



การวิเคราะห์ถดถอย (Regression Analysis)

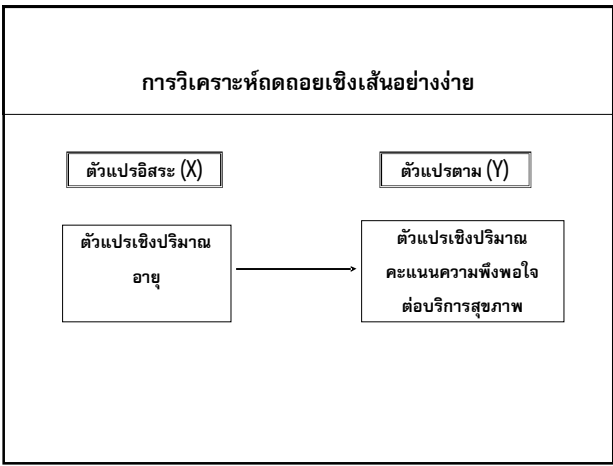
ผศ.ดร.ปิยะนุช จิตตุนนท์
คณะพยาบาลศาสตร์ ม.สงขลานครินทร์

ประเภทการวิเคราะห์ถดถอย

1. การวิเคราะห์ถดถอยเชิงเส้นอย่างง่าย
(simple linear regression analysis)
2. การวิเคราะห์ถดถอยเชิงพหุ
(multiple regression analysis)

การวิเคราะห์ถดถอยเชิงเส้นอย่างง่าย (simple linear regression analysis)

เป็นเทคนิคการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร 2 ตัว ซึ่งประกอบด้วยตัวแปรตาม (dependent variable) 1 ตัว และตัวแปรอิสระ (independent variable) 1 ตัว ที่มีความสัมพันธ์อยู่ในรูปเชิงเส้น โดยตัวแปรทั้ง 2 เป็นตัวแปรเชิงปริมาณ



- วัตถุประสงค์ของการวิเคราะห์ถดถอยเชิงเส้นอย่างง่าย**
1. ศึกษาความสัมพันธ์ระหว่างตัวแปรตาม (Y) และตัวแปรต้น/ตัวแปรอิสระ (X)
 2. ประมาณ/ทำนายค่าตัวแปรตาม (Y) เมื่อทราบค่าตัวแปรอิสระ (X)

- เงื่อนไขของการวิเคราะห์ถดถอยเชิงเส้นอย่างง่าย
(Assumptions of SLR)**
1. ตัวแปรตามและตัวแปรต้นมีความสัมพันธ์กันเชิงเส้นตรง (linearity)
 2. ตัวแปรที่มีการแจกแจงแบบปกติ (normality)
 3. ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติ (normality)
 4. ค่าแปรปรวนของค่าคลาดเคลื่อนคงที่ (constant variance/homoscedasticity)
 5. ค่าคลาดเคลื่อนเป็นอิสระต่อกัน (autocorrelation)



สมการการถดถอยเชิงเส้นอย่างง่าย

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

Y คือ ตัวแปรตาม (DV)

X คือ ตัวแปรอิสระ (IV)

a คือ ค่าคงที่ (intercept): ค่าของ Y เมื่อ X เป็น 0

b คือ สัมประสิทธิ์ความถดถอย (regression coefficient)/ความชันของเส้นตรง (slope)

ขั้นตอนการวิเคราะห์ถดถอยเชิงเส้นอย่างง่าย

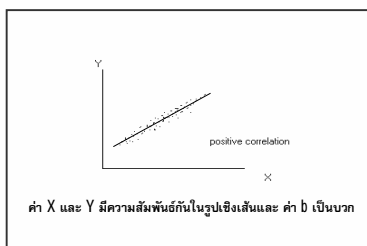
1. ศึกษารูปแบบความสัมพันธ์ระหว่างตัวแปรอิสระกับตัวแปรตาม

