Case) และ Exposure (0 = Unexpose, 1 = Expose) กรณีเป็น Cross-sectional study หรือ Cohort study ผลการศึกษาที่ได้คือค่า Prevalence rate ratio หรือ Risk ratio ตามลำดับดังนี้

```
. cs dep_code21 ess_code

                 |   ess_code             |
                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+------------
           Cases |        36           3  |         39
        Noncases |       199         242  |        441
-----------------+------------------------+------------
           Total |       235         245  |        480
                 |                        |
            Risk |  .1531915    .0122449  |     .08125
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
 Risk difference |         .1409466       |    .0928822     .189011
      Risk ratio |         12.51064       |    3.905801    40.07272
 Attr. frac. ex. |          .920068       |    .7439705    .9750454
 Attr. frac. pop |         .8492936       |
                +-------------------------------------------------
                                chi2(1) =    31.92  Pr>chi2 = 0.0000
```

กรณีไม่ทราบข้อมูลดิบอาจใช้คำสั่งดังนี้

```
. csi 36 3 199 242
```

<u>แปลผล</u> : นิสิตแพทย์ที่มีความง่วง (ess_code) มากกว่าปกติจะเกิดภาวะซึมเศร้า (dep_code21) มากกว่านิสิตแพทย์ปกติเป็น 12.5 เท่า อย่างมีนัยสำคัญทางสถิติ

ในกรณี Case-Control study ใช้คำสั่ง " cc (ตัวแปร case) (ตัวแปร exposure) " จากคำสั่งดังกล่าว cc ตามด้วย Case (0 = No Case, 1 = Case) และ Exposure (0 = Unexpose, 1 = Expose) ผลการศึกษาที่ได้คือค่า odds ratio ดังนี้

```
. cc dep_code21 ess_code
                                                      Proportion
                 |   Exposed   Unexposed  |      Total     Exposed
-----------------+------------------------+------------------------
           Cases |        36           3  |         39       0.9231
        Controls |       199         242  |        441       0.4512
-----------------+------------------------+------------------------
           Total |       235         245  |        480       0.4896
                 |                        |
                 |      Point estimate    |    [95% Conf. Interval]
                 |------------------------+------------------------
      Odds ratio |         14.59296       |    4.489618    74.83922 (exact)
 Attr. frac. ex. |         .9314738       |    .7772639     .986638 (exact)
 Attr. frac. pop |          .859822       |
                +-------------------------------------------------
                                chi2(1) =    31.92  Pr>chi2 = 0.0000
```

กรณีไม่ทราบข้อมูลดิบอาจใช้คำสั่งดังนี้

```
. cci 36 3 199 242
```

<u>แปลผล</u> : นิสิตแพทย์ที่มีความง่วง (ess_code) มากกว่าปกติจะเกิดภาวะซึมเศร้า (dep_code21) มากกว่านิสิตแพทย์ปกติเป็น 14.6 เท่า อย่างมีนัยสำคัญทางสถิติ

ในกรณีที่เปรียบเทียบข้อมูลเชิงคุณภาพในคนกลุ่มเดียวกันใช้ McNemar's Test หรือ Matched case-control สำหรับโปรแกรม Stata ใช้คำสั่ง " mcc (ตัวแปร case) (ตัวแปร control) " ค่าสถิติที่ได้คือ McNemar's chi square

ตัวอย่างต่อไปนี้จะเปรียบเทียบความอ้วน (BMI>23kg/m^2) ระหว่างภาคการศึกษาที่ 1 และ 2 ในนิสิตแพทย์กลุ่มเดียวกัน

```
. mcc bmi_code11 bmi_code21

                 |   Controls             |
Cases            |   Exposed   Unexposed  |      Total
-----------------+------------------------+------------
         Exposed |       128          65  |        193
       Unexposed |        83         340  |        423
-----------------+------------------------+------------
           Total |       211         405  |        616

McNemar's chi2(1) =      2.19    Prob > chi2 = 0.1390
Exact McNemar significance probability       = 0.1621

Proportion with factor
        Cases       .3133117
        Controls    .3425325     [95% Conf. Interval]
                   ---------     --------------------
        difference -.0292208     -.0694831   .0110415
        ratio       .9146919      .8127556   1.029413
        rel. diff. -.0444444     -.1046126   .0157237

        odds ratio  .7831325       .557165   1.096677   (exact)
```

