Multinomial Logistic Regression

by

Kanlaya Vanichbuncha: Ph.D

กำหนดการอบรม

วันที่ 17 ธันวาคม 2562

9.00 -12.00 Multiple Linear Regression

13.00 – 16.00 : Exploratory Factor Analysis

วันที่ 18 ธันวาคม 2562

9.00 -10.00 Multiple Linear Regression and Factor Analysis

10.00 -12.00 Binary Logistic Regression

13.00 – 16.00 Binary and Multinomial Logistic Regression

วันที่ 19 ธันวาคม 2562

9.00 -12.00 Multinomial Logistic Regression

13.00 – 16.00 :Log-Linear Model

Note : Coffee break :10.30-10.50 , 14.30 – 14.50

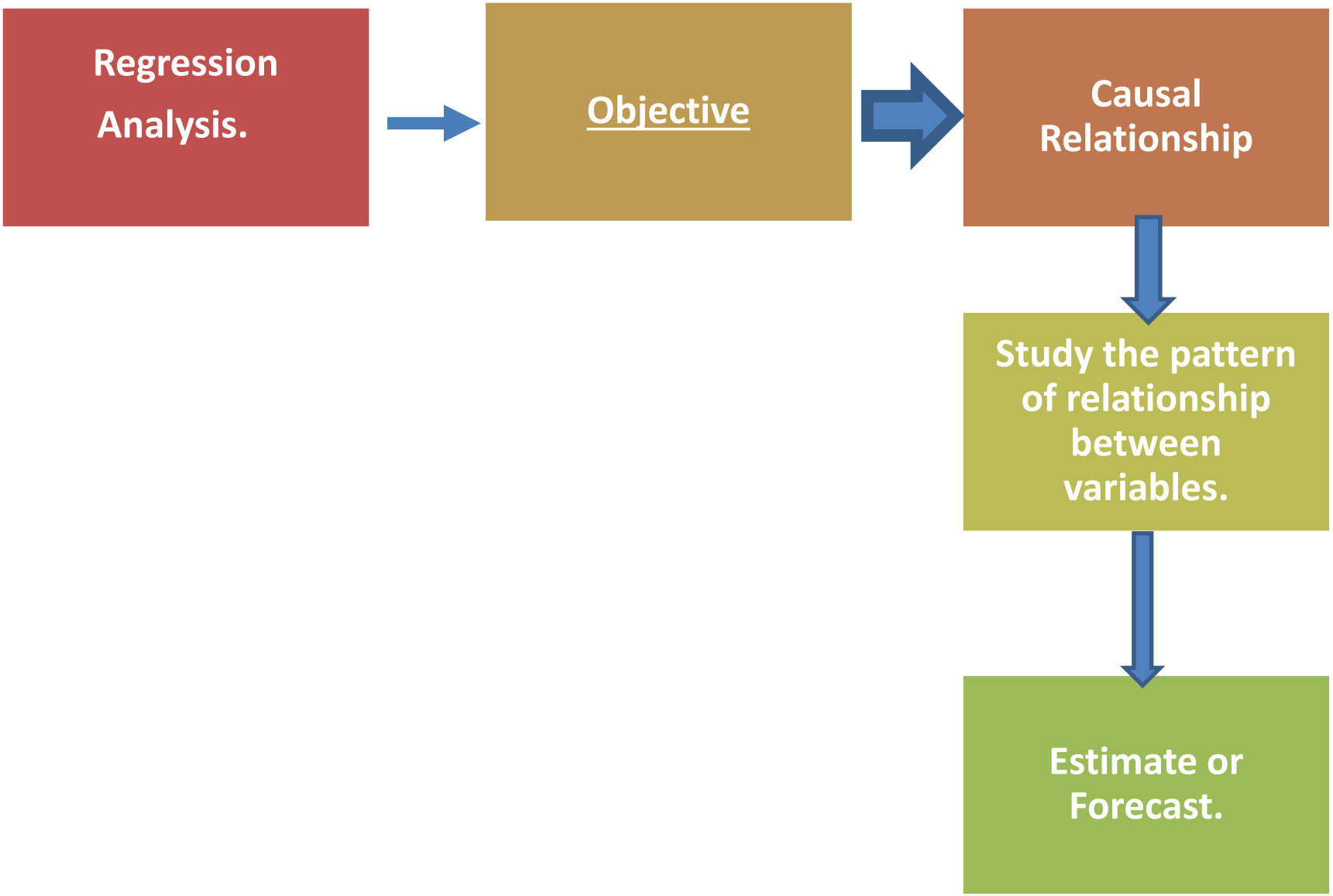
Regression Analysis.

- Analyze the relationship among variables.
- Causal Relationship

Objective

1. Study the pattern of relationship between variables.
2. Estimate or Forecast.

- Simple Linear Regression
- Multiple Linear Regression


Regression Analysis.
Objective
Causal Relationship
Study the pattern of relationship between variables.
Estimate or Forecast.

Type of Data

1

Cross-sectional data

2

Panel data (Time – series data + Cross-sectional data)

3

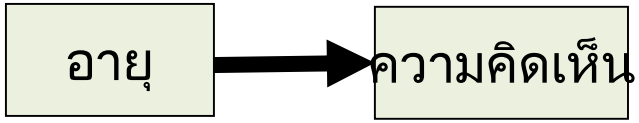
Panel data

(Longitudinal data + Cross-sectional data)

Simple Regression

One dependent and one independent variable. Both of them are quantitative.

Independent (X)	Dependent (Y)
อายุ	ความคิดเห็น



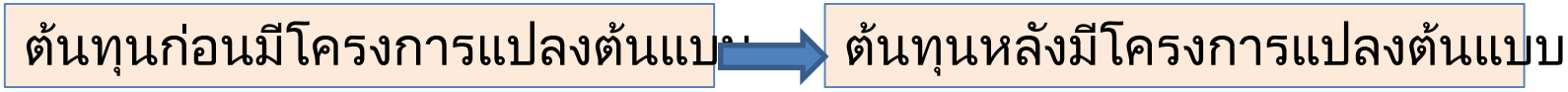


Figure 1 : Positive linear relationship

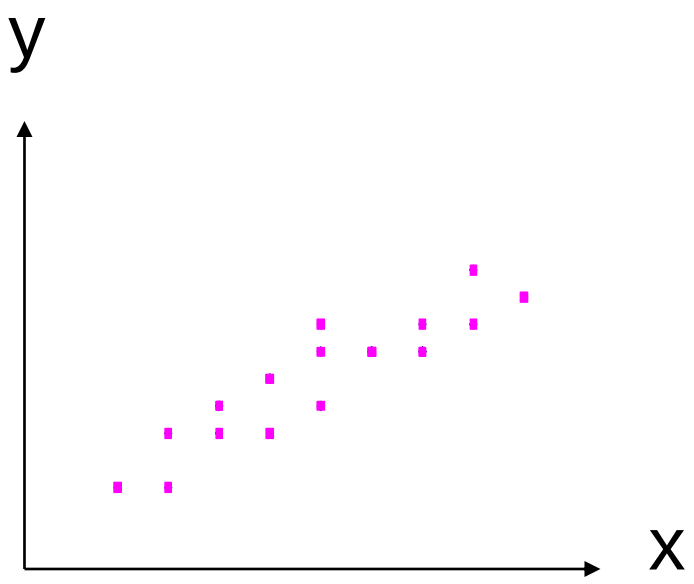


Figure 2 : Negative linear relationship

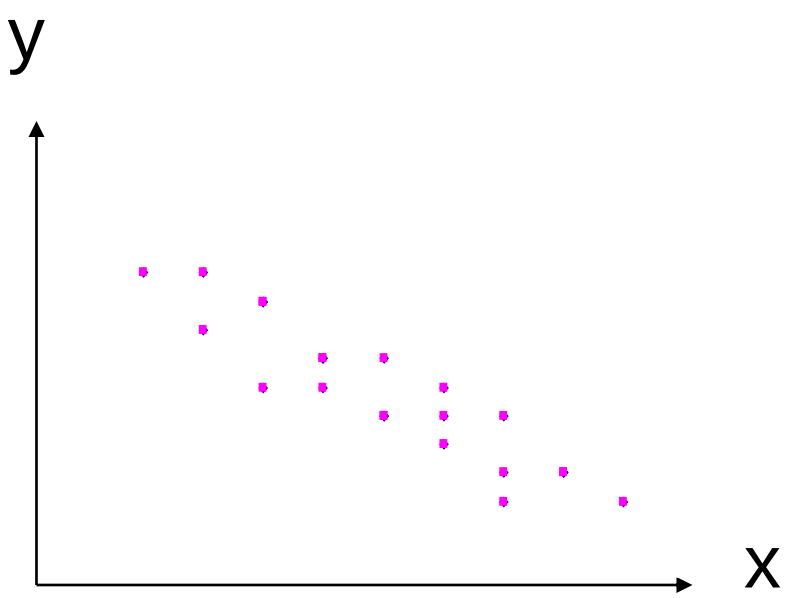


Figure 3

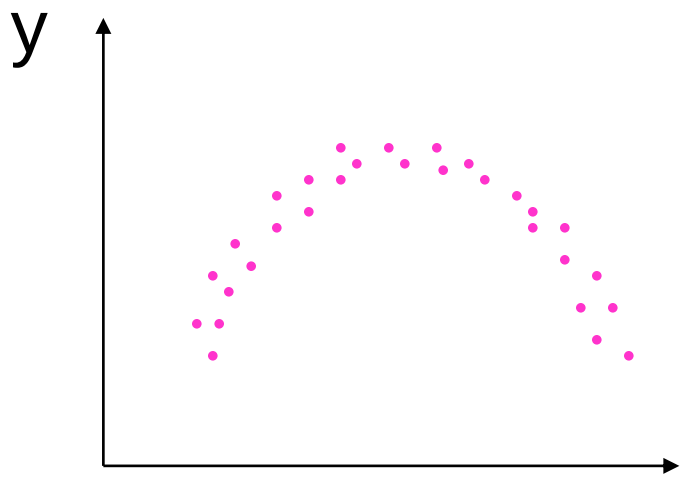


Figure 4

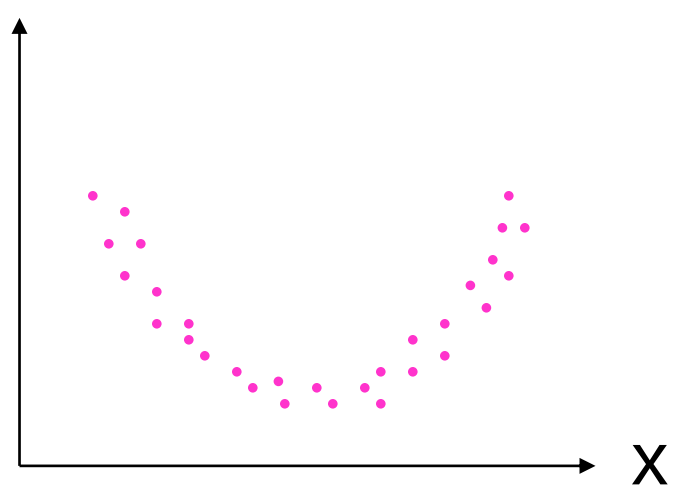


Figure 5

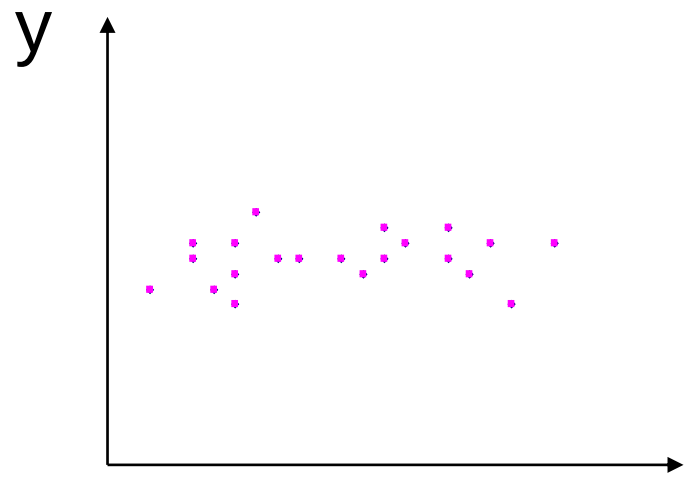
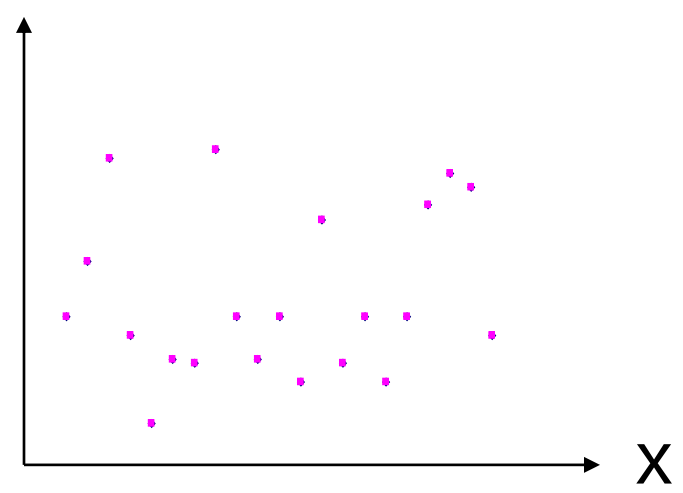


Figure 6



Simple Linear Regression Analysis

The relationship between dependent (Y) and independent (X) is linear form.

$$Y = \beta_0 + \beta_1 X + e \qquad \text{i = 1,2,..., N}$$

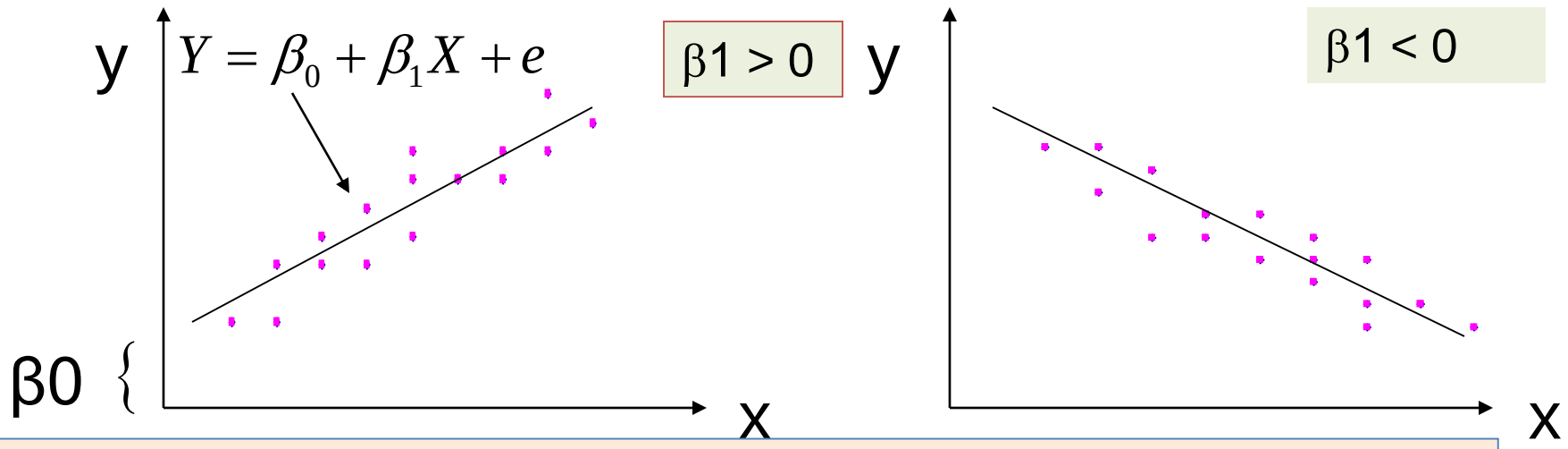
Y = dependent variable – ความพอใจ

X = independent variable – ขนาดพื้นที่

β_1 = slope of regression line

e = error – difference between actual value (Y) and Predicated ($\hat{Y}$) value

Predicted by regression line ($\hat{Y}$)



β1 = slope of the line or the average changed in Y for each change of one unit (either increase or decrease) in X

β1 = regression coefficients

β0 = Y- intercept or the estimated value of Y when X = 0

The β0 unit and β1 are the same of dependent variable Y.

