ประเด็นเพิ่มเติมสำหรับการวิเคราะห์ถดถอย

สถิติขั้นกลางสำหรับจิตวิทยา

สันทัด พรประเสริฐมานิต

โครงร่างการนำเสนอ

- การทำนายด้วยตัวแปรจัดกลุ่ม (Categorical Independent Variables)
- การวิเคราะห์ถดถอยแบบลำดับขั้น (Hierarchical Regression)
- ความสัมพันธ์บางส่วน
- การใช้โปรแกรมคัดเลือกตัวแปรอิสระ
- การเปลี่ยนศูนย์กลาง (Centering)
- ข้อตกลงเบื้องต้นก่อนการใช้สถิติ
- การวิเคราะห์กำลังและการกำหนดขนาดกลุ่มตัวอย่าง

การทำนายด้วยตัวแปรจัดกลุ่ม

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

การทำนายด้วยตัวแปรจัดกลุ่ม

- ในการวิเคราะห์ถดถอย ตัวแปรอิสระจะมีการกระจายเป็นรูปแบบใดก็ได้
  - กล่าวคือ การวิเคราะห์ถดถอยจะมีข้อตกลงเบื้องต้นว่า ค่าความผิดพลาด (error) มีการกระจายเป็นโค้งปกติ แต่ไม่ได้มีข้อจำกัดเรื่องการกระจายของตัวแปรตามหรือตัวแปรอิสระ
  - ต่างกับการทดสอบทางสถิติของค่าสหสัมพันธ์
- คราวที่แล้ว กล่าวถึงตัวแปรอิสระที่เป็นตัวแปรแบบระดับอันตรภาค (Interval scale) หรือสัดส่วน (Ratio scale) เท่านั้น
- ในที่นี่ จะกล่าวถึง การทำนายด้วยตัวแปรจัดกลุ่มแบบไม่มีอันดับ (Nominal scale) และตัวแปรจัดกลุ่มแบบมีอันดับ (Ordinal scale) ตามลำดับ

การทำนายด้วยตัวแปรจัดกลุ่ม

- ตัวแปรแบบนามบัญญัติ (Nominal scale) สามารถใช้เป็นตัวแปรอิสระได้ แต่ต้องแปลงให้เป็นตัวแปรแบบดัมมี่ (Dummy variable)
- ตัวแปรดัมมี่ คือ ตัวแปรที่มีค่า 1 หรือ 0 บ่งชี้ว่าเป็นสมาชิกกลุ่มที่กำหนดหรือไม่
- จำนวนของตัวแปรดัมมี่สำหรับตัวแปรแบบนามบัญญัติหนึ่ง จะเท่ากับจำนวนกลุ่มลบหนึ่ง

การทำนายด้วยตัวแปรจัดกลุ่ม

- ตัวแปรแบบมี 2 กลุ่ม หรือที่เรียกว่าตัวแปรทวินาม (Dichotomous variable)

ตัวแปรนามบัญญัติ

เพศ
ชาย
หญิง

ตัวแปรดัมมี่

ชาย
1
0

- จำนวนตัวแปรดัมมี่ = 2 – 1 = 1 ตัวแปร
- กลุ่มอ้างอิง คือ กลุ่มที่ถูกกำหนดให้มีค่า 0 ในที่นี้คือ หญิง

การทำนายด้วยตัวแปรจัดกลุ่ม

- ตัวแปรมีมากกว่า 2 กลุ่ม หรือที่เรียกว่าตัวแปรพหุนาม (Polytomous variable)

ตัวแปรนามบัญญติ

ศาสนา
พุทธ
คริสต์
อิสลาม
อื่นๆ

ตัวแปรดัมมี่

พุทธ	คริสต์	อิสลาม
1	0	0
0	1	0
0	0	1
0	0	0

- จำนวนตัวแปรดัมมี่ = 4 – 1 = 3 ตัวแปร
- กลุ่มอ้างอิง คือ กลุ่มที่ถูกกำหนดให้มีค่า 0 ในที่นี้คือ กลุ่มศาสนาอื่นๆ

การทำนายด้วยตัวแปรจัดกลุ่ม

- ใน SPSS สามารถใช้คำสั่ง Transform → Recode into Different Variables… ในการสร้างตัวแปรดัมมี่
- กลุ่มอ้างอิง สามารถเป็นกลุ่มใดก็ได้ตามที่ผู้วิจัยกำหนด โดยส่วนใหญ่แล้วจะใช้กลุ่มควบคุม เพื่อช่วยในการแปลความหมายสัมประสิทธิ์ถดถอย
  - จุดตัดแกน *Y* คือ ค่าเฉลี่ยของตัวแปรตามในกลุ่มอ้างอิง
  - ความชัน คือ ความแตกต่างระหว่างกลุ่มที่ใช้ในการกำหนดตัวแปรดัมมี่นั้น และกลุ่มอ้างอิง

ตัวแปรนามบัญญัติ

เพศ
ชาย
หญิง

ตัวแปรดัมมี่

ชาย
1
0

$$\hat{Y} = a + bX$$

ค่าที่ทำนายได้ ซึ่งค่าที่ดีที่สุด คือค่าเฉลี่ยของกลุ่ม

ใส่ตัวแปรดัมมี่

$$\bar{Y}_{\text{ชาย}} = a + b(1) = a + b$$

$$\bar{Y}_{\text{หญิง}} = a + b(0) = a$$

$$a = \bar{Y}_{\text{หญิง}}$$

$$b = \bar{Y}_{\text{ชาย}} - \bar{Y}_{\text{หญิง}}$$

ตัวแปรนามบัญญัติ

ศาสนา
พุทธ
คริสต์
อิสลาม
อื่นๆ

ตัวแปรดัมมี่

พุทธ	คริสต์	อิสลาม
1	0	0
0	1	0
0	0	1
0	0	0

$$\hat{Y} = a + \underbrace{b_1X_1 + b_2X_2 + b_3X_3}_{\text{ใส่ตัวแปรดัมมี่}}$$

ค่าเฉลี่ยของกลุ่ม

$$a = \bar{Y}_{\text{อื่นๆ}}$$
$$b_1 = \bar{Y}_{\text{พุทธ}} - \bar{Y}_{\text{อื่นๆ}}$$
$$b_2 = \bar{Y}_{\text{คริสต์}} - \bar{Y}_{\text{อื่นๆ}}$$
$$b_3 = \bar{Y}_{\text{อิสลาม}} - \bar{Y}_{\text{อื่นๆ}}$$

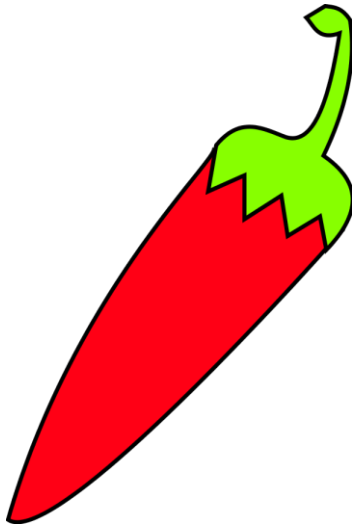
$$\bar{Y}_{\text{พุทธ}} = a + b_1(1) + b_2(0) + b_3(0) = a + b_1$$
$$\bar{Y}_{\text{คริสต์}} = a + b_1(0) + b_2(1) + b_3(0) = a + b_2$$
$$\bar{Y}_{\text{อิสลาม}} = a + b_1(0) + b_2(0) + b_3(1) = a + b_3$$
$$\bar{Y}_{\text{อื่นๆ}} = a + b_1(0) + b_2(0) + b_3(0) = a$$

การทำนายด้วยตัวแปรจัดกลุ่ม

- แท้จริงแล้ว หากนำตัวแปรดัมมี่เป็นตัวแปรอิสระเพียงอย่างเดียว ผลที่ได้จะเทียบเท่ากับการวิเคราะห์ความแปรปรวนแบบทางเดียว
  - ค่า *F* จากการวิเคราะห์ถดถอยและความแปรปรวนทางเดียวจะเท่ากัน
  - หากมี 2 กลุ่ม ค่า *t* ในการทดสอบความชัน จะเท่ากับค่า *t* ในการทดสอบทีแบบอิสระจากกัน
  - ค่า R^2 เท่ากับค่า Eta-squared

การทำนายด้วยตัวแปรจัดกลุ่ม

หาปัจจัยที่อธิบายความแตกต่างระหว่างการกินเผ็ดโดยให้กลุ่มตัวอย่างปรุงรสก๋วยเตี๋ยวที่ตนเองทานเป็นประจำ แล้ววัดน้ำหนักของพริกป่นที่ใส่ โดยปัจจัยที่ทดสอบมีดังนี้



ภาค
(เหนือ กลาง อีสาน ใต้)

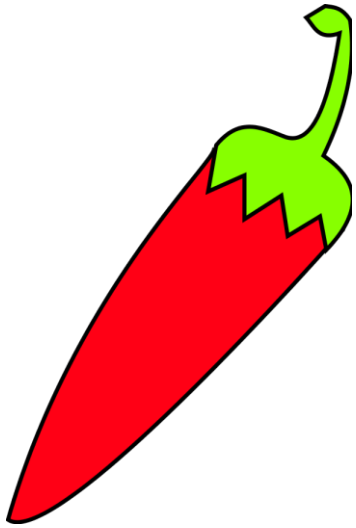
เพศ
(หญิง ชาย)

อายุ
(18 – 60 ปี)

ระดับการกินเผ็ด

การทำนายด้วยตัวแปรจัดกลุ่ม

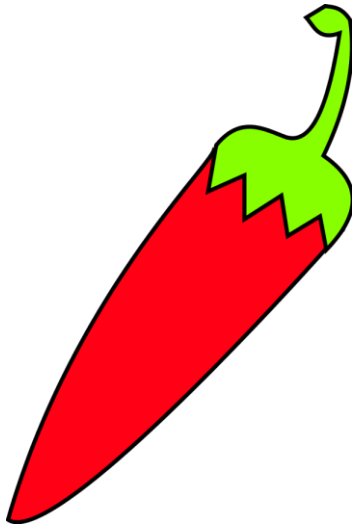
ตัวอย่างการแปลงข้อมูลใน SPSS

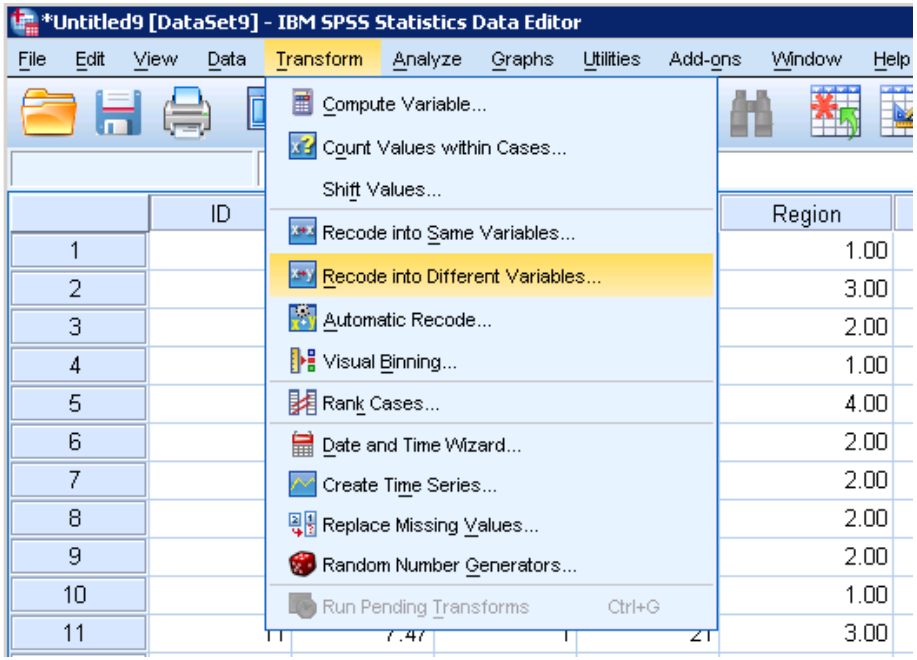


	ID	Chili_mg	Sex	Age	Region
1	1	4.08	1	23	1.00
2	2	11.15	2	43	3.00
3					
4					
5					
6	6	6.78	2	30	2.00
7	7	2.36	1	59	2.00
8	8	6.18	1	45	2.00
9	9	1.81	2	43	2.00
10	10	.00	1	52	1.00
11	11	7.47	1	21	3.00
12	12	.05	2	56	2.00
13	13	1.24	1	18	2.00
14	14	.00	2	50	2.00
15	15	6.89	2	20	1.00
16	16	4.81	2	41	2.00
17	17	9.87	1	26	3.00
18	18	8.20	2	26	3.00
19	19	13.90	2	48	4.00
20	20	.73	2	58	1.00

การทำนายด้วยตัวแปรจัดกลุ่ม

ตัวอย่างการแปลงข้อมูลใน SPSS



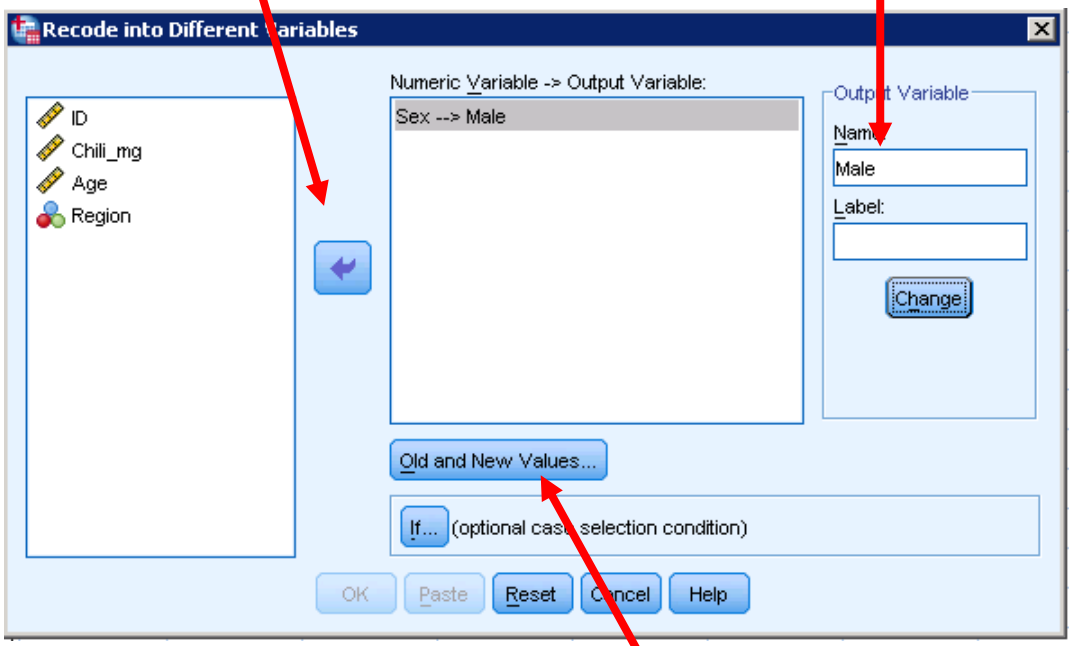


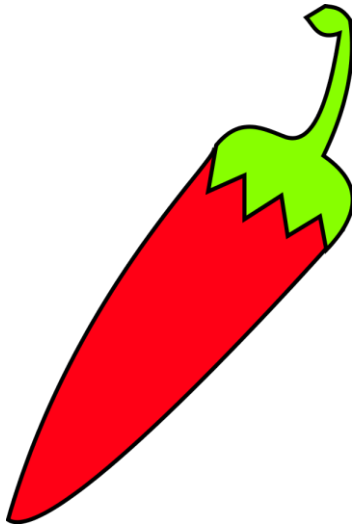
เลือกแปลงค่าเป็นตัวแปรใหม่ เพราะการวิเคราะห์ข้อมูล มักมีผิดพลาดเสมอ หากไม่เก็บข้อมูลเก่าไว้ อาจทำให้แก้ไขข้อผิดพลาดไม่ได้

การทำนายด้วยตัวแปรจัดกลุ่ม

ใส่ตัวแปร Sex เข้าไป

ใส่ชื่อตัวแปรใหม่



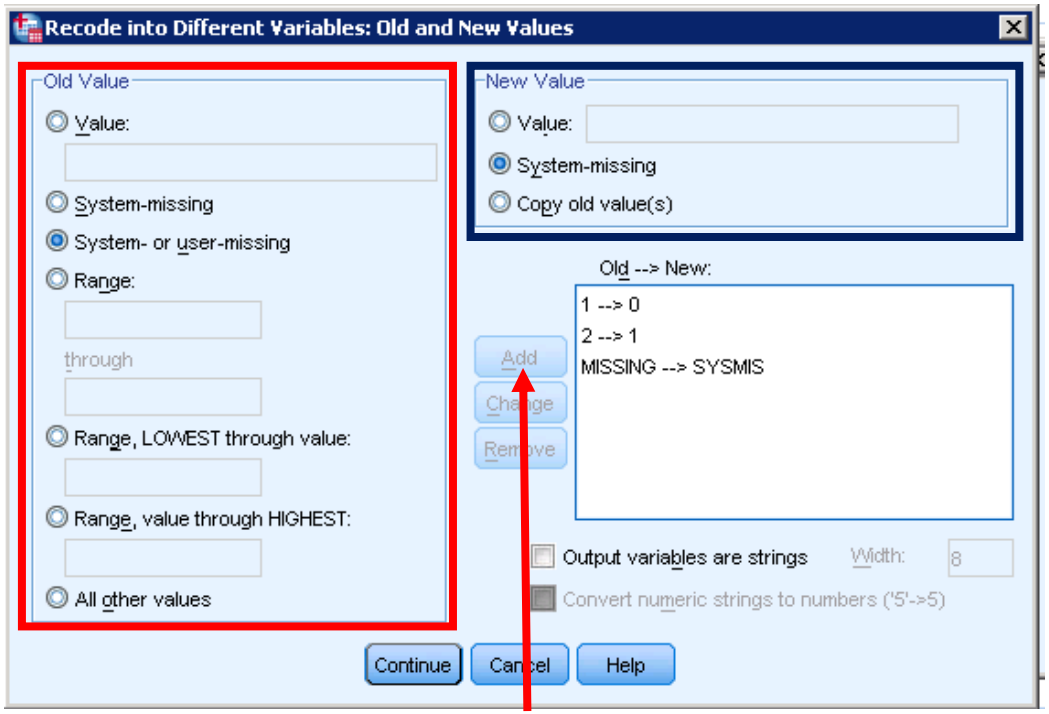


กดปุ่มนี้ เพื่อบอกว่าจะแปลงค่าอะไรในตัวแปรเก่า เป็นค่าอะไรในตัวแปรใหม่

การทำนายด้วยตัวแปรจัดกลุ่ม

ใส่ค่าตัวแปรเก่า

ใส่ค่าตัวแปรใหม่



กด Add เพื่อยอมรับการแปลงค่า

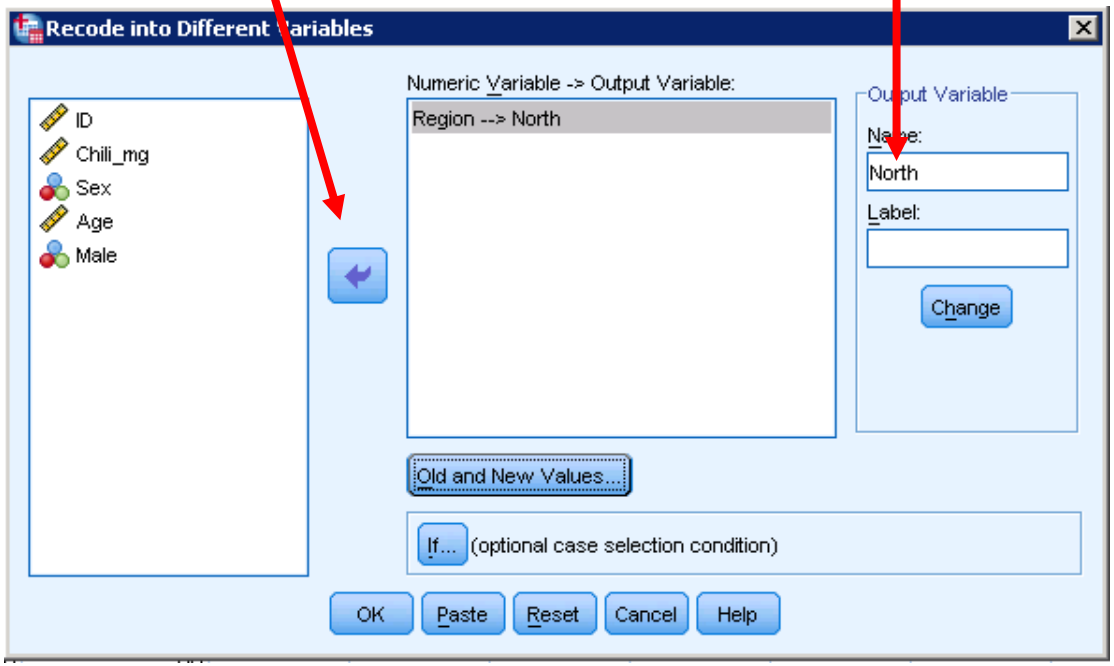
Sex	Male
1 (Female)	0
2 (Male)	1
System- or user-missing	System-missing

