

Example of propensity score analysis:

```
obs:          3,154
vars:          8          30 Mar 2011 08:08
size:         53,618 (99.5% of memory free)
```

```
-----
      storage  display  value
variable name  type    format  label    variable label
-----
age            byte    %8.0g           Age (years) at baseline
bmi            float   %9.0g           Body mass index (kg/m^2)
chd69          byte    %8.0g    yesno    Incident CHD by 1969
chol           int     %8.0g           Total cholesterol (mg/dL) at baseline
sbp            int     %8.0g           Systolic blood pressure (mmHg) at baseline
smokcat        byte    %12.0g    smokcat  Baseline smoking
herb_A         byte    %8.0g    yesno    Treatment = hearb A
-----
```

Step 1: การสร้าง logistic model predictor for "treatment"

Concept 1: การเลือก variable เข้าใน model ควรเลือกตัวที่เป็น Confounding by indication คือมี association กับ treatment (คือ herb_A ในที่นี้) และเป็นเหตุให้เกิด outcome (CHD) โดยไม่ใช่ mediator เช่น พบว่าผู้ที่สูบบุหรี่ ใช้ herbA มากกว่าผู้ที่ไม่สูบ เนื่องจากกลิ่นของ herbA คล้ายยาสูบ กรณีนี้เป็น confounder, แต่ herb_A มีผลลดความดันโลหิต sbp กรณีนี้เป็น mediator จึงไม่ถูกนำเข้ามาใน model (สังเกต variable ตัวแรกแทนที่จะเป็น outcome กลายเป็น treatment แทน)

```
. logistic herb_A i.smokcat bmi age chol
```

Logistic regression	Number of obs	=	3142
	LR chi2(5)	=	48.82
	Prob > chi2	=	0.0000
Log likelihood = -2153.3531	Pseudo R2	=	0.0112

-----+-----							
herb_A		Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----							
smokcat							
1		1.078689	.1283486	0.64	0.524	.8543096	1.362
2		1.346144	.1061582	3.77	0.000	1.15336	1.571151
bmi		1.024752	.0146114	1.71	0.086	.9965103	1.053793
age		1.032278	.0067966	4.82	0.000	1.019042	1.045685
chol		1.001889	.0008432	2.24	0.025	1.000238	1.003543

Concept 2: ประเมิน adequacy of logistic model: GOF, linearity, omitted interaction

1.1.1 ประเมิน goodness of fit of model ด้วย Hosmer-Lemeshow

```
. estat gof, group(10)
```

Logistic model for herb_A, goodness-of-fit test

(Table collapsed on quantiles of estimated probabilities)

```

number of observations =      3142
      number of groups =         10
Hosmer-Lemeshow chi2(8) =         9.95
      Prob > chi2 =       0.2683

```

//comment: the Hosmer-Lemeshow test for goodness of fit is negative (p=0.2683). However this test is not sensitive therefore the negative test does not assure nonlinearity.

1.1.2 Test the departure from linearly of continuous predictors (chol bmi, age) by transform to restricted cubic splin

การทำ restricted cubic splin จินตนาการเหมือนนำกราฟมาตัดเป็นท่อนๆ แล้วดูว่าแต่ละท่อนมี slope (coef) เหมือนกันไหม เพราะถ้าเป็นเส้นตรง slope ควรเท่ากัน การใช้คำสั่ง testparm ก็เป็นการทดสอบทางสถิติว่า coef homogenize หรือไม่

```
. mkspline agesp = age, cubic
```

```
. mkspline cholsp = chol, cubic
```

```
. mkspline bmisp = bmi, cubic
```

```
. logistic herb_A i.smokcat agesp* cholsp* bmisp*
```

```
Logistic regression               Number of obs   =       3142
                                LR chi2(14)       =       58.68
                                Prob > chi2        =       0.0000
Log likelihood = -2148.4188       Pseudo R2      =       0.0135
```