For Sample Data

$$\hat{Y}_i = b_0 + b_1 X_i \ldots(1)$$

$$Y_i = \beta_0 + \beta_1 X_i + e_i \ldots(2)$$

(2) - (1) $Y_i - \hat{Y}_i = e_i$

$$\hat{\beta}_0 = b_0, \hat{\beta}_1 = b_1$$

Testing Hypothesis about β0 and β1

1. Test about β (slope of linear line)

$H_0 : \beta 1 = 0$

$H_1 : \beta 1 \neq 0$

or H_0 : There is no linear relationship between Y and X.

H_1: X and Y are linearly related.

Test Statistics

$$t_{n-2} = \frac{b_1 - 0}{SE(b_1)}$$

Reject H_0 if $t > t_{1-\alpha/2,n-2}$ or $t < -t_{1-\alpha/2,n-2}$

The coefficient of Determination : R^2

To the proportion of total variation in dependent variable Y that is explained by the variation of independent variable X.

SS total = SS regression + SS Error

SST = SS Reg + SSE

$$R^2 = \frac{SS\,\mathrm{Re}\,g}{SST} =$$

$0 \le R^2 \le 1$ or $0\% \le R^2 \le 100\%$

$$\text{SSTotal} = \text{SST} = \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2$$

$$\text{SSReg} = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2$$

$$\text{SSE} = \text{SSError} = \text{SSResidual} = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

For Sample Data

Regression equation

$$\hat{Y}_i = b_0 + b_1 X_i$$

คะแนนความพอใจ = 0.75 + 0.578 ขนาดพื้นที่

ถ้าขนาดพื้นที่เพิ่มขึ้น 10 rai (1 unit = 10 rai) ➔ คะแนนความพอใจจะเพิ่มขึ้นโดยเฉลี่ย 0.578 คะแนน

R Square = 65.5%

ขนาดของพื้นที่มีอิทธิพลต่อคะแนนความผันแปรของคะแนนความพอใจ 65.5 % อีก 3 เป็นอิทธิพลของปัจจัยอื่นๆ

The Coefficient of Correlation

Pearson correlation:

Describes the strength of the relationship between two sets of quantitative variable.

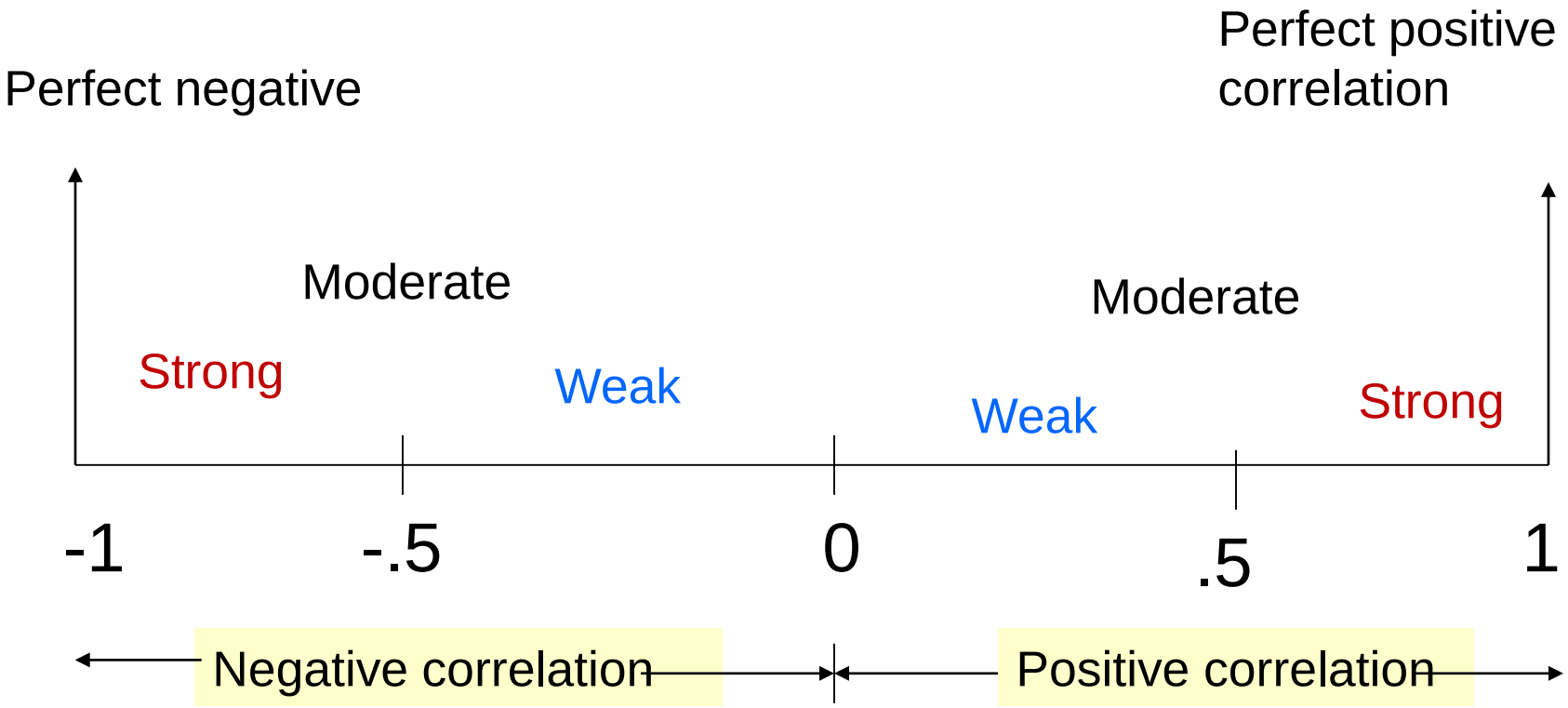
$$-1 \leq r \leq +1$$

r = +1, or r = -1 : indicate perfect correlation.

r = +1 : X and Y are perfectly related in a positive linear.

r = -1 : X and Y are perfectly related in a negative linear.

r = 0 : There is absolutely no relationship between two variables.


Perfect negative
Perfect positive correlation
Moderate
Moderate
Strong
Weak
Weak
Strong
-1
-.5
0
.5
1
Negative correlation
Positive correlation

Conditions or Assumptions

1. error e_i ~ normal (0, σ^2)
2. V(e) = σ^2 is constant
3. e_t and e_{t+1} are independent

Properties of a and b for Least Square Method

1. $\sum e_i = \sum (Y_i - \hat{Y}_i) = 0$
2. $(\bar{X}, \bar{Y})$ is the point on regression line
3. $\sum e_i^2$ is minimum

Examining Condition

1. e is normal

   - Kolmogorov-Smirnov Test (any sample size : n)

   - Shapiro-Wilk ($n \leq 50$)

2. V(e) is constant

   (if V(e) is not constant, Heteroscedastic Problem)

   - Plot graph between e and $\hat{Y}$ or X

3. e_t and e_{t+1} are independent

   - Durbin-Watson

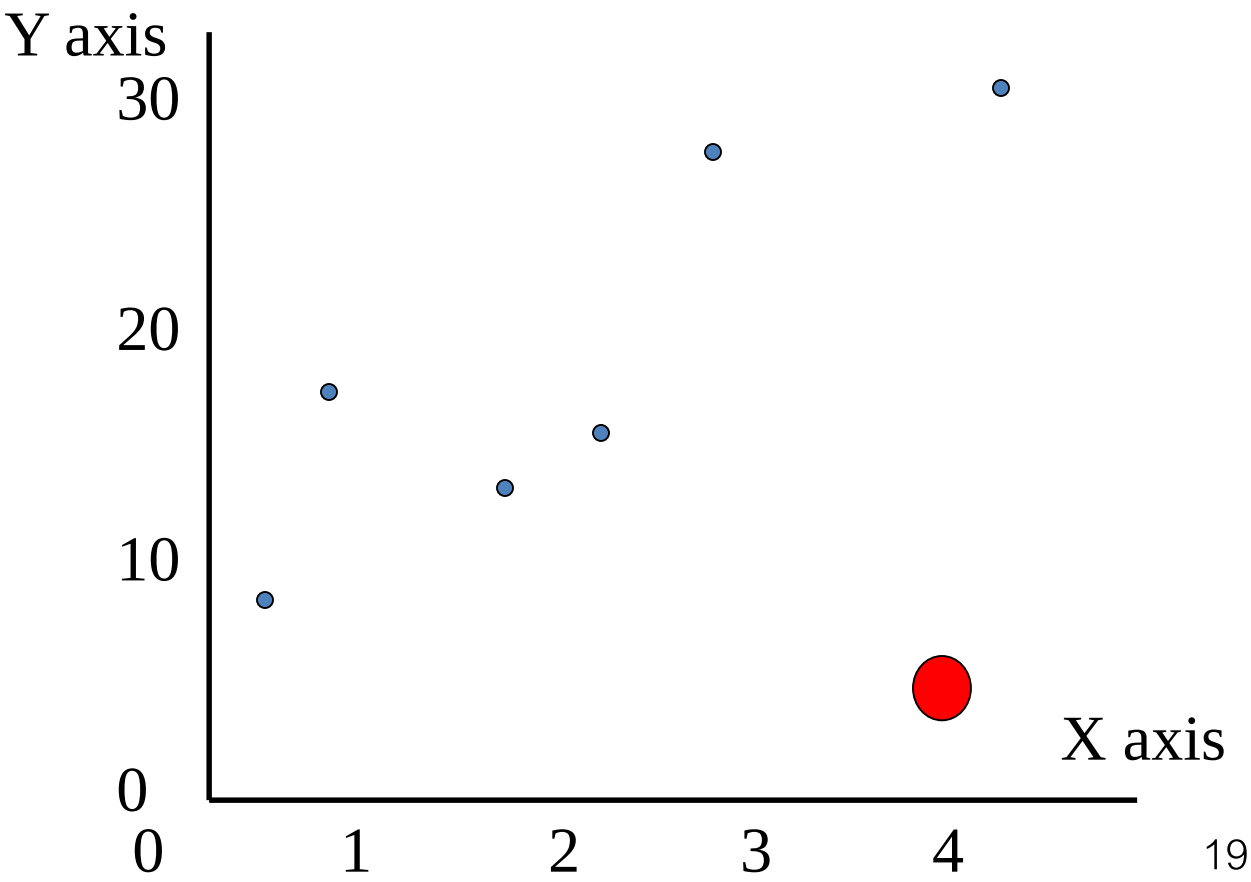
Regression: Outliers

- Choose “influence” and “distance” statistics such as Cook’s Distance, DFFIT, standardized residual
- High values signal potential outliers
- Note: This is less useful if you have a VERY large dataset, because you have to look at each case value.

Scatterplots

- Example: Study time and student achievement.
  - X variable: Average # hours spent studying per day
  - Y variable: Score on reading test

Case	X	Y
1	2.6	28
2	1.4	13
3	.65	17
4	4.1	31
5	.25	8
6	1.9	16
7	3.5	6

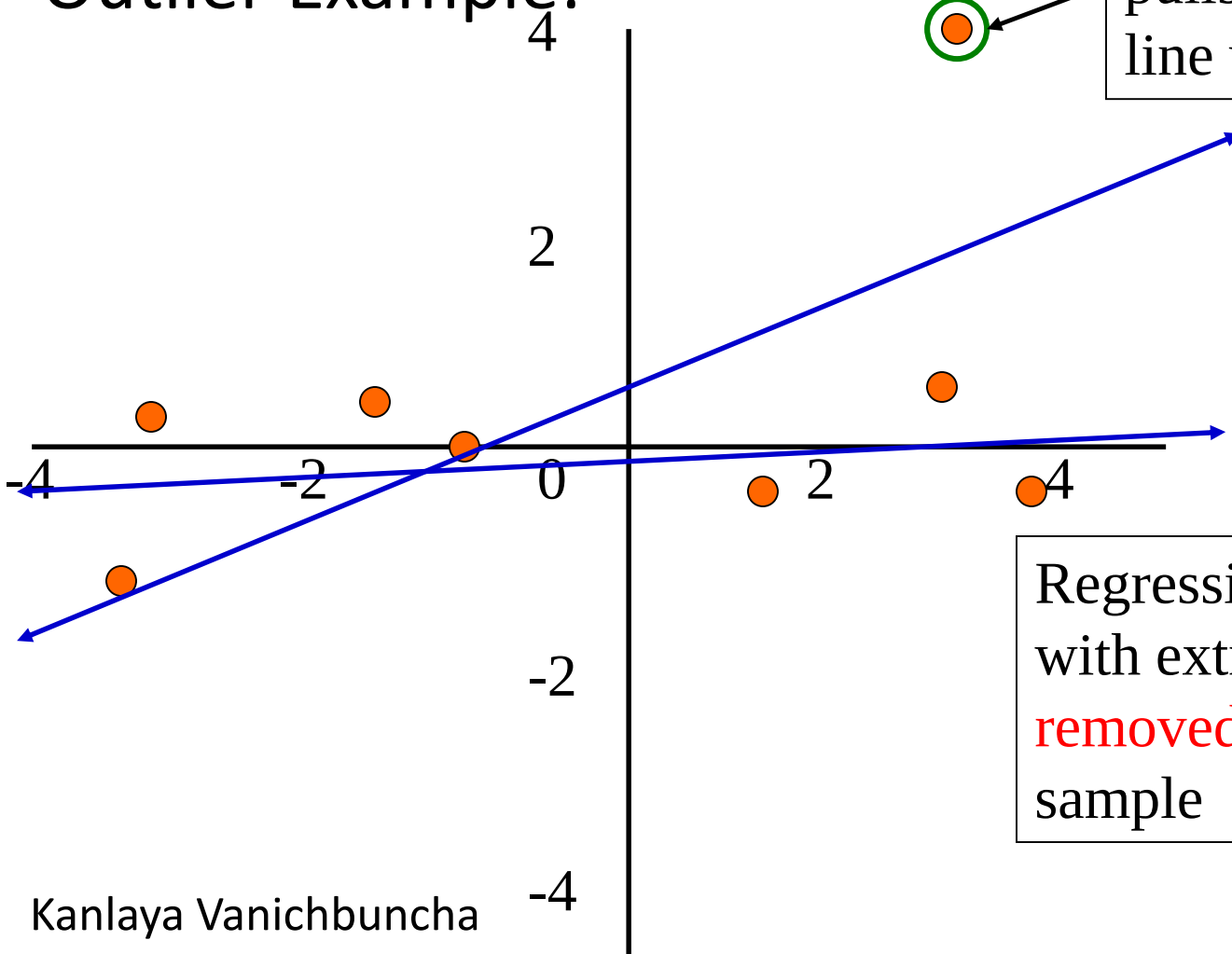


Regression: Outliers

- Note: Even if regression assumptions are met, slope estimates can have problems
- Example: Outliers -- cases with extreme values that differ greatly from the rest of your sample
  - More formally: "influential cases"
- Outliers can result from:
  - Errors in coding or data entry
  - Highly unusual cases
  - Or, sometimes they reflect important "real" variation
- Even a few outliers can dramatically change estimates of the slope, **especially if N is small**.