กรณีไม่ทราบข้อมูลดิบอาจใช้คำสั่งดังนี้

```
. mcci 128 65 83 340
```

<u>แปลผล</u> : นิสิตแพทย์ภาคเรียนที่ 1 มีความอ้วนเป็น 0.783 เท่าของภาคเรียนที่ 2 มีค่า 95% confidence interval = 0.557 – 1.097 ค่า p-value = 0.139 สรุปผลว่านิสิตแพทย์มีความอ้วนไม่แตกต่างกันในภาคเรียนที่ 1 และภาคเรียนที่ 2 (หรือแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ)

ในกรณี Cohort study แบ่งเป็น 2 กรณี ได้แก่ กรณี Cumulative Cohort study ใช้คำสั่งเดียวกับ Cross-sectional study แต่ถ้าเป็นกรณี Incidence Density (ใช้หน่วย person-time) ใช้คำสั่งดังนี้

```
. ir ess_code dep_code21 time
```

: time หมายถึงระยะเวลาที่ expose แล้วเกิด disease หน่วยเป็น person-time

การเปรียบเทียบค่าเฉลี่ยของข้อมูลเชิงปริมาณ

การเปรียบเทียบค่าเฉลี่ยตัวแปร 2 กลุ่มที่ไม่เป็นอิสระต่อกันด้วย Paired t test

```
. ttest bmi1 = bmi2

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    bmi1 |     445    20.77876    .1488599    3.140204    20.48621    21.07132
    bmi2 |     445    20.89191    .1471435    3.103996    20.60273    21.18109
---------+--------------------------------------------------------------------
    diff |     445   -.1131461    .0564163    1.190103   -.2240222     -.00227
------------------------------------------------------------------------------
```

```
                    Ho: mean(bmi1 - bmi2) = mean(diff) = 0

  Ha: mean(diff) < 0         Ha: mean(diff) ~= 0        Ha: mean(diff) > 0
       t =  -2.0056                t =  -2.0056              t =  -2.0056
   P < t =   0.0228          P > |t| =   0.0455          P > t =   0.9772
```

การเปรียบเทียบค่าเฉลี่ยตัวแปร 2 กลุ่มที่เป็นอิสระต่อกันด้วย t test

ทดสอบว่าทั้งสองกลุ่มมี Variance แตกต่างกันหรือไม่โดยใช้ Variance ratio test ดังนี้

```
. sdtest age,by(gender)

Variance ratio test
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |     243    20.71193     .120497    1.878362    20.47458    20.94929
       2 |     348    20.66379    .0959472    1.789872    20.47508     20.8525
---------+--------------------------------------------------------------------
combined |     591    20.68359     .075085    1.825354    20.53612    20.83105
------------------------------------------------------------------------------
    ratio = sd(1) / sd(2)                                         f =   1.1013

Ho: ratio = 1                                    degrees of freedom = 242, 347

    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.7949         2*Pr(F > f) = 0.4103           Pr(F > f) = 0.2051
```

การเปรียบเทียบค่าเฉลี่ยตัวแปร 2 กลุ่มที่เป็นอิสระต่อกันด้วย t test equal variance
(ถ้าข้อมูลทั้งสองกลุ่ม Variance ไม่เท่ากันให้พิมพ์ unequal ต่อท้ายคำสั่ง)

```
. ttest age,by(gender)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       1 |     243    20.71193     .120497    1.878362    20.47458    20.94929
       2 |     348    20.66379    .0959472    1.789872    20.47508     20.8525
---------+--------------------------------------------------------------------
combined |     591    20.68359     .075085    1.825354    20.53612    20.83105
---------+--------------------------------------------------------------------
    diff |            .0481411    .1527143               -.2517898    .3480719
------------------------------------------------------------------------------
    diff = mean(1) - mean(2)                                      t =   0.3152

Ho: diff = 0                                     degrees of freedom =      589

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.6237         Pr(|T| > |t|) = 0.7527          Pr(T > t) = 0.3763
```