2. ศึกษาความสัมพันธ์ระหว่างตัวแปร X กับ Y ในรูปเชิงเส้น

3. ทดสอบความสัมพันธ์ระหว่างตัวแปร X กับ Y

4. ศึกษาระดับความสัมพันธ์ระหว่างตัวแปร X กับ Y

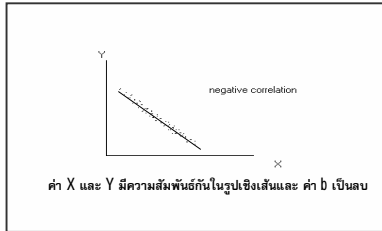
1. ศึกษารูปแบบความสัมพันธ์ระหว่างตัวแปร X และ Y



ความสัมพันธ์ในรูปเส้นตรงและเป็นบวก

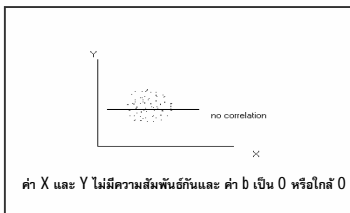


รูปแบบความสัมพันธ์ระหว่างตัวแปร X และ Y



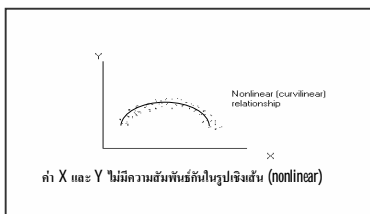
ความสัมพันธ์ในรูปเส้นตรงและเป็นลบ

รูปแบบความสัมพันธ์ระหว่างตัวแปร X และ Y



ไม่มีความสัมพันธ์กัน

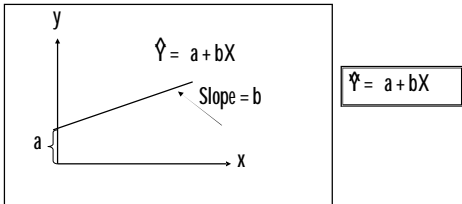
รูปแบบความสัมพันธ์ระหว่างตัวแปร X และ Y



ความสัมพันธ์ในรูปเส้นโค้งและเป็นลบ



2. การศึกษาความสัมพันธ์ระหว่าง X และ Y ในรูปเชิงเส้น



3. ทดสอบความสัมพันธ์ระหว่างตัวแปร X กับ y

สมมติฐานการทดสอบ

$H_0: \beta = 0$ หรือ ตัวแปรตามไม่มีความสัมพันธ์กับตัวแปรอิสระในรูปแบบเชิงเส้น
 $H_1: \beta \neq 0$ หรือ ตัวแปรตามมีความสัมพันธ์กับตัวแปรอิสระในรูปแบบเชิงเส้น

$H_0: \beta = 0$ หรือ ความพึงพอใจต่อบริการไม่ขึ้นกับอายุในรูปแบบเชิงเส้น
 $H_1: \beta \neq 0$ หรือ ความพึงพอใจต่อบริการขึ้นกับอายุในรูปแบบเชิงเส้น

3. ทดสอบความสัมพันธ์ระหว่างตัวแปร X กับ y

สถิติทดสอบ

$$F\text{-test} : F = \frac{\text{Mean square regression (MSR)}}{\text{Mean square error (MSE)}}$$

$$t\text{-test} : t = \frac{b}{SE_b}$$



4. ศึกษาระดับความสัมพันธ์ระหว่างตัวแปร X กับ Y

สถิติที่ใช้วัดระดับความสัมพันธ์คือ ค่าสัมประสิทธิ์การตัดสินใจ หรือค่า R^2

$$R^2 = \frac{\text{Sum square regression (SSR)}}{\text{Sum square total (SST)}}$$

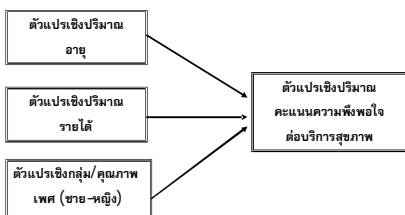
ถ้า R^2 มีค่าใกล้ 1 แสดงว่า ตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้มาก

ถ้า R^2 มีค่าน้อย หรือใกล้ 0 แสดงว่าตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้น้อย

การวิเคราะห์ถดถอยเชิงพหุ (Multiple regression analysis)