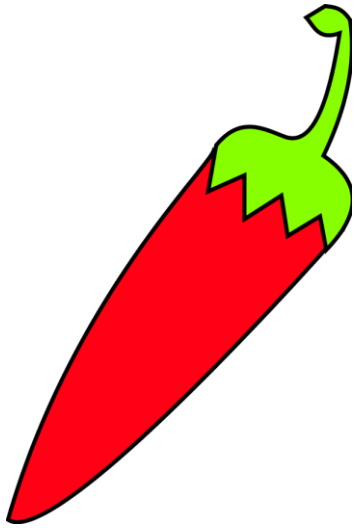
หลังจากนี้ กด Continue และ OK

การทำนายด้วยตัวแปรจัดกลุ่ม

ใส่ตัวแปร Region เข้าไป

ใส่ชื่อตัวแปรใหม่





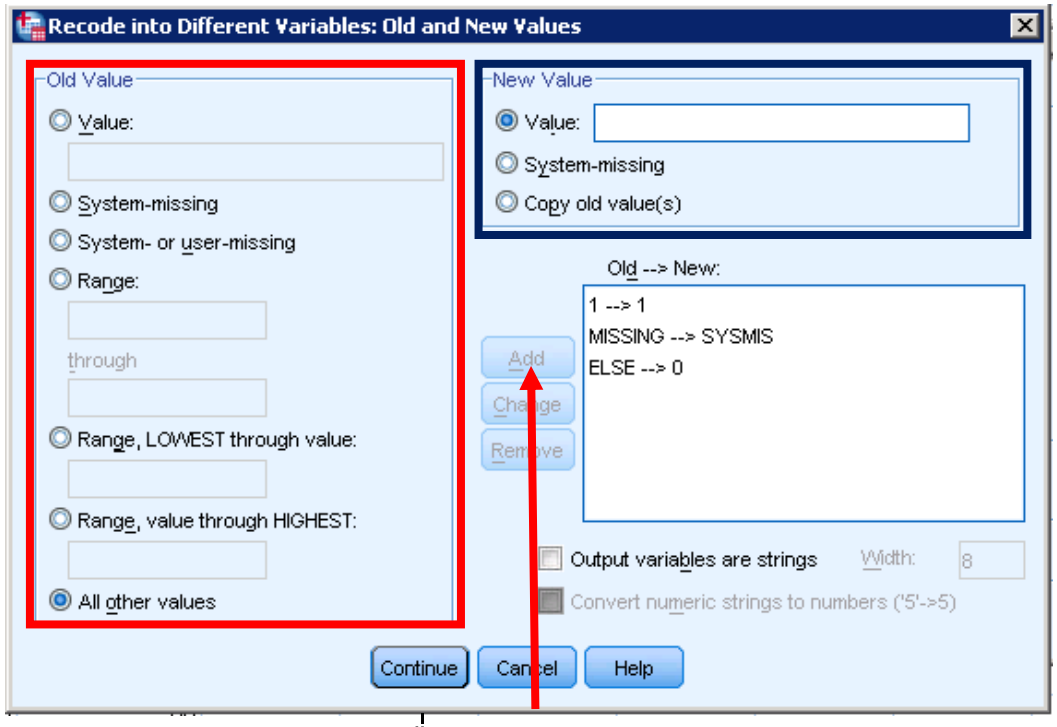
ทำแบบนี้ 3 ครั้ง เพื่อสร้างตัวแปรดัมมี่ 3 ตัวแปร

แทนภาคเหนือ กลาง อีสาน (ภาคใต้เป็นกลุ่มอ้างอิง)

การทำนายด้วยตัวแปรจัดกลุ่ม

ใส่ค่าตัวแปรเก่า

ใส่ค่าตัวแปรใหม่



กด **Add** เพื่อยอมรับการแปลงค่า

Region	North
1 (North)	1
System- or user-missing	System-missing
All other values	0

หลังจากนี้ กด **Continue** และ **OK**

รูปแบบข้อมูลใหม่

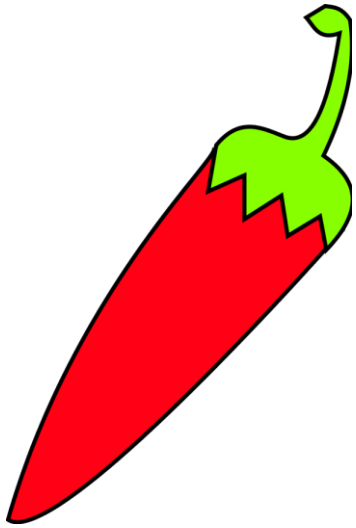
แปลงจาก Sex

แปลงจาก Region

	ID	Chili_mg	Sex	Age	Region	Male	North	Center	Northeast
1	1	4.08	1	23	1.00	.00	1.00	.00	.00
2	2	11.15	2	43	3.00	1.00	.00	.00	1.00
3	3	7.75	2	36	2.00	1.00	.00	1.00	.00
4	4	.00	1	39	1.00	.00	1.00	.00	.00
5	5	20.09	2	24	4.00	1.00	.00	.00	.00
6	6	6.78	2	30	2.00	1.00	.00	1.00	.00
7	7	2.36	1	59	2.00	.00	.00	1.00	.00
8	8	6.18	1	45	2.00	.00	.00	1.00	.00
9	9	1.81	2	43	2.00	1.00	.00	1.00	.00
10	10	.00	1	52	1.00	.00	1.00	.00	.00
11	11	7.47	1	21	3.00	.00	.00	.00	1.00
12	12	.05	2	56	2.00	1.00	.00	1.00	.00
13	13	1.24	1	18	2.00	.00	.00	1.00	.00
14	14	.00	2	50	2.00	1.00	.00	1.00	.00
15	15	6.89	2	20	1.00	1.00	1.00	.00	.00
16	16	4.81	2	41	2.00	1.00	.00	1.00	.00
17	17	9.87	1	26	3.00	.00	.00	.00	1.00
18	18	8.20	2	26	3.00	1.00	.00	.00	1.00
19	19	13.90	2	48	4.00	1.00	.00	.00	.00
20	20	.73	2	58	1.00	1.00	1.00	.00	.00

การทำนายด้วยตัวแปรจัดกลุ่ม

ทำนายด้วยเพศ
เพียงตัวแปรเดียว



Regression: Male → Chili

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.046[a]	.002	.000	5.38673

a. Predictors: (Constant), Male

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	24.339	1	24.339	.839	.360[b]
	Residual	11548.690	398	29.017		
	Total	11573.029	399			

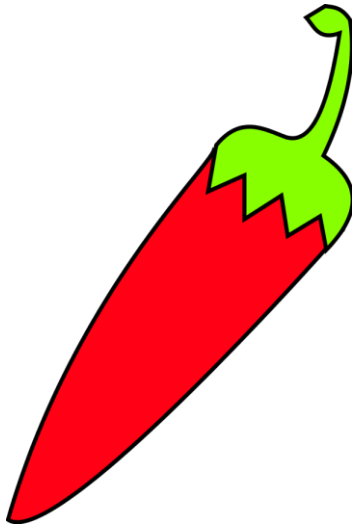
a. Dependent Variable: Chili_mg

b. Predictors: (Constant), Male

$R^2 = .002$, $F(1, 398) = .84$, $p = .36$

การทำนายด้วยตัวแปรจัดกลุ่ม

ทำนายด้วยเพศ
เพียงตัวแปรเดียว



Regression: Male → Chili

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.956	.381		18.262	.000
	Male	.493	.539	.046	.916	.360

a. Dependent Variable: Chili_mg

$$\hat{Y} = 6.956 + 0.493(\text{Male})$$

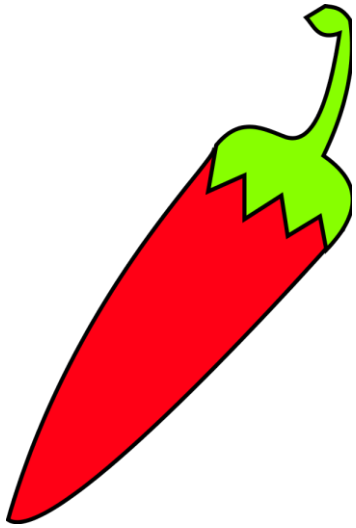
$$\bar{Y}_{\text{Male}} = 6.956 + 0.493 = 7.449$$

$$\bar{Y}_{\text{Female}} = 6.956$$

การทำนายด้วยตัวแปรจัดกลุ่ม

Regression: Male → Chili

ทำนายด้วยเพศ
เพียงตัวแปรเดียว



Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	6.956	.381		18.262	.000
	Male	.493	.539	.046	.916	.360

a. Dependent Variable: Chili_mg

$a = 6.956$, $t(398) = 18.262$, $p < .001$

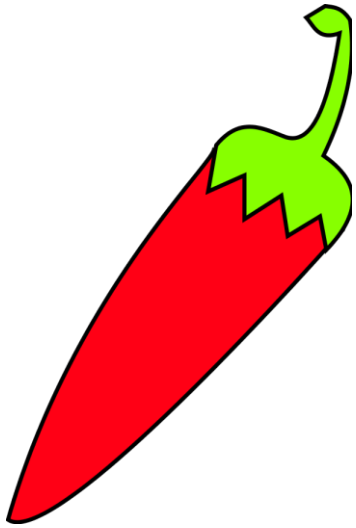
ค่าเฉลี่ยการใส่พริกป่นของเพศหญิงแตกต่างจาก 0 อย่างมีนัยสำคัญ

$b = .002$, $t(398) = .916$, $p = .36$

เพศชายและเพศหญิงใส่พริกป่นไม่แตกต่างกันอย่างมีนัยสำคัญ

การทำนายด้วยตัวแปรจัดกลุ่ม

ทำนายด้วยเพศ
เพียงตัวแปรเดียว



Independent *t*-test: Sex → Chili

Group Statistics

	Sex	N	Mean	Std. Deviation	Std. Error Mean
Chili_mg	Female	200	6.9560	5.53732	.39155
	Male	200	7.4494	5.23180	.36994

ค่าเฉลี่ยเท่ากับที่คำนวณจากสมการทำนาย

Independent Samples Test

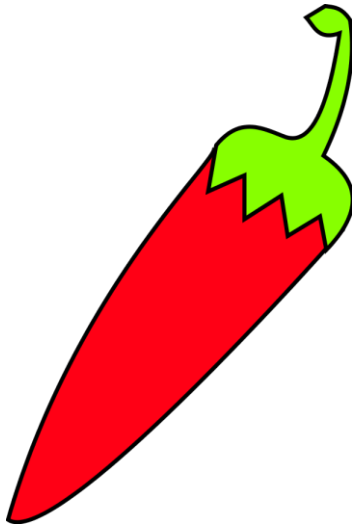
		Levene's Test for Equality of Variances				
		F	Sig.	t	df	Sig. (2-tailed)
Chili_mg	Equal variances assumed	1.826	.177	-.916	398	.360
	Equal variances not assumed			-.916	396.725	.360

ค่า *t* เหมือนกับค่าในการทดสอบความชัน

การทำนายด้วยตัวแปรจัดกลุ่ม

Regression: North, Center, Northeast → Chili

ทำนายด้วยภาค
เพียงตัวแปรเดียว



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.795[a]	.633	.630	3.27698

a. Predictors: (Constant), Northeast, Center, North

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7320.535	3	2440.178	227.234	.000[b]
	Residual	4252.494	396	10.739		
	Total	11573.029	399			

a. Dependent Variable: Chili_mg

b. Predictors: (Constant), Northeast, Center, North

$R^2 = .633$, $F(3, 396) = 227.23$, $p < .001$

ภาคเป็นปัจจัยทำนายการใส่พริกป่นอย่างมีนัยสำคัญ

การทำนายด้วยตัวแปรจัดกลุ่ม

Regression: North, Center, Northeast → Chili

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.942	.328		42.544	.000
	North	-10.875	.463	-.875	-23.466	.000
	Center	-9.951	.463	-.801	-21.472	.000
	Northeast	-6.130	.463	-.493	-13.226	.000

a. Dependent Variable: Chili_mg

$$\hat{Y} = 13.942 - 10.875(\text{North}) - 9.951(\text{Center}) - 6.130(\text{NE})$$

$$\bar{Y}_{\text{North}} = 13.942 - 10.875 = 3.067$$
$$\bar{Y}_{\text{Center}} = 13.942 - 9.951 = 3.991$$
$$\bar{Y}_{\text{Northeast}} = 13.942 - 6.130 = 7.812$$
$$\bar{Y}_{\text{South}} = 13.942$$

การทำนายด้วยตัวแปรจัดกลุ่ม

Regression: North, Center, Northeast → Chili

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.942	.328		42.544	.000
	North	-10.875	.463	-.875	-23.466	.000
	Center	-9.951	.463	-.801	-21.472	.000
	Northeast	-6.130	.463	-.493	-13.226	.000

a. Dependent Variable: Chili_mg

$a = 13.942$, $p < .001$ — ค่าเฉลี่ยการใส่พริกป่นของภาคใต้แตกต่างจาก 0 อย่างมีนัยสำคัญ

$b_1 = -10.875$, $p < .001$ — ภาคเหนือกินเผ็ดน้อยกว่าภาคใต้อย่างมีนัยสำคัญ

$b_2 = -9.951$, $p < .001$ — ภาคกลางกินเผ็ดน้อยกว่าภาคใต้อย่างมีนัยสำคัญ

$b_3 = -6.130$, $p < .001$ — ภาคอีสานกินเผ็ดน้อยกว่าภาคใต้อย่างมีนัยสำคัญ

การทำนายด้วยตัวแปรจัดกลุ่ม

One-way ANOVA: Region → Chili

Descriptives

Chili_mg

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1.00	100	3.0667	2.67541	.26754	2.5358	3.5976	.00	11.17
2.00	100	3.9908	3.01594	.30159	3.3924	4.5892	.00	10.88
3.00	100	7.8119	3.90368	.39037	7.0373	8.5865	.00	17.78
4.00	100	13.9415	3.38556	.33856	13.2697	14.6133	6.23	24.94
Total	400	7.2027	5.38564	.26928	6.6733	7.7321	.00	24.94

ค่าเฉลี่ยเหมือนกับค่าที่ทำนายได้จากสมการถดถอย

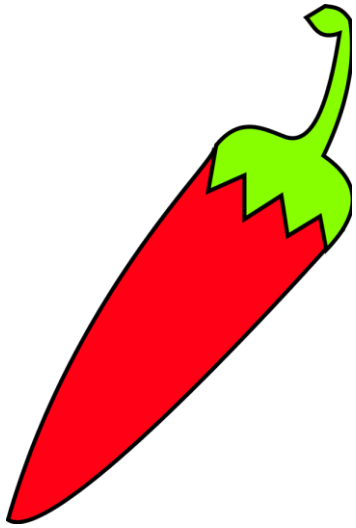
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7320.535	3	2440.178	227.234	.000
Within Groups	4252.494	396	10.739		
Total	11573.029	399			

ค่า F เหมือนกับค่าในการทดสอบ R^2

การทำนายด้วยตัวแปรจัดกลุ่ม

Regression: Male, North, Center, NE, Age → Chili

ทำนายด้วยตัวแปรอิสระทั้งหมด



Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.800[a]	.640	.636	3.25102

a. Predictors: (Constant), Northeast, Male, Age, North, Center

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	7408.781	5	1481.756	140.196	.000[b]
	Residual	4164.248	394	10.569		
	Total	11573.029	399			

a. Dependent Variable: Chili_mg

b. Predictors: (Constant), Northeast, Male, Age, North, Center

$R^2 = .64$, $F(5, 394) = 140.196$, $p < .001$

การทำนายด้วยตัวแปรจัดกลุ่ม

Regression: Male, North, Center, NE, Age → Chili

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	14.896	.609		24.465	.000
	Age	-.030	.012	-.074	-2.459	.014
	Male	.499	.325	.046	1.535	.126
	North	-10.891	.460	-.877	-23.686	.000
	Center	-9.898	.460	-.797	-21.504	.000
	Northeast	-6.139	.460	-.494	-13.351	.000

a. Dependent Variable: Chili_mg

$$\hat{Y} = 14.896 - 0.03(\text{Age}) + 0.499(\text{Male}) - 10.891(\text{North}) - 9.898(\text{Center}) - 6.139(\text{NE})$$

ในที่นี้ การหาค่าเฉลี่ยของกลุ่ม (เพศหรือภาค) จะไม่ได้เป็นค่าเฉลี่ยโดยรวมแล้ว แต่หากจะเป็นค่าเฉลี่ยในแต่ระดับของค่าตัวแปรอื่น

$$\hat{Y} = 14.896 - 0.03(\text{Age}) + 0.499(\text{Male}) - 10.891(\text{North}) - 9.898(\text{Center}) - 6.139(\text{NE})$$

เช่น หาค่าเฉลี่ยของแต่ละภาค เมื่ออายุเท่ากับ 40 ปี ในเพศชาย

$$\hat{Y} = 14.896 - 0.03(40) + 0.499(1) - 10.891(\text{North}) - 9.898(\text{Center}) - 6.139(\text{NE})$$

$$\hat{Y} = 14.195 - 10.891(\text{North}) - 9.898(\text{Center}) - 6.139(\text{NE})$$

$$\bar{Y}_{\text{North}} = 14.195 - 10.891 = 3.304$$
$$\bar{Y}_{\text{Center}} = 14.195 - 9.898 = 4.297$$
$$\bar{Y}_{\text{Northeast}} = 14.195 - 6.139 = 8.056$$
$$\bar{Y}_{\text{South}} = 14.195$$

ตรงนี้เป็นค่าเฉลี่ยของแต่ละภาค ในกลุ่มเพศชายอายุ 40 ปี

การทำนายด้วยตัวแปรจัดกลุ่ม

- ค่าเฉลี่ยที่คำนวณได้จากสมการถดถอย เมื่อแทนค่าตัวแปรอื่น อาจเรียกว่า ค่าเฉลี่ยที่ปรับแล้ว (Adjusted means)
- แตกต่างกับ ค่าเฉลี่ยที่หาจากคะแนนของแต่ละกลุ่มโดยตรง ที่เรียกว่า ค่าเฉลี่ยที่ยังไม่ได้ปรับ (Unadjusted means)

Regression: Male, North, Center, NE, Age → Chili

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	14.896	.609		24.465	.000
	Age	-.030	.012	-.074	-2.459	.014
	Male	.499	.325	.046	1.535	.126
	North	-10.891	.460	-.877	-23.686	.000
	Center	-9.898	.460	-.797	-21.504	.000
	Northeast	-6.139	.460	-.494	-13.351	.000

a. Dependent Variable: Chili_mg

$a = 14.896$, $p < .001$ — ค่าเฉลี่ยการใส่พริกป่นของหญิงใต้ที่อายุเท่ากับ 0 แตกต่างจาก 0 อย่างมีนัยสำคัญ

$b_{AGE} = -0.03$, $p = .014$ — ผลของอายุเมื่อควบคุมเพศและภาค มีผลอย่างมีนัยสำคัญ

$b_{MALE} = 0.499$, $p = .126$ — เพศชายและหญิงกินเผ็ดไม่แตกต่างกันอย่างมีนัยสำคัญ เมื่อควบคุมอายุและภาค

$b_{NORTH} = -10.891$, $p < .001$ — ภาคเหนือกินเผ็ดน้อยกว่าภาคใต้อย่างมีนัยสำคัญ เมื่อควบคุมอายุและเพศ

$b_{CENTER} = -9.898$, $p < .001$ — ภาคกลางกินเผ็ดน้อยกว่าภาคใต้อย่างมีนัยสำคัญ เมื่อควบคุมอายุและเพศ

$b_{NE} = -6.139$, $p < .001$ — ภาคอีสานกินเผ็ดน้อยกว่าภาคใต้อย่างมีนัยสำคัญ เมื่อควบคุมอายุและเพศ

การทำนายด้วยตัวแปรจัดกลุ่ม

- การจัดการตัวแปรจัดอันดับ (Ordinal scale) นั้น สามารถทำได้ 2 รูปแบบ
  - ใส่ตัวเลข แทนแต่ละอันดับ แล้วนำตัวเลขใช้เป็นตัวแปรอิสระเลย