Regression: Outliers

- Outlier Example:



Extreme case that pulls regression line up

Regression line with extreme case removed from sample

Outlier Diagnostics

- Residuals: The numerical value of the **error**
  - Error = distance that points falls from the line
  - Cases with unusually large error may be outliers
  - Note: residuals have many other uses!
- Standardized residuals
  - Z-score of residuals… converts to a neutral unit
  - Often, standardized residuals larger than 3 are considered worthy of scrutiny
    - But, it isn't the best outlier diagnostic.

Outlier Diagnostics

- Cook's D: Identifies cases that are strongly influencing the regression line
  - SPSS calculates a value for each case
    - Go to "Save" menu, click on Cook's D
- How large of a Cook's D is a problem?
  - Rule of thumb: Values greater than: $4 / (n - k - 1)$
  - Example: N=7, K = 1: Cut-off = 4/5 = .80
  - Cases with higher values should be examined.

Outlier Diagnostics

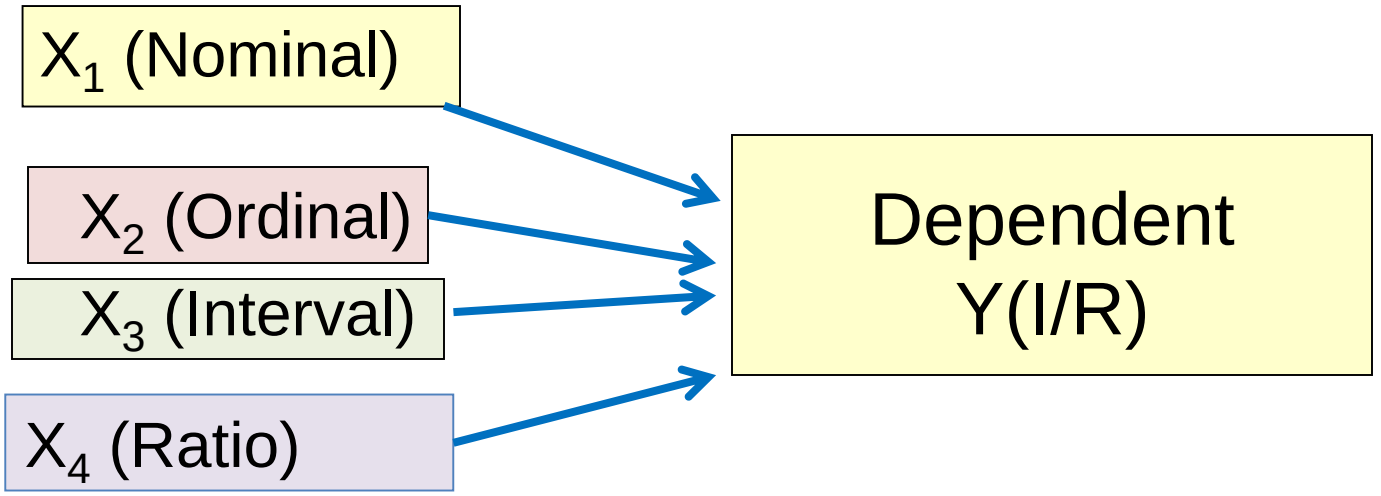
- Example: Outlier/Influential Case Statistics

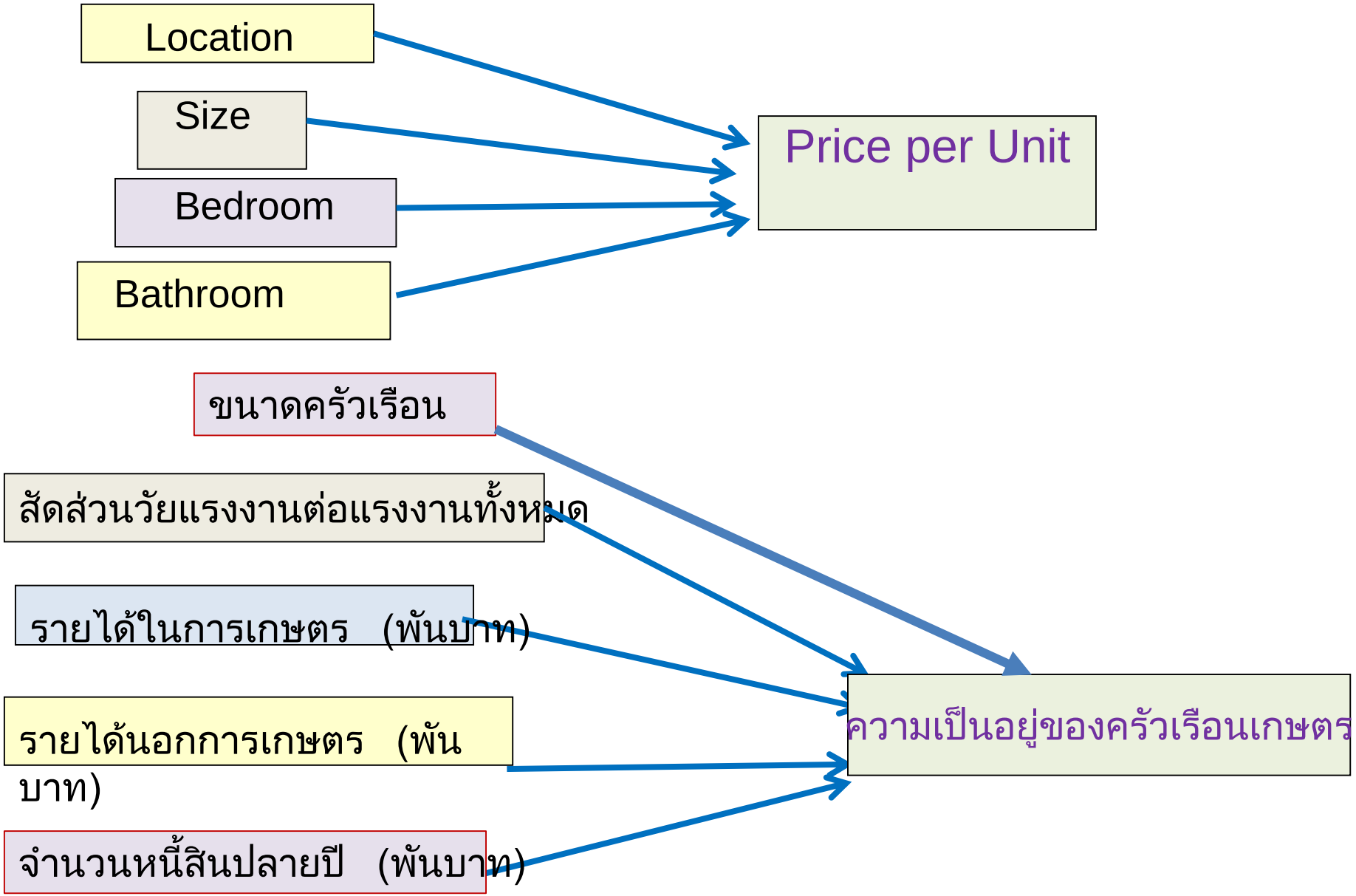
Hours	Score	Residual	Std Residual	Cook's D
2.60	28	9.32	1.01	.124
1.40	13	-1.97	-.215	.006
.65	17	4.33	.473	.070
4.10	31	7.70	.841	.640
.25	8	-3.43	-.374	.082
1.90	16	-.515	-.056	.0003
3.50	6	-15.4	-1.68	.941

Multiple Linear Regression

One dependent and at least 2 independent variables (k independent variables are potentially related to the dependent variable).

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + e \qquad k \geq 2$$




Location
Size
Bedroom
Bathroom
Price per Unit
ขนาดครัวเรือน
สัดส่วนวัยแรงงานต่อแรงงานทั้งหมด
รายได้ในการเกษตร (พันบาท)
รายได้นอกการเกษตร (พันบาท)
จำนวนหนี้สินปลายปี (พันบาท)
ความเป็นอยู่ของครัวเรือนเกษตร

Estimate Y by $\hat{Y}$

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_k X_k$$

conditions

1. Error ~ Normal (o,σ^2)
2. V(e) = σ^2 is constant (If not, Heteroscedastic)
3. e_t and e_{t+1} are independent
4. $X_1,\ldots,X_k$ are independent (If not, Multicollinearity)

For k = 3 $\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3$

Y = ความเป็นอยู่ของครัวเรือนเกษตร

X_1 = ขนาดครัวเรือน (คน)

X_2 = ร้อยละวัยแรงงานต่อแรงงานทั้งหมด

X_3 = รายได้ในการเกษตร (พันบาท)

ถ้า b_2 = 0.85 => ถ้าร้อยละวัยแรงงานต่อแรงงานทั้งหมดเพิ่มขึ้น 1 %

จะทำให้ความเป็นอยู่ของครัวเรือนเกษตรเพิ่มขึ้นโดยเฉลี่ย 0.85 โดยที่ขนาดครัวเรือน และ รายได้ในการเกษตรเหมือนกัน

ฐานะความเป็นอยู่ของครัวเรือน (Welfare) ในที่นี้ใช้สัดส่วนของรายได้สุทธิครัวเรือนต่อคน เทียบกับ รายได้ ณ ระดับเส้นความยากจน

Testing Hypothesis

Step 1 : Test effects of independent variables set

$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

H_1 : At least one of $\beta_i \neq 0$; i = 1,2,…,k

Test statistic F = MSReg / MSE

SV	df	SS	MS=SS/df	F
Regression (X)	k	SSReg	MSReg	MSReg/MSE
Error	n-k-1	SSE	MSE	
Total	n-1	SST		

Conclusion :

1.1 Accept $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

-No linear relationship between Y and $X_1, X_2, \ldots, X_k$ ➡ Stop

1.2 Reject H_0 => There is at least one of X_i's that, which linear relationship with Y ➡ Test for each X or for each β in Step 2

Step 2 : Test effects for each independent variable

$H_0 : \beta_i = 0$

$H_1 : \beta_i \neq 0$

Test statistic

$$t = \frac{b_i - 0}{SE(b_i)}$$

$$SE(b_i) = \frac{\sqrt{MSE}}{\sqrt{\sum X_i^2}}$$

Reject H_0 if $|t| > t_{1-\alpha/2,n-k-1}$

Coefficient of Determination

$$R^2 = \frac{SS\,\mathrm{Re}\,g}{SST} = 1 - \frac{SSE}{SST}$$

$$R^2 = \frac{SST - SSE}{SST}$$

R^2 measure of fit of the regression line

Note: As more independent variables (X) are added to regression model => R^2 will never decrease, even if the additional variables are not related to Y.

Adjusted R^2

$$R^2_{adj} = 1 - \frac{SSE/(n-k-1)}{SST/(n-1)}$$

$$= 1 + \frac{(n-1)}{(n-k-1)}(R^2 - 1)$$

Multiple Correlation Coefficient (R)

$$R = +\sqrt{R^2}$$

Multiple correlation coefficient = simple correlation between the predicted $\hat{Y}$ and the actual Y.

Multicollinearity

Problem of Multicollinearity exist when independent variables (X's) are correlated with one another.

Problems high degree of Multicollinearity among the independent variables.

1. Contradiction

   Step 1 : Accept $H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$

   but in Step 2 : Reject $H_0 : \beta_i = 0$

- Or **step1: reject Ho but in step 2 :accept Ho**

2. b unstable

3. SE(b) will be very large.

   t-Test too small => accept H_0 even when X and Y are strong related with Y

Detecting Multicollinearity

1. Compute the pairwise correlation between X's. Multicollinearity may be a serious problem if any pairwise correlation is bigger than 0.5 (X_1 may be highly correlated with $a_2X_2 + a_3X_3$)

2. Tolerance

Tolerance (Xi) = 1 - $R^2(X_i)$

$$0 \le R^2(X_i) \le 1$$

$$0 \le Tolerance \le 1$$

$R^2(X_i)$ = Coefficient of determination of

$$X_i = a + b_1X_1 + b_2X_2 + \ldots + b_{i-1}X_{i-1} + b_{i+1}X_{i+1} + \ldots + b_kX_k$$

3.VIF (Variance Inflation Factor)

$$VIF(X_i) = \frac{1}{1 - R^2(X_i)}$$

$$1 \le VIF(X_i) < \infty$$

1. $VIF(X_i) > 10$ => multicollinearity may be influencing the least square estimate of regression coefficients (b_i)

2. $\overline{VIF} = \frac{\sum_{j=1}^{k} VIF(X_j)}{k}$ > 10 => serious problem of multicollinearity

$\overline{VIF}$ indicate how many times larger the error sum of squares for the regression is due to multicollinearity than it would be if the variables were uncorrelated.

Corrections for Multicollinearity

1. Remove variables that are highly correlated with others.
2. Use Factor Analysis
3. Ridge Regression

Variables Selection

1. Enter / Remove
2. Forward Selection
3. Backward Elimination
4. Stepwise Regression

Backward elimination procedure

The backward elimination procedure is basically a sequence of tests for significance of explanatory variables.
Starting out with the maximum model

$$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon,$$

- we remove (or, *eliminate*) the variable with the highest p – value for t he test of significance of the variable , conditioned on the - value being bigger than some pre-determined level (say, 0.10).

- Next, we fit the reduced model (having removed the variable from the maximum model), and remove from the *reduced* model the variable with the highest p – value for the test of significance of that variable if **p – value > 0.10 , and so on**.

- The procedure ends when no more variables can be removed from the model at significance level 10%.
Note that we use the F -test criterion in this procedure.

The **forward selection procedure**

- The **forward selection procedure** is a reversed version of the backward elimination procedure.
- Instead of starting with the maximum model, and eliminating variables one by one, we start with an `empty' model with no explanatory variables, and add variables one by one until we cannot improve the model significantly by adding another variable.

Stepwise regression procedure

- Once a variable is added to the model, in forward selection, it stays in the model-irrespective of which other variables are added later on.
- However, it can easily happen that a variable entered early in the procedure becomes superfluous because of its interrelationship with other variables added to the model later on in the procedure.

The **stepwise regression procedure** modifies the forward selection procedure in the following way.
Each time a new variable is added to the model, the significance of each of the variables already in the model is re-examined.

