Interaction:

ในการศึกษา clinical trial ยาสมุนไพรสมมติว่าชื่อ herbs A (ต่อไปจะเรียกว่า ht grop) เทียบกับ placebo ในการลดระดับ LDL เนื่องจาก participants ส่วนหนึ่งใช้ statin อยู่ จึงต้องการหาว่ามี "synergistic" ในการลด LDL ระหว่าง ht กับ statin หรือไม่

หรืออีกนัยหนึ่ง "การใช้หรือไม่ใช้ statin มีผลทำให้ "effect" ของ herbs A ต่อระดับ LDL แตกต่างกัน กล่าวคือ statin เป็น "effect modifier" (อีกชื่อของ interaction) ของ herbs ต่อ LDL หรือไม่

เราสามารถดูแบบเผินๆ ได้

```
. table ht statins, contents(mean ldlch)

----------------------------------------
random          |
assignment to   |    use of statin
hormone therapy |        no        yes
----------------+-----------------------
        placebo | -5.894632   -2.91579
herbs A therapy | -24.42815  -13.28753  <- สังเกตว่าในกลุ่ม herbs มีผลลด LDL แตกต่างกัน
----------------------------------------
```

Concept 1: Interaction ต้องสร้าง Product term

```
. gen statinht = statins*ht.

. reg ldlch statins ht statinht

      Source |       SS       df       MS              Number of obs =    2597
-------------+------------------------------           F(  3,  2593) =   58.25
       Model |  197600.198     3  65866.7326           Prob > F      =  0.0000
    Residual |  2932194.28  2593  1130.81152           R-squared     =  0.0631
-------------+------------------------------           Adj R-squared =  0.0621
       Total |  3129794.48  2596  1205.62191           Root MSE      =  33.628

------------------------------------------------------------------------------
       ldlch |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     statins |   2.978842   1.901594     1.57   0.117    -.7499543    6.707639
          ht |  -18.53352   1.675653   -11.06   0.000    -21.81927   -15.24776
    statinht |   8.161779   2.720817     3.00   0.003     2.826585    13.49697
       _cons |  -5.894632   1.188171    -4.96   0.000    -8.224491   -3.564773
------------------------------------------------------------------------------
```

เราสามารถใช้วิธีอื่นที่รวดเร็วขึ้น ด้วยการ fit ด้วย i.modifier##i.predictor

หมายเหตุ i. นำหน้า variable ที่เป็น categorical

c. นำหน้า variable ที่เป็น continuous

```
. reg ldlch i.statins##i.ht
```

Source	SS	df	MS		
Model	197600.198	3	65866.7326	Number of obs	2597
Residual	2932194.28	2593	1130.81152	F(3, 2593)	58.25
Total	3129794.48	2596	1205.62191	Prob > F	0.0000
				R-squared	0.0631
				Adj R-squared	0.0621
				Root MSE	33.628

ldlch	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
1.statins	2.978842	1.901594	1.57	0.117	-.7499543	6.707639
1.ht	-18.53352	1.675653	-11.06	0.000	-21.81927	-15.24776
statins#ht						
1 1	8.161779	2.720817	3.00	0.003	2.826585	13.49697
cons	-5.894632	1.188171	-4.96	0.000	-8.224491	-3.564773

ตอนนี้แถวที่ 3 ในที่นี้คือ product term เพื่อทดสอบ effect modier

p value ได้จ ากการทดสอบ เป็นของ null hypothesis คือไม่มี effect modification

จากตวั อยา่ งนี้ p = 0.003 จึง reject null hypothesis สรุปว่า statin น่าจะเป็น effect modifier ของ ht ตอ่ LDL

อย่าลืมว่าต้องมี ## เพราะถ้าใส่ # อันเดียว จะเหมือนการ fit ด้วย categorical predictor ธรรมดา

```
. reg ldlch i.statins#i.ht
```

Source	SS	df	MS		
Model	197600.198	3	65866.7326	Number of obs	2597
Residual	2932194.28	2593	1130.81152	F(3, 2593)	58.25
Total	3129794.48	2596	1205.62191	Prob > F	0.0000
				R-squared	0.0631
				Adj R-squared	0.0621
				Root MSE	33.628

ldlch	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
statins#ht						
0 1	-18.53352	1.675653	-11.06	0.000	-21.81927	-15.24776
1 0	2.978842	1.901594	1.57	0.117	-.7499543	6.707639
1 1	-7.392894	1.949993	-3.79	0.000	-11.21659	-3.569194
_cons	-5.894632	1.188171	-4.96	0.000	-8.224491	-3.564773