ความพึงพอใจในสินค้า	น้อยมาก	น้อย	ปานกลาง	สูง	สูงมาก
ตัวเลขที่ใช้แทน	1	2	3	4	5

ตัวเลขที่ใช้ ขึ้นอยู่กับวิจารณญาณของผู้วิจัย

  - ไม่สนใจอันดับ แปลงเป็นตัวแปรดัมมี่เลย
- ถ้าเป็นไปได้ ให้ใช้วิธีแรก แต่หากไม่แน่ใจว่าการใส่ตัวเลขแทนถูกต้องหรือไม่ ให้ใช้การแปลงตัวแปรดัมมี่

การวิเคราะห์ถดถอยแบบลำดับขั้น

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

การวิเคราะห์ถดถอยแบบลำดับขั้น

- ตัวแปรอิสระไม่จำเป็นต้องใส่พร้อมกันทีเดียว อาจจะใส่ทีละชุดได้
- เช่น นักวิจัยต้องการทดสอบอิทธิพลของบุคลิกภาพ ต่อผลการปฏิบัติงาน แต่ทราบว่าอายุงานและระดับการศึกษามีผลต่อผลการปฏิบัติงานด้วย
  - ขั้นที่ 1 ใส่อายุงานและระดับการศึกษาเป็นตัวแปรอิสระ
  - ขั้นที่ 2 ใส่บุคลิกภาพเป็นตัวแปรอิสระเพิ่มเติม
- วิธีการแบบนี้ ทำให้ผู้วิจัยสามารถเปรียบเทียบสัมประสิทธิ์การทำนายที่เพิ่มขึ้นได้
- วิธีการแบบนี้ จะเรียกว่า การวิเคราะห์ถดถอยแบบลำดับขั้น (Hierarchical Regression, Sequential Regression)

การวิเคราะห์ถดถอยแบบลำดับขั้น

ขั้นที่ 1

ตัวแปรชุดที่ 1 → ตัวแปรตาม

R_A^2

ขั้นที่ 2

ตัวแปรชุดที่ 1 → ตัวแปรตาม

ตัวแปรชุดที่ 2 → ตัวแปรตาม

R_{A+B}^2

ΔR^2

ใช้ *F* test ทดสอบว่าตัวแปรที่ใส่เข้าไปมีผลทำนายตัวแปรตามเพิ่มขึ้นหรือไม่

$H_0: \Delta P^2 = 0$

$H_1: \Delta P^2 > 0$

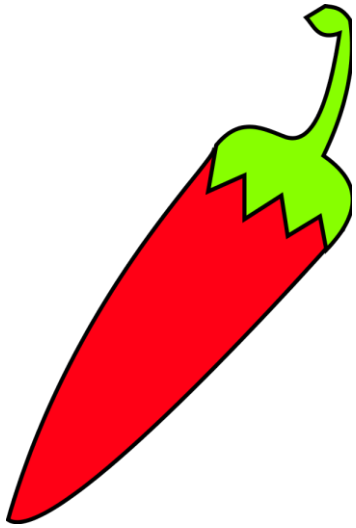
การวิเคราะห์ถดถอยแบบลำดับขั้น

ประโยชน์ของการวิเคราะห์ถดถอยแบบลำดับขั้น

- ผู้วิจัยสามารถใส่ตัวแปรควบคุม หรือตัวแปรที่มีงานวิจัยสนับสนุนจำนวนมากแล้วในขั้นที่ 1 หลังจากนั้น ให้ใส่ตัวแปรที่ผู้วิจัยสนใจทดสอบในขั้นที่ 2
  - เช่น เชาวน์ปัญญาเป็นตัวแปรอิสระที่ใช้ในการคัดเลือกบุคลากรเพื่อเข้าทำงาน ผู้วิจัยสร้างวิธีการสัมภาษณ์ใหม่ขึ้นมา เพื่อดูว่าสามารถทำนายได้มากขึ้นกว่าเดิมหรือไม่
- หากผู้วิจัย มีตัวแปรอิสระที่สามารถจับกลุ่มเป็นชุดเดียวกัน แล้วต้องการทดสอบว่าตัวแปรอิสระชุดนี้มีอิทธิพลต่อตัวแปรตามหรือไม่ หากควบคุมตัวแปรอิสระอื่น
  - เช่น ตัวแปรบุคลิกภาพ 5 องค์ประกอบ ตัวแปรดัมมี่จากตัวแปรนามบัญญัติเดียวกัน

การวิเคราะห์ถดถอยแบบลำดับขั้น

ต้องการทดสอบว่าภาคถิ่นกำเนิด สามารถทำนายการทานอาหารเผ็ดเพิ่มขึ้นจากเพศและอายุหรือไม่



สมมติฐาน คือ ภาคสามารถอธิบายความแปรปรวนได้สูงขึ้น

H_0: $\Delta P^2 = 0$

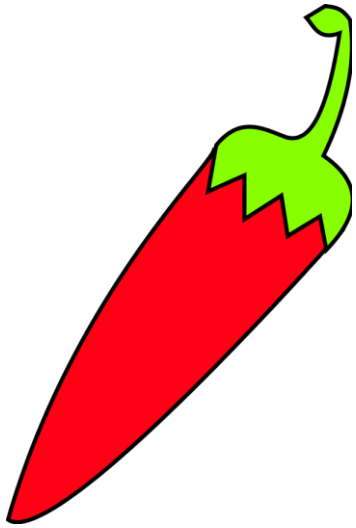
H_1: $\Delta P^2 > 0$

ขั้นที่ 1: เพศ, อายุ

ขั้นที่ 2: เพศ, อายุ, และภาค

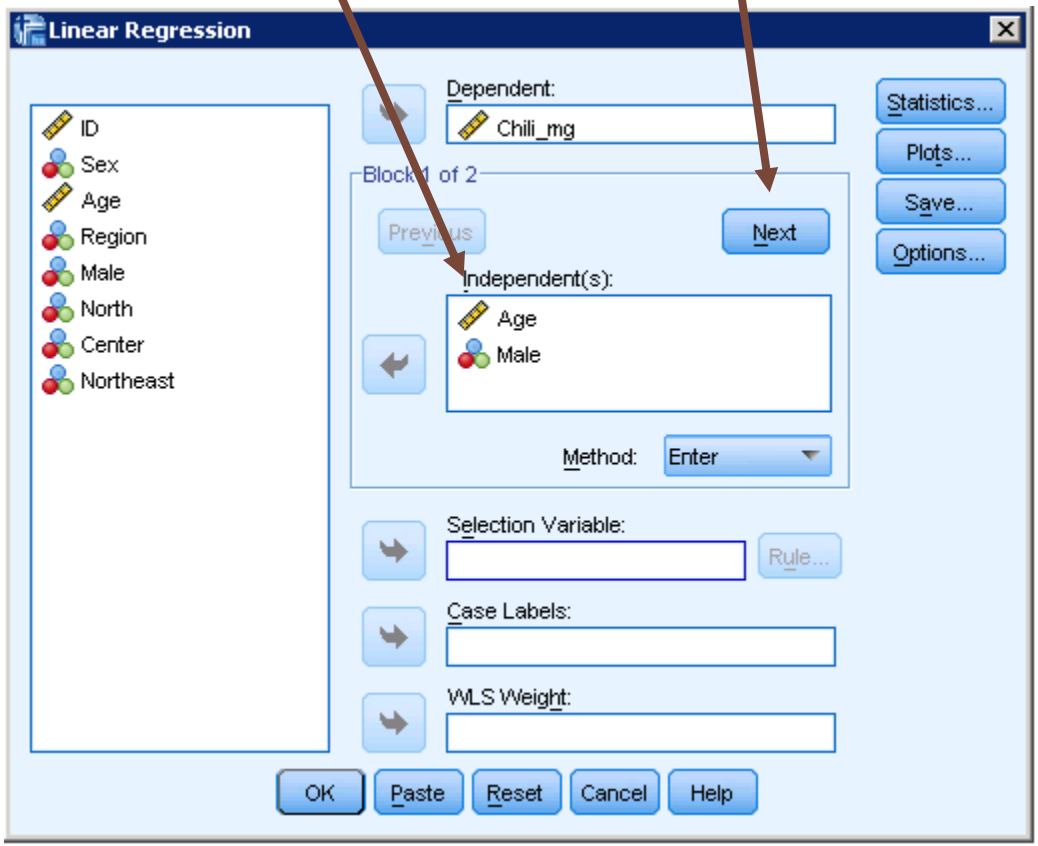
การวิเคราะห์ถดถอยแบบลำดับขั้น

ต้องการทดสอบว่าภาคถิ่นกำเนิด สามารถทำนายการทานอาหารเผ็ดเพิ่มขึ้นจากเพศและอายุหรือไม่



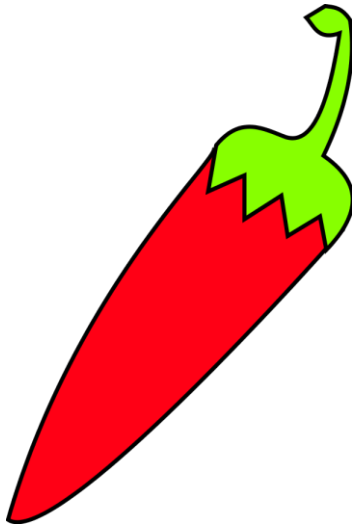
ใส่ตัวแปรอิสระชุดที่ 1

กด Next

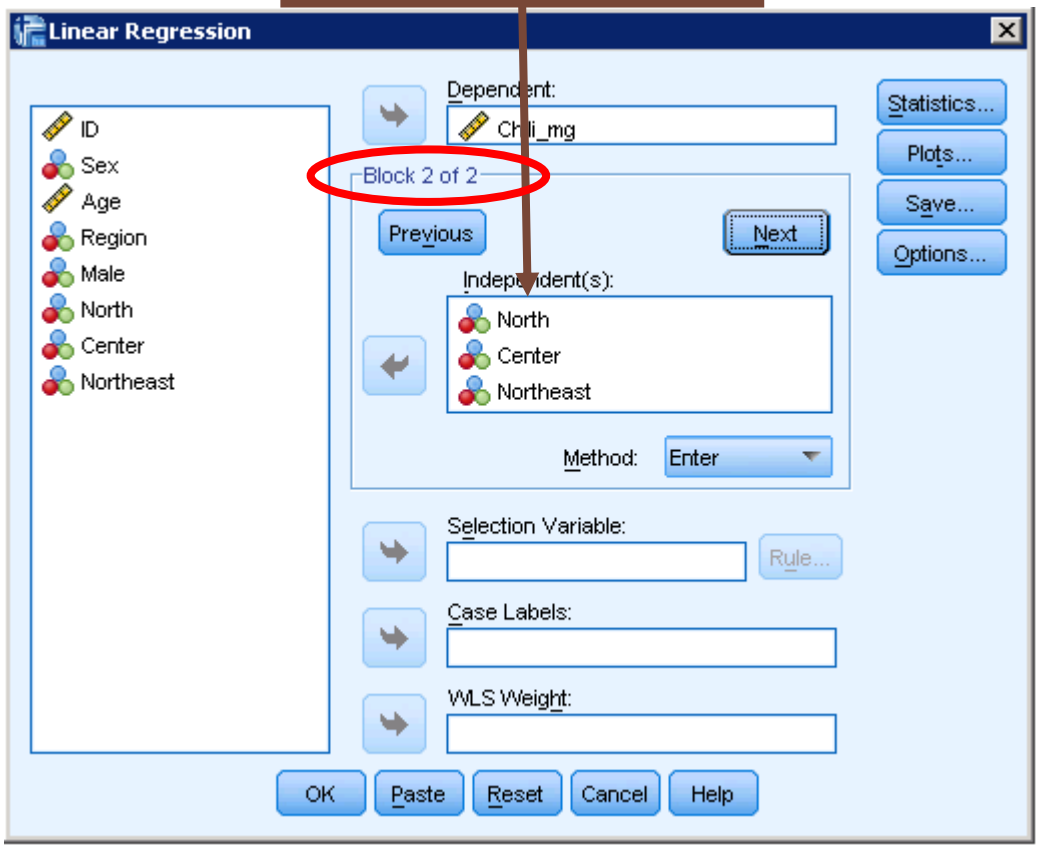


การวิเคราะห์ถดถอยแบบลำดับขั้น

ต้องการทดสอบว่าภาคถิ่นกำเนิด สามารถทำนายการทานอาหารเผ็ดเพิ่มขึ้นจากเพศและอายุหรือไม่

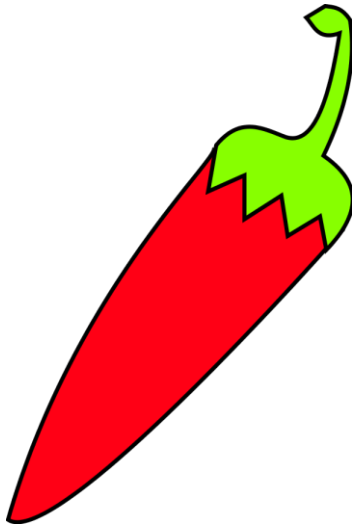


ใส่ตัวแปรอิสระชุดที่ 2

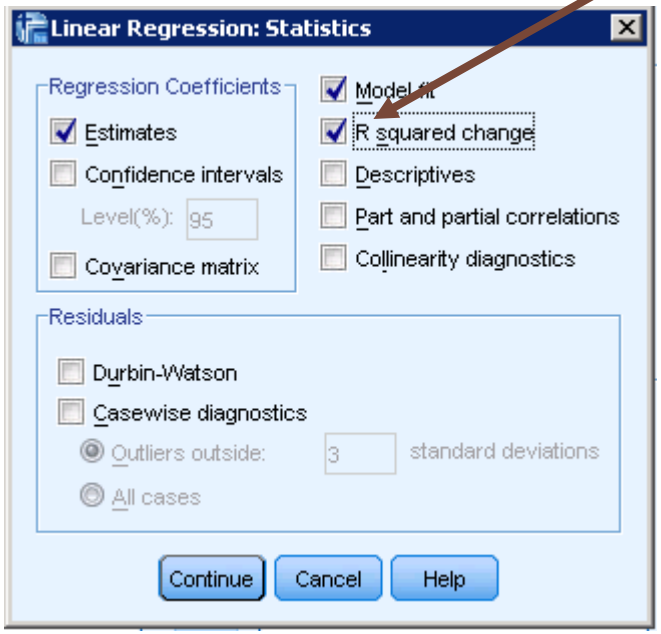


การวิเคราะห์ถดถอยแบบลำดับขั้น

ต้องการทดสอบว่าภาคถิ่นกำเนิด สามารถทำนายการทานอาหารเผ็ดเพิ่มขึ้นจากเพศและอายุหรือไม่



เลือกเพื่อทดสอบ ΔR^2



การวิเคราะห์ถดถอยแบบลำดับขั้น

ผลของโมเดลที่ 1 ที่ใส่อายุและเพศ

ΔR^2 เปรียบเทียบเมื่อไม่มีตัวแปรกับใส่ตัวแปรชุดที่ 1

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.098[a]	.010	.005	5.37318	.010	1.926	2	397	.147
2	.800[b]	.640	.636	3.25102	.631	230.153	3	394	.000

ผลของโมเดลที่ 2 ที่ใส่อายุ, เพศ, และภาค

ΔR^2 เปรียบเทียบเมื่อใส่ตัวแปรชุดที่ 1 และใส่ตัวแปรชุดที่ 1 และชุดที่ 2

ΔR^2 = .63, $F(3, 394) = 230.15$, $p < .001$

ภาคสามารถอธิบายการกินเผ็ดได้เพิ่มเติมจากอายุและเพศอย่างมีนัยสำคัญ

การวิเคราะห์ถดถอยแบบลำดับขั้น

ผลการทดสอบ R^2 เมื่อใส่ตัวแปรชุดที่ 1

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	111.224	2	55.612	1.926	.147[b]
	Residual	11461.805	397	28.871		
	Total	11573.029	399			
2	Regression	7408.781	5	1481.756	140.196	.000[c]
	Residual	4164.248	394	10.569		
	Total	11573.029	399			

a. Dependent Variable: Chili_mg

b. Predictors: (Constant), Male, Age

c. Predictors: (Constant), Male, Age, Northeast, North, Center

ผลการทดสอบ R^2 เมื่อใส่ตัวแปรชุดที่ 2

การวิเคราะห์ถดถอยแบบลำดับขั้น

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	8.361	.895		9.345	.000
	Age	-.035	.020	-.087	-1.735	.084
	Male	.500	.537	.046	.931	.353
2	(Constant)	14.896	.609		24.465	.000
	Age	-.030	.012	-.074	-2.459	.014
	Male	.499	.325	.046	1.535	.126
	North	-10.891	.460	-.877	-23.686	.000
	Center	-9.898	.460	-.797	-21.504	.000
	Northeast	-6.139	.460	-.494	-13.351	.000

a. Dependent Variable: Chili_mg

ผลของ Model 2 เหมือนกับผลที่ใส่ตัวแปรทุกตัวพร้อมกัน

$$\text{Model 1: } \hat{Y} = 8.361 - 0.035(\text{Age}) + 0.500(\text{Male})$$

$$\text{Model 2: } \hat{Y} = 14.896 - 0.030(\text{Age}) + 0.499(\text{Male}) - 10.891(\text{North}) - 9.898(Center) - 6.139(\text{NE})$$

การวิเคราะห์ถดถอยแบบลำดับขั้น

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.361	.895		9.345	.000
	Age	-.035	.020	-.087	-1.735	.084
	Male	.500	.537	.046	.931	.353
2	(Constant)	14.896	.609		24.465	.000
	Age	-.030	.012	-.074	-2.459	.014
	Male	.499	.325	.046	1.535	.126
	North	-10.891	.460	-.877	-23.686	.000
	Center	-9.898	.460	-.797	-21.504	.000
	Northeast	-6.139	.460	-.494	-13.351	.000

a. Dependent Variable: Chili_mg

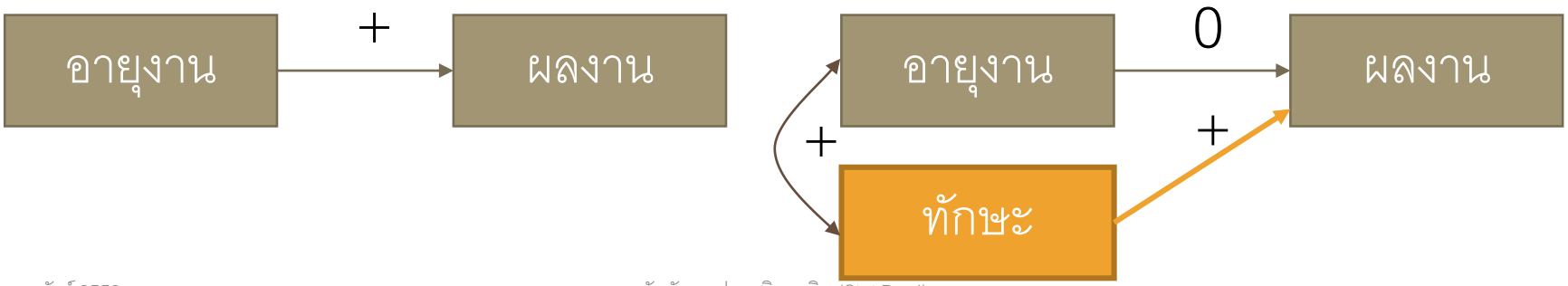
ผู้วิจัยบางคน แปลความหมายค่าสัมประสิทธิ์ถดถอยที่เปลี่ยนแปลง

ค่าความชันมากขึ้น → เกิดผลของการกด (Suppression Effect)

ค่าความชันลดลง → อิทธิพลของตัวแปรลดลง เป็นเหตุการณ์ที่เกิดตามปกติ

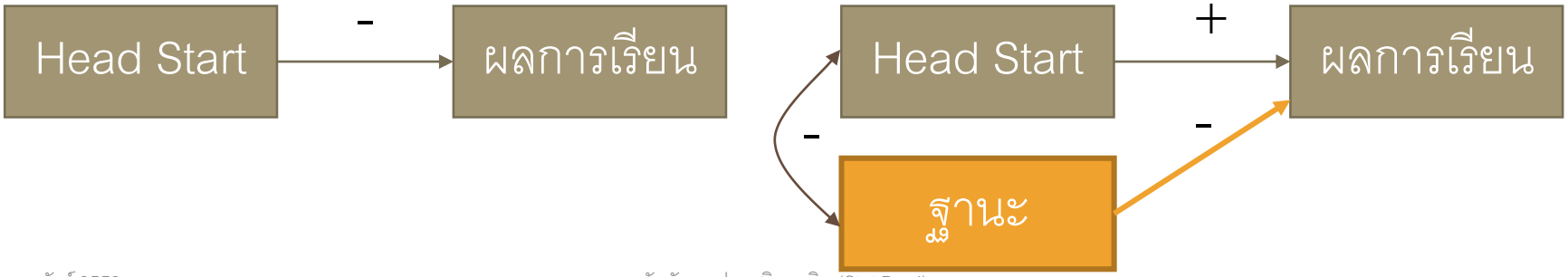
การวิเคราะห์ถดถอยแบบลำดับขั้น

- ค่าของความชันลดลง เมื่อมีตัวแปรอื่นร่วมในการถดถอย เป็นเหตุการณ์ที่เกิดขึ้นเป็นปกติในการวิเคราะห์ถดถอย
  - เช่น ท่านพบว่าอายุงานมีผลต่อผลการปฏิบัติงาน
  - แต่คนที่อายุงานมาก มักจะมีทักษะในการทำงานมากกว่า
  - ดังนั้น หากทำนายผลการปฏิบัติงาน ด้วยอายุงานและทักษะพร้อมกัน ค่าสัมประสิทธิ์ถดถอยของอายุงานน้อยลง (จนบางครั้ง ไม่ถึงระดับนัยสำคัญ)