herb_A	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
smokcat						
1	1.090735	.1303817	0.73	0.467	.8629194	1.378695
2	1.350093	.1069418	3.79	0.000	1.155951	1.576841
agesp1	.9555508	.0646946	-0.67	0.502	.8368045	1.091148
agesp2	1.408054	1.075757	0.45	0.654	.3149913	6.29419
agesp3	.6682795	1.325552	-0.20	0.839	.0136954	32.60926
agesp4	.8331028	1.361786	-0.11	0.911	.0338321	20.51486
cholsp1	1.004159	.0043189	0.96	0.335	.9957296	1.01266
cholsp2	.9926813	.0267778	-0.27	0.785	.9415611	1.046577
cholsp3	.9990209	.1414442	-0.01	0.994	.7569358	1.31853
cholsp4	1.046889	.2012986	0.24	0.812	.7181736	1.526061

bmispl		.930967	.0597617	-1.11	0.265	.8209051	1.055785
bmispl2		1.91327	.6921351	1.79	0.073	.9415637	3.887789
bmispl3		.0274355	.057726	-1.71	0.087	.0004439	1.695484
bmispl4		143.0575	456.5112	1.56	0.120	.274941	74435.81

-----.

```
. testparm agesp2-agesp4
```

```
( 1) [herb_A]agesp2 = 0
```

```
( 2) [herb_A]agesp3 = 0
```

```
( 3) [herb_A]agesp4 = 0
```

```
      chi2( 3) =      5.17
```

```
      Prob > chi2 =      0.1595
```

```
. testparm cholsp2-cholsp4
```

```
( 1) [herb_A]cholsp2 = 0
```

```
( 2) [herb_A]cholsp3 = 0
```

```
( 3) [herb_A]cholsp4 = 0
```

```
      chi2( 3) =      1.36
```

```
      Prob > chi2 =      0.7155
```

```
. testparm bmispl2-bmispl4
```

```
( 1) [herb_A]bmispl2 = 0
```

```
( 2) [herb_A]bmispl3 = 0
```

```
( 3) [herb_A]bmispl4 = 0
```

```
      chi2( 3) =      3.23
```

```
      Prob > chi2 =      0.3579.
```

```
estat gof, group(10) table
```

```
Logistic model for herb_A, goodness-of-fit test
```

```
(Table collapsed on quantiles of estimated probabilities)
```

```
      number of observations =      3142
```

```

        number of groups =          10

Hosmer-Lemeshow chi2(8) =          3.80

        Prob > chi2 =          0.8751

```

```

//comment: From the test departure from linearity after transform to cubic spline. There is no
evidence of departure from linearity for each continuous variable. age p=0.1457 chol
p=0.6657 bmi =0.3579

```

1.1.3 Screen for omitted interaction

```

display "P-values for pairwise interactions:"

P-values for pairwise interactions:

.foreach x in i.smokcat c.bmi c.age c.chol {
  foreach y in i.smokcat c.bmi c.age c.chol {
    if ("`y'">"`x'") {
      quietly logistic herb_A i.smokcat c.bmi c.age c.chol `x'#`y'

      * overall test for interaction

      . quietly testparm `x'#`y'

      display "`x' and `y': " round(r(p), .0001)

    }
  }
}

c.bmi and i.smokcat: .2418

c.bmi and c.chol: .0639

c.age and i.smokcat: .4793

c.age and c.bmi: .476

c.age and c.chol: .2248

c.chol and i.smokcat: .1359

//comment there is no important omitted interaction(all have p> 0.05)

```

Step 2 estimate propensity score

```
. predict prop_score  
  
(option pr assumed; Pr(herb_A))  
  
(12 missing values generated)  
  
. predict logit_ps, xb  
  
(12 missing values generated)  
  
. xtile ps_quintile = prop_score, nq(5)  
  
. label variable ps_quintile "Propensity score quintile"
```

2.2.1 Evaluate the degree of balance achieved by the estimated scores in the distribution of confounders between behavior type groups. การประเมิน balance หมายถึงเมื่อดูทีละ covariate ภายใน quintile ของ propensity score เดียวกัน ควรมี no (กรณี categorical) หรือ mean (กรณี continuous) พอๆ กันในกลุ่ม treatment และ non treatment

```
. foreach x in age bmi chol{  
2.      tab herb_A, sum(`x')  
3.      table herb_A ps_quintile, c(mean `x') format(%9.2f)  
4. }
```