- That is, at each step in the forward selection procedure, we test for significance of each of the variables currently in the model, and remove the one with the highest p-value .
- If p -value is above some threshold value, say 0.10).
- **The stepwise regression procedure continues until no more variables can be added or removed.**

Regression with Qualitative or Categorical Independent Variables

Y = dependent variable (quantitative variables)

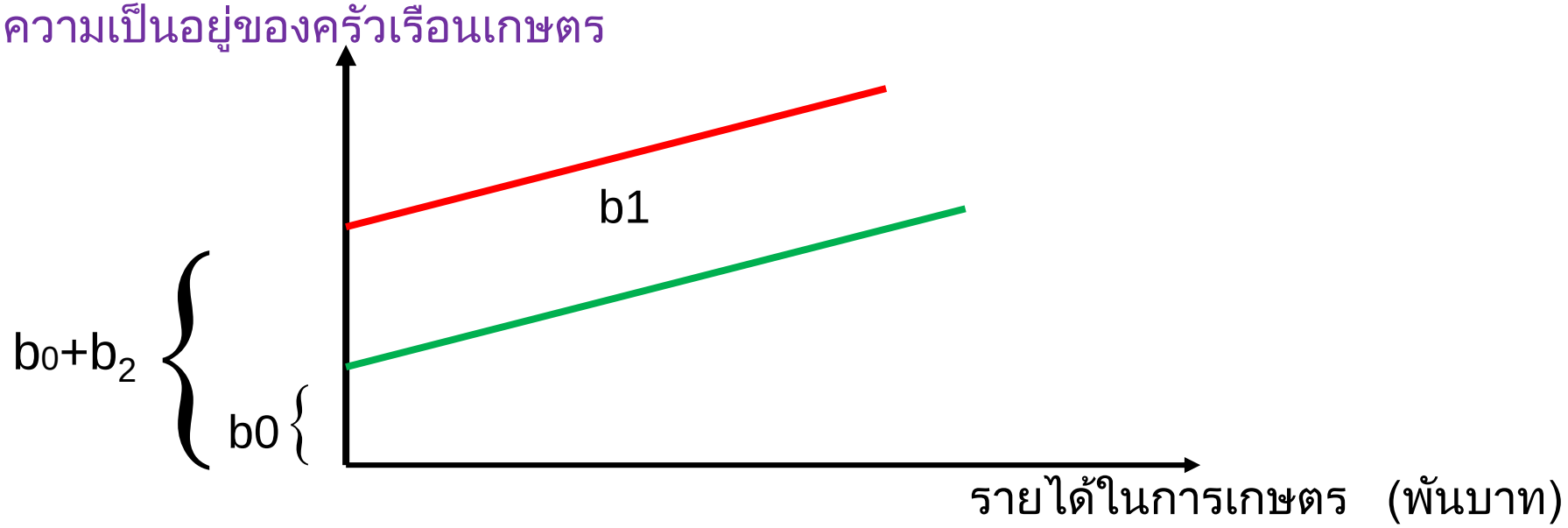
$X_1,\ldots, X_k$ = independent variable (quantitative and qualitative variables)

Y = ความเป็นอยู่ของครัวเรือนเกษตร

X_1 = รายได้ในการเกษตร (พันบาท) X2 = ภาค

- Number of Indicator Variables = Number of group -1

➢ Use Indicator variable or dummy variable for qualitative(Categorical Variable)variable.



Testing Hypothesis

Step 1 : Use ANOVA for

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

$$H_1 : \text{At least one of } \beta_i \neq 0 \; ; \; i = 1,2,\ldots,k$$

If Accept H_0 => stop

If Reject H_0 => go to step 2

Step 2

Use t-Test for

$H_0 : \beta_i = 0 \quad ; i = 1,2,\ldots,k$

$H_1 : \beta_i \neq 0$

β_2 = regression coefficient of Region

$H_0 : \beta_2 = 0$

$H_1 : \beta_2 \neq 0$

(3)-(4) => b_2 or β_2

b_2 = ผลต่างของความเป็นอยู่ของครัวเรือนเกษตร เฉลี่ยที่เกิดจากอิทธิพลของระหว่างภาคกลางและภาคอื่นๆ

ถ้ายอมรับ $H_0 : \beta_2 = 0$ => ไม่มีความแตกต่างของคะแนนพอใจเฉลี่ยที่เกิดจากอิทธิพลของรายได้ในการเกษตร (พันบาท)ระหว่างภาคตะวันออกเฉียงเหนือและภาคอื่นๆ

Example Y = ความเป็นอยู่ของครัวเรือนเกษตร

X_1 = รายได้ในการเกษตร (พันบาท)

X_2 = ภาค

ความเป็นอยู่ของครัวเรือนเกษตร $= \beta_0 + \beta_1$ รายได้ในการเกษตร + β_2ภาค + e

ความเป็นอยู่ของครัวเรือนเกษตร $= b_0 + b_1$ รายได้ในการเกษตร + b2 ภาค

$$\text{ภาค} = \begin{cases} 1 & \text{ถ้าเป็นภาคตะวันออกเฉียงเหนือ} \\ 0 & \text{ถ้าเป็นภาคอื่นๆ} \end{cases}$$

ฐานะความเป็นอยู่ของครัวเรือน (Welfare) ในที่นี้ใช้สัดส่วนของรายได้สุทธิครัวเรือนต่อคน เทียบกับ รายได้ ณ ระดับเส้นความยากจน

สำหรับภาคตะวันออกเฉียงเหนือ(ภาค = 1)

$\widehat{\text{ความเป็นอยู่ของครัวเรือนเกษตร}}$ = b0 + b1 รายได้ในการเกษตร + b2 (1)

= (b0 + b2) + b1 รายได้ในการเกษตร …(1)

สำหรับภาคอื่นๆ(ภาค = 0)

$\widehat{\text{ความเป็นอยู่ของครัวเรือนเกษตร}}$ = b0 + b1 รายได้ในการเกษตร + b2 (0)

= b0 + b1 รายได้ในการเกษตร …(2)

(1) – (2) = b2

ถ้า b2 = 3 หมายถึง ครัวเรือนเกษตรในภาคตะวันออกเฉียงเหนือ มีความเป็นอยู่ของครัวเรือนเกษตรโดยเฉลี่ย ดี กว่า ภาคอื่นๆ 3 เท่า เมื่อมีรายได้ในการเกษตรเท่ากัน

Y = ความเป็นอยู่ของครัวเรือนเกษตร X_1 = ขนาดพื้นที่ , X_2 = ภาค

$$X_2 = \begin{cases} 1; & \text{ภาคกลาง} \\ 2; & \text{ภาคเหนือ} \\ 3; & \text{ภาคอีสาน} \end{cases}$$

สร้างตัวแปร indicator 2 ตัว คือ

$$\text{ภาคกลาง} = \begin{cases} 1 ; & \text{ภาคกลาง} \\ 0 ; & \text{ภาคอื่นๆ} \end{cases}$$

$$\text{ภาคเหนือ} = \begin{cases} 1; & \text{ภาคเหนือ} \\ 0; & \text{ภาคอื่นๆ} \end{cases}$$

ภาคกลาง = ภาคเหนือ = 0 =>ภาคอีสาน

คะแนนพอใจ = b0 + b1 ขนาดพื้นที่ + b2 ภาคกลาง + b3 ภาคเหนื

1. ภาคกลาง =ภาคเหนือ = 0 => ภาคอีสาน

ความเป็นอยู่ของครัวเรือนเกษตร = b0 + b1ขนาดพื้นที่ + b2(0) + b3(0)

ความเป็นอยู่ของครัวเรือนเกษตร = b0 + b1ขนาดพื้นที่ …(3)

2. ภาคกลาง= 1 ภาคเหนือ = 0

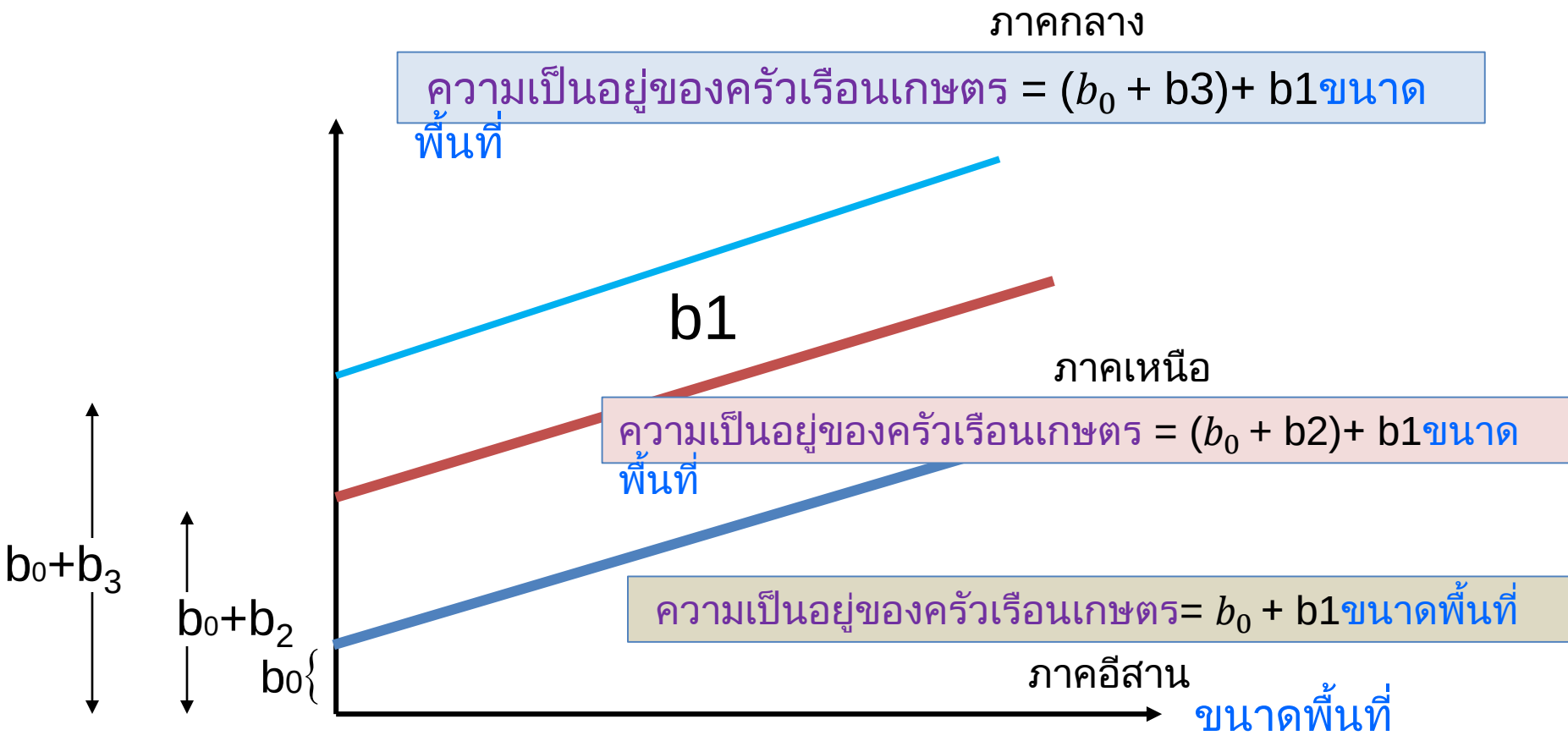
ความเป็นอยู่ของครัวเรือนเกษตร = b0 + b1ขนาดพื้นที่ + b2(1) + b3(0)

ความเป็นอยู่ของครัวเรือนเกษตร = (b0 + b2)+ b1ขนาดพื้นที่ …(4)

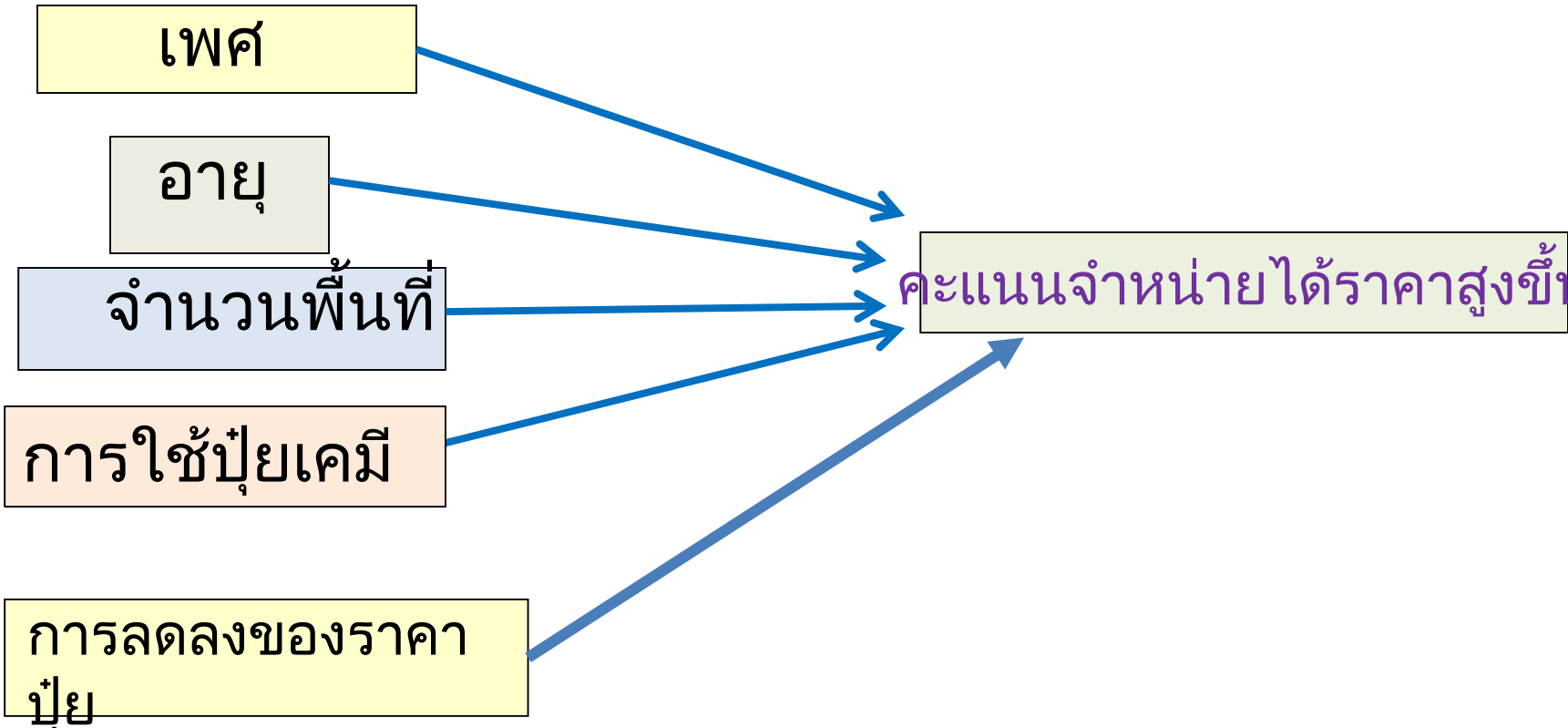
3. ภาคกลาง = 0 , ภาคเหนือ = 1

ความเป็นอยู่ของครัวเรือนเกษตร= b0 + b1ขนาดพื้นที่ + b2(0) + b3(1)