การวิเคราะห์ถดถอยแบบลำดับขั้น

- ผลของการกด (Suppression Effect) คือ ผลของตัวแปรเด่นชัดขึ้น ถ้ามีตัวแปรอื่นร่วมในการถดถอย
  - เช่น คนที่เข้าร่วมโปรแกรม Head Start มีแนวโน้มที่จะมีผลการเรียนน้อยกว่าคนที่ไม่เข้าร่วม (ดูแต่เพียงค่าสหสัมพันธ์ระหว่าง 2 ตัวแปร)
  - แต่คนที่เข้าร่วมโปรแกรม Head Start ก็มักจะมีสถานะทางเศรษฐกิจไม่ดี
  - ดังนั้น ผลของโปรแกรม Head Start มีผลทางบวกต่อผลการเรียน หากควบคุมสถานทางเศรษฐกิจ



การวิเคราะห์ถดถอยแบบลำดับขั้น

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	8.361	.895		9.345	.000
	Age	-.035	.020	-.087	-1.735	.084
	Male	.500	.537	.046	.931	.353
2	(Constant)	14.896	.609		24.465	.000
	Age	-.030	.012	-.074	-2.459	.014
	Male	.499	.325	.046	1.535	.126
	North	-10.891	.460	-.877	-23.686	.000
	Center	-9.898	.460	-.797	-21.504	.000
	Northeast	-6.139	.460	-.494	-13.351	.000

a. Dependent Variable: Chili_mg

ในที่นี้ ความชันไม่ได้ปลี่ยนแปลง จนอธิบายผลของอายุและเพศเพิ่มเติมได้

การวิเคราะห์ถดถอยแบบลำดับขั้น

- แท้จริงแล้ว หากตัวแปรที่ใส่เพิ่มในขั้นที่ 2 มีเพียงแค่ตัวแปรเดียว ค่า *t* ในการทดสอบความชันของตัวแปรนั้น หากนำมายกกำลังสองแล้วจะเท่ากับค่า *F* เลย และค่า *p* ที่ได้จะเท่ากันด้วย
- ด้วยเหตุนี้ การวิเคราะห์ถดถอยแบบลำดับขั้น จะใช้ก็ต่อเมื่อ ตัวแปรอิสระที่เพิ่มมีมากกว่า 1 ตัว
- ในบางกรณี ผู้วิจัยต้องการใส่ตัวแปรอิสระมากกว่า 2 ลำดับขั้น เช่น
  - ขั้นที่ 1 ใส่ตัวแปรเพศลงไป เพื่อทำนายความเผ็ด
  - ขั้นที่ 2 ใส่ตัวแปรอายุลงไปทำนายเพิ่มเติม
  - ขั้นที่ 3 ใส่ตัวแปรแสดงปฏิสัมพันธ์ระหว่างเพศและอายุลงไป

การวิเคราะห์ถดถอยแบบลำดับขั้น (การเขียนรายงาน)

- งานวิจัยนี้ต้องการทดสอบว่าเพศ (ชาย 50%, หญิง 50%), อายุ (M = 39.73 ปี, SD = 13.16 ปี), และภาค (เหนือ 25%, กลาง 25%, อีสาน 25%, ใต้ 25%) เป็นปัจจัยในการทำนายการกินเผ็ดของคนไทย (M = 7.2 กรัม, SD = 5.39 กรัม) หรือไม่ จึงเก็บกลุ่มตัวอย่างจำนวน 400 คน และวิเคราะห์ด้วยการวิเคราะห์ถดถอยแบบพหุ โดยแปลงเพศและภาค เป็นตัวแปรแบบดัมมี่ ตารางที่ 1 แสดงค่าสหสัมพันธ์ระหว่างตัวแปรในการวิเคราะห์ถดถอย ผลการทดสอบพบว่าตัวแปรทั้งหมดสามารถทำนายการกินเผ็ดได้อย่างมีนัยสำคัญ, $R^2 = .64$, $F(5, 364) = 140.2$, $p < .001$

การวิเคราะห์ถดถอยแบบลำดับขั้น (การเขียนรายงาน)

- ตารางที่ 1 แสดงค่าสหสัมพันธ์ระหว่างตัวแปรในการวิเคราะห์ถดถอย (N = 400)

ตัวแปร	1	2	3	4	5	6
1. การกินเผ็ด						
2. อายุ	-.086*					
3. เพศชาย	.046	.007				
4. ภาคเหนือ	-.444**	-.034	.000			
5. ภาคกลาง	-.345**	.067	.000	-.333**		
6. ภาคอีสาน	.065	-.023	.000	-.333**	-.333**	
M	7.20	39.74	0.50	0.25	0.25	0.25
SD	5.39	13.16	0.50	0.43	0.43	0.43

หมายเหตุ *$p < .05$, **$p < .01$, กลุ่มอ้างอิงคือเพศหญิง และภาคใต้

การวิเคราะห์ถดถอยแบบลำดับขั้น (การเขียนรายงาน)

- โดยอายุเป็นปัจจัยทางลบอย่างมีนัยสำคัญ, b = -0.03, p = .014, ปัจจัยเพศไม่ถึงระดับนัยสำคัญ, b = 0.50, p = .126, ส่วนภาค เนื่องจากว่าต้องทดสอบตัวแปรดัมมี่ 3 ตัวแปรพร้อมกัน จึงทดสอบด้วยการวิเคราะห์ถดถอยแบบลำดับขั้น พบว่าภาคเป็นปัจจัยทำนายการกินเผ็ดอย่างมีนัยสำคัญ, ΔR^2 = .64, $F(5, 364) = 140.2$, $p < .001$, โดยภาคเหนือ, ภาคกลาง, และภาคอีสานกินเผ็ดน้อยกว่าภาคใต้อย่างมีนัยสำคัญ ($p < .001$) ผลการวิเคราะห์ถดถอยแสดงไว้ดังตารางที่ 2

การวิเคราะห์ถดถอยแบบลำดับขั้น (การเขียนรายงาน)

- ตารางที่ 2 ผลการวิเคราะห์ถดถอยแบบลำดับขั้น ในการทำนายการกินเผ็ดของกลุ่มตัวอย่าง

ตัวแปร	ขั้นที่ 1		ขั้นที่ 2	
	b	β	*b*	β
ค่าคงที่	8.361		14.896	
อายุ	-0.035	-.09	-0.030*	-.07
เพศชาย	0.500	.05	0.499	.05
ภาคเหนือ			-10.891**	-.88
ภาคกลาง			-9.898**	-.80
ภาคอีสาน			-6.139**	-.49
R^2	.01		.64**	
R^2 ที่เพิ่มขึ้น			.63**	

หมายเหตุ *$p < .05$, **$p < .01$

ความสัมพันธ์บางส่วน

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

ความสัมพันธ์บางส่วน

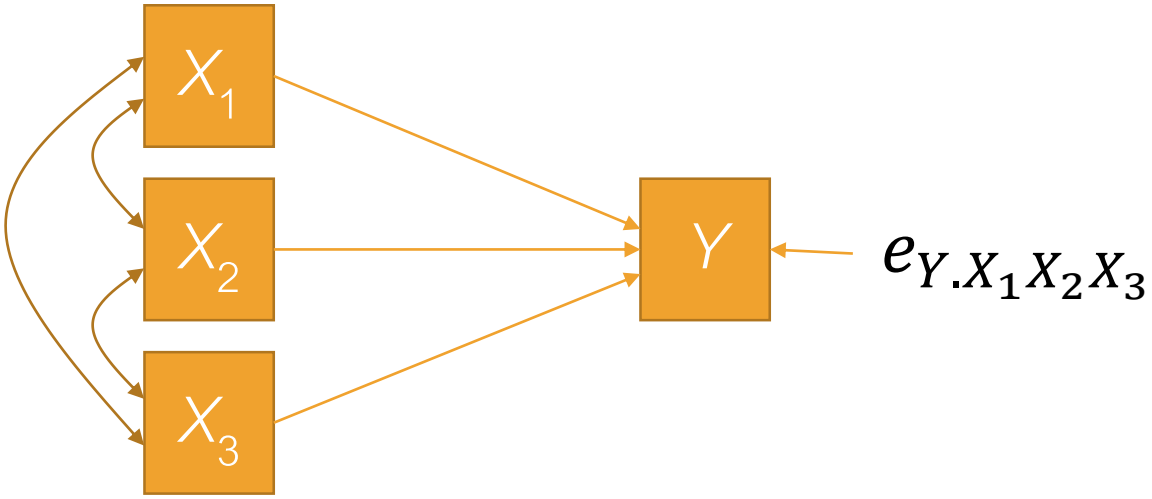
- ในกรณีที่มีตัวแปรอิสระมากกว่า 1 ตัวแปร ค่าสัมประสิทธิ์สหสัมพันธ์ (R) หรือสัมประสิทธิ์การทำนาย (R^2) บอกอำนาจทำนายของตัวแปรอิสระทั้งหมด ไม่สามารถบอกความสำคัญของตัวแปรแต่ละตัวได้
- นอกเหนือจาก ค่าสัมประสิทธิ์ถดถอยแบบมาตรฐาน (β) เราสามารถหาความสัมพันธ์ระหว่างตัวแปรอิสระตัวแปรหนึ่งกับตัวแปรตาม เมื่อควบคุมตัวแปรอิสระอื่นให้คงที่ได้

ความสัมพันธ์บางส่วน

- ยกตัวอย่างเช่น นักวิจัยต้องการทำนายตัวแปรตามด้วยตัวแปรต้น 3 ตัว

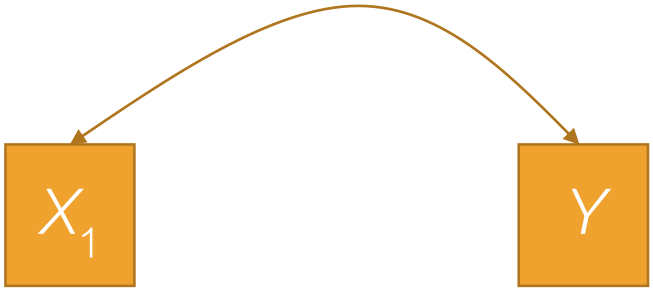
เขียนสมการในรูปค่า Y ที่แท้จริง

$$Y = a + b_1X_1 + b_2X_2 + b_3X_3 + e_{Y.X_1X_2X_3}$$



ความสัมพันธ์บางส่วน

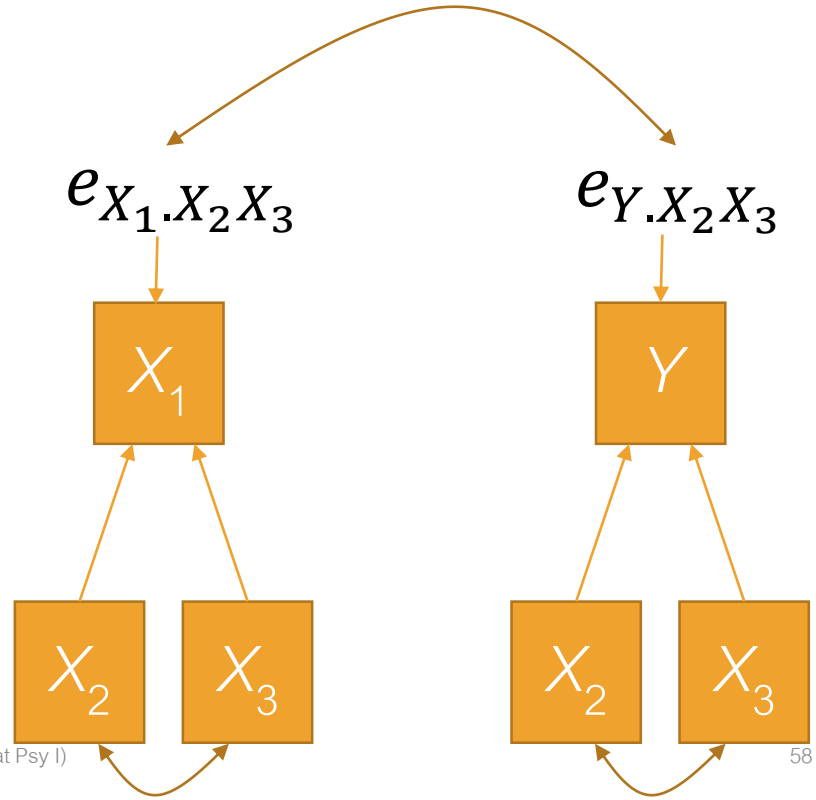
- ความสัมพันธ์แบบ Pearson หรือความสัมพันธ์แบบ Zero-order (Zero-order correlation) คือการหาความสัมพันธ์ระหว่างตัวแปรทั้งสองโดยตรง



ความสัมพันธ์บางส่วน

- ความสัมพันธ์บางส่วน (Partial Correlation) คือความสัมพันธ์ระหว่างความผิดพลาดของตัวแปรสองตัว เมื่อนำอิทธิพลของตัวแปรควบคุมออกจากตัวแปรทั้งสอง

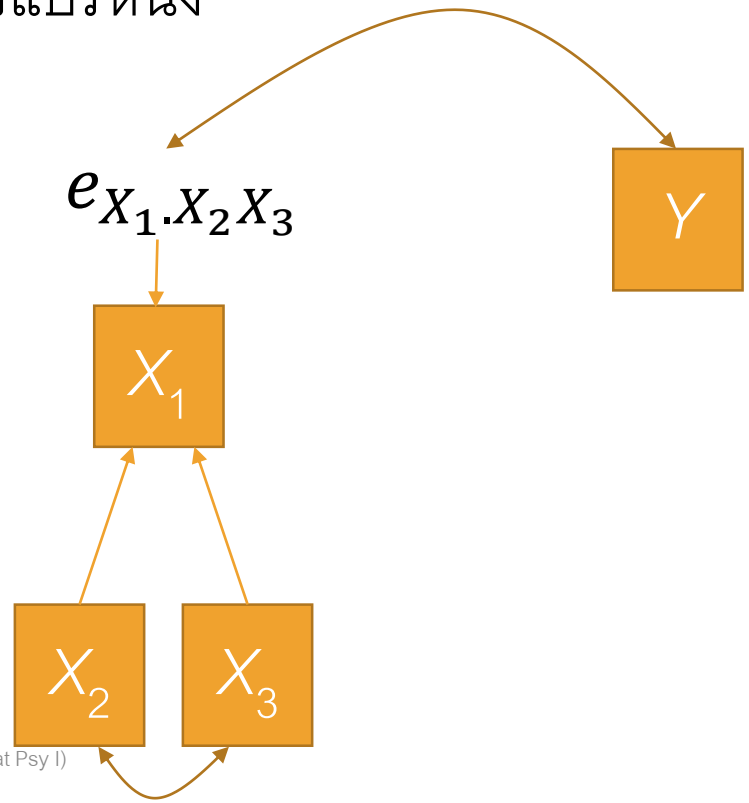
1. นำ X_1 ไปทำนายด้วย X_2 และ X_3 แล้วบันทึกความผิดพลาดในการทำนาย X_1
2. นำ Y ไปทำนายด้วย X_2 และ X_3 แล้วบันทึกความผิดพลาดในการทำนาย Y
3. หาความสัมพันธ์ของความผิดพลาดทั้งสอง



ความสัมพันธ์บางส่วน

- ความสัมพันธ์กึ่งบางส่วน (Part Correlation หรือ Semipartial Correlation) คือความสัมพันธ์ระหว่างความผิดพลาดของตัวแปรหนึ่งเมื่อนำอิทธิพลของตัวแปรควบคุมออก และคะแนนดิบของอีกตัวแปรหนึ่ง

1. นำ X_1 ไปทำนายด้วย X_2 และ X_3 แล้วบันทึกความผิดพลาดในการทำนาย X_1
2. หาความสัมพันธ์ของความผิดพลาดกับ Y



ความสัมพันธ์บางส่วน

- ความสัมพันธ์แบบ Pearson คือการมองความสัมพันธ์ระหว่างตัวแปรสองตัว โดยไม่สนค่าของตัวแปรอื่นเลย
- ความสัมพันธ์บางส่วน คือ การมองความสัมพันธ์ระหว่างตัวแปรสองตัว หากตัวแปรควบคุมคงที่
- ความสัมพันธ์กึ่งบางส่วน ไม่สามารถแปลความหมายตรงๆ ได้ แต่หากนำค่านี้มายกกำลังสอง สามารถแปลความหมายอย่างอ้อมได้ว่า หากนำตัวแปรอิสระตัวนี้ออกจากสมการถดถอยแล้ว ค่าสัมประสิทธิ์ในการทำนายจะลดลงเท่าไร
  - ซึ่งก็คือค่า ΔR^2 เปรียบเทียบระหว่างมีและไม่มีตัวแปรดังกล่าวนั่นเอง

ความสัมพันธ์บางส่วน

- ตัวอย่างเช่น ทำนายผลการปฏิบัติงาน (Y) ด้วยความรับผิดชอบ (X_1) และความสามารถ (X_2)
- ความสัมพันธ์แบบ Pearson ระหว่างความรับผิดชอบและผลการปฏิบัติงาน คือ ความสัมพันธ์ระหว่างตัวแปรสองตัว โดยไม่สนใจตัวแปรความสามารถ ราวกับว่าเราไม่ได้วัดตัวแปรความสามารถมา
- ความสัมพันธ์บางส่วน ระหว่างความรับผิดชอบและผลการปฏิบัติงาน คือ หากควบคุมระดับความสามารถให้คงที่ (เช่น ระหว่างคนที่เก่งมากๆ เหมือนกัน) ความสัมพันธ์ระหว่างความรับผิดชอบและผลการปฏิบัติงานจะมีระดับเท่าไร

ความสัมพันธ์บางส่วน

- ความสัมพันธ์กึ่งบางส่วน ระหว่างความรับผิดชอบและผลการปฏิบัติงาน หากนำค่านี้มายกกำลังสองแล้ว หมายความว่า หากนำตัวแปรความรับผิดชอบออกจากการทำนายผลการปฏิบัติงานแล้ว (เหลือความสามารถเป็นตัวแปรทำนายเพียงตัวเดียว) ค่าสัมประสิทธิ์การทำนายจะลดลงเท่าไร

ความสัมพันธ์บางส่วน

- ใน SPSS สามารถหาได้โดยเข้าที่ Statistics… แล้วเลือก Part and partial corelations
  - เช่น จากตัวอย่างเรื่องการทำนายความน่ารัก

ค่า Pearson Correlation กับตัวแปรตาม

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part	Part-squared
1	(Constant)	2.368	10.908		.217	.829				
	face	.808	.158	.667	5.104	.000	.655	.606	.563	.317
	shape	.065	.140	.059	.462	.646	.306	.069	.051	.003
	habit	.237	.168	.195	1.407	.166	.116	.205	.155	.024
	close	-.118	.162	-.105	-.726	.472	.230	-.108	-.080	.006

a. Dependent Variable: cute

ค่าสหสัมพันธ์บางส่วนกับตัวแปรตาม

ค่าสหสัมพันธ์กึ่งบางส่วนกับตัวแปรตาม

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Part-squared
		B	Std. Error	Beta			Zero-order	Partial	Part	
1	(Constant)	2.368	10.908		.217	.829				
	face	.808	.158	.667	5.104	.000	.655	.606	.563	.317
	shape	.065	.140	.059	.462	.646	.306	.069	.051	.003
	habit	.237	.168	.195	1.407	.166	.116	.205	.155	.024
	close	-.118	.162	-.105	-.726	.472	.230	-.108	-.080	.006

a. Dependent Variable: cute

ตัวอย่างการแปลความหมายสำหรับตัวแปรรูปร่าง

ค่าสหสัมพันธ์ระหว่างรูปร่างและความน่ารักอยู่ในทางบวกเล็กน้อย

หากควบคุมหน้าตา นิสัย และความใกล้ชิดให้คงที่แล้ว ค่าสหสัมพันธ์ระหว่างรูปร่างและความน่ารักจะใกล้ 0 (คือมีความสัมพันธ์กันน้อยมาก)