Herb A			
treatment1 Summary of Age (years) at baseline			
status	Mean	Std. Dev.	Freq.
-----+-----			
No	45.78722	5.296946	1565
Yes	46.762744	5.6993059	1589
-----+-----			
Total	46.278694	5.5240451	3154

Herb A					
treatment	Propensity score quintile				
status	1	2	3	4	5
-----+-----					
No	41.94	43.70	45.25	48.03	52.26
Yes	41.39	43.80	45.59	48.08	52.57

Herb A			
treatment1	Summary of Body mass index (kg/m^2)		
status	Mean	Std. Dev.	Freq.
-----+-----			
No	24.445088	2.5284337	1565
Yes	24.590554	2.604175	1589
-----+-----			
Total	24.518375	2.5674957	3154

Herb A					
treatment	Propensity score quintile				
status	1	2	3	4	5
-----+-----					
No	23.79	24.36	24.62	24.55	25.17
Yes	23.83	24.23	24.45	24.87	25.23

Herb A	Summary of Total cholesterol		
treatment1	(mg/dL) at baseline		
status	Mean	Std. Dev.	Freq.
-----+-----			
No	223.87677	42.418052	1558
Yes	228.82702	44.260532	1584
-----+-----			

Total		226.37237	43.420426	3142
-------	--	-----------	-----------	------

Herb A						
treatment		Propensity score quintile				
status		1	2	3	4	5

-----+-----

No		197.59	218.87	226.25	236.78	251.57
Yes		200.27	217.50	226.51	236.16	252.54

```
. table type_A ps_quintile smokcat
```

Type A		Baseline smoking and Propensity score quintile														
behaviora		----- Non-smoker -----					----- <1 pack/day -----					----- >=1 pack/day -----				
l pattern		1	2	3	4	5	1	2	3	4	5	1	2	3	4	5

