

BiostatIIlab3_categorical predictors

Categorical predictors:

หมายเหตุ: หากมีข้อแก้ไข หรือแนะนำ กรุณา email: pgomutbu@med.cmu.ac.th

. ชุดข้อมูลที่นำมาวิเคราะห์ (ข้อมูลสมมติ)

- > 1. PaCO2 - partial pressure of carbon dioxide from blood gas (normal 35-45 mm Hg)
- > 2. Dyspnea - Dyspnea severity, coded 1-5
- > 3. age10 - age in units of 10 years
- > 4. smoking - an indicator of current smoking (1 = yes, 0 = no).
- > 5. drink - an indicator of any current alcohol use, coded like smoking.

. เรากำลังสนใจวิเคราะห์ 5-level Dyspnea severity ในการเป็น ordinal categorical predictor ของ PaCO2. ตั้งที่เรทแรกแล้วว่า multivariable linear regression ใช้ในกรณี outcome เป็น continuous

Concept 1: อย่าลืมใส่ i.prefix (หรือ xi ใน STATA 10 หรือก่อนหน้า)

เพื่อให้แสดงแต่ category เราต้องใส่ i.หน้าตัวแปร หากไม่ใส่ STATA จะปฏิบัติต่odysp เป็น dichotomous หนึ่งตัวแปร (1= dyspnea, 0 = no dyspnea)

หลังจากใส่ i. แล้ว "แต่ละ" category จะถูกปฏิบัติเป็น dichotomous (1= dyspnea cat2, 0= reference)

ในกรณี 1 predictor

. reg PaCO2 i.dysp

Source	SS	df	MS	Number of obs =	551
Model	1203.77413	4	300.943533	F(4, 546) =	11.41
Residual	14398.4676	546	26.3708197	Prob > F =	0.0000
Total	15602.2417	550	28.3677121	R-squared =	0.0772
				Adj R-squared =	0.0704
				Root MSE =	5.1353

PaCO2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
dysp					
2	1.699789	.9111373	1.87	0.063	-.0899745 3.489553
3	.4579446	.8402237	0.55	0.586	-1.192522 2.108411
4	-1.715916	.8452219	-2.03	0.043	-3.3762 -.0556309
5	-2.931698	1.033544	-2.84	0.005	-4.961907 -.9014894
_cons	29.21766	.7490536	39.01	0.000	27.74628 30.68904

ในกรณี adjust ด้วย predictors อื่นด้วย

. reg PaCO2 i.dyspnea age10 smoking drink

Source	SS	df	MS	Number of obs =	551
Model	1865.4298	7	266.489971	F(7, 543) =	10.53
Residual	13736.8119	543	25.2979961	Prob > F =	0.0000
Total	15602.2417	550	28.3677121	R-squared =	0.1196
				Adj R-squared =	0.1082
				Root MSE =	5.0297

BiostatIILab3_categorical predictors

Concept 3: แตกต่างระหว่างกลุ่ม? -> แตกต่างอย่างเป็นระบบ? -> แบบเส้นตรง?

เราไม่สามารถคิด dyspnea level แบบ continuous variable เพราะ แต่ละ category ที่ต่างกัน 1 ช่วง ไม่มี (หรือวัดลำบาก) ถึงความเป็น "interval" กล่าวคือ ความทุกข์ทรมานของ dyspnea ระหว่างระดับ 1 กับ 2 กับ ระดับ 4 กับ 5 ไม่เท่ากัน การที่เราอนุรักษ์ความเป็น Categorical ของ ordinal variable ไว้ไม่ collapse ให้กลายเป็น dichotomous ก็เพื่อสงวน "Information" ไว้ กล่าวคือ "Trend" หรือ "Dose effect"

ขั้นตอนทดสอบ เพื่อนำไปสู่ข้อสรุปดังกล่าว ประกอบด้วย

** 3.1 การทดสอบ Heterogeneity ใช้คำสั่ง testparm ซึ่งเป็นการจับคู่เปรียบเทียบแล้วดูภาพรวม มี null hypothesis คือแต่ละกลุ่มที่เปรียบเทียบ "ไม่แตกต่างกัน"

```
. testparm i(2/5).dysp  
  
( 1) 2.dysp = 0  
( 2) 3.dysp = 0  
( 3) 4.dysp = 0  
( 4) 5.dysp = 0  
  
F( 4, 546) = 11.41  
Prob > F = 0.0000
```

P value ในที่นี้ (prob>F) มีค่า <0.05 แสดงว่า reject null hypothesis นั่นคือแต่ละกลุ่มที่เปรียบเทียบ "แตกต่างกัน" (มี Heterogeneity)

**3.2 การทดสอบ trend ใช้สมการ ซึ่งแตกต่างตามจำนวน category ดังนี้ (ปรับปรุงจาก Vittinghoff et al. Regression method in Biostatistic p.82)

```
3 level test +3.val = 0  
4 level test -2.val + 3.val + 3*4.val = 0  
5 level test -2.val + 4.val + 2*5.val = 0  
6 level test -3* 2.val - 3.val + 4.val + 3*5.val = 0  
มี null hypothesis คือ แต่ละกลุ่มที่เปรียบเทียบ "ไม่แตกต่างกัน-อย่างเป็นระบบ"
```

. ในกรณีนี้เรามี dyspnea 5 level

```
. test -2.dysp + 4.dysp + 2*5.dysp=0  
  
( 1) - 2.dysp + 4.dysp + 2*5.dysp = 0  
  
F( 1, 546) = 18.34  
Prob > F = 0.0000
```

P value ในที่นี้ (prob>F) มีค่า <0.05 แสดงว่า reject null hypothesis นั่นคือแต่ละกลุ่มที่เปรียบเทียบ มีระดับแตกต่างกัน อย่างเป็นระบบ หรือมี trend (ซึ่งอาจเป็นแบบ linear หรือไม่ก็ได้)

**3.3 การทดสอบ Departure from linearity

หากพบว่ามี trend และอยากทราบว่าเป็น linear หรือไม่ วิธีการหนึ่งคือ "ทดสอบ heterogeneity กับ linear -> colinearity?" กล่าวคือ null hypothesis ไม่เป็น linear

```
. reg PaCO2 dysp i.dysp  
note: 5.dysp omitted because of collinearity
```

Source	SS	df	MS	Number of obs =	551
Model	1203.77413	4	300.943533	F(4, 546) =	11.41
Residual	14398.4676	546	26.3708197	Prob > F =	0.0000
-----				R-squared =	0.0772
-----				Adj R-squared =	0.0704
Total	15602.2417	550	28.3677121	Root MSE =	5.1353

BiostatIILab3_categorical predictors

PaCO2	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
dysp	-.7329245	.2583859	-2.84	0.005	-1.240477 - .2253723
dysp 2	2.432714	.7851076	3.10	0.002	.8905125 3.974915
3	1.923794	.6418315	3.00	0.003	.6630322 3.184555
4	.4828578	.6882206	0.70	0.483	-.8690264 1.834742
5	(omitted)				
_cons	29.95058	.9530924	31.42	0.000	28.07841 31.82276

. testparm i(2/5).physact

- (1) 2.physact = 0
- (2) 3.physact = 0
- (3) 4.physact = 0

F(3, 546) = 4.72
 Prob > F = 0.0029

P value ในที่นี้ (prob>F) มีค่า<0.05 แสดงว่าreject null hypothesis นั่นคือ trend ที่มันจะเป็น linear

closed on: 10 Mar 2011, 18:16:34