. oneway ess_score class, bonferroni: การเปรียบเทียบตัวแปร 3 กลุ่มมีตัวแปรอิสระ 1 ตัว

```
                        Analysis of Variance
    Source              SS         df      MS            F     Prob > F
------------------------------------------------------------------------
Between groups      129.642456      5   25.9284912      1.64     0.1480
 Within groups         7511.58    475   15.8138526
------------------------------------------------------------------------

    Total           7641.22245    480   15.9192134

Bartlett's test for equal variances:  chi2(5) =   9.8425  Prob>chi2 = 0.080
```

```
                      Comparison of ESS_Score by Class
                                (Bonferroni)
Row Mean-|
Col Mean |          1          2          3          4          5
---------+------------------------------------------------------------
       2 |    .204187
         |      1.000
         |
       3 |    .995726    .791539
         |      1.000      1.000
         |
       4 |    .239225    .035038   -.756501
         |      1.000      1.000      1.000
         |
       5 |      .3409    .136712   -.654827    .101674
         |      1.000      1.000      1.000      1.000
         |
       6 |    2.07509     1.8709    1.07937    1.83587    1.73419
         |      0.202      0.403      1.000      0.488      0.850
```

การหาความสัมพันธ์ระหว่างตัวแปร (Correlation)

. pwcorr bmi2 dep_score2 ess_score, sig bon: การหาความสัมพันธ์ของ 2 ตัวแปรเชิงปริมาณ

```
             |     bmi2 dep_sc~2 ess_sc~e
-------------+---------------------------
        bmi2 |   1.0000
             |
             |
  dep_score2 |   0.0202   1.0000
             |   1.0000
             |
   ess_score |  -0.0545   0.4154   1.0000
             |   0.7254   0.0000
```

. spearman bmi_code2 dep_code2: การหาความสัมพันธ์ของ 2 ตัวแปรเชิงคุณภาพ

```
 Number of obs =     499
Spearman's rho =      -0.0621
Test of Ho: bmi_code2 and dep_code2 are independent
    Prob > |t| =       0.1660
```

. replace dep_code1=1 if dep_code1==2 : แสดงการแทนค่าตัวแปร

```
(28 real changes made)
```

. generate age_code=age : แสดงการเพิ่มตัวแปร คำสั่งให้ age_code มีค่าเท่ากับตัวแปร age

```
(17 missing values generated)
```

. replace age_code=1 if age_code<=20 :แสดงการแทนค่า age_code เป็น 1 เมื่ออายุ 20 ปีลงมา

```
(276 real changes made)
```

. replace age_code=2 if age_code>20 : แสดงการแทนค่า age_code เป็น 2 เมื่ออายุ 20ปีขึ้นไป

```
(340 real changes made)
```

<u>การวิเคราะห์ปัจจัยเสี่ยงโดยคำนึงถึงผลกระทบจากปัจจัยอื่น (Multivariate analysis)</u>

การวิเคราะห์ multivariate โดยใช้ logistic regression ที่มีตัวแปรต้นเป็นค่าต่อเนื่อง

ขั้นตอนแรก : ตรวจสอบ Multicollinearity ซึ่งเกิดจากตัวแปรต้นมีความสัมพันธ์กันเองซึ่งไม่ตรงกับนิยามของตัวแปรอิสระหรือตัวแปรต้น หากพบต้องพิจารณาตัดตัวแปรใดตัวแปรหนึ่งออก การตรวจสอบใช้คำสั่งดังนี้

. collin dep_score2 bmi2 bmi1

```
  Collinearity Diagnostics
                        SQRT                           Cond        R-
  Variable       VIF    VIF    Tolerance  Eigenval     Index    Squared
----------------------------------------------------------------------
dep_score2      1.00    1.00    0.9950     1.9376     1.0000     0.0050
      bmi2      7.15    2.67    0.1399     0.9898     1.3991     0.8601
      bmi1      7.15    2.67    0.1399     0.0726     5.1666     0.8601
----------------------------------------------------------------------
  Mean VIF      5.10              Condition Number    5.1666
                 Determinant of correlation matrix    0.1392
```