เป็นเทคนิคการวิเคราะห์ความสัมพันธ์ของตัวแปรหลายตัว ซึ่งประกอบด้วยตัวแปรตาม (dependent variable) 1 ตัว ซึ่งเป็นตัวแปรเชิงปริมาณ สเกลแบบช่วง หรือ สเกลอัตราส่วน (interval/ratio scale) และ ตัวแปรต้น/ตัวแปรอิสระ (independent variable) หลายตัว ซึ่งเป็นตัวแปรเชิงปริมาณ หรือตัวแปรเชิงกลุ่ม/เชิงคุณภาพ

การวิเคราะห์ถดถอยเชิงพหุ





วัตถุประสงค์ของการวิเคราะห์ถดถอยเชิงพหุ

1. เพื่อศึกษาว่าปัจจัย/ตัวแปรอิสระใดบ้างที่มีอิทธิพลต่อตัวแปรตาม

2. เพื่อประมาณหรือพยากรณ์ค่าตัวแปรตาม (Y) เมื่อทราบค่าตัวแปรอิสระ (X)

หลักการของการวิเคราะห์ถดถอยเชิงพหุ

1. ประเภทของข้อมูล

- ตัวแปร ประกอบด้วยตัวแปรอิสระ/ต้น ≥ 2 ตัว และตัวแปรตาม 1 ตัว
- ทั้งตัวแปรอิสระและตัวแปรตามเป็นตัวแปรเชิงปริมาณที่มีสเกลการวัดเป็นช่วง (interval scale) หรือ อัตราส่วน (ratio scale)
 - กรณีตัวแปรอิสระเป็นตัวแปรเชิงคุณภาพที่มีสเกลการวัดเป็นนามมาตรา (nominal scale) หรือ อันดับ (ordinal scale) ต้องมีการแปลงตัวแปรให้เป็นตัวแปรหุ่นก่อน (dummy variable)
 - กรณีตัวแปรตามเป็นตัวแปรเชิงคุณภาพ/กลุ่ม ต้องใช้สถิติ logistic regression

หลักการของการวิเคราะห์ถดถอยเชิงพหุ

2. ขนาดของกลุ่มตัวอย่าง

- จำนวนต่ำสุดที่ยอมรับได้คือ 5 รายต่อ 1 ตัวแปรอิสระ/ตัวแปรทำนาย
- จำนวนที่เหมาะสมอยู่ระหว่าง 15-20 รายต่อ 1 ตัวแปรอิสระ/ตัวแปรทำนาย
- กรณีใช้เทคนิค stepwise ใช้ตัวอย่าง 50 รายต่อ 1 ตัวแปรอิสระ
- ใช้การวิเคราะห์อำนาจการทดสอบ (power analysis) กำหนดขนาดกลุ่มตัวอย่าง



เงื่อนไขของการวิเคราะห์ถดถอยเชิงพหุ
(Assumption of regression analysis)

1. ตัวแปรตามและตัวแปรต้นมีความสัมพันธ์กันเชิงเส้นตรง (linearity)
2. ตัวแปรที่มีการแจกแจงแบบปกติ (normality)
3. ค่าความคลาดเคลื่อนมีการแจกแจงแบบปกติด้วยค่าเฉลี่ยที่เป็นศูนย์ (normality)
4. ค่าแปรปรวนของค่าคลาดเคลื่อนคงที่ (constant variance/homoscedasticity)

เงื่อนไขของการวิเคราะห์ถดถอยเชิงพหุ
(Assumption of regression analysis)

5. ค่าความคลาดเคลื่อนและค่าประมาณตัวแปรตามมีความสัมพันธ์กันเชิงเส้นตรง (linearity)
6. ค่าความคลาดเคลื่อนเป็นอิสระต่อกัน (autocorrelation)
7. ตัวแปรอิสระไม่มีความสัมพันธ์กันเองสูง (multicollinearity)

การตรวจสอบเงื่อนไขของการวิเคราะห์ถดถอยเชิงพหุ

Assumption	วิธีการตรวจสอบ
1. Normality	1. Histogram (ตัวแปรแต่ละตัวและ ZRESID) 2. Normal probability plot (ระหว่าง ZRESID กับ ZPRED) 3. ใช้สถิติ Shapiro-Wilk test (กลุ่มตัวอย่าง < 50) Kolmogorov-Smirnov test (กลุ่มตัวอย่าง ≥ 50)