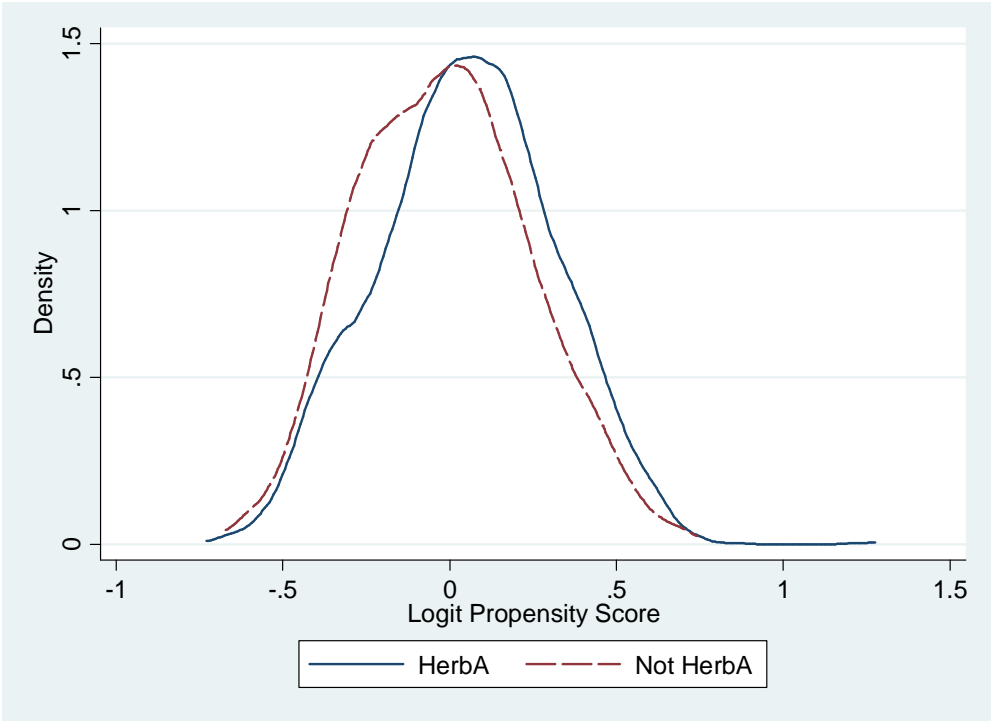
-----+-----

No		339	223	138	100	64	30	64	31	36	15	1	59	137	142	179
Yes		242	177	145	125	94	16	54	42	41	22	1	51	136	184	254

Comment: There are fairly balance between herbA and no herbA group. There are more balance within quintile of propensity score.

2.2.2 Ggraphical approach based on the estimated propensity scores, evaluate the degree of overlap

```
. twoway (kdensity logit_ps if herbA==1, lpattern(solid)) ///
(kdensity logit_ps if herbA==0, lpattern(longdash)), ///
ytitle("Density") xtitle("Logit Propensity Score") ///
legend(order(1 "HerbA" 2 "Not HerbA")) name(pscores)
```

//comment: The density plot of propensity scores by treatment show good overlap (minimal poor overlap area) of covariate between herbA and no herbA

2.2.3 Check event in each propensity quintile

```
. table herb_A ps_quintile, c(count chd69)
```

Herb A						
treatment		Propensity score quintile				
status		1	2	3	4	5
-----+-----						
No		370	346	306	278	258
Yes		259	282	323	350	370

Step 3 การนำ propensity score มาใช้ ในที่นี้ยกตัวอย่างการนำไปเป็น single covaribale ในการ adjuste ผลของ **herbA** ต่อการเกิด CHD

Conditional effect

```
. cs chd69 type_A, by(ps_quintile) or
```

Propensity score	OR	[95% Conf. Interval]		M-H Weight
-----+-----				
1	3.674699	1.204277	11.1906	1.583466 (Cornfield)
2	1.426423	.693242	2.934592	5.929936 (Cornfield)
3	1.410544	.7851403	2.533463	9.348172 (Cornfield)
4	3.309917	1.688749	6.481462	5.394904 (Cornfield)
5	2.130104	1.356909	3.342874	13.80573 (Cornfield)
-----+-----				
Crude	2.372929	1.804034	3.121147	
M-H combined	2.072189	1.565935	2.742109	

```
Test of homogeneity (M-H)      chi2(4) =    5.346  Pr>chi2 = 0.2536
```

หรือ

```
. logistic chd69 type_A i.ps_quintile
```

```
Logistic regression              Number of obs   =       3142
                                LR chi2(5)        =       137.98
                                Prob > chi2        =       0.0000
Log likelihood = -820.60997      Pseudo R2      =       0.0776
```

chd69	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
type_A	2.069802	.2955071	5.10	0.000	1.564596	2.738139

ps_quintile						
2		2.15134	.7086272	2.33	0.020	1.128061 4.102846
3		3.472767	1.074431	4.02	0.000	1.893759 6.368348
4		3.678128	1.129672	4.24	0.000	2.01463 6.715192
5		8.470685	2.464822	7.34	0.000	4.788877 14.98316

//Comment: Conditional OR = 2.07. หมายความว่า เมื่อเทียบกับผู้ไม่ได้รับ HerbA กลุ่มที่ได้รับ HerbA มี odds เกิด CHD 2.07 เท่า

Marginal effect

```

margins herb_A
. matrix b = r(b)
. scalar EY0 = b[1, 1]
. scalar EY1 = b[1, 2]
. dis EY1/(1-EY1)*(1-EY0)/EY0
2.0223713

```

//comment : marginal OR = 2.022 หมายความว่า หากเราให้ Herb A แก่ประชากรที่เราศึกษา ทุกคน ในกลุ่มที่ได้รับ HerbA มี odds เกิด CHD 2.022 เท่า ซึ่งจะเห็นว่าไม่แตกต่างจาก conditional OR มากนัก