สำหรับ Stata version 7 ใช้คำสั่ง regress (ตามด้วยตัวแปรตามและตัวแปรต้น) แล้วตามด้วยคำสั่ง vif ดังนี้

```
. regress ess_code dep_score2 bmi2 bmi1

      Source |       SS       df       MS              Number of obs =     408
-------------+------------------------------           F(  3,   404) =   23.02
       Model |  14.8801595     3  4.96005316           Prob > F      =  0.0000
    Residual |   87.058566   404    .2154915           R-squared     =  0.1460
-------------+------------------------------           Adj R-squared =  0.1396
       Total |  101.938725   407  .250463699           Root MSE      =  .46421

------------------------------------------------------------------------------
    ess_code |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  dep_score2 |    .025588   .0032241     7.94   0.000     .0192499     .031926
        bmi2 |    .044805   .0190704     2.35   0.019     .0073154    .0822945
        bmi1 |  -.0461701   .0188293    -2.45   0.015    -.0831857   -.0091545
       _cons |   .1730347   .1701269     1.02   0.310    -.1614097    .5074792
------------------------------------------------------------------------------

. vif

    Variable |       VIF       1/VIF
-------------+----------------------
        bmi2 |      5.95    0.168033
        bmi1 |      5.95    0.168036
  dep_score2 |      1.00    0.997836
-------------+----------------------
    Mean VIF |      4.30
```

ผลการตรวจสอบพบว่าค่า VIF ของตัวแปร bmi1 และ bmi2 มีค่าสูงกว่าตัวแปรอื่นซึ่งน่าจะเกิด multicollinearity ระหว่างตัวแปรทั้งสอง จึงควรตัดตัวแปรใดตัวแปรหนึ่งออก ในที่นี้ตัด bmi1 ออกเนื่องจากเป็นค่า bmi ของเทอมที่ 1 ซึ่งตัวแปรอื่นนั้นวัดผลในเทอมที่ 2 ทั้งหมด นำมาตรวจสอบซ้ำดังนี้ (สำหรับ version 7 ทำซ้ำเช่นเดิม)

```
. collin dep_score2 bmi2

  Collinearity Diagnostics

                        SQRT                           Cond        R-
  Variable       VIF    VIF    Tolerance  Eigenval     Index    Squared
----------------------------------------------------------------------
dep_score2      1.00    1.00    0.9996     1.0202     1.0000     0.0004
      bmi2      1.00    1.00    0.9996     0.9798     1.0204     0.0004
----------------------------------------------------------------------
  Mean VIF      1.00              Condition Number    1.0204
                 Determinant of correlation matrix    0.9996
```

ขั้นตอนที่สอง : การสร้างสมการที่เหมาะสมด้วยวิธี Stepwise logistic regression มักกำหนดเกณฑ์คัดเข้าสู่โมเดล (pe) ที่ 0.2 และเกณฑ์คัดออกจากโมเดล (pr) ที่ 0.25 ซึ่งใช้ได้หลายคำสั่ง ตัวแปรตาม (0 = No

Disease, 1 = Disease) จะอยู่ติดคำสั่ง (ในที่นี้ให้ 0 = ความง่วงระดับปกติ, 1 = ความง่วงเกินปกติ) แล้วตามด้วยตัวแปรต้น ดังนี้

```
. sw logistic ess_code dep_score2 bmi2, pr(0.25) pe(0.2)

                                 begin with full model
p = 0.4298 >= 0.2500  removing bmi2

Logistic regression                               Number of obs   =        463
                                                  LR chi2(1)      =      70.92
                                                  Prob > chi2     =     0.0000
Log likelihood = -285.45905                       Pseudo R2       =     0.1105

------------------------------------------------------------------------------
    ess_code | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  dep_score2 |   1.131642   .0189613     7.38   0.000     1.095082    1.169422
------------------------------------------------------------------------------
```

```
. sw logistic ess_code dep_score2 bmi2, pr(0.25) pe(0.2) forward

                                begin with empty model
p = 0.0000 <  0.2000  adding   dep_score2

Logistic regression                               Number of obs   =        463
                                                  LR chi2(1)      =      70.92
                                                  Prob > chi2     =     0.0000
Log likelihood = -285.45905                       Pseudo R2       =     0.1105

------------------------------------------------------------------------------
    ess_code | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  dep_score2 |   1.131642   .0189613     7.38   0.000     1.095082    1.169422
------------------------------------------------------------------------------
```