เมื่อนำตัวแปรรูปร่างออกจากสมการถดถอยแล้ว ค่าสัมประสิทธิ์การทำนายลดลงเพียง 0.3% เท่านั้น

การใช้โปรแกรมคัดเลือกตัวแปรอิสระ

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

การใช้โปรแกรมคัดเลือกตัวแปรอิสระ

- ในบางครั้ง ผู้วิจัยต้องการทำนายตัวแปรตาม โดยผู้วิจัยคิดว่าตัวแปรอิสระที่เป็นไปได้มีเยอะมาก และต้องการคัดเลือกตัวแปรอิสระออกมาเพื่อให้มีประสิทธิภาพในการทำนายตัวแปรตามสูงสุด
  - ใช้ตัวแปรอิสระน้อย ทำนายได้สูง
- โปรแกรมทางสถิติสามารถใช้ Computer Algorithm ในการคัดเลือกตัวแปรอิสระ โดยมีวิธีที่จะกล่าวถึง 4 วิธี คือ
  - Forward Selection
  - Backward Elimination
  - Stepwise Selection
  - All Possible Regression

การใช้โปรแกรมคัดเลือกตัวแปรอิสระ

Forward Selection

- ตัวแปรอิสระจะถูกใส่เข้าไปในโมเดลทีละตัว
- จากตัวแปรอิสระทั้งหมด ตัวที่ถูกคัดเลือกจะเป็นตัวที่มี *p* value น้อยที่สุด จากการทดสอบ *F* $(H_0: \Delta \mathrm{P}^2 = 0)$
- ตัวแปรจะถูกคัดเลือกเข้าไปเรื่อยๆ จนกระทั่งไม่มีตัวแปรใดที่มี *p* value ถึงระดับนัยสำคัญ

การใช้โปรแกรมคัดเลือกตัวแปรอิสระ

Backward Elimination

- ตัวแปรอิสระจะถูกนำเข้าไปในโมเดลทั้งหมด แล้วคัดออกทีละตัว
- จากตัวแปรอิสระทั้งหมด ตัวที่ถูกคัดเลือกจะเป็นตัวที่มี *p* value มากที่สุด จากการทดสอบ F เมื่อคัดตัวแปรนั้นทิ้ง $(H_0: \Delta \mathrm{P}^2 = 0)$
- ตัวแปรจะถูกคัดออกไปเรื่อยๆ จนกระทั่งทุกตัวแปรมี *p* value ถึงระดับนัยสำคัญ

การใช้โปรแกรมคัดเลือกตัวแปรอิสระ

Stepwise Selection

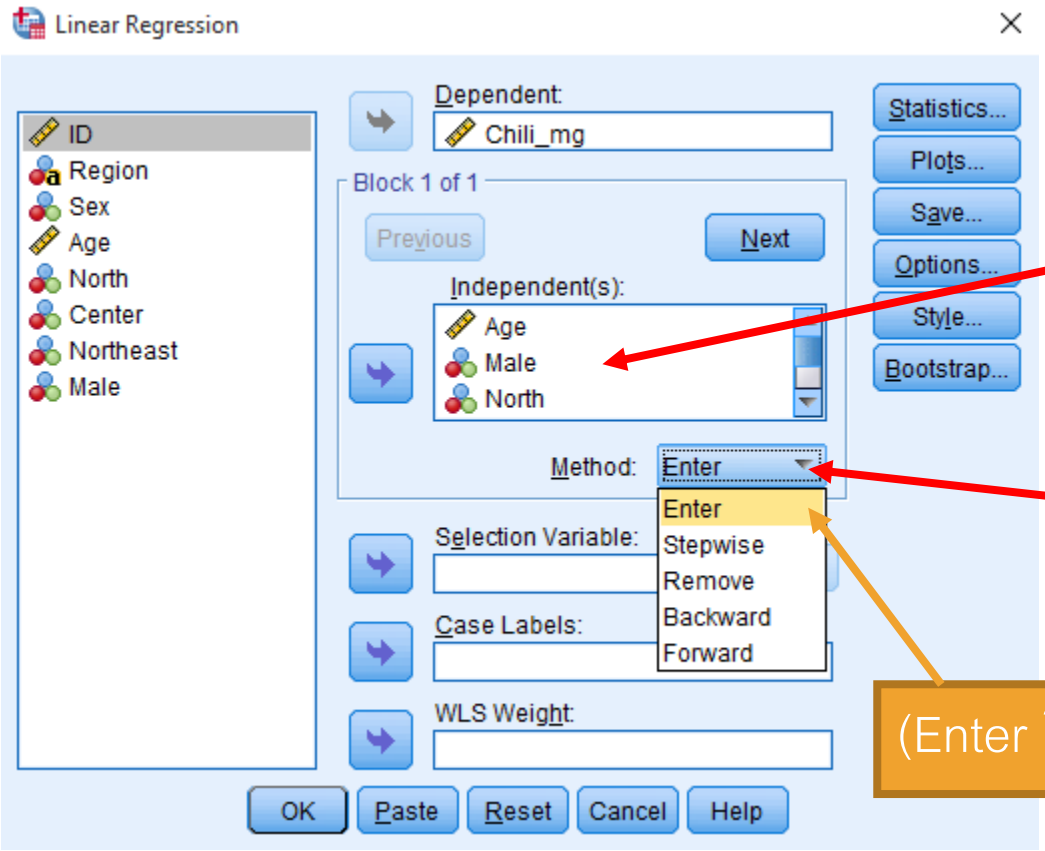
- เป็นการผสมผสานกันระหว่าง Forward Selection และ Backward Elimination
- กล่าวคือ เป็นการทำ Forward Selection แต่หลังจากที่ใส่ตัวแปรเข้าไปแล้ว จะมีการตรวจสอบว่าตัวแปรทุกตัว ยังคงถึงระดับนัยสำคัญ ถ้าไม่ใช่จะคัดทิ้ง
- หลักการแบบนี้ จะทำไปเรื่อยๆ จนกระทั่งตัวแปรอิสระทุกตัวในโมเดลถึงระดับนัยสำคัญ และตัวแปรอิสระที่ไม่ได้อยู่ในโมเดลไม่ถึงระดับนัยสำคัญ

การใช้โปรแกรมคัดเลือกตัวแปรอิสระ

All possible regression

- เป็นการหา Subset ที่เป็นไปได้ทั้งหมดในการคัดเลือกตัวแปรอิสระ
- เช่น หากตัวแปร X_1, X_2, X_3 เป็นตัวแปรอิสระ Subset ที่เป็นไปได้ทั้งหมดมี 7 รูปแบบ
  - $\{X_1\}$, $\{X_2\}$, $\{X_3\}$, $\{X_1, X_2\}$, $\{X_1, X_3\}$, $\{X_2, X_3\}$, $\{X_1, X_2, X_3\}$
- หลังจากนั้นนำผลที่ได้มาเปรียบเทียบ Adjusted R^2 ว่าชุดใดมีมากที่สุด
- หากมีตัวแปรอิสระ p ตัว จะมีความเป็นไปได้ของ Subset เท่ากับ $2^p - 1$ รูปแบบ

การใช้โปรแกรมคัดเลือกตัวแปรอิสระ



ใส่ตัวแปรอิสระ
ที่ต้องการทดสอบทั้งหมด

เลือกวิธีในการคัดเลือก
ตัวแปร

(Enter คือใส่ทั้งหมดโดยไม่มีการคัดเลือก)

การใช้โปรแกรมคัดเลือกตัวแปรอิสระ

ผลการทำวิธี Stepwise ในการคัดเลือกตัวแปร Age, Male, North, Center, Northeast

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	8.581	.279		30.761	.000
	North	-5.515	.558	-.444	-9.884	.000
2	(Constant)	10.877	.278		39.142	.000
	North	-7.810	.481	-.629	-16.227	.000
	Center	-6.886	.481	-.554	-14.307	.000
3	(Constant)	13.942	.328		42.544	.000
	North	-10.875	.463	-.875	-23.466	.000
	Center	-9.951	.463	-.801	-21.472	.000
	Northeast	-6.130	.463	-.493	-13.226	.000
4	(Constant)	15.140	.589		25.716	.000
	North	-10.891	.461	-.877	-23.646	.000
	Center	-9.898	.461	-.797	-21.468	.000
	Northeast	-6.139	.461	-.494	-13.329	.000
	Age	-.030	.012	-.074	-2.444	.015

a. Dependent Variable: Chili_mg

ตัวแปรที่ถูกคัดเลือก คือ

Age, North, Center, Northeast

การใช้โปรแกรมคัดเลือกตัวแปรอิสระ

- การคัดเลือกโดยใช้คอมพิวเตอร์มีข้อเสียจำนวนมาก จนผมคิดว่าไม่ควรใช้
  - ได้รับผลกระทบจาก Sampling error สูงมาก หากเกิดการแกว่งของค่า *b* หรือ R^2 เพียงแค่เล็กน้อย อาจทำให้ผลการคัดเลือกตัวแปรเปลี่ยนอย่างมาก ซึ่งทำให้ต้องใช้กลุ่มตัวอย่างสูงมากๆ
  - ค่า R^2 ที่ได้รับ จะมีค่าเฟ้อกว่าความเป็นจริง จากกระบวนการคัดเลือกตัวแปร
  - วิธี Forward, Backward, Stepwise จะใช้วิธีการตัดสินตาม *p* value ซึ่งอาจทำให้คัดเลือกตัวแปรที่มีค่าสหสัมพันธ์บางส่วนน้อยมากในกรณีที่มีจำนวนกลุ่มตัวอย่างสูง แต่ในกรณีที่จำนวนกลุ่มตัวอย่างน้อย ตัวแปรที่มีค่าสหสัมพันธ์บางส่วนสูงอาจถูกคัดทิ้ง

การใช้โปรแกรมคัดเลือกตัวแปรอิสระ

- หากคุณต้องใช้จริงๆ ให้ใช้วิธี All Possible Regression ที่ไม่สามารถทำได้โดยอัตโนมัติใน SPSS
- หากหลีกเลี่ยงไม่ได้จริงๆ ให้ใช้วิธี Stepwise และนำตัวแปรที่คัดเลือกมาทั้งหมดมาตรวจสอบค่าสหสัมพันธ์บางส่วนอีกทีหนึ่ง แล้วคัดตัวแปรที่มีค่าสหสัมพันธ์บางส่วนน้อย (เช่น น้อยกว่า .1) ออกทีละตัว จนไม่มีเหลือ
- ใช้จำนวนกลุ่มตัวอย่างสูงมากๆ เช่น จำนวนกลุ่มตัวอย่างอย่างน้อย 40 เท่าของจำนวนตัวแปรอิสระที่นำมาคัดเลือก

การเปลี่ยนศูนย์กลาง

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

การเปลี่ยนศูนย์กลาง

- การเปลี่ยนศูนย์กลาง (Centering) คือการนำตัวแปรอิสระไปลบด้วยค่าคงที่ เพื่อให้จุดตัดแกน *Y* มีความหมายในเชิงปฏิบัติ

$$\boxed{\hat{Y} = a + b_1X_1 + b_2X_2}$$

ค่าที่ทำนายได้ เมื่อ X_1 และ X_2 เท่ากับ 0

$$\boxed{\hat{Y} = a' + b_1(X_1 - c) + b_2(X_2 - d)}$$

ค่าที่ทำนายได้ เมื่อ $X_1 = c$ และ $X_2 = d$

การเปลี่ยนศูนย์กลาง

- เช่น ทำนายผลการปฏิบัติงาน (*M* = 50, *SD* = 10) ด้วยคะแนนเชาวน์ปัญญา (*M* = 100, *SD* = 15) และคะแนนสอบ TOEIC (*M* = 550, *SD* = 100)

$$\hat{Y} = 11.167 + 0.33(\text{IQ}) + 0.01(\text{TOEIC})$$

ค่าผลการปฏิบัติงานที่ทำนายได้ เมื่อ IQ = 0 และ TOEIC = 0

$$\hat{Y} = 49.667 + 0.33(\text{IQ} - 100) + 0.01(\text{TOEIC} - 550)$$

ค่าผลการปฏิบัติงานที่ทำนายได้ เมื่อ IQ = 100 และ TOEIC = 550

จากตรงนี้ สามารถทดสอบว่าจุดตัดแกน *Y* แตกต่างจาก 0 อย่างมีนัยสำคัญหรือไม่

การเปลี่ยนศูนย์กลาง

- เราสามารถแปลงตัวแปรอิสระได้ใน SPSS โดยใช้คำสั่ง Transform → Compute Variables … แล้วหลังจากนั้นให้นำตัวแปรอิสระไปบวกหรือลบด้วยค่าคงที่
- การเปลี่ยนศูนย์กลางจะสำคัญมากขึ้น หากผู้วิจัยต้องการใช้การวิเคราะห์ถดถอย ในการหาปฏิสัมพันธ์ระหว่างตัวแปร (Interaction) หรือการทำนายแบบไม่เป็นเส้นตรง (Nonlinear regression) ซึ่งจะไม่กล่าวถึงในวิชานี้

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

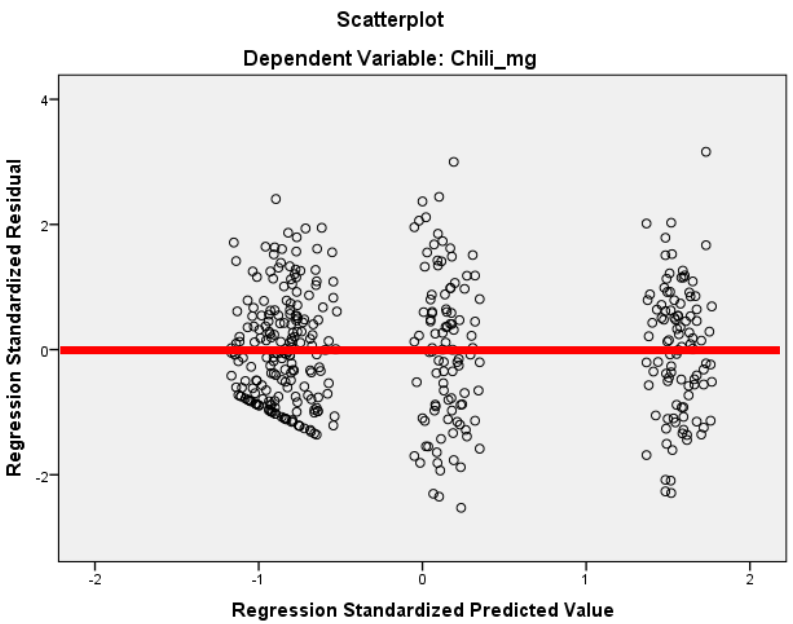
สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

- ข้อตกลงเบื้องต้นทางสถิติเกือบทั้งหมด สามารถทดสอบได้โดยการสร้างกราฟที่แกน *X* เป็นค่าที่ทำนายได้ และแกน *Y* เป็นค่าความผิดพลาดในการทำนาย เรียกสั้นๆ ว่ากราฟความผิดพลาดในการทำนาย (Residual Plot)

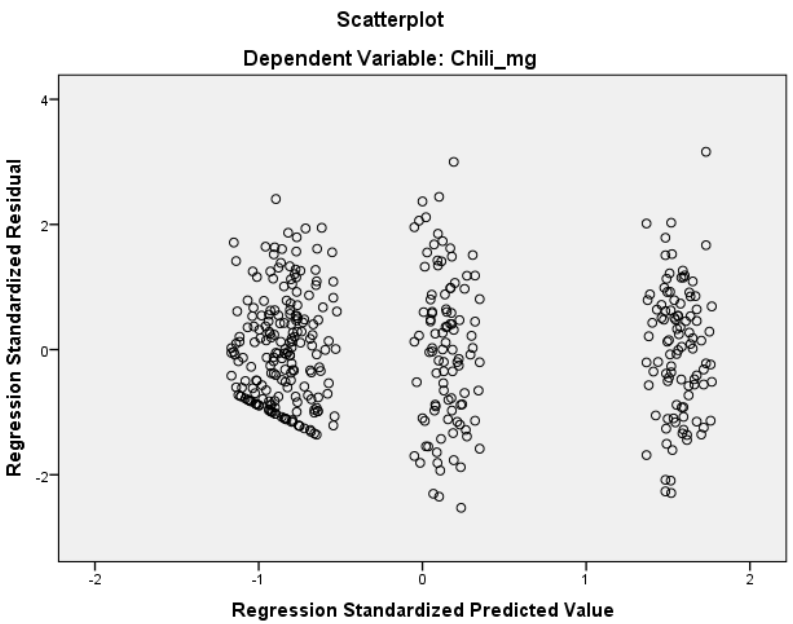
ค่าความผิดพลาดในการทำนาย
อยู่ในรูปคะแนนมาตรฐาน

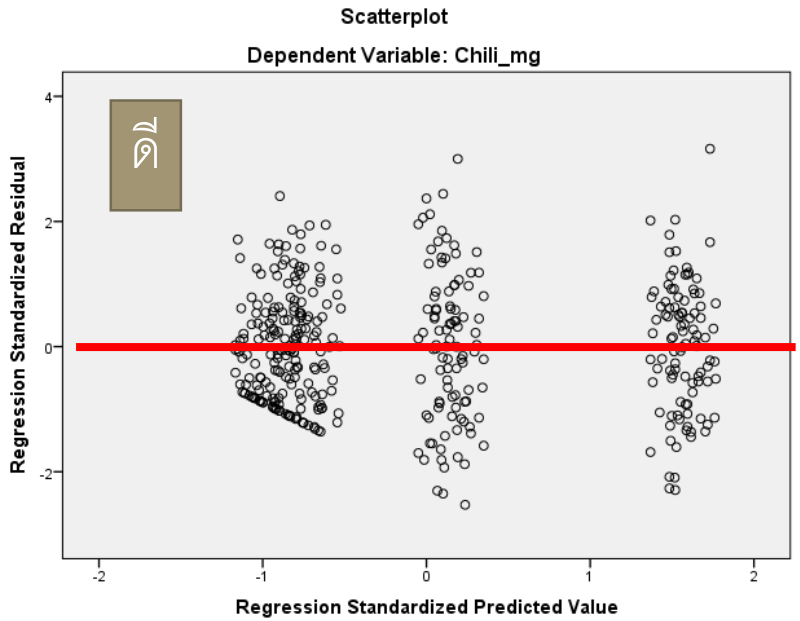


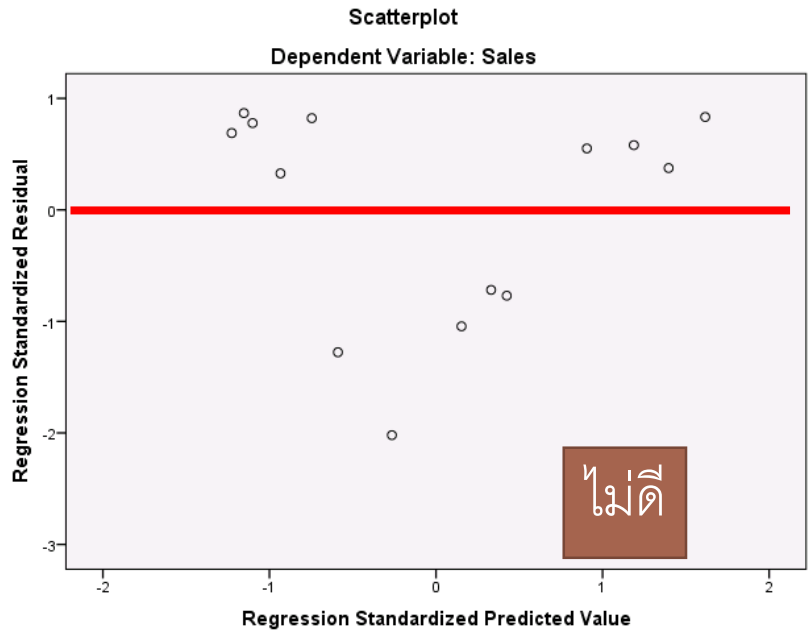
ค่าที่ทำนายได้อยู่ในรูปคะแนนมาตรฐาน

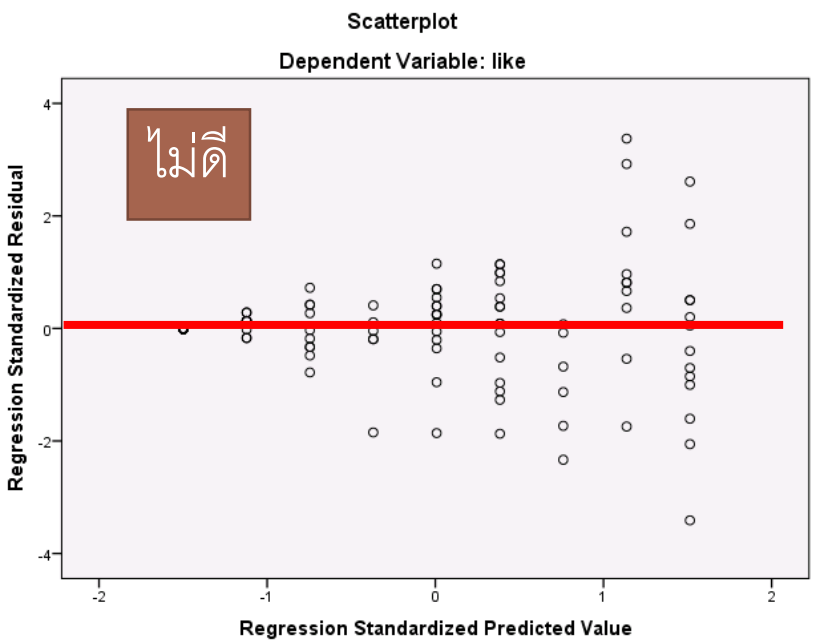
ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