แถว 3 ในที่นี้ กลายเป็น category 3

p value เป็นของ null hypothesis ว่า category3 นี้มี ldlch ไม่ต่างจาก reference gr

(คือ 0 0 ไม่ใช้ทั้ง herbs และ statin)

****

Concept 2: Baseline effect + Product term effect = subgroup analysis

ในที่นี้ coef ก็คือ "effect" ของปัจจัยต่อการลด LDL
ถ้า 18.5 คือ effect ของ herbs ต่อการลด LDL ในกลุ่มผู้ "ไม่ใช้" statin
แล้ว effect ของ herbs ต่อการลด LDL ในกลุ่มผู้ใช้ statin ล่ะ? ตอนนี้แหละเป็นการนำ coef ของ product term มาใช้

```
. lincom 1.ht + 1.statins#1.ht

 ( 1)  1.ht + 1.statins#1.ht = 0

------------------------------------------------------------------------------
       ldlch |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -10.37174   2.143603    -4.84   0.000    -14.57508   -6.168391
------------------------------------------------------------------------------
```

ผลคือ 10.37 (เท่ากับคิดในใจจากตารางด้านบน -18.53 + 8.16) เป็น effect ของ herbs ต่อการลด LDL ในกลุ่มผู้ใช้ statin
ซึ่งแปลว่า Effect อ่อนกว่า
แปลแบบเต็มๆ คือ..
ในกลุ่มผู้ไม่ใช้ statin: การใช้ herbs มีผลลด LDL 18.53 เมื่อเทียบกับการไม่ใช้ herbs
ในกลุ่มผู้ใช้ statin : การใช้ herbs มีผลลด LDL 10.37 เมื่อเทียบกับการไม่ใช้ herbs

สังเกตว่า ถ้า coef ของ product term = 0 นั่นคือไม่มี interaction : "effect" ของ ht ในกลุ่มผู้ใช้ statin และ ไม่ใช้ statin มีค่าเท่ากัน

เช่นเดียวกัน ถ้า 2.97 คือ effect ของ statin ต่อการลด LDL ในกลุ่มผู้ไม่ใช้ Herbs
แล้ว effect ของ statin ต่อการลด LDL ในกลุ่มผู้ใช้ Herbs ก็หาได้จาก

```
. lincom 1.statins + 1.statins#1.ht

 ( 1)  1.statins + 1.statins#1.ht = 0

------------------------------------------------------------------------------
       ldlch |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |   11.14062   1.945967     5.72   0.000     7.324816    14.95643
------------------------------------------------------------------------------
```

ผลคือ 11.14 (=2.97 + 8.16) เป็น effect ของ statin ต่อ การลด LDL ในกลุ่ม ผู้ใช้ herbs อ่อนลง ทำไม ? ตรงนี้ต้อ งตัง สติให้ดี ณ ขณะนี้เราต้อ งการ "ลด" LDL นั่นคือยิ่ ติดลบมากยิ่งดี แปลแบบเต็มๆ คือ..
ในกลุ่ม ผู้ไม่ใช้ Herbs : การใช้ statin มีผลลด LDL 2.98 "ได้น้อยกว่า" เมื่อเทียบกับการไม่ใช้ statin
ในกลุ่ม ผู้ใช้ Herbs : การใช้ statin มีผลลด LDL 11.14 "ได้น้อยกว่า" เมื่อเทียบกับการไม่ใช้ statin
(หมายถึง นาย ก. และ นาย ข. ต่างก็ใช้ Herbs นาย ก. ซึ่งใช้ statin ลด LDL 10 นาย ข. ซึ่งไม่ใช้ statin ลด 21.14)

Concept 3 : Interaction ใน **continuous variable** ควร **Centering** และ **categorized**