ขั้นสุดท้าย: ทดสอบว่าสมการมีความเหมาะสมหรือไม่ โดยพิสูจน์สมมติฐานหลักคือ “สมการมีความเหมาะสม” (เฉพาะ Stata version 10)

```
. estat gof

Logistic model for ess_code, goodness-of-fit test

       number of observations =       463
 number of covariate patterns =        40
             Pearson chi2(38) =        32.02
                  Prob > chi2 =         0.7416
```

สรุป : สมการมีความเหมาะสม

<u>แปลผล</u>: นิสิตแพทย์มีความง่วงเกินปกติเป็น 1.13 เท่าต่อคะแนนซึมเศร้า 1 หน่วย

การวิเคราะห์ Multivariate ที่มีตัวแปรต้นเป็นค่าไม่ต่อเนื่อง

โดยวิธีสร้าง Dummy variable กับตัวแปรที่เป็น categorical dataโดยใช้ i.ตามด้วยตัวแปรดังกล่าว โดยค่าของตัวแปรที่มีค่าน้อยที่สุดจะถูกกำหนดเป็น Reference group

```
. xi: sw logistic ess_code i.dep_code2 age, pr(0.25) pe(0.2)
i.dep_code2       _Idep_code2_0-2     (naturally coded; _Idep_code2_0 omitted)
                      begin with full model
p < 0.2500            for all terms in model

Logistic regression                               Number of obs   =        467
                                                  LR chi2(3)      =      34.14
                                                  Prob > chi2     =     0.0000
Log likelihood = -306.54167                       Pseudo R2       =     0.0528
```

```
------------------------------------------------------------------------------
    ess_code | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_Idep_code~1 |   11.39614   8.550008     3.24   0.001     2.619028     49.5879
_Idep_code~2 |   17.18714   17.87494     2.73   0.006     2.238403    131.9681
         age |     1.0665   .0574384     1.20   0.232      .959661    1.185234
------------------------------------------------------------------------------

. estat gof


. Logistic model for ess_code, goodness-of-fit test

       number of observations =       467
 number of covariate patterns =        22
             Pearson chi2(18) =        13.16
                  Prob > chi2 =         0.7822
```

สรุป : สมการมีความเหมาะสม

<u>หมายเหตุ</u>: `Idep_code~0 =` ไม่มีภาวะซึมเศร้า

`Idep_code~1 =` อารมณ์ซึมเศร้าระยะแรก

`Idep_code~2 =` อารมณ์ซึมเศร้ารุนแรง

<u>แปลผล</u>: นิสิตแพทย์ที่มีอารมณ์ซึมเศร้าระยะแรกมีความง่วงเกินปกติเป็น 11.4 เท่าเมื่อเทียบกับนิสิตแพทย์ที่ไม่มีภาวะซึมเศร้า

นิสิตแพทย์ที่มีอารมณ์ซึมเศร้ารุนแรงมีความง่วงเกินปกติเป็น 17.2 เท่าเมื่อเทียบกับนิสิตแพทย์ที่ไม่มีภาวะซึมเศร้า

ปัจจัยด้านอายุไม่สัมพันธ์กับความง่วง

การวิเคราะห์ค่าความถูกต้อง (Validity) **ในการทำนายสำหรับตัวแปรตามที่มีค่าเป็น** 0 **กับ** 1 (Binary outcome) **และตัวแปรต้นที่เป็นค่าต่อเนื่องโดยใช้** Receiver Operating Characteristic Curve (ROC Curve)