การตรวจสอบเงื่อนไขของการวิเคราะห์ถดถอยเชิงพหุ

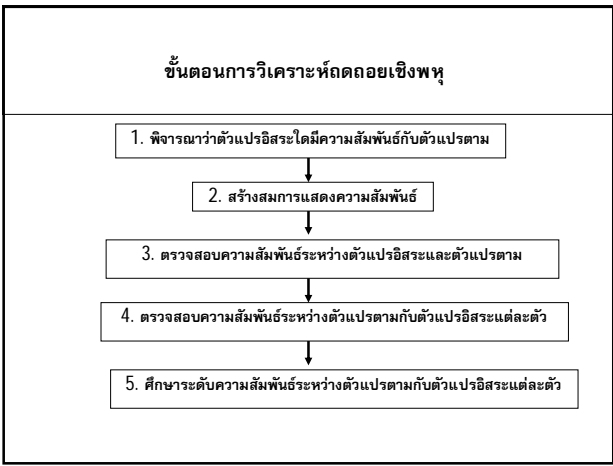
Assumption	วิธีการตรวจสอบ
2. linearity	1. Scatter plot (ZRESID & Predicted DV หรือ ZPRED) 2. Partial regression plot ระหว่าง DV และ IV แต่ละตัว

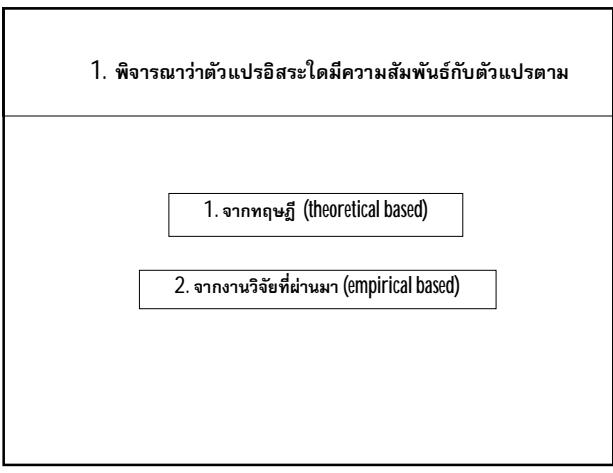
การตรวจสอบเงื่อนไขของการวิเคราะห์ถดถอยเชิงพหุ

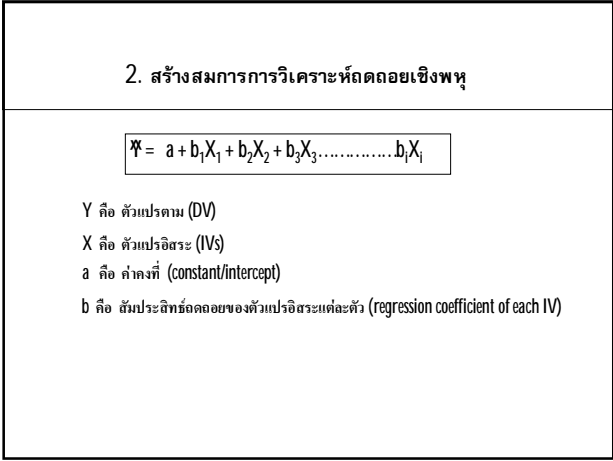
Assumption	วิธีการตรวจสอบ
3. Homoscedasticity	1. Scatter plot (ZRESID on the y axis กับ predicted DV on the x axis)

การตรวจสอบเงื่อนไขของการวิเคราะห์ถดถอยเชิงพหุ

Assumption	วิธีการตรวจสอบ
4. Autocorrelation	1. สถิติทดสอบ Durbin-Watson มีค่าเข้าใกล้ 2 (ค่าระหว่าง 1.5-2.5)
5. Multicollinearity	1. ดูจากค่าสัมประสิทธิ์สหสัมพันธ์ $r > .80$ หรือ $> .65$ 2. ดูจากค่า VIF > 5 ; TOF เข้าใกล้ 0









3. การตรวจสอบความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระ

Testing the significance of overall model (R^2)

ทดสอบว่าตัวแปรอิสระสามารถอธิบาย/ทำนายตัวแปรตามได้อย่างมีนัยสำคัญหรือไม่ และอธิบายได้มากน้อยเพียงใด