เราสามารถใช้วิธีพื้นฐานในการทดสอบ **interaction** ได้เหมือนกัน แต่จะมีปัญหาในการตีความ ยกตัวอย่าง

```
. regress sbp glucose bmi glucosebmi
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
glucose	.274853	.0534465	5.14	0.000	.1700537	.3796523
bmi	.9605177	.2085822	4.60	0.000	.5515243	1.369511
glucosebmi	-.0073418	.0017137	-4.28	0.000	-.0107021	-.0039816
_cons	100.7322	6.261295	16.09	0.000	88.45487	113.0095

0.27 คือ **mean SBP** ที่เพิ่มขึ้น ในทุก **1 unit** ของ **blood glucose** เมื่อ **BMI = 0?**
0.96 คือ **mean SBP** ที่เพิ่มขึ้น ในทุก **1 unit** ของ **bmi** เมื่อ **blood glucose =0 ?**
0.007 คือ **mean SBP** ที่เพิ่มขึ้น ในทุก **1 unit** ของ **bmi** ในกลุ่มผู้ที่ **blood glucose** ไม่เท่ากับ **0?**

การ **Centering** คือ การปรับจุด **"อ้างอิง"** ของ **continuous data** แทนที่จะเป็น **0** ให้เป็นค่ากลาง หรือค่าของคนส่วนใหญ่ **(** ใครจะมี **BMI 0?)** เช่น **mean** หรือ **median** หรือ **significant clinical cutpoint value**

```
. sum bmi
```

Variable	Obs	Mean	Std. Dev.	Min	Max
bmi	2758	28.57925	5.517783	15.21	54.13

จะได้ว่า **mean** ของ **bmi** คือ **28.6**

```
gen cbmi = bmi-28.6
gen cglucose = glucose-126

gen cglucose_cbmi = cglucose* cbmi

. regress sbp glucose cbmi cglucose_cbmi
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
cglucose	.0648764	.0106964	6.07	0.000	.0439026	.0858501
cbmi	.035446	.0701359	0.51	0.613	-.1020782	.1729702
cglucose_c~i	-.0073418	.0017137	-4.28	0.000	-.0107021	-.0039816
_cons	136.3774	.4109187	331.88	0.000	135.5717	137.1831

0.06 คือ **mean SBP** ที่เพิ่มขึ้น ในทุก **1 unit** ของ **blood glucose** ในผู้ที่ **BMI = 28.6**
0.03 คือ **mean SBP** ที่เพิ่มขึ้น ในทุก **1 unit** ของ **bmi** ในผู้ที่ **blood glucose = 126**
0.007 คือ **mean SBP** ที่เพิ่มขึ้น ในทุก **1 unit** ของ **bmi** ในกลุ่มผู้ที่ **blood glucose** ไม่เท่ากับ **126**
เราสรุปได้ว่ามี Interaction **โดยดูที่** p value **ของ** product term < 0.01 **แต่แปลผลอะไรไม่ได้มากกว่านั้น**

หากเราต้องการให้ **interaction** นำไปสู่ **subgroup analysis** ในกรณีที่ตัวแปรเป็น **continuous** การแบ่งเป็นกลุ่มที่มีหรือไม่มีเท่าค่า **centering** ไม่ให้ข้อมูลอะไรนัก เราควร **categorized**
เช่นเราอยากรู้ **effect** ของ **BMI** ต่อ **SBP** แยกตามกลุ่ม **blood glucose -> normal,IFG,DM**

```
. recode glucose min/99=1 100/125=2 126/max=3, gen(dmgrp)

. label define dmgrpvals 1 "normal" 2 "IFG" 3 "diabetes"

. label values dmgrp dmgrpvals
```

```
. tab dmgrp

  (mg/dl) ) |      Freq.     Percent        Cum.
------------+-----------------------------------
     normal |      1,332       48.21       48.21
        IFG |        700       25.33       73.54
   diabetes |        731       26.46      100.00
------------+-----------------------------------
      Total |      2,763      100.00
```

ตอนนี้เราสามารถวิเคราะห์แยกตามกลุ่ม

```
. reg sbp i.dmgrp##c.cbmi

         sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dmgrp |
          2  |    3.31619   .8950856     3.70   0.000     1.561083    5.071297
          3  |   8.081379   .9335075     8.66   0.000     6.250933    9.911825
        cbmi |   .1896834   .1060169     1.79   0.074    -.0181973     .397564
dmgrp#c.cbmi |
          2  |  -.0342154   .1690438    -0.20   0.840    -.3656809    .2972501
          3  |  -.4930951   .1610166    -3.06   0.002    -.8088208   -.1773694
       _cons |   132.4312     .54383   243.52   0.000     131.3648    133.4975
------------------------------------------------------------------------------
```