ในที่นี้จากการวิเคราะห์ Multivariate ข้างต้นต้องการทำนายความง่วง (ess_code) โดยการเปรียบเทียบสมการข้างต้นทั้งสองสมการว่าสมการใดเหมาะสมกว่ากัน โดยการดูพื้นที่ใต้กราฟ ROC Curve ถ้าสมการใดหรือวิธีใดมีพื้นที่ใต้กราฟมากกว่า วิธีนั้นเหมาะสมกว่าตามหลักเกณฑ์ของ ROC Curve [แกน Y เป็น Sensitivity (True positive) แกน X เป็น 1-Specificity (False negative)] ใช้คำสั่ง predict ตามด้วยตัวอักษรใดก็ได้เป็นการสั่งให้จำสมการนั้นไว้แล้วเปรียบเทียบพื้นที่ใต้กราฟด้วยคำสั่ง roccomp ตามด้วยตัวแปรตามที่ต้องการทำนายในที่นี้คือ ess_code แล้วตามด้วยสมการที่ 1 และสมการที่ 2 ในที่นี้คือ p1 และ p2 ตามลำดับ

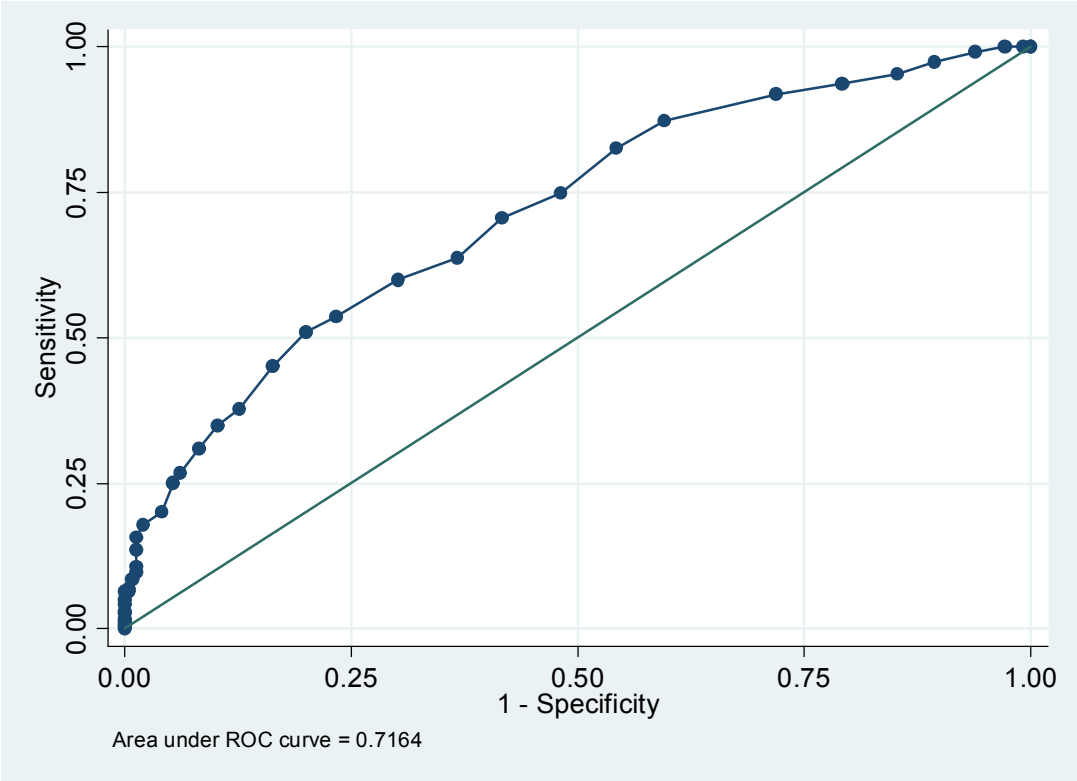
```
. sw logistic ess_code dep_score2 bmi2, pr(0.25) pe(0.2) forward

                                 begin with empty model
p = 0.0000 <  0.2000  adding   dep_score2

Logistic regression                               Number of obs   =        463
                                                  LR chi2(1)      =      70.92
                                                  Prob > chi2     =     0.0000
Log likelihood = -285.45905                       Pseudo R2       =     0.1105

------------------------------------------------------------------------------
    ess_code | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  dep_score2 |   1.131642   .0189613     7.38   0.000     1.095082    1.169422
------------------------------------------------------------------------------

. predict p1
(option pr assumed; Pr(ess_code))
(93 missing values generated)

. xi: sw logistic ess_code i.dep_code2 age, pr(0.25) pe(0.2)
i.dep_code2       _Idep_code2_0-2     (naturally coded; _Idep_code2_0 omitted)
                      begin with full model
p < 0.2500            for all terms in model

Logistic regression                               Number of obs   =        467
                                                  LR chi2(3)      =      34.14
                                                  Prob > chi2     =     0.0000
Log likelihood = -306.54167                       Pseudo R2       =     0.0528

------------------------------------------------------------------------------
    ess_code | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_Idep_code~1 |   11.39614   8.550008     3.24   0.001     2.619028     49.5879
_Idep_code~2 |   17.18714   17.87494     2.73   0.006     2.238403    131.9681
         age |     1.0665   .0574384     1.20   0.232      .959661    1.185234
------------------------------------------------------------------------------

. predict p2
(option pr assumed; Pr(ess_code))
(106 missing values generated)

. roccomp ess_code p1 p2

                                ROC                    -Asymptotic Normal--
                     Obs       Area     Std. Err.      [95% Conf. Interval]
-------------------------------------------------------------------------
p1                   467     0.7176       0.0231        0.67230     0.76282
p2                   467     0.5904       0.0247        0.54193     0.63880
-------------------------------------------------------------------------
Ho: area(p1) = area(p2)
    chi2(1) =    15.73       Prob>chi2 =   0.0001
```