สมมติฐาน

H_0 : ตัวแปรตามไม่ขึ้นกับตัวแปรอิสระทั้ง k ตัว	$H_0: b_1 = b_2 = \dots = 0$
H_1 : ตัวแปรตามขึ้นกับตัวแปรอิสระอย่างน้อย 1 ตัว	$H_1: b_j \neq 0$ อย่างน้อย 1 ค่า

สถิติทดสอบ

$$F\text{-test} : F = \frac{\text{Mean square regression (MSR)}}{\text{Mean square error (MSE)}}$$

สถิติที่ได้จากการวิเคราะห์ถดถอยเชิงพหุ

1. Coefficient of determination (R^2): สัมประสิทธิ์การตัดสินใจพหุ

เป็นค่าที่บอกว่าตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้เท่าไรโดยค่าจะอยู่ระหว่าง 0-1

การอ่านค่าจะอ่านเป็นร้อยละ เช่น R^2 เท่ากับ .32 แสดงว่าตัวแปรอิสระสามารถอธิบายความผันแปรของตัวแปรตามได้ร้อยละ 32

สถิติที่ได้จากการวิเคราะห์ถดถอยเชิงพหุ

2. Adjusted Coefficient of determination (Adjusted R^2)

เป็นค่า R^2 ที่มีการปรับโดยนำจำนวนตัวแปรอิสระและจำนวนตัวอย่างเข้ามาพิจารณาด้วย

ค่า adjusted R^2 จะมีค่าต่ำกว่าค่า R^2 เล็กน้อย



4. การตรวจสอบความสัมพันธ์และระดับความสัมพันธ์

Testing the significance of regression coefficient

ทดสอบว่าตัวแปรอิสระตัวใดที่สามารถอธิบาย/ทำนายตัวแปรตามได้อย่างมีนัยสำคัญ และตัวแปรใดมีความสัมพันธ์/มีอิทธิพลต่อตัวแปรตามมาก -น้อยที่สุด

สมมติฐาน

H_0 : ตัวแปรตามไม่ขึ้นกับตัวแปรอิสระตัวที่ i หรือ $H_0: \beta_i = 0$
 H_1 : ตัวแปรตามขึ้นกับตัวแปรอิสระตัวที่ i หรือ $H_1: \beta_i \neq 0$

สถิติทดสอบ

t-test : $t = b/SE_b$

สถิติที่ได้จากการวิเคราะห์ถดถอยเชิงพหุ

Unstandardized Regression Coefficient (b):
สัมประสิทธิ์ถดถอยไม่ปรับมาตรฐาน

Standardized Regression Coefficient (β:
Beta): สัมประสิทธิ์ถดถอยปรับมาตรฐาน

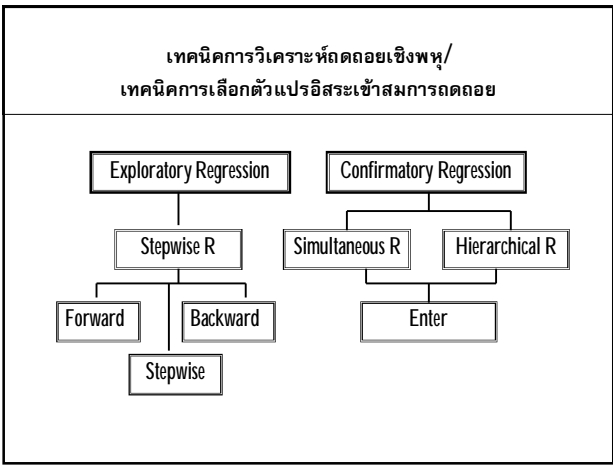
- ⊗ เป็นค่าที่บอกถึงปริมาณการเปลี่ยนแปลงของตัวแปรตามเมื่อตัวแปรอิสระเปลี่ยนแปลงไป 1 หน่วย
- ⊗ เครื่องหมาย + - บอกทิศทางความสัมพันธ์ระหว่างตัวแปรตามกับตัวแปรอิสระแต่ละตัว