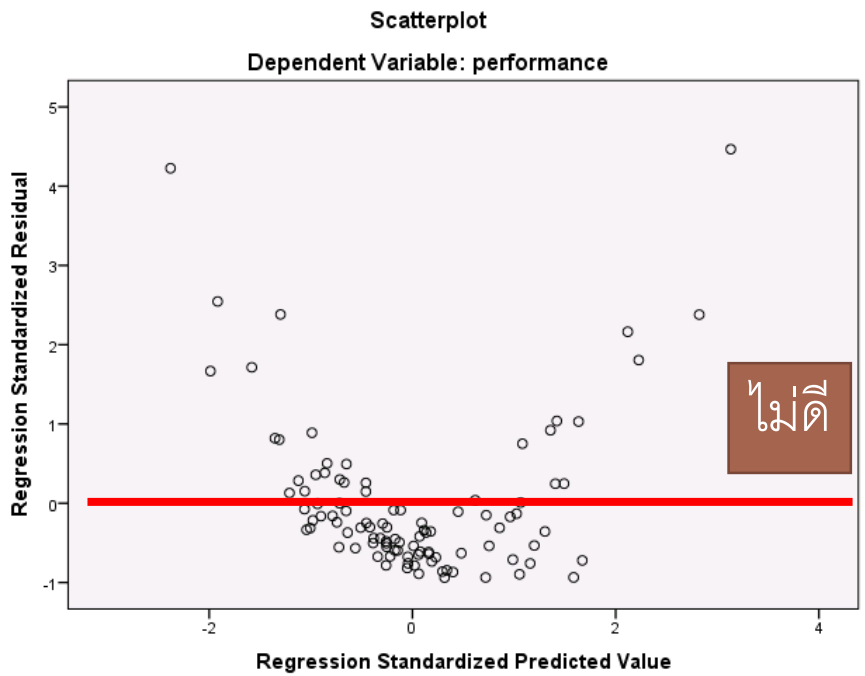
- ผลของ Residual Plot ที่ดี คือความสัมพันธ์ของค่าที่ทำนายได้และค่าความผิดพลาดเท่ากับ 0 มีการกระจายคงที่ ไม่สามารถลากเส้นอธิบายความสัมพันธ์ได้




Scatterplot
Dependent Variable: Chili_mg
ดี
Regression Standardized Residual
Regression Standardized Predicted Value

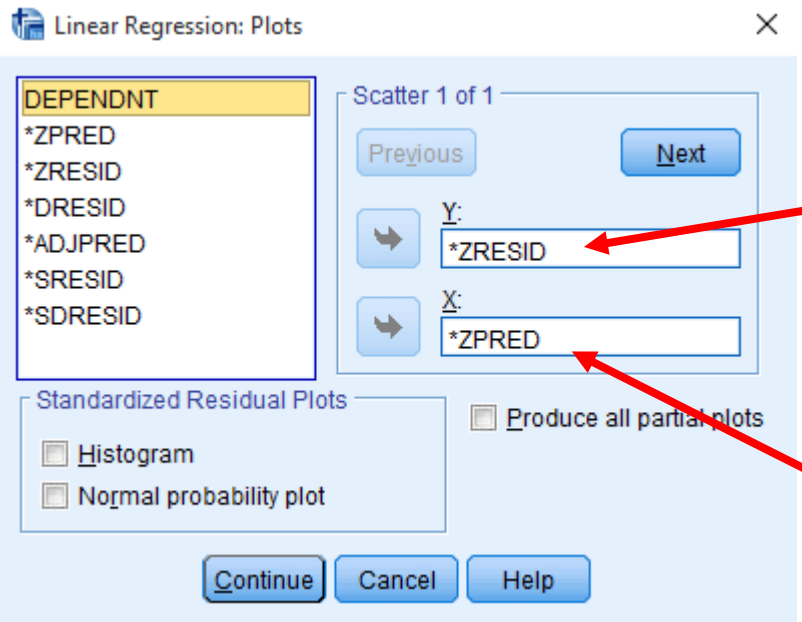

Scatterplot
Dependent Variable: Sales
ไม่ดี
Regression Standardized Residual
Regression Standardized Predicted Value


Scatterplot
Dependent Variable: like
ไม่ดี
Regression Standardized Residual
Regression Standardized Predicted Value


Scatterplot
Dependent Variable: performance
ไม่ดี
Regression Standardized Residual
Regression Standardized Predicted Value

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

- สามารถหาได้โดย เลือก Plot…



ค่าความผิดพลาดในการทำนาย
อยู่ในรูปคะแนนมาตรฐาน

ค่าที่ทำนายได้อยู่ในรูปคะแนนมาตรฐาน

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

1. ความผิดพลาดในการทำนาย (error) มีการกระจายเป็นโค้งปกติ
   - ทดสอบด้วยการสร้าง Residual plot โดยดูว่าการกระจายเหนือเส้น 0 และใต้เส้น 0 พอกันหรือไม่
   - อาจหา Histogram ของค่าความผิดพลาดในการทำนาย (**Plot…** แล้วเลือก **Histogram**) แล้วเทียบกับการกระจายโค้งปกติ
   - การละเมิด ไม่มีผลต่อการทดสอบทางสถิติ แต่มีผลต่อความแม่นยำในการทำนายค่าของตัวแปรตาม ที่ความผิดพลาดในด้านบวกไม่เหมือนกับความผิดพลาดในด้านลบ

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

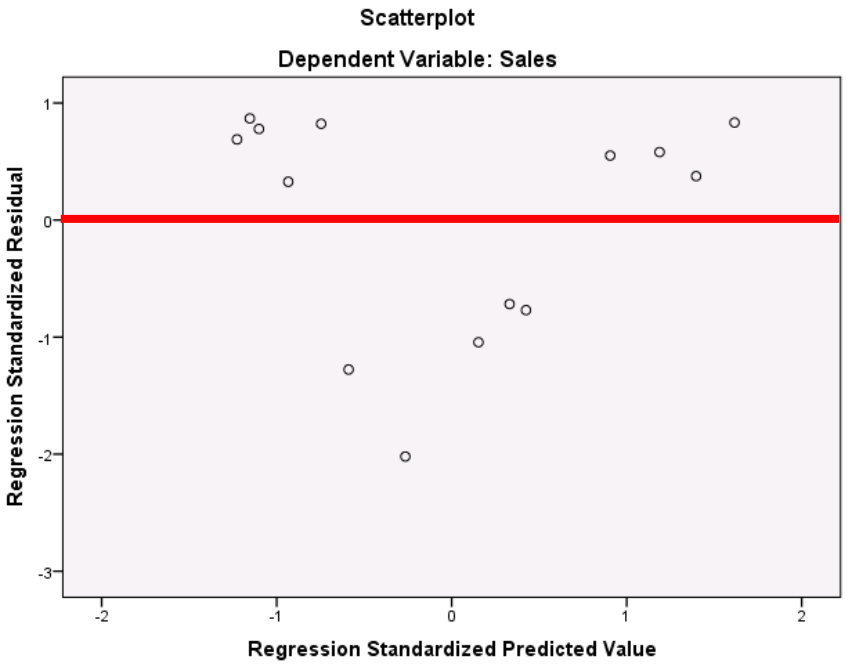
2. ความผิดพลาดในการทำนายเป็นอิสระจากกัน (Independence of correlated errors)
   - หากพูดอย่างเคร่งครัด คือ ค่าความผิดพลาดในการทำนาย จะต้องไม่มีความสัมพันธ์กัน ค่าความผิดพลาดของหน่วยหนึ่งจะต้องไม่มีอิทธิพลต่อความผิดพลาดของอีกหน่วยหนึ่ง

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

2. ความผิดพลาดในการทำนายเป็นอิสระจากกัน (Independence of correlated errors)

- ตัวอย่างการละเมิดแบบที่ 1 ค่าความผิดพลาดจะมีความสัมพันธ์กัน หากวัดในช่วงเวลาที่ใกล้กัน หรือที่เรียกว่า ความสัมพันธ์กับตนเอง (autocorrelation)
- เช่น ทำนายรายได้ของบริษัทด้วย GDP ในช่วงปี 2000-2013 ที่ความผิดพลาดในการทำนายในปีที่ใกล้กัน (เช่น 2005 และ 2006) จะมีแนวโน้มไปในทางเดียวกัน
- สามารถตรวจสอบได้ด้วย Residual Plot
- หรือ สามารถตรวจสอบได้ด้วยค่า Durbin-Watson (สามารถเลือกได้ โดยเข้าไปที่ Statistics…) ค่าที่เป็นไปได้อยู่ในช่วง 0-4 ค่าที่มีปัญหาจะน้อยกว่า 1 หรือมากกว่า 3 แต่ไม่ได้หมายความว่าค่าอยู่ในช่วง 1-3 จะไม่มีปัญหา

ค่าความผิดพลาดสามารถอธิบายได้ด้วย
ค่าที่ทำนายได้อย่างเป็นระบบ



Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.952[a]	.907	.899	61.401	.695

a. Predictors: (Constant), GDP

b. Dependent Variable: Sales

ค่า Durbin-Watson น้อยกว่า 1

จากตัวอย่างนี้ การนำข้อมูลแต่ละปีมาใช้ในสมการถดถอย ทำให้ค่าความผิดพลาดในปีใกล้เคียงกันมีความสัมพันธ์กันทางบวก

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

2. ความผิดพลาดในการทำนายเป็นอิสระจากกัน (Independence of correlated errors)
   - ในโปรแกรม R ค่า Durbin-Watson จะสามารถทดสอบทางสถิติว่าแตกต่างจาก 2 (ไม่มี autocorrelation) อย่างมีนัยสำคัญหรือไม่ หากใช้การทดสอบนี้ (ในกรณีที่มีจำนวนกลุ่มตัวอย่างสูง) ประกอบกับพิจารณาตัวแปรที่ใช้ในการวิจัยว่าเกี่ยวข้องกับช่วงเวลาหรือไม่ จะช่วยตรวจสอบว่าข้อตกลงนี้ถูกละเมิดหรือไม่
   - หากข้อตกลงนี้ถูกละเมิด ค่า Standard Error จะไม่ถูกต้อง ทำให้ผลการทดสอบทางสถิติผิด (*p* value ไม่ถูกต้องหรือช่วงเชื่อมั่นไม่ถูกต้อง)
   - ต้องใส่ลักษณะความสัมพันธ์ระหว่างความผิดพลาดในการทำนายตามเงื่อนเวลานี้ไว้ในการวิเคราะห์ถดถอย ซึ่งจะไม่กล่าวถึงในวิชานี้

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

2. ความผิดพลาดในการทำนายเป็นอิสระจากกัน (Independence of correlated errors)

- ตัวอย่างการละเมิดแบบที่ 2 ค่าความผิดพลาดจะจับกลุ่มกันใน Residual plot ราวกับว่าข้อมูลถูกเก็บมาเป็นกลุ่มๆ แต่ไม่ได้นำตัวแปรกลุ่มนี้เข้าไปในสมการ
- เช่น ผู้วิจัยทำนายผลการเรียนด้วยความถี่ในการเข้าสายของนักเรียน แต่การเก็บข้อมูลนั้น ผู้วิจัยเก็บข้อมูลจากนักเรียนหลายๆ ห้องเรียน แล้วผู้วิจัยไม่ได้นำห้องเรียนมาใช้ในการอธิบายความแตกต่าง
- หากละเมิด จะทำให้ค่า Standard error ไม่ถูกต้อง
- แก้ไขได้ด้วยการวิเคราะห์พหุระดับ (Multilevel models) หรือหากจำนวนกลุ่ม (เช่น ห้องเรียน) มีน้อย อาจนำกลุ่มไปแปลงเป็นตัวแปรดัมมี่แล้วใส่ไปเป็นตัวแปรอิสระ

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

3. ความแปรปรวนของความผิดพลาดในการทำนาย ไม่ขึ้นอยู่กับค่าของตัวแปรอิสระ หรือที่เรียกว่า ความเท่าเทียมกันของการกระจาย (Homoscedasticity)
   - การกระจายของความผิดพลาดในการทำนายจะต้องใกล้เคียงกันตลอดใน Residual Plot
   - ตัวอย่างการละเมิด เช่น คัดเลือกนักเรียนชายระดับมัธยมปลาย ไปคุยกับนักเรียนหญิงในรุ่นเดียวกัน หลังจากนั้นให้นักเรียนชายประเมินว่าชอบนักเรียนหญิงมากน้อยเพียงใด และประเมินว่าลักษณะภายนอกของนักเรียนหญิงที่คุยด้วย ดีมากน้อยเพียงใด

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.033	1.417		-.023	.981
	appearance	9.077	.251	.964	36.099	.000

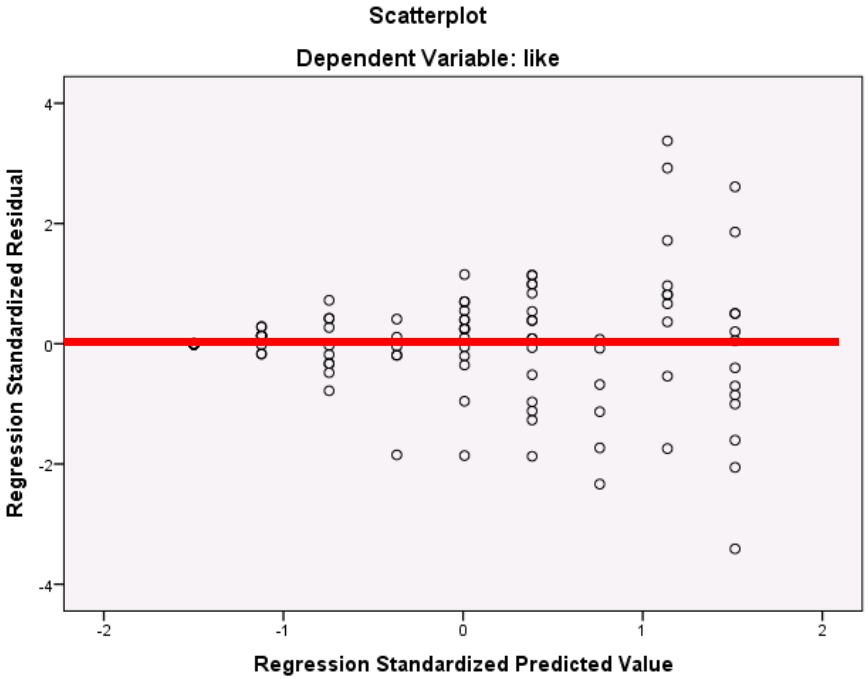
a. Dependent Variable: like

หน้าตาทำนายความชอบในทางบวกได้อย่างมีนัยสำคัญ

อำนาจในการทำนายไม่เท่าเทียมกัน

สังเกตได้ว่า การกระจายของค่าความผิดพลาด

สูงขึ้นเมื่อค่าที่ทำนายได้สูงขึ้น



ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

3. ความแปรปรวนของความผิดพลาดในการทำนาย ไม่ขึ้นอยู่กับค่าของตัวแปรอิสระ หรือที่เรียกว่า ความเท่าเทียมกันของการกระจาย (Homoscedasticity)
   - หากละเมิดข้อตกลงนี้ จะทำให้สัมประสิทธิ์ในการทำนายค่าเดียว ไม่สามารถใช้ได้ตลอดค่าของค่าที่ทำนายได้
   - ค่า Standard error ของการทดสอบทางสถิติจะผิดพลาด
   - แก้ไขได้ด้วยการพิจารณาว่าเป็นไปได้ไหม ที่จะเกิดผลปฏิสัมพันธ์บางอย่างที่ผู้วิจัยไม่ได้นึกถึง
   - หากไม่มีผลปฏิสัมพันธ์ดังกล่าว อาจแก้ไขได้โดยการวิเคราะห์ถดถอยแบบถ่วงน้ำหนัก (Weighted Regression) เพื่อปรับ Standard Error ให้ถูกต้อง

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

4. ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามเป็นเชิงเส้น (Linear Relationship)
   - ตัวอย่างการละเมิด เช่น ทำนายผลการปฏิบัติงานด้วยความเครียดและความสามารถ ซึ่งในความเป็นจริง ความเครียดมีผลเชิงเส้นโค้งต่อผลการปฏิบัติงาน เครียดน้อยเกินไปหรือมากเกินไปจะทำให้ผลการปฏิบัติงานน้อย แต่ความเครียดระดับปานกลางทำให้ผลการปฏิบัติงานดี
   - หากใส่ความเครียดเป็นตัวแปรอิสระ โดยไม่คำนึงว่าความเครียดมีผลเชิงเส้นโค้ง จะทำให้ค่าในสมการถดถอยผิดพลาดทั้งหมด
   - การตรวจสอบสามารถทำได้ด้วย Residual Plot แล้วดูว่ามีการเกาะกลุ่มของค่าความผิดพลาดในรูปเส้นโค้ง หรือเส้นรูปแบบอื่นที่ต่างจากปกติหรือไม่

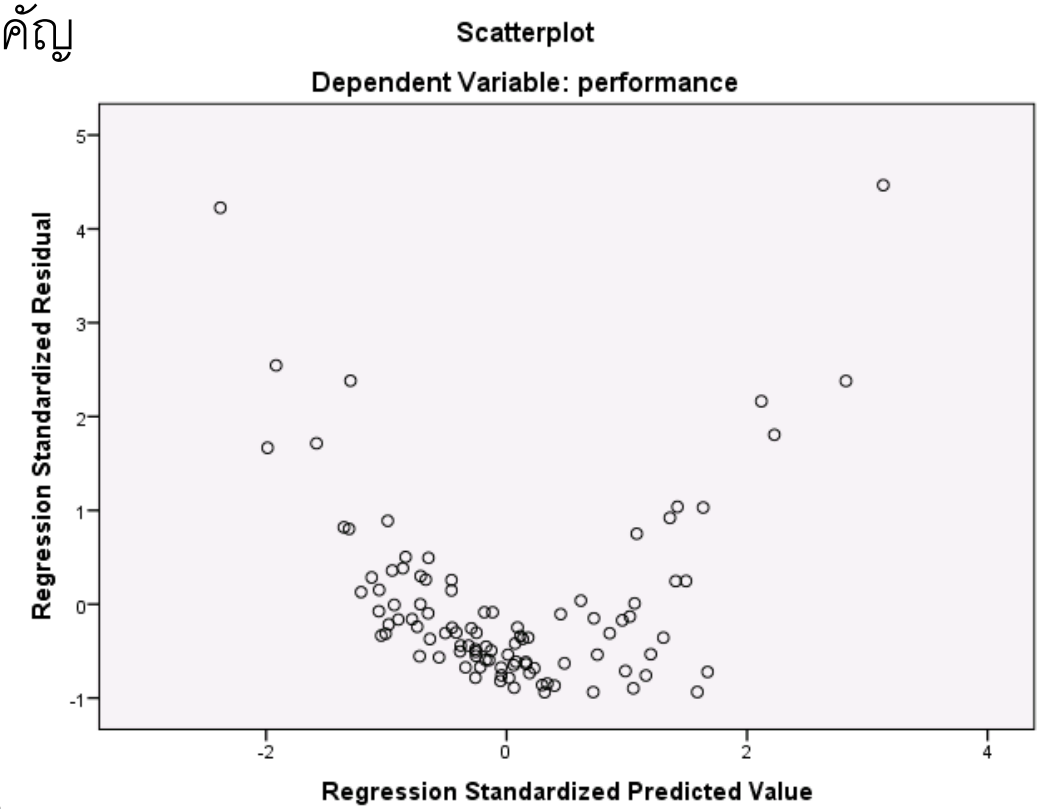
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	234.189	45.261		5.174	.000
	stress	-1.468	.925	-.186	-1.588	.116
	ability	-1.001	.935	-.125	-1.071	.287

a. Dependent Variable: performance

หากตีความหมายจากการใส่ตัวแปรความเครียดแบบปกติ
พบว่าไม่มีตัวแปรใดถึงระดับนัยสำคัญ

พบรูปแบบค่าความผิดพลาดเป็นเส้นโค้ง



ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

4. ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามเป็นเชิงเส้น (Linear Relationship)
   - การแก้ไข สามารถทำได้ด้วยการเปลี่ยนรูปแบบอิทธิพลให้ถูกต้อง เช่น ใส่ตัวแปรความเครียดยกกำลังสองเข้าไปในสมการเพิ่มเติม เพื่อวิเคราะห์ถดถอยแบบไม่เป็นเส้นตรง เรื่องนี้ไม่ได้สอนในวิชานี้