จากผลการวิเคราะห์ข้างต้นด้วยคำสั่ง roccomp พบว่าพื้นที่ใต้กราฟของวิธีที่ 1 เท่ากับ 0.7176 ส่วนวิธีที่ 2 เท่ากับ 0.5904 มีค่า p-value เท่ากับ 0.001 สรุปว่าพื้นที่ใต้กราฟของทั้งสองวิธีแตกต่างอย่างมีนัยสำคัญทางสถิติโดยวิธีที่ 1 เหมาะสมในการทำนายความง่วงหรือ ess_code มากกว่าวิธีที่ 2 หมายความว่า dep_score2 เพียงอย่างเดียวสามารถทำนายความง่วงได้ดีกว่าการใช้ dep_code และ age

ถ้าต้องการดูรูปกราฟใช้คำสั่ง roctab พิมพ์ option graph ถ้าต้องการดู Sensitivity และ Specificity ให้พิมพ์ option detail เพื่อหาจุดตัดที่เหมาะสมคือค่าของ dep_score2 ที่ให้ค่า Sensitivity และ Specificity สูงที่สุดหรือมีค่า Accuracy มากที่สุด

ค่า Accuracy คำนวณจาก Sensitivity x prevalence + Specificity x (1-prevalence) ซึ่งต้องทราบค่า prevalence ในกลุ่มประชากรที่ทำการศึกษาแล้วแทนค่าในสูตร

```
. roctab ess_code dep_score2, graph
```



จากกราฟจะเห็นว่าเส้นที่ลากทะแยงมุมทำมุม 45 องศา คือเส้นมาตรฐานของ ROC Curve แต่ละจุดบนเส้นมาตรฐานจะมีค่า True positive = False positive ดังนั้นถ้ากราฟที่ได้ต่ำกว่าเส้นมาตรฐานนี้หรือมีพื้นที่ใต้กราฟต่ำกว่าหรือเท่ากับ 0.5 ถือว่าวิธีนั้นหรือสมการนั้นหรือตัวแปรนั้นไม่เหมาะสมที่จะทำนายตัวแปรตาม เพราะแต่ละจุดบนกราฟนั้นจะมีค่า False positive มากกว่าค่า True positive กราฟที่เหมาะสมจึงควรมีพื้นที่ใต้กราฟมากกว่า 0.5 หรือเส้นกราฟอยู่เหนือเส้นมาตรฐานดังที่ได้อธิบายข้างต้น และจุดตัดที่เหมาะสมคือจุดบนเส้นกราฟที่มีระยะห่างจากจุด (1,0) คือ True positive มีค่ามากที่สุดซึ่งเท่ากับ 1 และ False positive มีค่าน้อยที่สุดซึ่งเท่ากับ 0

```
. roctab ess_code dep_score2, detail

Detailed report of Sensitivity and Specificity
```