ความเป็นอยู่ของครัวเรือนเกษตร = (b0 + b3)+ b1ขนาดพื้นที่ …(5)


ภาคกลาง
ความเป็นอยู่ของครัวเรือนเกษตร = (b_0 + b3)+ b1ขนาดพื้นที่
b1
ภาคเหนือ
ความเป็นอยู่ของครัวเรือนเกษตร = (b_0 + b2)+ b1ขนาดพื้นที่
ความเป็นอยู่ของครัวเรือนเกษตร= b_0 + b1ขนาดพื้นที่
b0+b3
b0+b2
b0
ภาคอีสาน
ขนาดพื้นที่

Case Study 1


เพศ
อายุ
จำนวนพื้นที่
การใช้ปุ๋ยเคมี
การลดลงของราคาปุ๋ย
คะแนนจำหน่ายได้ราคาสูงขึ้น

ราคาสูงขึ้น $= \beta_0 + \beta_1$เพศ $+ \beta_2$อายุ $+ \beta_3$พื้นที่ $+ \beta_4$การใช้ปุ๋ยเคมี

$+ \beta_5$ต้นทุนปุ๋ยลดลง $+$ e

ราคาสูงขึ้น $= b_0 + b_1$เพศ $+ b_2$อายุ $+ b_3$พื้นที่ $+ b_4$การใช้ปุ๋ยเคมี

$+ b_5$ร้อยละต้นทุนปุ๋ยลดลง

เพศ 0 ; หญิง

การไม่ใช้ปุ๋ยเคมี = 0

Descriptive Statistics

	Mean	Std. Deviation	N
จำหน่ายได้ราคาสูงขึ้น	3.69	1.104	216
อายุ	52.44	10.550	216
เพศ	.56	.498	216
ขนาดพื้นที่(ไร่)	25.153	25.7575	216
การใช้ปุ๋ยเคมี	.76	.426	216
ร้อยละปุ๋ยเคมีที่ลดลง	48.258	27.112	216

Variables Entered/Removed[a]			
Model	Variables Entered	Variables Removed	Method
1	การใช้ปุ๋ยเคมี, ขนาดพื้นที่(ไร่), ร้อยละปุ๋ยเคมีที่ลดลง, เพศ, อายุ[b]	.	Enter

a. Dependent Variable: จำหน่ายได้ราคาสูงขึ้น

b. All requested variables entered.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.391[a]	.153	.133	1.028

a. Predictors: (Constant), ร้อยละปุ๋ยเคมีที่ลดลง, ขนาดพื้นที่(ไร่), การใช้ปุ๋ยเคมี, อายุ, เพศ

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	40.018	5	8.004	7.577	.000[b]
	Residual	221.815	210	1.056		
	Total	261.833	215			

a. Dependent Variable: จำหน่ายได้ราคาสูงขึ้น

b. Predictors: (Constant), ร้อยละปุ๋ยเคมีที่ลดลง, ขนาดพื้นที่(ไร่), การใช้ปุ๋ยเคมี, อายุ, เพศ

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	2.794	.403		6.935	.000		
	อายุ	.002	.007	.020	.302	.763	.963	1.038
	เพศ	-.138	.144	-.062	-.958	.339	.957	1.045
	ขนาดพื้นที่(ไร่)	.008	.003	.179	2.801	.006	.993	1.007
	การใช้ปุ๋ยเคมี	.011	.167	.004	.067	.946	.970	1.031
	ร้อยละปุ๋ยเคมีที่ลดลง	.014	.003	.341	5.286	.000	.972	1.029

a. Dependent Variable: จำหน่ายได้ราคาสูงขึ้น

แบบที่ 1 : .ให้ตัวแปร อายุ เพศ การใช้ปุ๋ยเคมี เป็นตัวแบ

ราคาสูงขึ้น = 2.794 -0.138เพศ + 0.002อายุ + 0.008พื้นที่
+ 0.011การใช้ปุ๋ยเคมี + 0.014 ร้อยละต้นทุนปุ๋ยลดลง

แบบที่ 2 : .นำตัวแปร อายุ เพศ การใช้ปุ๋ยเคมี ออกจากส

ราคาสูงขึ้น = b_0 + b_3พื้นที่ + b_5ร้อยละต้นทุนปุ๋ยลดลง

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.390[a]	.152	.144	1.021

a. Predictors: (Constant), ร้อยละปุ๋ยเคมีที่ลดลง, ขนาดพื้นที่(ไร่)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	40.855	2	20.428	19.594	.000[b]
	Residual	228.316	219	1.043		
	Total	269.171	221			

a. Dependent Variable: จำหน่ายได้ราคาสูงขึ้น

b. Predictors: (Constant), ร้อยละปุ๋ยเคมีที่ลดลง, ขนาดพื้นที่(ไร่)

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	2.801	.158		17.70	.000		
	ขนาดพื้นที่(ไร่)	.008	.003	.183	2.936	.004	.999	1.001
	ร้อยละปุ๋ยเคมีที่ลดลง	.014	.003	.348	5.593	.000	.999	1.001
a. Dependent Variable: จำหน่ายได้ราคาสูงขึ้น								

ราคาสูงขึ้น $= 2.801 + 0.008$ ขนาดพื้นที่ $+ 0.014$ร้อยละต้นทุนปุ๋ยลดลง

Adjusted R Square = 14.4 %

Case study 2 : data file พอใจเกษตรอินทรีย์.sav

1.Enter Method

พอใจโดยรวม $= b_0 + b_1$ประสานงาน $+ b_2$การรับสมัคร $+ \ldots\ldots + b_9$สามัคคี

Descriptive Statistics			
	Mean	Std. Deviation	N
พอใจโดยรวม	3.99	1.961	296
พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	4.80	1.048	296
พอใจการรับสมัคร/การคัดเลือกเกษตรกร	4.75	1.005	296
พอใจด้านการติดตามกิจกรรมการเกษตรของเจ้าหน้าที่ในพื้นที่	4.81	1.166	296
พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	4.13	1.500	296
พอใจการเชื่อมโยงตลาด	4.03	1.325	296
ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	3.92	1.172	296
พอใจการพบปะแลกเปลี่ยนความคืบหน้า อุปสรรค	3.89	1.153	296
พอใจความเข้มแข็งของผู้นำ	3.83	1.203	296
พอใจความสามัคคีของสมาชิกในกลุ่ม	3.35	1.167	296

Variables Entered/Removed[a]			
Model	Variables Entered	Variables Removed	Method
1	พอใจความสามัคคีของสมาชิกในกลุ่ม, พอใจด้านการติดตามกิจกรรมการเกษตรของเจ้าหน้าที่ในพื้นที่, พอใจการเชื่อมโยงตลาด, ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์, พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ, พอใจความเข้มแข็งของผู้นำ, พอใจการรับสมัคร/การคัดเลือกเกษตรกร, พอใจความพร้อมของแหล่งจำหน่ายผลผลิต, พอใจการพบปะแลกเปลี่ยนความคืบหน้า อุปสรรค[b]	.	Enter
a. Dependent Variable: พอใจโดยรวม			
b. All requested variables entered.			

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.870[a]	.756	.749	.983

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	857.721	9	95.302	98.661	.000[b]
	Residual	276.265	286	.966		
	Total	1133.986	295			
a. Dependent Variable: พอใจโดยรวม						

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-3.590	.307		-11.69	.000		
	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.524	.113	.280	4.644	.000	.234	4.274
	พอใจการรับสมัคร/การคัดเลือกเกษตรกร	.124	.122	.063	1.009	.314	.216	4.626
	พอใจด้านการติดตามกิจกรรมการเกษตรของเจ้าหน้าที่ในพื้นที่	-.027	.074	-.016	-.369	.712	.442	2.263
	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.430	.089	.329	4.819	.000	.183	5.469
	พอใจการเชื่อมโยงตลาด	-.014	.097	-.010	-.148	.882	.197	5.086
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.370	.149	.221	2.479	.014	.107	9.372
	พอใจการพบปะแลกเปลี่ยนความคืบหน้า อุปสรรค	.271	.157	.159	1.731	.085	.100	9.966
	พอใจความเข้มแข็งของผู้นำ	.074	.081	.045	.910	.363	.345	2.899
	พอใจความสามัคคีของสมาชิกในกลุ่ม	.031	.066	.018	.467	.641	.547	1.828

a. Dependent Variable: พอใจโดยรวม

1.Enter : 3 Independent variables

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.866[a]	.750	.747	.986

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	850.232	3	283.411	291.646	.000[b]
	Residual	283.755	292	.972		
	Total	1133.986	295			

a. Dependent Variable: พอใจโดยรวม

b. Predictors: (Constant), พอใจความพร้อมของแหล่งจำหน่ายผลผลิต, ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์, พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ

R Square = 75.6% Adjusted R Square = 74.9%

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-3.387	.281		-12.048	.000		
	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.603	.066	.322	9.112	.000	.685	1.459
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.684	.059	.409	11.656	.000	.696	1.436
	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.437	.048	.334	9.174	.000	.646	1.548
a. Dependent Variable: พอใจโดยรวม								

พอใจโดยรวม = -3.387 + 0.603ประสานงาน + 0.684ปรับเปลี่ยน + 0.437ความพร้อม

2.Backward

Variables Entered/Removed[a]			
Model	Variables Entered	Variables Removed	Method
1	พอใจความสามัคคีของสมาชิกในกลุ่ม, พอใจด้านการติดตามกิจกรรมการเกษตรของเจ้าหน้าที่ในพื้นที่, พอใจการเชื่อมโยงตลาด, ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์, พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ, พอใจความเข้มแข็งของผู้นำ, พอใจการรับสมัคร/การคัดเลือกเกษตรกร, พอใจความพร้อมของแหล่งจำหน่ายผลผลิต, พอใจการพบปะแลกเปลี่ยนความคืบหน้าอุปสรรค[b]	.	Enter
2	.	พอใจการเชื่อมโยงตลาด	Backward (criterion: Probability of F-to-remove >= .100).
3	.	พอใจด้านการติดตามกิจกรรมการเกษตรของเจ้าหน้าที่ในพื้นที่	Backward (criterion: Probability of F-to-remove >= .100).
4	.	พอใจความสามัคคีของสมาชิกในกลุ่ม	Backward (criterion: Probability of F-to-remove >= .100).
5	.	พอใจการรับสมัคร/การคัดเลือกเกษตรกร	Backward (criterion: Probability of F-to-remove >= .100).
6	.	พอใจความเข้มแข็งของผู้นำ	Backward (criterion: Probability of F-to-remove >= .100).

a. Dependent Variable: พอใจโดยรวม

b. All requested variables entered.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.870	.756	.749	.983
2	.870	.756	.750	.981
3	.870	.756	.750	.980
4	.870	.756	.751	.978
5	.869	.755	.751	.978
6	.868	.754	.751	.979

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-3.590	.307		-11.69	.000		
	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.524	.113	.280	4.644	.000	.234	4.274
	พอใจการรับสมัคร/การคัดเลือกเกษตรกร	.124	.122	.063	1.009	.314	.216	4.626
	พอใจด้านการติดตามกิจกรรมการเกษตรของเจ้าหน้าที่ในพื้นที่	-.027	.074	-.016	-.369	.712	.442	2.263
	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.430	.089	.329	4.819	.000	.183	5.469
	พอใจการเชื่อมโยงตลาด	-.014	.097	-.010	-.148	.882	.197	5.086
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.370	.149	.221	2.479	.014	.107	9.372
	พอใจการพบปะแลกเปลี่นความคืบหน้า อุปสรรค	.271	.157	.159	1.731	.085	.100	9.966
	พอใจความเข้มแข็งของผู้นำ	.074	.081	.045	.910	.363	.345	2.899
	พอใจความสามัคคีของสมาชิกในกลุ่ม	.031	.066	.018	.467	.641	.547	1.828

2	(Constant)	-3.599	.301		-11.967	.000		
	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.526	.112	.281	4.682	.000	.236	4.243
	พอใจการรับสมัคร/การคัดเลือกเกษตรกร	.122	.122	.063	1.002	.317	.218	4.587
	พอใจด้านการติดตามกิจกรรมการเกษตรของเจ้าหน้าที่ในพื้นที่	-.027	.074	-.016	-.367	.714	.442	2.262
	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.419	.048	.320	8.722	.000	.629	1.590
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.370	.149	.221	2.480	.014	.107	9.365
	พอใจการพบปะแลกเปลี่ยนความคืบหน้าอุปสรรค	.272	.156	.160	1.741	.083	.101	9.950
	พอใจความเข้มแข็งของผู้นำ	.072	.080	.044	.901	.368	.349	2.863
	พอใจความสามัคคีของสมาชิกในกลุ่ม	.031	.066	.019	.471	.638	.547	1.827

	Constant)	-3.614	.297		-12.14	.000		
	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.519	.111	.278	4.689	.000	.242	4.138
	พอใจการรับสมัคร/การคัดเลือกเกษตรกร	.106	.113	.054	.933	.351	.251	3.991
3	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.419	.048	.321	8.743	.000	.629	1.589
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.366	.149	.219	2.463	.014	.107	9.311
	พอใจการพบปะแลกเปลี่ยนความคืบหน้า อุปสรรค	.273	.156	161	1.751	.081	.101	9.946
	พอใจความเข้มแข็งของผู้นำ	.072	.080	.044	.902	.368	.349	2.863
	พอใจความสามัคคีของสมาชิกในกลุ่ม	.032	.066	.019	.487	.626	.548	1.824

4	(Constant)	-3.599	.296		-12.17	.000		
	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.518	.111	.277	4.683	.000	.242	4.134
	พอใจการรับสมัคร/การคัดเลือกเกษตรกร	.110	.113	.056	.974	.331	.252	3.968
	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.420	.048	.321	8.775	.000	.630	1.588
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.360	.148	.215	2.437	.015	.108	9.261
	พอใจการพบปะแลกเปลี่นความคืบหน้า อุปสรรค	.283	.155	.166	1.831	.068	.102	9.782
	พอใจความเข้มแข็งของผู้นำ	.088	.073	.054	1.212	.227	.421	2.377

5	(Constant)	-3.518	.284		-12.39	.000		
	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.604	.066	.323	9.102	.000	.672	1.489
	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.425	.048	.325	8.928	.000	.637	1.570
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.368	.148	.220	2.488	.013	.108	9.238
	พอใจการพบปะแลกเปลี่ยนความคืบหน้าอุปสรรค	.278	.154	.164	1.803	.072	.102	9.774
	พอใจความเข้มแข็งของผู้นำ	.088	.073	.054	1.202	.230	.421	2.377
6	(Constant)	-3.484	.283		-12.332	.000		
	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.614	.066	.328	9.311	.000	.682	1.467
	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.424	.048	.325	8.909	.000	.637	1.570
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.380	.147	.227	2.580	.010	.109	9.190
	พอใจการพบปะแลกเปลี่ยนความคืบหน้าอุปสรรค	.332	.148	.195	2.242	.026	.112	8.964

a. Dependent Variable: พอใจโดยรวม

ความพอใจโดยรวม = -3.484 + 0.614ประสานงาน + 0.424 ความพร้อม + 0.38 ปรับเปลี่ยน + 0.332แล