สถิติที่ได้จากการวิเคราะห์ถดถอยเชิงพหุ

Unstandardized Regression Coefficient (b):
สัมประสิทธิ์ถดถอยไม่ปรับมาตรฐาน

Standardized Regression Coefficient (β):
สัมประสิทธิ์ถดถอยปรับมาตรฐาน

- ⊗ การอ่านค่าจะอ่านเช่นเดียวกับค่าสัมประสิทธิ์สหสัมพันธ์ (r) โดยเปรียบเทียบค่าสัมประสิทธิ์ถดถอยของตัวแปรอิสระแต่ละตัว เช่น X_1 มีค่า Beta = 0.65; X_2 มีค่า Beta = 0.33; X_3 มีค่า Beta = -0.217 สรุปได้ว่าตัวแปรตามจะขึ้นอยู่กับตัวแปร X_1 มากที่สุด รองลงมาได้แก่ตัวแปร X_2 และ X_3 ตามลำดับ



Confirmatory Regression

<p style="text-align: center;">การวิเคราะห์ถดถอยแบบมาตรฐาน (Simultaneous Regression)</p> <ul style="list-style-type: none"> ⊗ ใส่ตัวแปรอิสระทุกตัวเข้าในการวิเคราะห์ในครั้งเดียวกัน/พร้อมกัน ⊗ ใช้ในกรณีที่ตัวแปรที่ต้องการศึกษามีความสำคัญพอ ๆ กัน 	<p style="text-align: center;">การวิเคราะห์ถดถอยแบบเชิงชั้น (Hierarchical Regression)</p> <ul style="list-style-type: none"> ⊗ ใส่ตัวแปรอิสระเข้าในการวิเคราะห์ทีละตัว หรือ ทีละกลุ่มตามลำดับ ⊗ ความสำคัญ มาก → น้อย หรือ น้อย → มาก
--	---

**การวิเคราะห์ถดถอยแบบเป็นขั้นตอน
(Stepwise Regression)**

<p style="text-align: center;">Forward</p> <ul style="list-style-type: none"> ⊗ ตัวแปรอิสระจะถูกใส่เข้าไปทีละตัว ⊗ โดยโปรแกรมคอมพิวเตอร์จะเป็นตัวเลือกลำดับของตัวแปรอิสระที่จะใส่เข้าไปในการวิเคราะห์ ตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามสูงสุดจะถูกใส่เข้าไปก่อน 	<p style="text-align: center;">Backward</p> <ul style="list-style-type: none"> ⊗ ตัวแปรอิสระทุกตัวถูกใส่เข้าไปในการวิเคราะห์พร้อมกันและถูกเอาออกจากการวิเคราะห์ทีละตัว <p style="text-align: center;">Stepwise</p> <ul style="list-style-type: none"> ⊗ การเลือกตัวแปรอิสระเข้าในการวิเคราะห์โดยใช้หลักการทั้ง Forward และ backward
---	---



คำสั่งโปรแกรม SPSS ที่ใช้ในการวิเคราะห์ความถดถอย

Analyze → Regression → Linear

Computer output of MR

Table 1: Variables entered/removed

Model	Variables entered	Variables removed	Method
1	Education level (yr), Months since hire, Minority classification Previous experience (mo), Beginning salary		Enter

a. All requested variable entered
b. Dependent variable: Current salary

Computer output MR

Table 2: Model Summary

Model	R	R ²	Adjusted R ²	SEE
1	.901 ^a	.811	.809	7461.92

a. Predictors: (constant), education level, months since hire, minority classification
previous experience, beginning salary



Computer output MR

Table 3: ANOVA^b

Model	SS	df	MS	F	Sig
1 Regression		5		401.787	.000
Residual		468			
Total		473			

a. Predictors: (constant), education level, months since hire, minority classification
previous experience, beginning salary
b. Dependent variable: Current salary

Table 4: Coefficients^a

Model	Unstandardized coefficients		Standardized coefficients	t	Sig
	B	SE	Beta		
(Constant)	-16010.296			-4.917	.000
Beginning sal	1.759		.811	29.714	.000
Previous exp	-16.682		-.102	-4.679	.000
Minority classif	-1006.986		-.024	-1.185	.237
Months since hir	163.352		.096	4.767	.000
Educational level	671.83		.114	4.059	.000

a. Dependent variable: Current salary

สวัสดิ์ค่ะ

ผศ.ดร.ปิยะนุช จิตตานุนท์
คณะพยาบาลศาสตร์ ม.สงขลานครินทร์