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

5. ตัวแปรอิสระทั้งหมดไม่มีความผิดพลาดในการวัด (No measurement error)
   - การวัดที่ดี หากวัดซ้ำในเวลาเดิม สถานการณ์เดิม ค่าที่ได้จะต้องเหมือนเดิม
   - ในการวัดทางจิตวิทยาในส่วนใหญ่ การวัดซ้ำในเวลาเดิม สถานการณ์เดิม อาจได้คะแนนไม่เหมือนเดิม เช่น การสอบวัดความสามารถ บางครั้งอาจมีปัจจัยภายนอกที่ทำให้ผลการสอบเปลี่ยนแปลงไป ทั้งที่ความสามารถเหมือนเดิม
   - หากมีความผิดพลาดในการวัดสูง จะทำให้ค่าที่ได้ในสมการถดถอยไม่ถูกต้อง
   - การแก้ไข ทำได้ด้วยการวิเคราะห์สถิติที่คำนึงถึงความผิดพลาดในการวัด เช่น การวิเคราะห์สมการเชิงโครงสร้าง (Structural Equation Modeling)

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

6. ไม่สามารถทำนายค่าของตัวแปรตามได้ด้วยค่าของตัวแปรอิสระนอกเหนือจากช่วงของข้อมูลที่เก็บมาได้
   - เช่น ท่านต้องการทำนายผลการสอบด้วยชั่วโมงการอ่านหนังสือ หากช่วงโมงการอ่านหนังสือที่ท่านเก็บข้อมูลจากกลุ่มตัวอย่าง อยู่ในช่วง 0-8 ชั่วโมง ท่านไม่สามารถทำนายตัวแปรตามได้ด้วยชั่วโมงการอ่านหนังสือที่สูงกว่า 8 ชั่วโมง
   - เนื่องจากว่า ผลของสมการถดถอยจะบอกว่า ข้อมูลเป็นเชิงเส้นตรงดังสมการ หากจำนวนชั่วโมงในการอ่านหนังสืออยู่ในช่วง 0-8 ชั่วโมง หากเก็บข้อมูลด้วยช่วงที่กว้างกว่านี้ ความสัมพันธ์อาจเป็นเส้นตรงหรือเส้นโค้งก็ได้
   - การทำนายค่าของตัวแปรตามด้วยตัวแปรอิสระนอกช่วงสามารถทำได้ แต่ผู้วิจัยต้องตระหนักว่าความถูกต้องจะต่ำกว่าค่าของตัวแปรอิสระในช่วงมาก

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

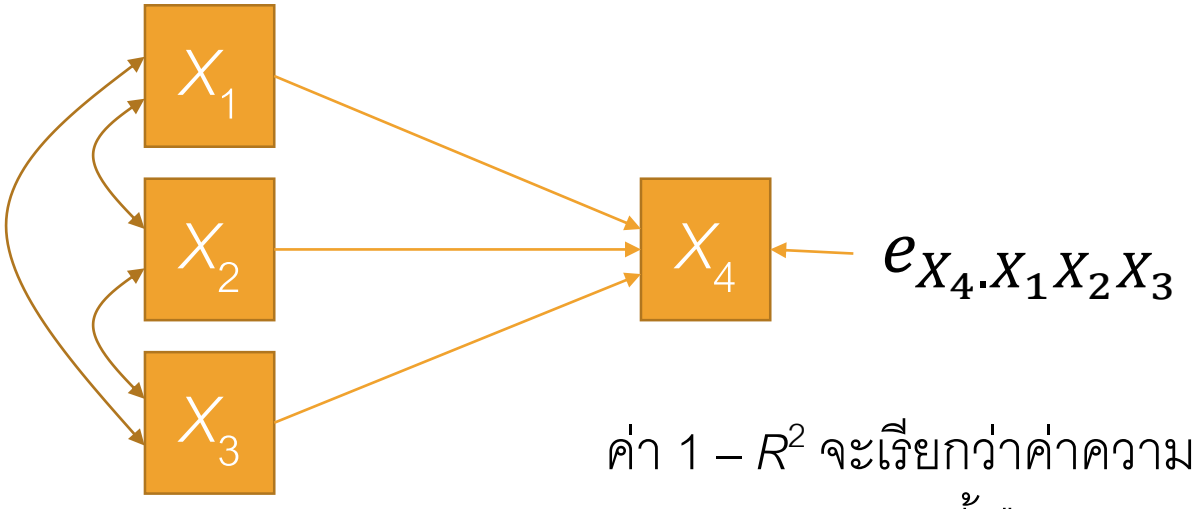
7. ไม่มีข้อมูลสุดโต่ง (Outliers)
   - ในการวิเคราะห์ถดถอยมีเทคนิคในการวิเคราะห์ข้อมูลสุดโต่งหลายวิธี
   - ในวิชานี้ ให้ใช้การสังเกต Residual plot แล้วดูว่ามีข้อมูลสุดโต่งหรือไม่

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

8. ตัวแปรอิสระไม่มีความสัมพันธ์ซึ่งกันและกันสูงมากจนเกินไป เรียกภาวะนี้ว่าการเป็นเส้นตรงเดียวกัน (Multicollinearity)
   - หากตัวแปรอิสระหนึ่งมีความแปรปรวนร่วมกับตัวแปรอิสระอื่นสูงมาก ทำให้ตัวแปรอิสระนี้แทบไม่มีข้อมูลใหม่ในการทำนายตัวแปรตามเลย
   - ภาวะนี้สามารถตรวจสอบได้ด้วย ค่าสหสัมพันธ์ระหว่างตัวแปรอิสระด้วยกัน ควรไม่เกิน .95 แต่วิธีนี้เป็นเพียงการตรวจสอบความสัมพันธ์รายคู่เท่านั้น

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

8. ตัวแปรอิสระไม่มีความสัมพันธ์ซึ่งกันและกันสูงมากจนเกินไป เรียกภาวะนี้ว่าการเป็นเส้นตรงเดียวกัน (Multicollinearity)
   - อีกวิธีหนึ่ง คือตรวจสอบว่าตัวแปรอิสระดังกล่าว ถูกอธิบายได้ด้วยตัวแปรอิสระอื่นมากน้อยเพียงใด



ค่า 1 – R^2 จะเรียกว่าค่าความคงทน (Tolerance)

ในตัวอย่างนี้ คือค่าความคงทนของ X_4

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

8. ตัวแปรอิสระไม่มีความสัมพันธ์ซึ่งกันและกันสูงมากจนเกินไป เรียกภาวะนี้ว่าการเป็นเส้นตรงเดียวกัน (Multicollinearity)
   - หากค่าความคงทน ต่ำกว่า .10 ถือว่ามีปัญหาดังกล่าว
   - ปัจจัยการเฟ้อของความแปรปรวน (Variance Inflation Factor; VIF) จะคำนวณโดย 1/Tolerance หากค่านี้สูงกว่า 10 ถือว่ามีปัญหา
   - ใน SPSS หาได้ด้วยการเลือก **Statistics…** แล้วกด **Collinearity diagnostics**

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	234.189	45.261		5.174	.000					
	stress	-1.468	.925	-.186	-1.588	.116	-.256	-.159	-.155	.693	1.443
	ability	-1.001	.935	-.125	-1.071	.287	-.229	-.108	-.104	.693	1.443

a. Dependent Variable: performance

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

8. ตัวแปรอิสระไม่มีความสัมพันธ์ซึ่งกันและกันสูงมากจนเกินไป เรียกภาวะนี้ว่าการเป็นเส้นตรงเดียวกัน (Multicollinearity)
   - ในบางครั้ง ถึงแม้ว่าตัวแปรอิสระดังกล่าวจะถูกอธิบายด้วยตัวแปรอิสระอื่นมาก ส่วนที่เหลือยังคงมีความสำคัญอยู่ เช่น การควบคุมตนเองโดยอัตโนมัติ (Automized self-regulation) และการควบคุมตนเองโดยรู้ตัว (Conscious self-regulation) ทั้งสองตัวแปรมีความสัมพันธ์กันสูงกว่า .95 แต่การควบคุมตนเองโดยรู้ตัว มีอิทธิพลในการทำงานบางอย่าง ที่การควบคุมตนเองโดยอัตโนมัติไม่สามารถอธิบายได้

ข้อตกลงเบื้องต้นก่อนการใช้สถิติ

8. ตัวแปรอิสระไม่มีความสัมพันธ์ซึ่งกันและกันสูงมากจนเกินไป เรียกภาวะนี้ว่าการเป็นเส้นตรงเดียวกัน (Multicollinearity)
   - ปัญหาของภาวะการเป็นเส้นตรง คือ ค่าสัมประสิทธิ์ถดถอยที่คำนวณได้จะไม่แม่นยำ และค่า Standard error จะสูงมาก
   - วิธีการแก้ไข คือการใช้การปรับศูนย์กลางด้วยความผิดพลาดในการทำนาย (Residual centering) ซึ่งไม่กล่าวถึงในวิชานี้

การหากำลัง

สถิติขั้นนำสำหรับจิตวิทยา

อ. สันทัด พรประเสริฐมานิต

การหากำลัง

- มีทั้งหมด 3 สถานการณ์
  - หากำลังในการทดสอบความชัน
  - หากำลังในการทดสอบสัมประสิทธิ์ทำนาย
  - หากำลังในการทดสอบสัมประสิทธิ์ทำนายที่เพิ่มขึ้น
- การหากำลัง หรือกลุ่มตัวอย่าง ให้ทดสอบผ่านการหาสหสัมพันธ์ ในกรณีที่มีตัวแปรทำนายเพียงตัวแปรเดียว

การหากำลัง (ความชัน)

- หากำลังในการทดสอบความชัน หากมีตัวแปรอิสระเพียงแค่ตัวเดียว ให้แปลงค่าความชันให้เป็นค่าสหสัมพันธ์ และหากำลังผ่านการทดสอบสหสัมพันธ์
- อีกทางหนึ่ง คือการทดสอบความชันผ่านคำสั่งข้างล่าง ซึ่งผมเชื่อว่าพวกคุณสามารถทำเองได้ ผลที่ได้เหมือนกับกำลังการทดสอบสหสัมพันธ์ที่สอนไป
  - เลือกกลุ่มสถิติที่ต้องการทดสอบ คือ t-tests
  - เลือกสถิติที่ใช้ทดสอบ คือ Linear bivariate regression: One group, size of slope
  - เลือกว่าจะหากำลังในการทดสอบ คือ Post hoc: Compute achieved power – given α, sample size and effect size

การหากำลัง (ความชัน)

- หากำลังในการทดสอบความชัน หากมีตัวแปรอิสระมากกว่า 1 ตัว ให้ใช้คำสั่งข้างล่าง
  - เลือกกลุ่มสถิติที่ต้องการทดสอบ คือ t-tests
  - เลือกสถิติที่ใช้ทดสอบ คือ Linear multiple regression: Fixed model, single regression coefficient
  - เลือกว่าจะหากำลังในการทดสอบ คือ Post hoc: Compute achieved power – given α, sample size and effect size

การหากำลัง (ความชัน)

- ในการขนาดอิทธิพล f^2 จะต้องคำนวณผ่านสัมประสิทธิ์ทำนายบางส่วน (Partial squared multiple correlation; Partial R^2)
- ในกรณีต้องการหาความสัมพันธ์บางส่วนระหว่าง B และ Y โดยควบคุม A $(R^2_{YB\cdot A})$ สูตรการหาเป็นดังนี้

$$R^2_{YB\cdot A} = \frac{R^2_{Y\cdot A,B} - R^2_{Y\cdot A}}{1 - R^2_{Y\cdot A}}$$

- โดย $R^2_{Y\cdot A}$ คือสัมประสิทธิ์การทำนายเมื่อมี A เป็นตัวแปรอิสระตัวเดียว และ $R^2_{Y\cdot A,B}$ คือสัมประสิทธิ์การทำนายเมื่อมี A และ B เป็นตัวแปรอิสระ

การหากำลัง (ความชัน)

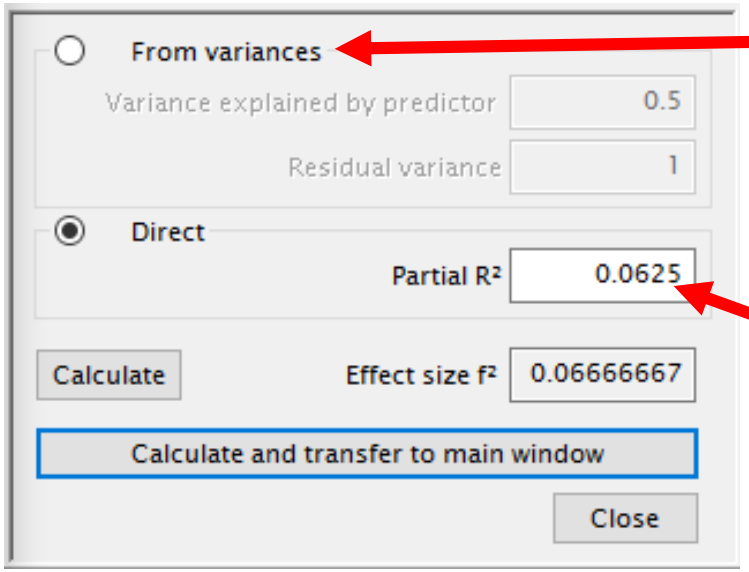
- หากต้องการกำลังในการทดสอบความชันของตัวแปร *X* แล้ว Partial R^2 ที่ใช้ในโปรแกรม สามารถคำนวณง่ายๆ ดังนี้

$$R^2_{YX\cdot\text{Others}} = \frac{R^2_{Y\cdot X,\text{Others}} - R^2_{Y\cdot\text{Others}}}{1 - R^2_{Y\cdot\text{Others}}}$$

- Others หมายถึง ตัวแปรอิสระตัวแปรอื่น
- $R^2_{Y\cdot X,\text{Others}}$ คือสัมประสิทธิ์การทำนายเมื่อมี *X* และตัวแปรอิสระตัวอื่น
- $R^2_{Y\cdot\text{Others}}$ คือสัมประสิทธิ์การทำนายเมื่อตัวแปรอิสระอื่น แต่ไม่มี *X*

การหากำลัง (ความชัน)

- เช่น ต้องการทดสอบความชันของการวิเคราะห์ถดถอยแบบพหุ ของตัวแปรหนึ่งที่หากใส่ตัวแปรนี้พร้อมกับตัวแปรอื่นอีก 2 ตัวจะมี R^2 = .25 แต่หากใส่แต่ตัวแปรอื่นจะมี R^2 = .20 จงหากำลังที่กลุ่มตัวอย่างเท่ากับ 120



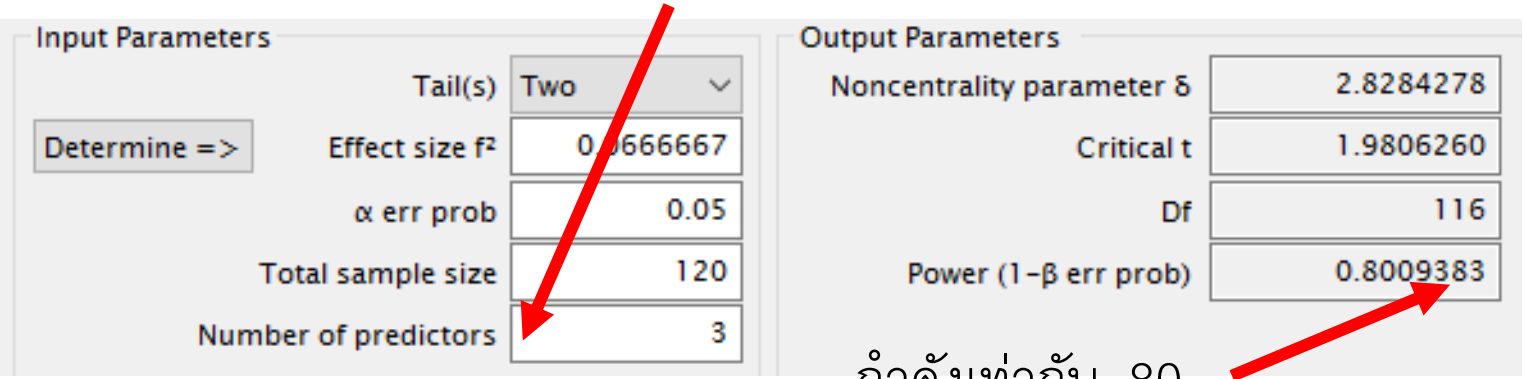
ไม่ควรใช้ มีโอกาสพลาดสูงมาก

$$\text{Partial } R^2 = \frac{.25 - .20}{1 - .20} = .0625$$

การหากำลัง (ความชัน)

- เช่น ต้องการทดสอบความชันของการวิเคราะห์ถดถอยแบบพหุ ของตัวแปรหนึ่งที่หากใส่ตัวแปรนี้พร้อมกับตัวแปรอื่นอีก 2 ตัวจะมี R^2 = .25 แต่หากใส่แต่ตัวแปรอื่นจะมี R^2 = .20 จงหากำลังที่กลุ่มตัวอย่างเท่ากับ 120

ตัวแปรอิสระรวม 3 ตัว



กำลังเท่ากับ .80

การหากำลัง (สัมประสิทธิ์การทำนาย)

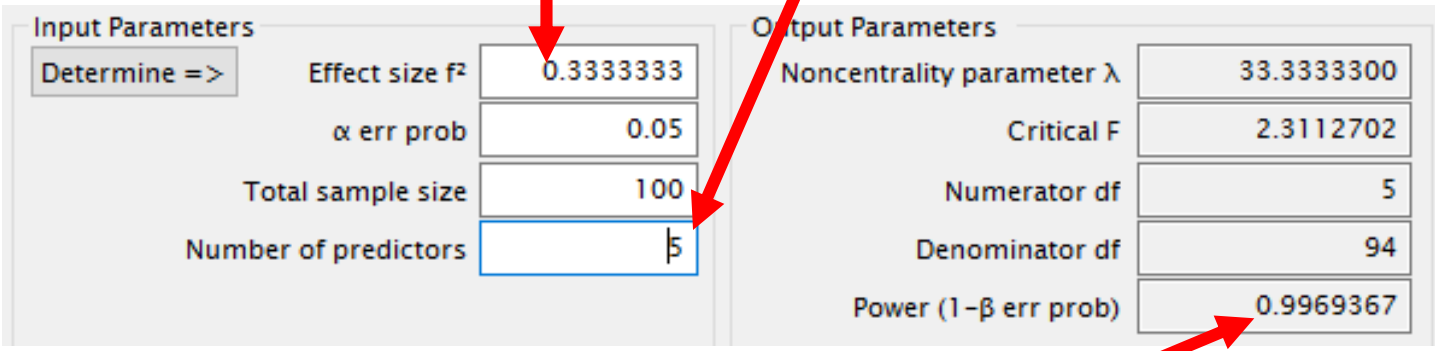
- หากมีตัวแปรอิสระเพียงแค่ตัวเดียว ให้ใช้การทดสอบค่าสหสัมพันธ์ เพราะค่าสหสัมพันธ์ยกกำลังสอง เท่ากับสัมประสิทธิ์การทำนาย
- หากมีตัวแปรอิสระมากกว่า 1 ตัว ให้ใช้คำสั่งดังต่อไปนี้
  - เลือกกลุ่มสถิติที่ต้องการทดสอบ คือ F tests
  - เลือกสถิติที่ใช้ทดสอบ คือ Linear multiple regression: Fixed model, R^2 deviation from zero
  - เลือกว่าจะหากำลังในการทดสอบ คือ Post hoc: Compute achieved power – given α, sample size and effect size

การหากำลัง (สัมประสิทธิ์การทำนาย)

- เช่น

ค่า Cohen's *f* สามารถกด Determine เพื่อคำนวณค่านี้ได้

จำนวนตัวแปรอิสระ



กำลังเท่ากับ .80

การหากำลัง (สัมประสิทธิ์การทำนาย)