R Square = 75.4% Adjusted R Square = 75.1%

3. Forward

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
2	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
3	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.	Forward (Criterion: Probability-of-F-to-enter <= .050)
4	พอใจการพบปะแลกเปลี่นความคืบหน้า อุปสรรค	.	Forward (Criterion: Probability-of-F-to-enter <= .050)

a. Dependent Variable: พอใจโดยรวม

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.723[a]	.522	.521	1.358
2	.824[b]	.679	.676	1.115
3	.866[c]	.750	.747	.986
4	.868[d]	.754	.751	.979

a. Predictors: (Constant), ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์

b. Predictors: (Constant), ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์, พอใจความพร้อมของแหล่งจำหน่ายผลผลิต

c. Predictors: (Constant), ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์, พอใจความพร้อมของแหล่งจำหน่ายผลผลิต, พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ

d. Predictors: (Constant), ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์, พอใจความพร้อมของแหล่งจำหน่ายผลผลิต, พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ, พอใจการพบปะแลกเปลี่นความคืบหน้า อุปสรรค

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	592.152	1	592.152	321.303	.000[b]
	Residual	541.834	294	1.843		
	Total	1133.986	295			
2	Regression	769.545	2	384.773	309.346	.000[c]
	Residual	364.441	293	1.244		
	Total	1133.986	295			
3	Regression	850.232	3	283.411	291.646	.000[d]
	Residual	283.755	292	.972		
	Total	1133.986	295			
4	Regression	855.051	4	213.763	223.008	.000[e]
	Residual	278.936	291	.959		
	Total	1133.986	295			

a. Dependent Variable: พอใจโดยรวม

b. Predictors: (Constant), ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์

c. Predictors: (Constant), ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์, พอใจความพร้อมของแหล่งจำหน่ายผลผลิต

d. Predictors: (Constant), ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์, พอใจความพร้อมของแหล่งจำหน่ายผลผลิต, พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ

e. Predictors: (Constant), ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์, พอใจความพร้อมของแหล่งจำหน่ายผลผลิต, พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ, พอใจการพบปะแลกเปลี่ยนความคืบหน้า อุปสรรค

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-.748	.276		-2.711	.007		
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	1.209	.067	.723	17.925	.000	1.000	1.000
2	(Constant)	-1.713	.241		-7.116	.000		
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.825	.064	.493	12.892	.000	.749	1.336
	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.597	.050	.457	11.942	.000	.749	1.336
3	(Constant)	-3.387	.281		-12.048	.000		
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.684	.059	.409	11.656	.000	.696	1.436
	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.437	.048	.334	9.174	.000	.646	1.548
	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.603	.066	.322	9.112	.000	.685	1.459
4	(Constant)	-3.484	.283		-12.332	.000		
	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.380	.147	.227	2.580	.010	.109	9.190
	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.424	.048	.325	8.909	.000	.637	1.570
	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.614	.066	.328	9.311	.000	.682	1.467
	พอใจการพบปะแลกเปลี่ยนความคืบหน้าอุปสรรค	.332	.148	.195	2.242	.026	.112	8.964

a. Dependent Variable: พอใจโดยรวม

Variables Entered/Removed[a]			
Model	Variables Entered	Variables Removed	Method
1	ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
4	พอใจการพบปะแลกเปลี่ยนความคืบหน้า อุปสรรค	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
a. Dependent Variable: พอใจโดยรวม			

1. Enter

R Square = 75.6% Adjusted R Square = 74.9%

พอใจโดยรวม = -3.387 + 0.603ประสานงาน + 0.684ปรับเปลี่ยน + 0.437ความพร้อม

2. Forward , Backward , Stepwise

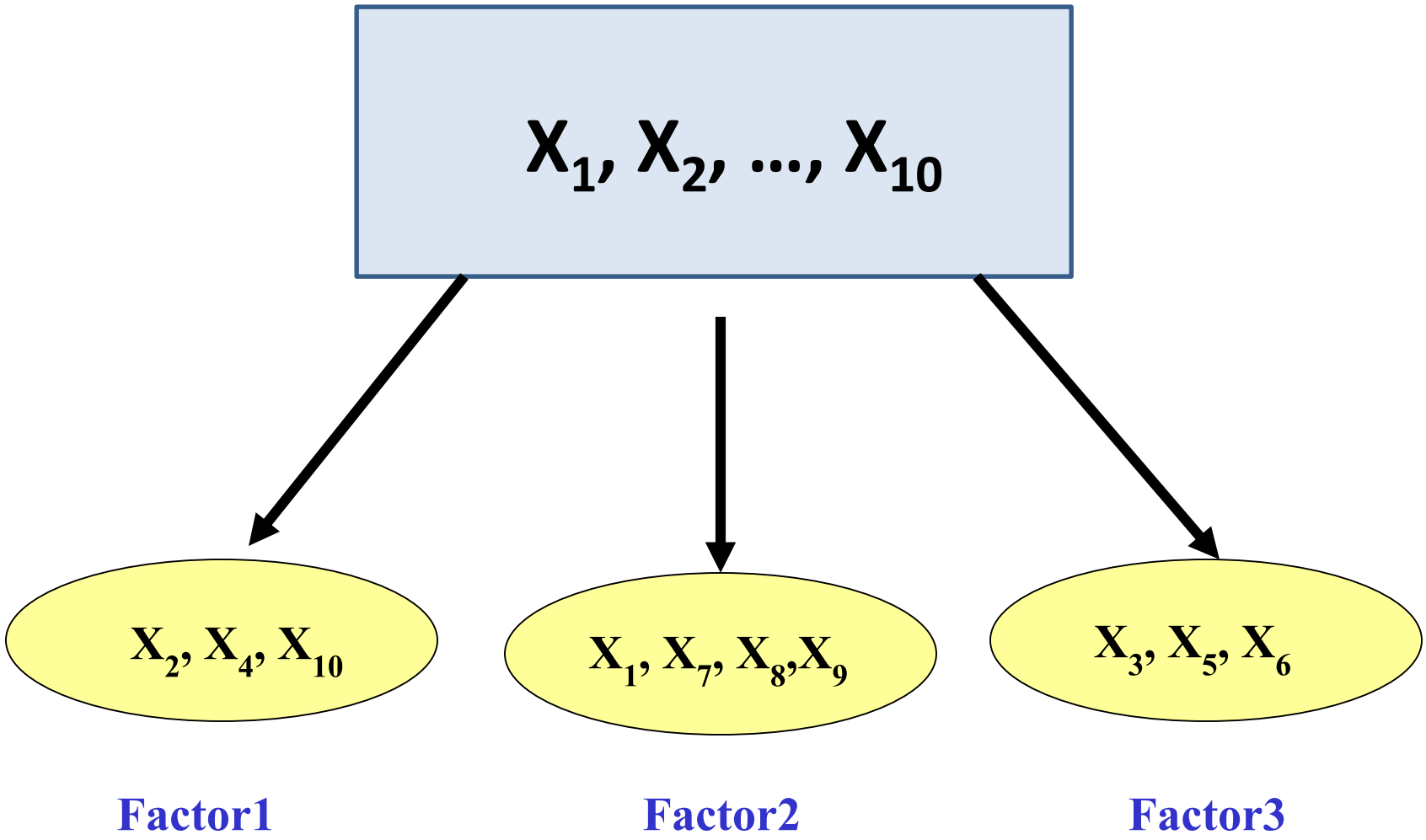
ความพอใจโดยรวม = -3.484 + 0.614ประสานงาน + 0.424 ความพร้อม + 0.38 ปรับเปลี่ยน + 0.332แ

R Square = 75.4% Adjusted R Square = 75.1%

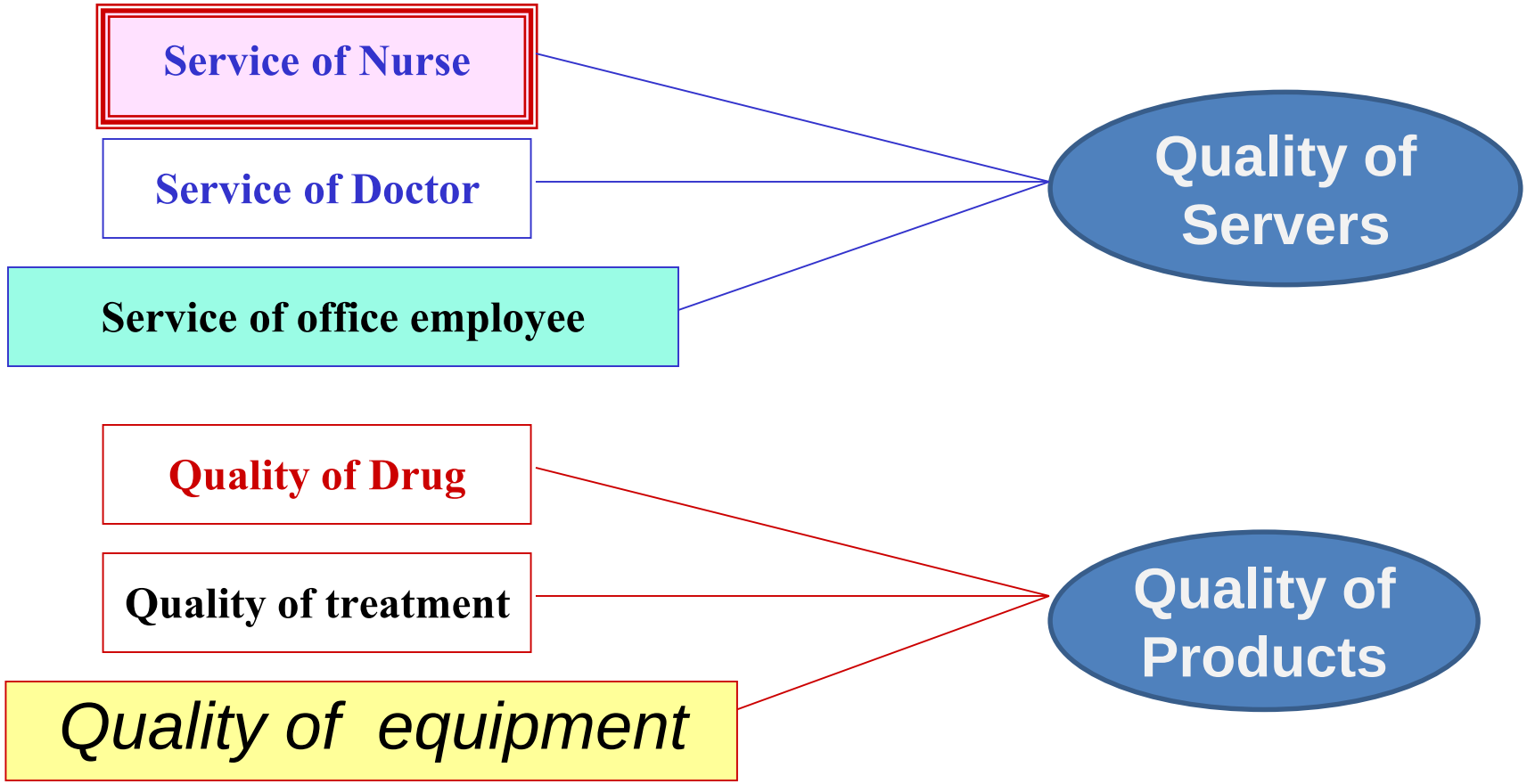
Factor Analysis

- FA is a multivariate statistical technique that is used to summarize the information contained in a large number of variable into a small number of subsets or factor.

- *FA is used to identify underlying dimensions or constructs in the data and reduce the number of variables by eliminating redundancy.*


$X_1, X_2, \ldots, X_{10}$
X_2, X_4, X_{10}
X_1, X_7, X_8, X_9
X_3, X_5, X_6
Factor1
Factor2
Factor3

Example


Service of Nurse
Service of Doctor
Service of office employee
Quality of Servers
Quality of Drug
Quality of treatment
Quality of equipment
Quality of Products

Factor Analysis

1. *Study of interrelationship among the variables in an effort to find a new set of variables.*
2. *Confirm hypothesis relationship structure.*

Types of Factor Analysis

1. Exploratory Factor Analysis

1.1 Principal Component Analysis

1.2 Common Factor Analysis

2. Confirmative Factor Analysis

Exploratory Factor Analysis

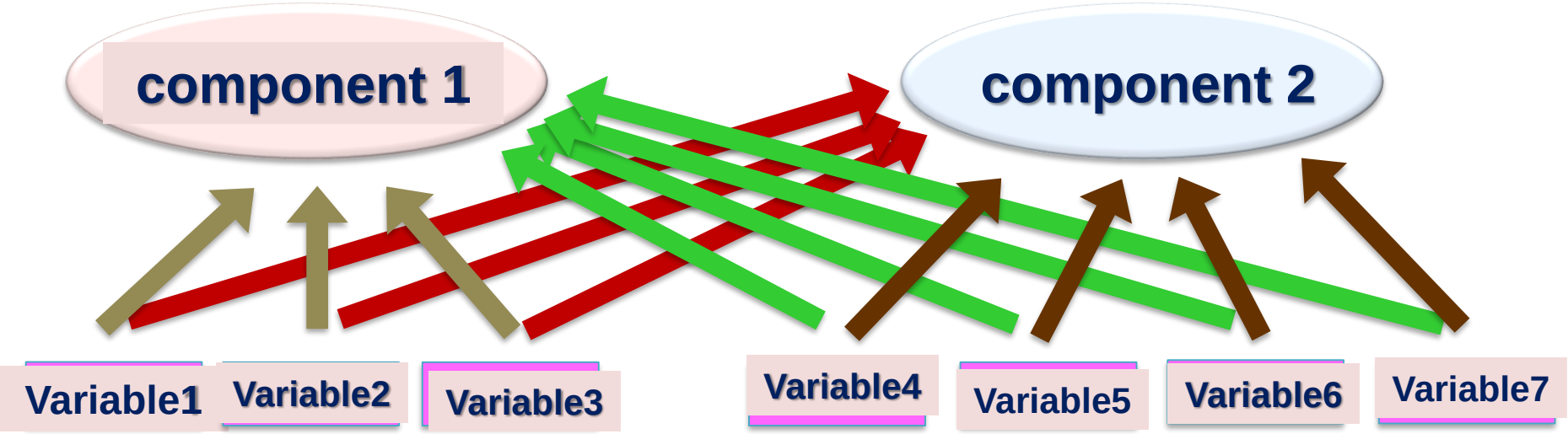
- Study structure of the factor model or the underlying theory is not know or specified a priori.

- Data are used to help reveal or identify the structure of the factor model.

- The researcher has no knowledge of the factor structure and is essentially seeking to identify the factor model that would account for the covariances among the variables.