Cut point	Sensitivity	Specificity	Correctly Classified	LR+	LR-
(>= 0)	100.00%	0.00%	48.96%	1.0000	
(>= 2)	100.00%	0.82%	49.38%	1.0082	0.0000
(>= 3)	100.00%	2.86%	50.42%	1.0294	0.0000
(>= 4)	99.15%	6.12%	51.67%	1.0562	0.1390
(>= 5)	97.45%	10.61%	53.12%	1.0902	0.2406
(>= 6)	95.32%	14.69%	54.17%	1.1174	0.3186
(>= 7)	93.62%	20.82%	56.46%	1.1823	0.3066
(>= 8)	91.91%	28.16%	59.38%	1.2795	0.2871
(>= 9)	87.23%	40.41%	63.33%	1.4639	0.3159
(>= 10)	82.55%	45.71%	63.75%	1.5207	0.3816
(>= 11)	74.89%	51.84%	63.12%	1.5550	0.4843
(>= 12)	70.64%	58.37%	64.38%	1.6967	0.5031
(>= 13)	63.83%	63.27%	63.54%	1.7376	0.5717
(>= 14)	60.00%	69.80%	65.00%	1.9865	0.5731
(>= 15)	53.62%	76.73%	65.42%	2.3046	0.6045
(>= 16)	51.06%	80.00%	65.83%	2.5532	0.6117
(>= 17)	45.11%	83.67%	64.79%	2.7628	0.6560
(>= 18)	37.87%	87.35%	63.13%	2.9931	0.7113
(>= 19)	34.89%	89.80%	62.92%	3.4196	0.7250
(>= 20)	31.06%	91.84%	62.08%	3.8053	0.7506
(>= 21)	26.81%	93.88%	61.04%	4.3787	0.7796
(>= 22)	25.11%	94.69%	60.63%	4.7316	0.7909
(>= 23)	20.00%	95.92%	58.75%	4.9000	0.8340
(>= 24)	17.87%	97.96%	58.75%	8.7574	0.8384
(>= 25)	15.74%	98.78%	58.13%	12.8582	0.8530
(>= 26)	13.62%	98.78%	57.08%	11.1206	0.8745
(>= 27)	10.64%	98.78%	55.63%	8.6880	0.9047
(>= 28)	9.79%	98.78%	55.21%	7.9929	0.9133
(>= 29)	8.51%	99.18%	54.79%	10.4255	0.9224
(>= 30)	6.81%	99.59%	54.17%	16.6811	0.9357
(>= 31)	6.38%	99.59%	53.96%	15.6385	0.9400
(>= 32)	6.38%	100.00%	54.17%		0.9362
(>= 33)	5.11%	100.00%	53.54%		0.9489
(>= 34)	4.26%	100.00%	53.12%		0.9574
(>= 35)	2.98%	100.00%	52.50%		0.9702
(>= 36)	2.55%	100.00%	52.29%		0.9745
(>= 42)	1.70%	100.00%	51.87%		0.9830
(>= 43)	1.28%	100.00%	51.67%		0.9872
(>= 44)	0.85%	100.00%	51.46%		0.9915
(>= 47)	0.43%	100.00%	51.25%		0.9957
(> 47)	0.00%	100.00%	51.04%		1.0000

Obs	ROC Area	Std. Err.	-Asymptotic Normal-- [95% Conf. Interval]	
480	0.7164	0.0231	0.67118	0.76159

จากค่าของ Sensitivity และ Specificity ถ้าทราบ prevalence ของประชากรที่ศึกษา ในที่นี้คือค่าความชุกของภาวะซึมเศร้าในนิสิตแพทย์ มศว ก็จะแทนค่าสูตรแล้วคำนวณหาค่า Accuracy สูงสุดได้ซึ่งจุดนั้นได้ ส่วนจุดที่มีค่า Depression score เหมาะสมในการแบ่งเกณฑ์ความง่วงที่มากเกินปกติและความง่วงที่อยู่ในเกณฑ์ปกติหาได้โดยใช้จุดที่ใกล้ตำแหน่ง (0,1) มากที่สุด จากกราฟข้างต้นคือค่าที่ Depressive score = 15