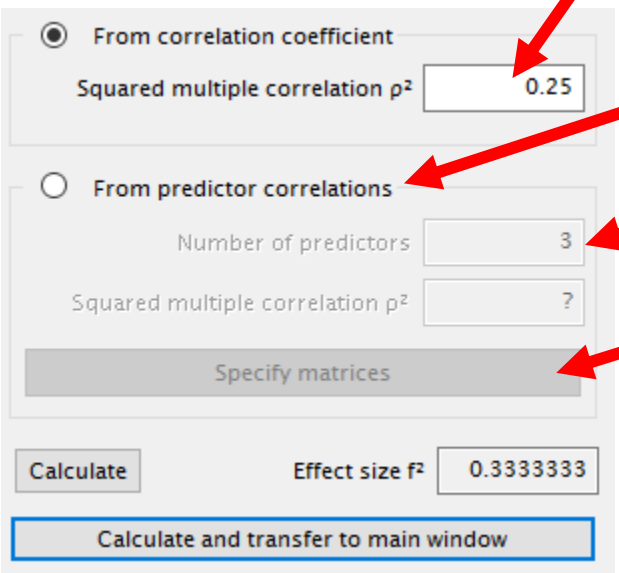
- เช่น

คำนวณผ่านค่า R^2

คำนวณผ่านค่า Pearson Correlations

ใส่จำนวนตัวแปรอิสระ

กดปุ่มนี้ เพื่อใส่ข้อมูลค่า Pearson Correlations แล้วโปรแกรมคำนวณค่า R^2 ลองกดแล้วใส่ค่าดู ใช้ไม่ยาก



การหากำลัง (สัมประสิทธิ์การทำนายที่เพิ่มขึ้น)

- หากมีตัวแปรอิสระในโมเดลแล้ว ใส่ชุดของตัวแปรอิสระเพิ่มจากชุดเดิม ได้ค่า R^2 เพิ่มขึ้น ในที่นี้ต้องการหากำลังในการทดสอบการเพิ่มขึ้นค่า R^2
  - เลือกกลุ่มสถิติที่ต้องการทดสอบ คือ F tests
  - เลือกสถิติที่ใช้ทดสอบ คือ Linear multiple regression: Fixed model, R^2 increase
  - เลือกว่าจะหากำลังในการทดสอบ คือ Post hoc: Compute achieved power – given α, sample size and effect size

การหากำลัง (สัมประสิทธิ์การทำนายที่เพิ่มขึ้น)

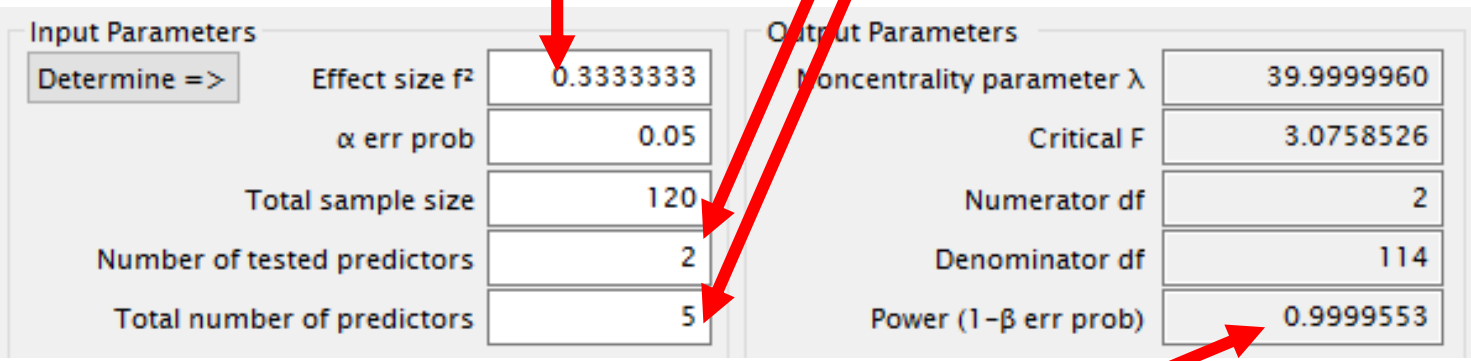
- เช่น

ค่า Cohen's *f* เทียบเท่ากับค่า R^2 ที่เพิ่มขึ้น

สามารถกด Determine เพื่อคำนวณค่านี้ได้

จำนวนตัวแปรอิสระที่ใส่เข้าไปเพิ่ม

จำนวนตัวแปรอิสระรวมทั้งหมด



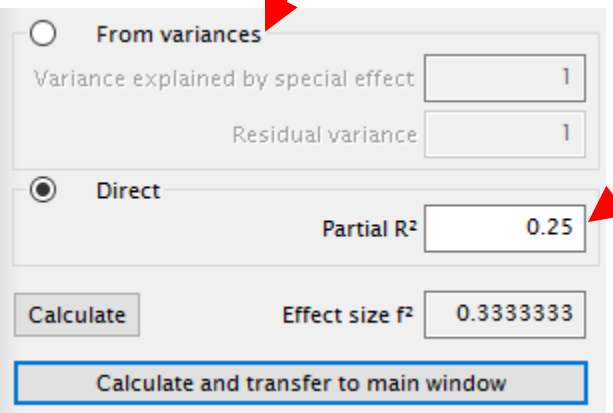
Input Parameters		Output Parameters	
Determine => Effect size f²	0.3333333	Noncentrality parameter λ	39.9999960
α err prob	0.05	Critical F	3.0758526
Total sample size	120	Numerator df	2
Number of tested predictors	2	Denominator df	114
Total number of predictors	5	Power (1-β err prob)	0.9999553

กำลังเท่ากับ .99

การหากำลัง (สัมประสิทธิ์การทำนายที่เพิ่มขึ้น)

- เช่น

ไม่ควรใช้ มีโอกาสพลาดสูงมาก



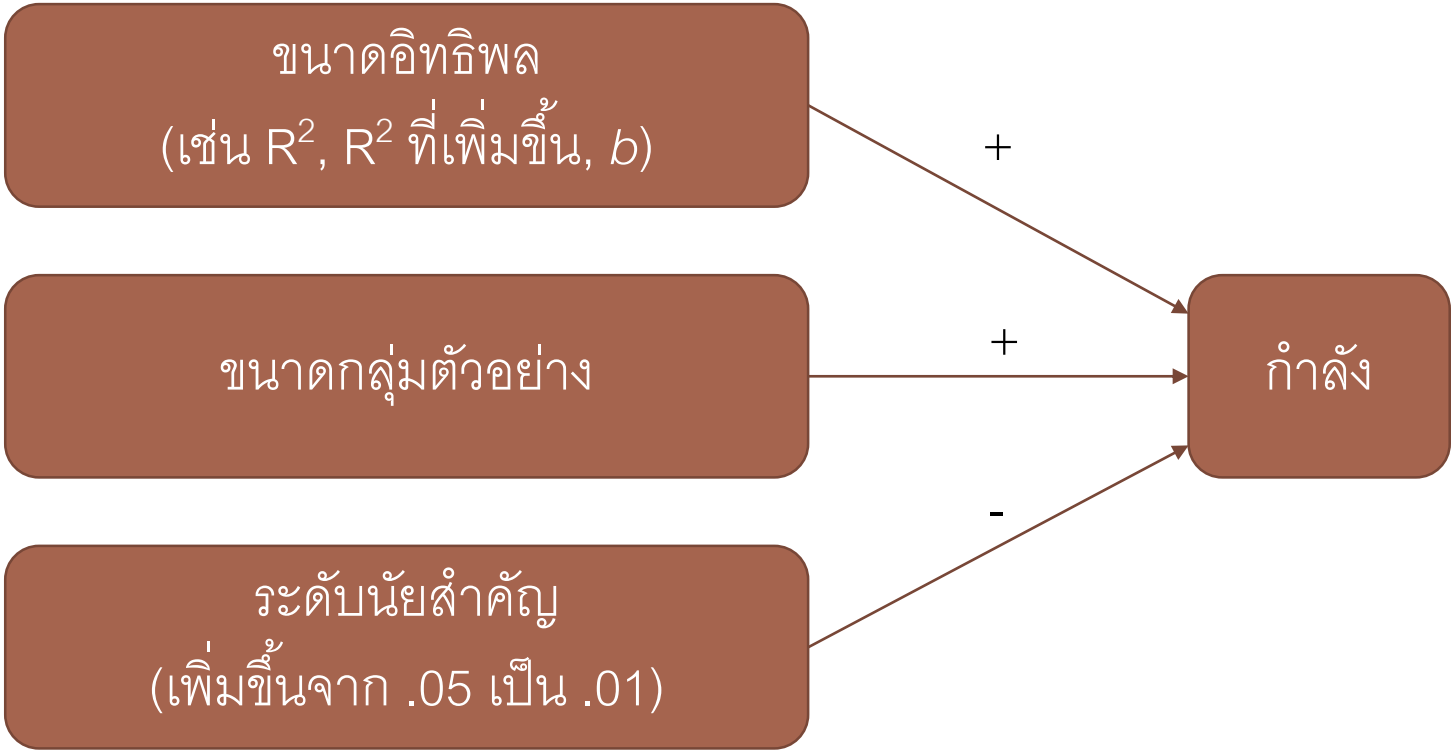
Partial R^2 สามารถคำนวณได้จากสูตรเดิม

เช่น ตัวแปรอิสระเดิม 3 ตัว: $R^2 = .20$

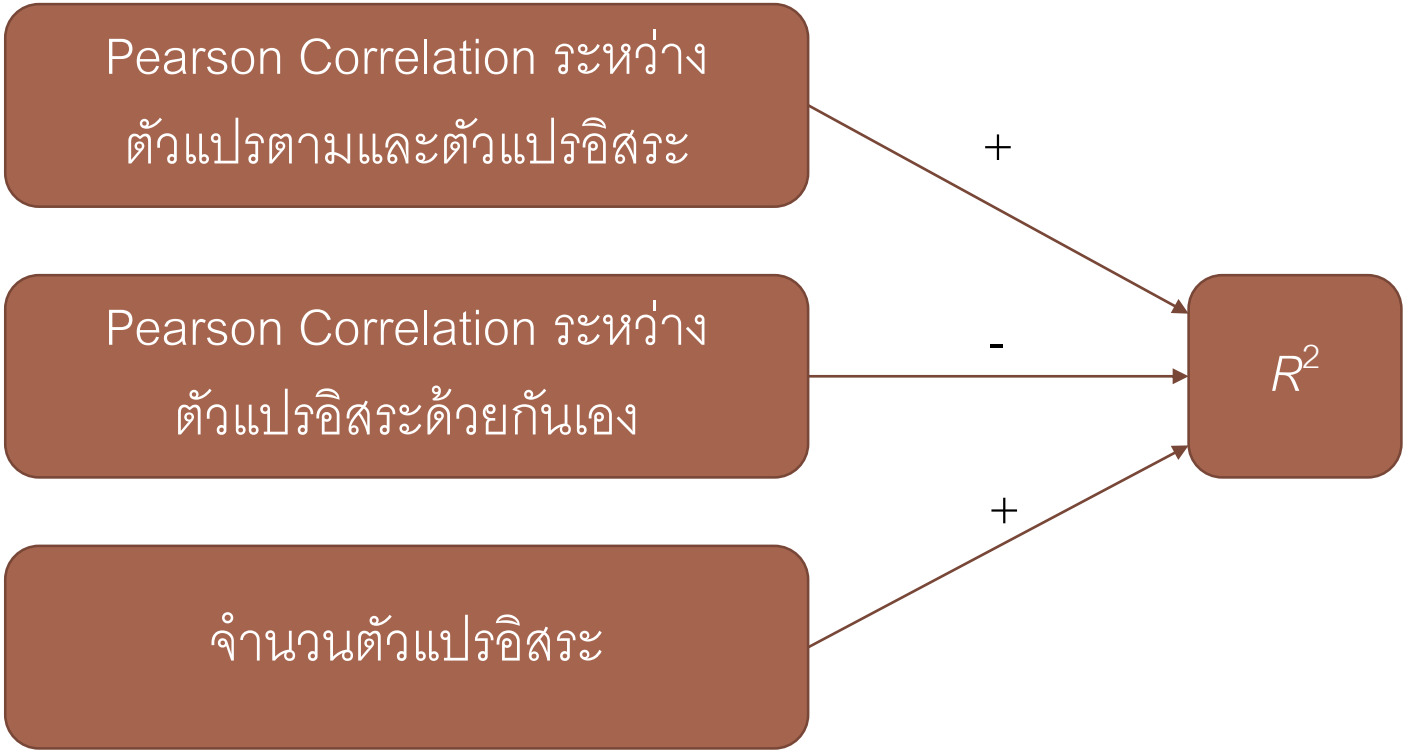
ตัวแปรอิสระเพิ่มขึ้นรวม 5 ตัว: $R^2 = .25$

$$\text{Partial } R^2 = \frac{.40 - .20}{1 - .20} = .25$$

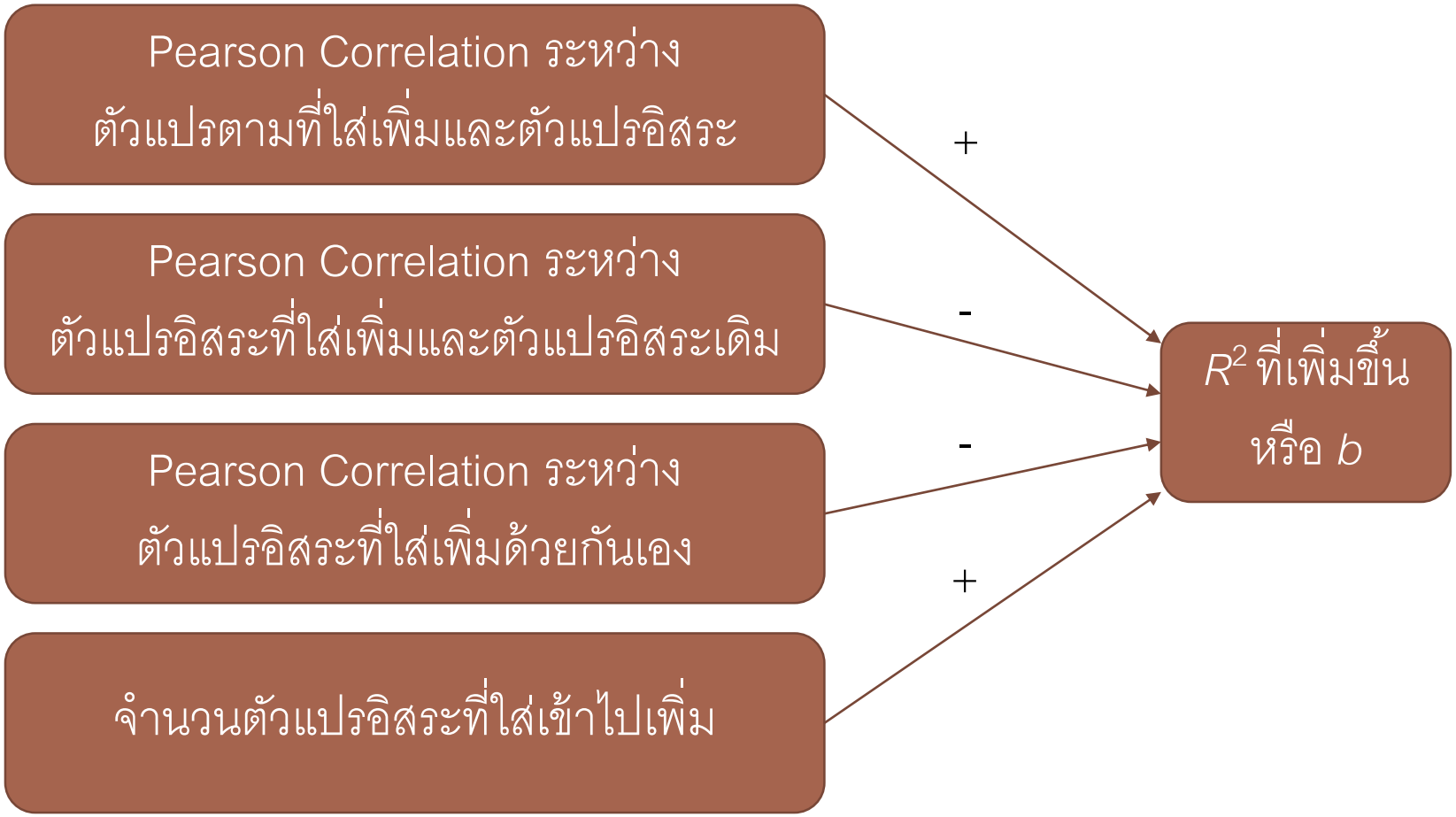
การหากำลัง


ขนาดอิทธิพล
(เช่น R^2, R^2 ที่เพิ่มขึ้น, *b*)
ขนาดกลุ่มตัวอย่าง
ระดับนัยสำคัญ
(เพิ่มขึ้นจาก .05 เป็น .01)
+
+
-
กำลัง

การหากำลัง


Pearson Correlation ระหว่างตัวแปรตามและตัวแปรอิสระ
Pearson Correlation ระหว่างตัวแปรอิสระด้วยกันเอง
จำนวนตัวแปรอิสระ
+
-
+
R^2

การหากำลัง

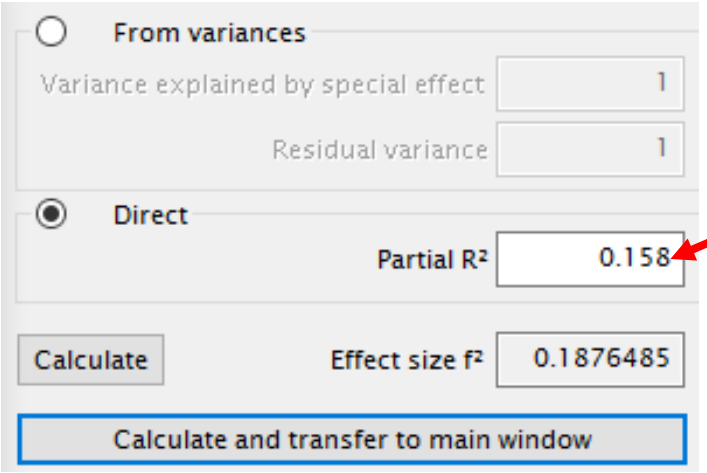

Pearson Correlation ระหว่าง
ตัวแปรตามที่ใส่เพิ่มและตัวแปรอิสระ
Pearson Correlation ระหว่าง
ตัวแปรอิสระที่ใส่เพิ่มและตัวแปรอิสระเดิม
Pearson Correlation ระหว่าง
ตัวแปรอิสระที่ใส่เพิ่มด้วยกันเอง
จำนวนตัวแปรอิสระที่ใส่เข้าไปเพิ่ม
+
-
-
+
R^2 ที่เพิ่มขึ้น
หรือ b

การกำหนดขนาดกลุ่มตัวอย่าง

- สามารถหาได้โดยวัตถุประสงค์เพื่อ
  - ให้ได้กำลังที่เหมาะสม
  - ให้ได้ความกว้างของช่วงเชื่อมั่นที่เหมาะสม
- การหาขนาดกลุ่มตัวอย่างเพื่อกำหนดความกว้างของช่วงเชื่อมั่น ให้ใช้ MBESS Package ใน R
- การหาขนาดกลุ่มตัวอย่างด้วยกำลัง วิธีการคล้ายคลึงกับการหากำลัง

การกำหนดขนาดกลุ่มตัวอย่าง

- เช่น ต้องการหากลุ่มตัวอย่างเพื่อทดสอบการใส่ตัวแปรอิสระเพิ่มขึ้น 4 ตัว จากเดิมที่มี 2 ตัว (รวมเป็น 6 ตัว) R^2 เดิมเท่ากับ .05 R^2 ใหม่คาดว่าจะเท่ากับ .20 หาก α = .05 แล้ว จำนวนกลุ่มตัวอย่างควรเป็นเท่าไรเพื่อให้ได้กำลัง .80



$$\text{Partial } R^2 = \frac{.20 - .05}{1 - .05} = .158$$

การกำหนดขนาดกลุ่มตัวอย่าง

- เช่น ต้องการหากลุ่มตัวอย่างเพื่อทดสอบการใส่ตัวแปรอิสระเพิ่มขึ้น 4 ตัว จากเดิมที่มี 2 ตัว (รวมเป็น 6 ตัว) R^2 เดิมเท่ากับ .05 R^2 ใหม่คาดว่าจะเท่ากับ .20 หาก α = .05 แล้ว จำนวนกลุ่มตัวอย่างควรเป็นเท่าไรเพื่อให้ได้กำลัง .80

ควรจะเก็บกลุ่มตัวอย่าง อย่างน้อย 69 คน

Input Parameters		Output Parameters	
Determine => Effect size f²	0.1876485	Noncentrality parameter λ	12.9477465
α err prob	0.05	Critical F	2.5201015
Power (1-β err prob)	0.80	Numerator df	4
Number of tested predictors	4	Denominator df	62
Total number of predictors	6	Total sample size	69
		Actual power	0.8023234

คาบต่อไป

- การบ้านที่ 7 (ส่งในสัปดาห์ที่ 7 – 11 มีนาคม)