แต่อย่าลืมว่า การแปลผลในกรณี **effect modifier** เป็น **categorical predictor** หากเราต้องการทดสอบ **heterogeneity** ต้องใช้ **F test**

```
. testparm dmgrp#c.cbmi

 ( 1)  2.dmgrp#c.cbmi = 0

 ( 2)  3.dmgrp#c.cbmi = 0

       F(  2,  2752) =    5.37

            Prob > F =    0.0047
```

P จาก **F test <0.01** คือมี **heterogeiniety** หรือมี **Interaction term**

ต่อไปเราก็สามารถทำ **subgroup analysis** ได้ตามใจต้องการด้วย **lincom** เช่น

กรณีอยากแบ่ง **subgroup** ตาม category ก็เลือกเปลี่ยน i ตามที่สนใจ

```
. lincom cbmi + 3.dmgrp#c.cbmi

 ( 1)  cbmi + 3.dmgrp#c.cbmi = 0

------------------------------------------------------------------------------
         sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -.3034117    .121189    -2.50   0.012    -.5410424    -.065781
------------------------------------------------------------------------------
```

0.303 เป็น **SBP** ที่ลดลง ต่อทุก **bmi** ที่เพิ่ม **1** หน่วยจาก **26.8** ในคนที่เป็น **DM (dmgrop 3)**
กรณีแบ่ง **subgroup** ตามค่าของ **continuous** ต้องทราบไวยากรณ์เพิ่มอีกนิดคือ
K*i.catvar#c.var <- k คือค่าของ **continuous** ที่เราต้องการ

```
. lincom 3.dmgrp + 30*3.dmgrp#c.cbmi

 ( 1)  3.dmgrp + 30*3.dmgrp#c.cbmi = 0

------------------------------------------------------------------------------
         sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |  -6.711473    4.80789    -1.40   0.163    -16.13891    2.715964
------------------------------------------------------------------------------
```

6.71 เป็น **SBP** ที่คนเป็น **dm (dmgrp3)** ต่างจากคนไม่เป็น **dm (dmgr1=Reference)**ในคนที่ **BMI = 30**

เมื่อวิเคราะห์เต็มรูปแบบด้วยการใช้คำสั่ง loop จะได้ผลดังต่อไปนี้
(**dmgroup** กลุ่มแรก ผลเป็น **omit** เนื่องจากมีจำนวนคนในกลุ่มนี้น้อยเกินไป)

```
. forvalues cbmi = 20(5)35 {
  2. lincom 1.dmgrp + `cbmi'*1.dmgrp#c.cbmi
  3. }

 ( 1)  1b.dmgrp + 20*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	(omitted)					

```
 ( 1)  1b.dmgrp + 25*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	(omitted)					

```
 ( 1)  1b.dmgrp + 30*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	(omitted)					

```
 ( 1)  1b.dmgrp + 35*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	(omitted)					

```
.
. forvalues cbmi = 20(5)35 {
  2. lincom 2.dmgrp + `cbmi'*1.dmgrp#c.cbmi
  3. }

 ( 1)  2.dmgrp + 20*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	3.31619	.8950856	3.70	0.000	1.561083	5.071297

```
 ( 1)  2.dmgrp + 25*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	3.31619	.8950856	3.70	0.000	1.561083	5.071297

```
 ( 1)  2.dmgrp + 30*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	3.31619	.8950856	3.70	0.000	1.561083	5.071297

```
 ( 1)  2.dmgrp + 35*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	3.31619	.8950856	3.70	0.000	1.561083	5.071297

```
.
. forvalues cbmi = 20(5)35 {
  2. lincom 3.dmgrp + `cbmi'*1.dmgrp#c.cbmi
  3. }

 ( 1)  3.dmgrp + 20*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	8.081379	.9335075	8.66	0.000	6.250933	9.911825

```
 ( 1)  3.dmgrp + 25*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	8.081379	.9335075	8.66	0.000	6.250933	9.911825

```
 ( 1)  3.dmgrp + 30*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	8.081379	.9335075	8.66	0.000	6.250933	9.911825

```
 ( 1)  3.dmgrp + 35*1b.dmgrp#co.cbmi = 0
```

sbp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	8.081379	.9335075	8.66	0.000	6.250933	9.911825