Software : SPSS, SAS, etc. :
Exploratory Factor Analysis

Exploratory Factor Analysis (Principal Component)



Confirmatory Factor Analysis

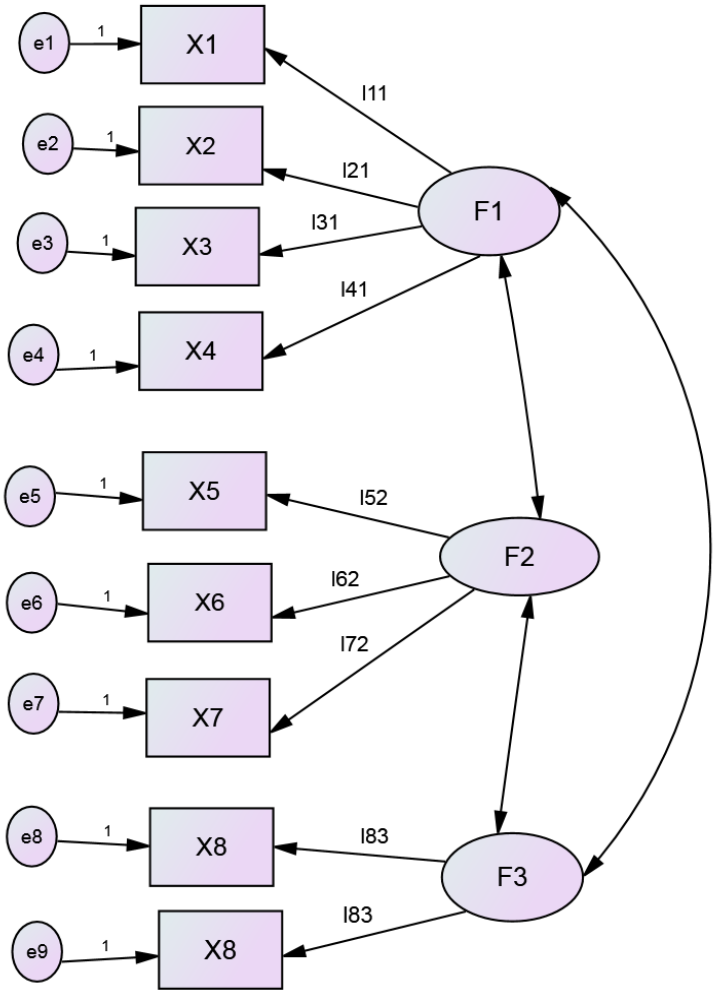
- To validate the hypothesized model and estimate model parameters.

- The precise structure of the model is know before.

Software : Amos ,EQS , MPLUS, LISREL

If we => use wish to confirm or negate the hypothesis Structure Confirmatory FA

Confirmatory Factor Analysis



- Factor F_1 is measured by the Observable variables x_1 - X4

- Factor F_2 is measured by X5 – X7
- Factor F_3 is measured only by X8,X9

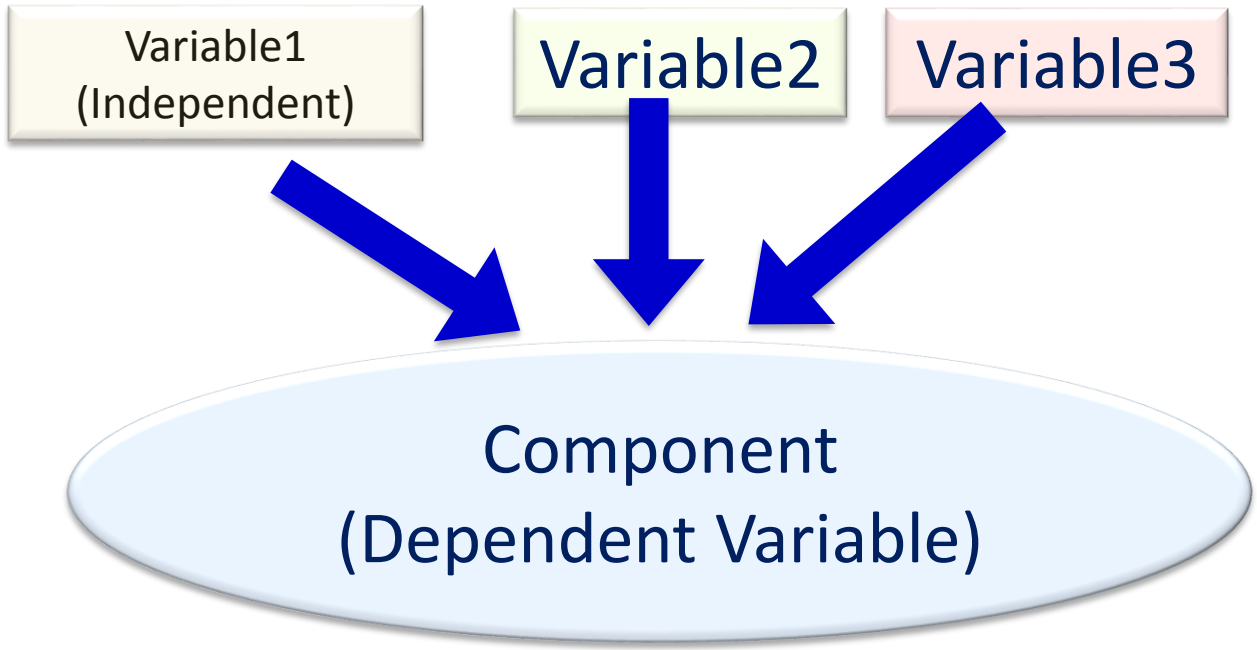
- F1-F3 can be correlated

$$X_1 = l_{11}F_1 + e_1$$
$$X_2 = l_{21}F_1 + e_2$$
$$X_3 = l_{31}F_1 + e_3$$
$$X_4 = l_{41}F_1 + e_4$$
$$X_5 = l_{52}F_2 + e_5$$
$$X_6 = l_{62}F_2 + e_6$$
$$X_7 = l_{72}F2 + e_7$$
$$X_8 = l_{93}F3 + e_8$$
$$X_9 = l_{93}F3 + e_9$$

I. Principal Component Analysis (PCA)

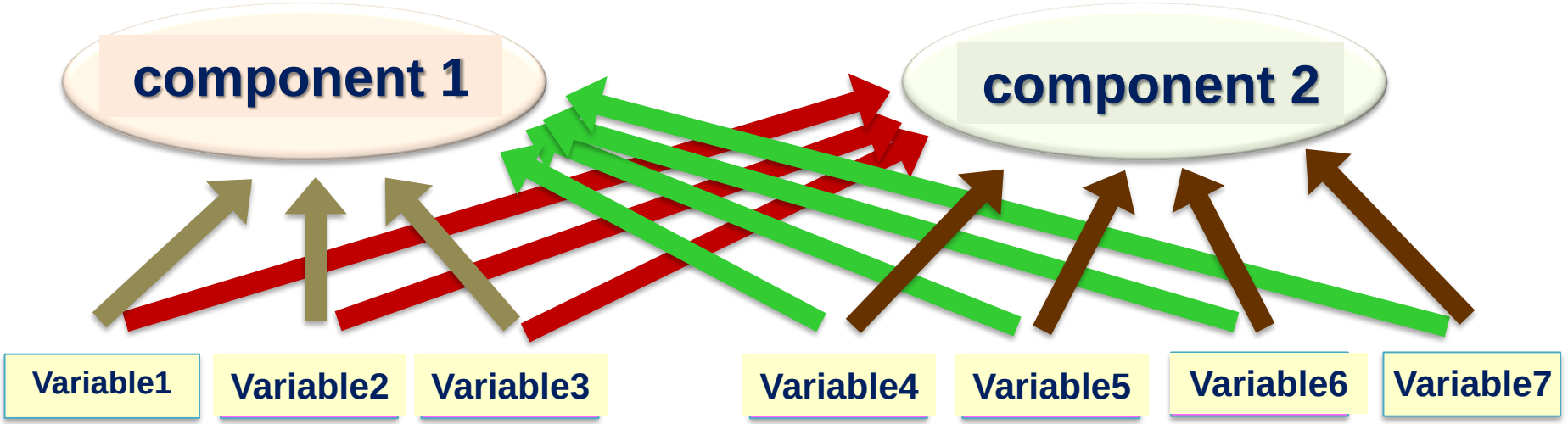
1. **Component (Factor) = Dependent variable**

   Observed / Measurement =Independent variable



Note Component arise from Measurement variables

Exploratory Factor Analysis (Principal Component)



I. Principal Component Analysis (PCA)

1. **Total variance in Data set.**

   How much of total variance they account for?

2. **Variance for each variable = 1**

3. **Full Solution**

number of component = number of measurement variable

4. **No unique error**

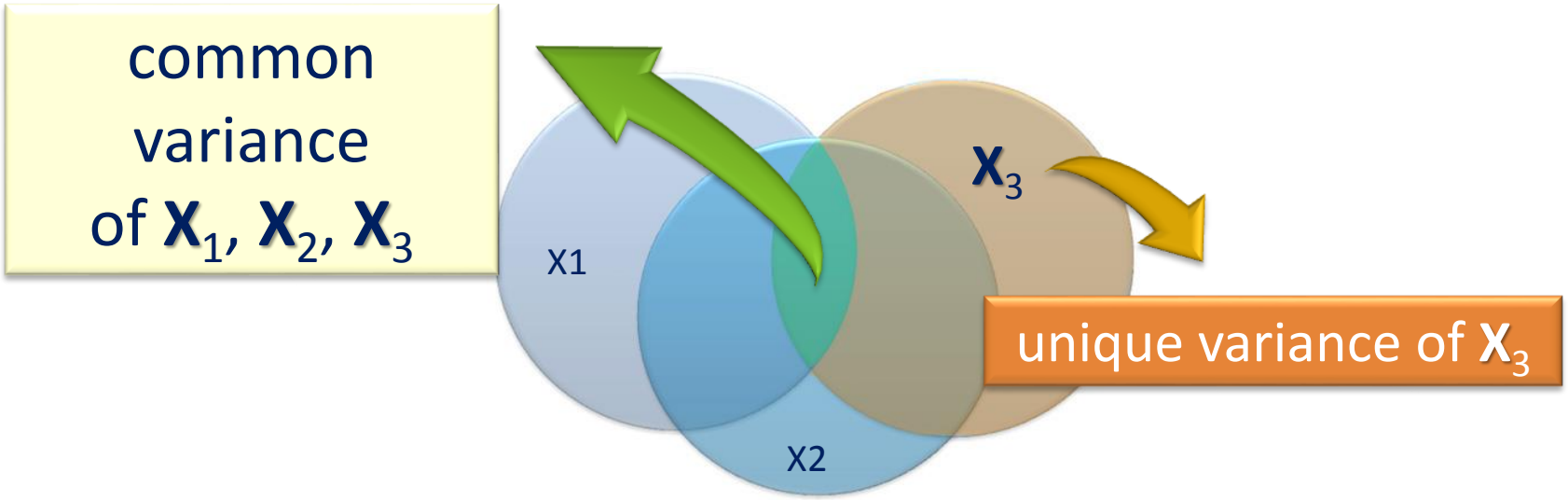
EXPLORATORY FACTOR ANALYSIS

II. Common Factor Analysis

(PAF: Principal Axis Factoring)

Note: Factor structure be descriptive of what the variables have in common

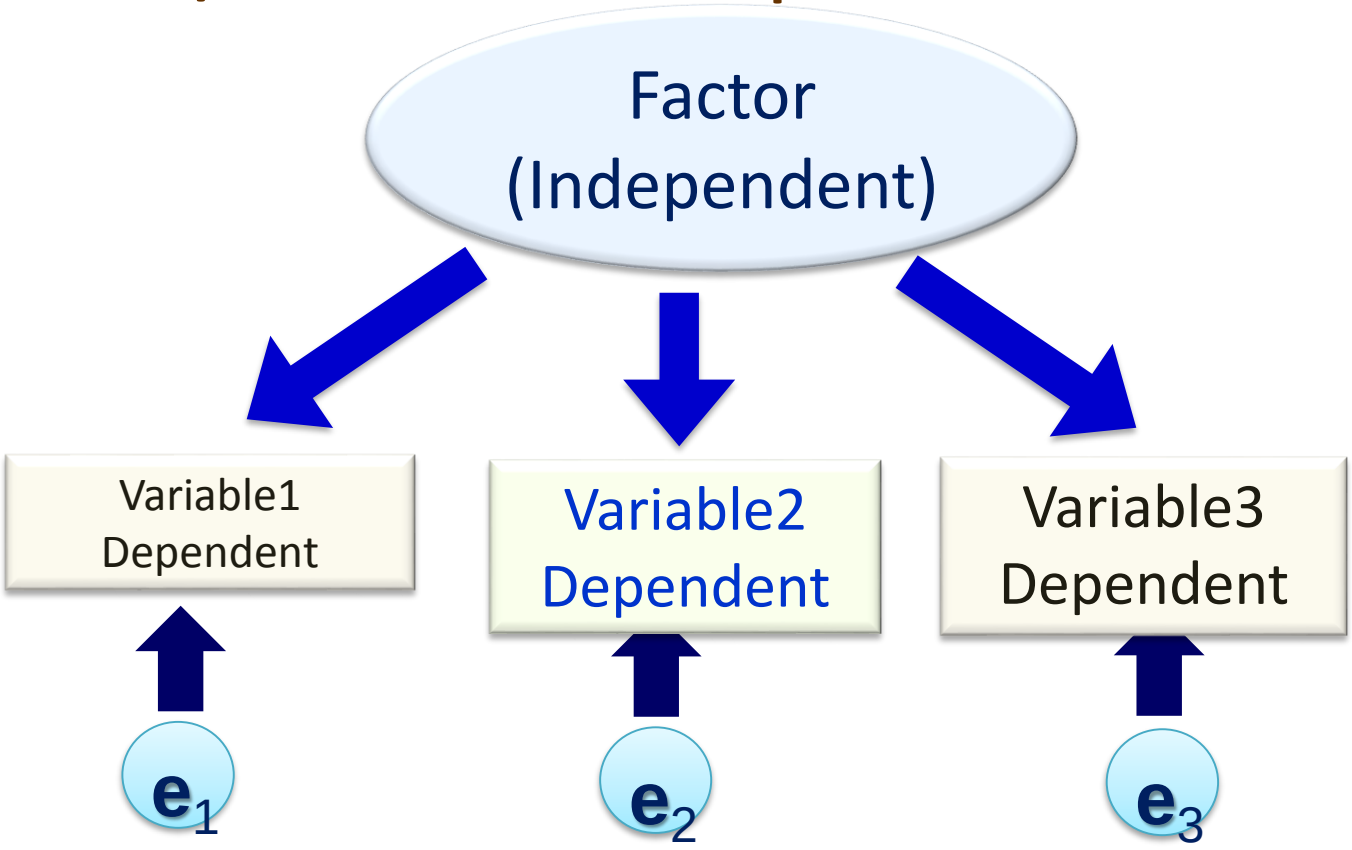
1. Total variance = common variance (shared)
   + unique variance



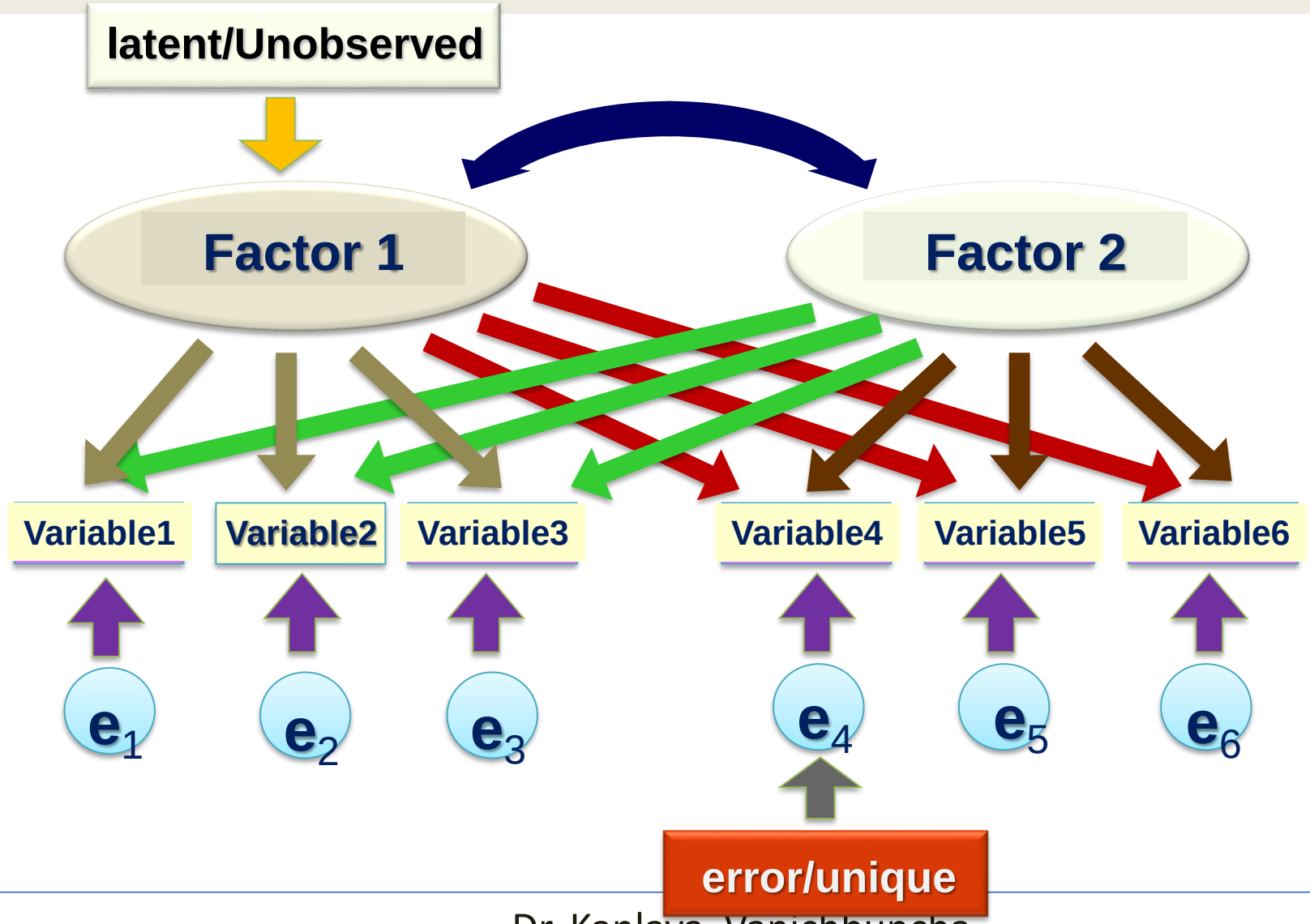
II. Common Factor Analysis
(PAF: Principal Axis Factoring)

1. **Factor= Independent variable**

   Indicator / Measurement = Dependent variable



II. Common Factor Analysis (PAF: Principal Axis Factoring)



Steps for Exploratory Factor Analysis

Step1. Check Appropriation of data for EFA

Ho : X1, X2 , …… , Xp are independent
H1: X1, X2 , …… , Xp are related

Use KMO (Kaiser-Meyer-Olkin)

$$KMO = \frac{\sum\sum_{i \neq j} r_{ij}^2}{\sum\sum_{i \neq j} r_{ij}^2 + \sum\sum_{i \neq j} a_{ij}^2}$$

$$0 \leq \text{KMO} \leq 1$$

r_{ij} = bivariate correlation between Xi and Xj

a_{ij} = Partial correlation between Xi and Xj when control others X

KMO	*Recommendation*
≥ 0.9	*Marvelous*
0.80 +	*Meritorious*
0.70 +	*Middling*
0.60 +	*Mediocre*
0.50 +	*Miserable*
Below 0.5	*Unacceptable*

Fundamental Steps for EFA

Step 2. Factor Extraction

1. Principal Component Analysis (PCA)

: no distributional assumptions.

2. Common Factor

- **Principal Axis Factoring (PAF)**
- **Maximum Likelihood** (MLE)

: assumes multivariate normality, but provide goodness of fit evaluation.

- **etc**

Principal Component Analysis (PCA)

Construct linear combinations of the original variables that account for a large part of the total variation.

1. p linear compounds are needed to account for the total variance of p variables.

2. The sum of the variances of all p principal components is equal to the sum of variances of the original variable.

Principal Component Analysis (PCA)

$$PC_1 = F_1 = \omega_{11}X_1 + \omega_{12}X_2 + \ldots + \omega_{1p}X_p$$

$$PC_2 = F_2 = \omega_{21}X_1 + \omega_{22}X_2 + \ldots + \omega_{2p}X_p$$

$$PC_m = F_m = \omega_{m1}X_1 + \omega_{m2}X_2 + \ldots + \omega_{mp}X_p$$

$$PC_p = F_p = \omega_{p1}X_1 + \omega_{p2}X_2 + \ldots + \omega_{pp}X_p$$

where $m < p$

$$\mathbf{Var(F_1) + \ldots + Var(F_p) = Var(X_1) + \ldots + Var(X_p)}$$

$F_1, F_2, \ldots, F_p$ are independent

Exploratory Factor Analysis

Use m Factors; m < p

$$X_1 = \ell_{11}F_1 + \ell_{12}F_2 + \ldots + \ell_{1m}F_m$$

$$X_2 = \ell_{21}F_1 + \ell_{22}F_2 + \ldots + \ell_{2m}F_m$$

$$\vdots$$

$$X_p = \ell_{p1}F_1 + \ell_{p2}F_2 + \ldots + \ell_{pm}F_m$$

ℓ_{ij} **= Factor loading of F_j associate with i^{th} variable**

Exploratory Factor Analysis : Common Factor

- **Common Factor assume that a variable consists of common and unique parts.**

- **The common part of a variable is that part of the variable's variation that is shared with other variable, whereas the unique part of the variable is that part of the variable's variation that is specific to that variable alone.**

Fundamental Steps for EFA

Step 3. Factor Selection

- Determine the appropriate number of factors by :

1. Scree plot of eigenvalues.

2. Goodness of model fit (χ^2, RMSEA etc.)

Step 4

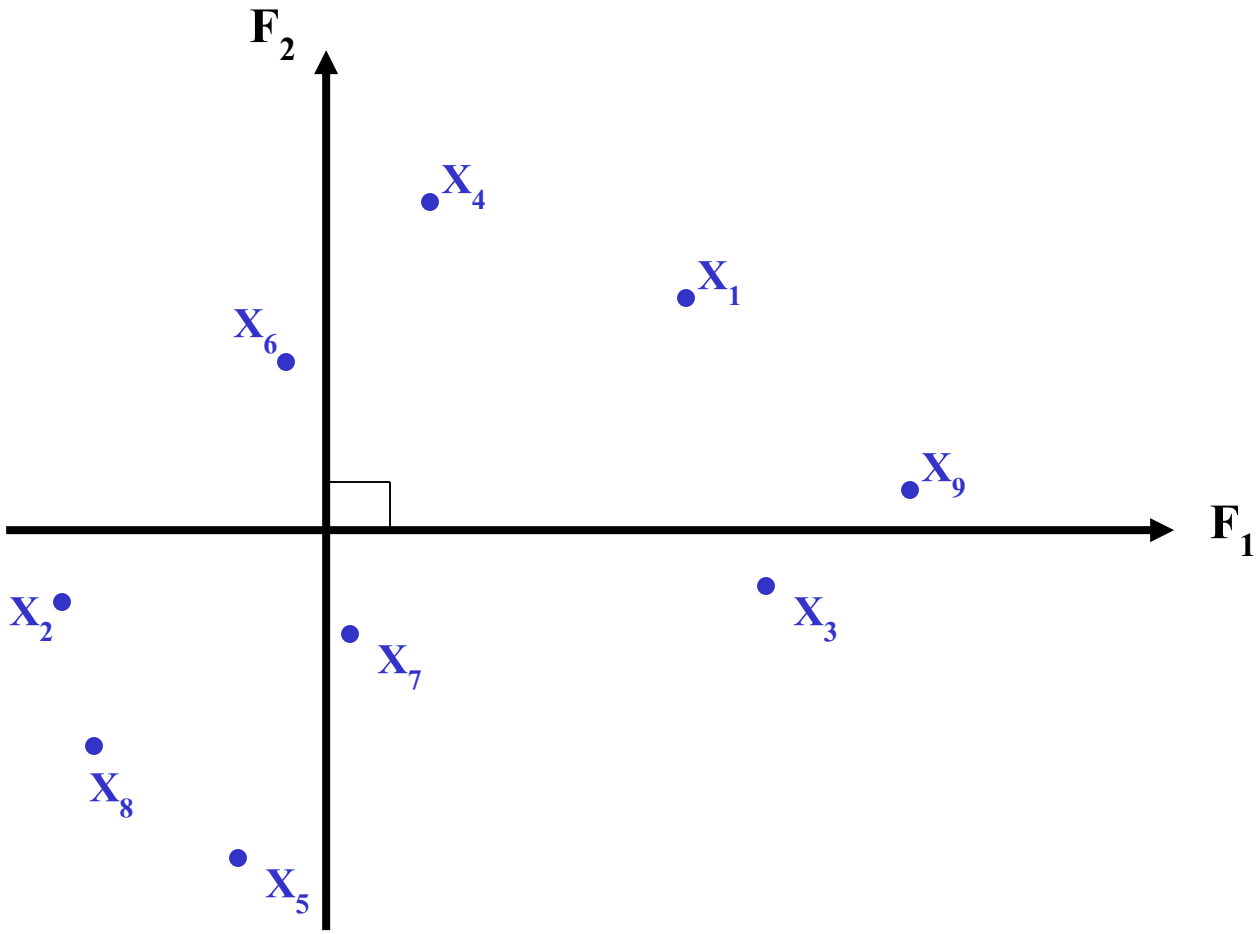
- Identify original variables for each factor by consider factor loading.

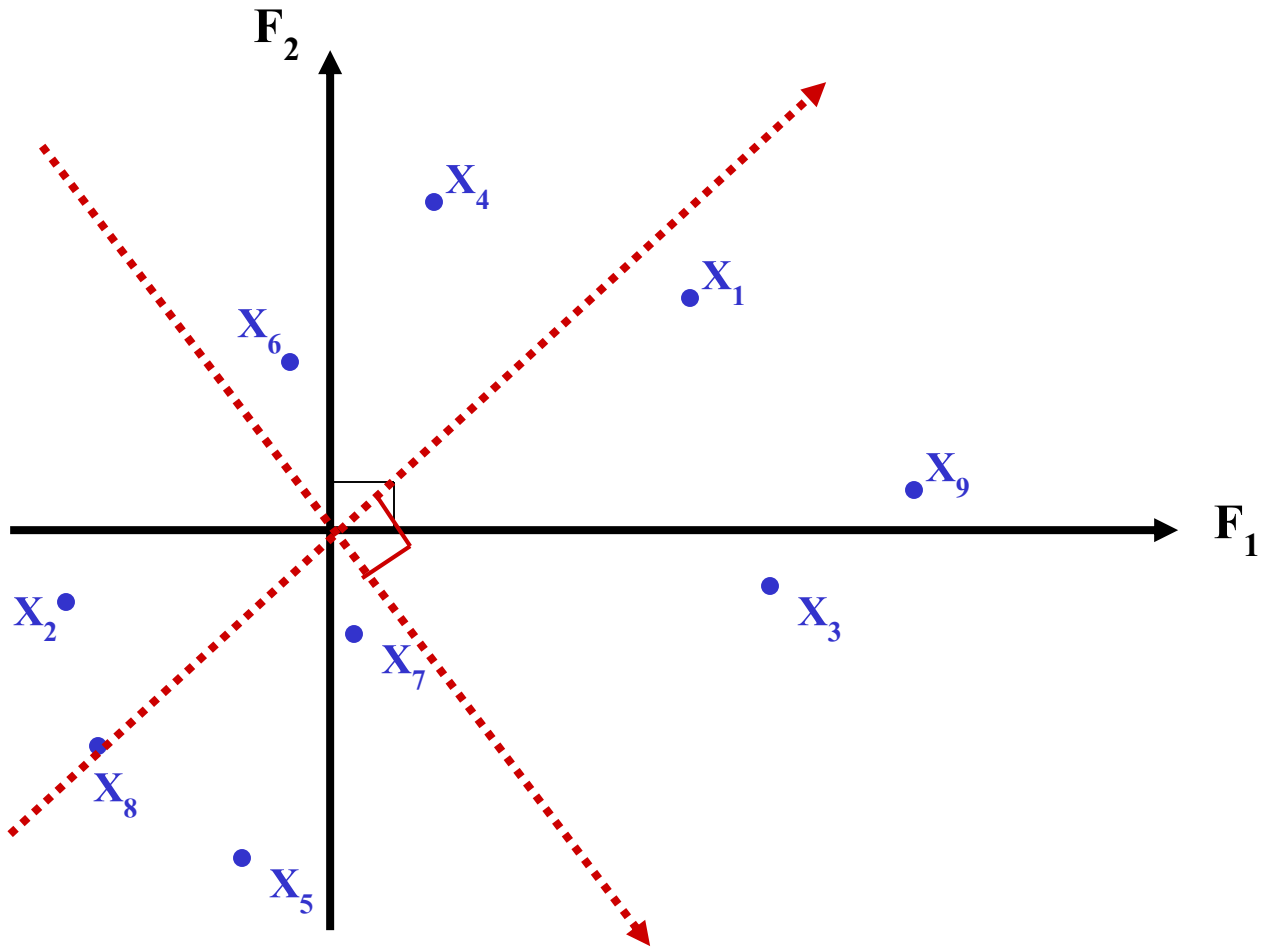
Step 5

Rotate axis of Factor

Step 6:

- Identify original variables for each factor by consider factor loading.


F2
F1
X4
X1
X6
X9
X2
X3
X7
X8
X5


F2
F1
X4
X1
X6
X9
X2
X3
X7
X8
X5

Rotation Method

I. Orthogonal Rotation

1. Varimax (Kaiser 1958)
2. Qaurtimax
3. Equimax

Note Give one Rotated Factor Matrix. (structural coefficient)

II. Oblique Factor Rotation

- Direct Oblimin

Note : Give 2 different Rotated factor Matrix

1. Pattern Matrix (Pattern coefficient)
2. Structure Matrix

2. Oblique Rotation Method

Factor axes can be rotated independently, so that they are not necessarily perpendicular to one another after rotation.

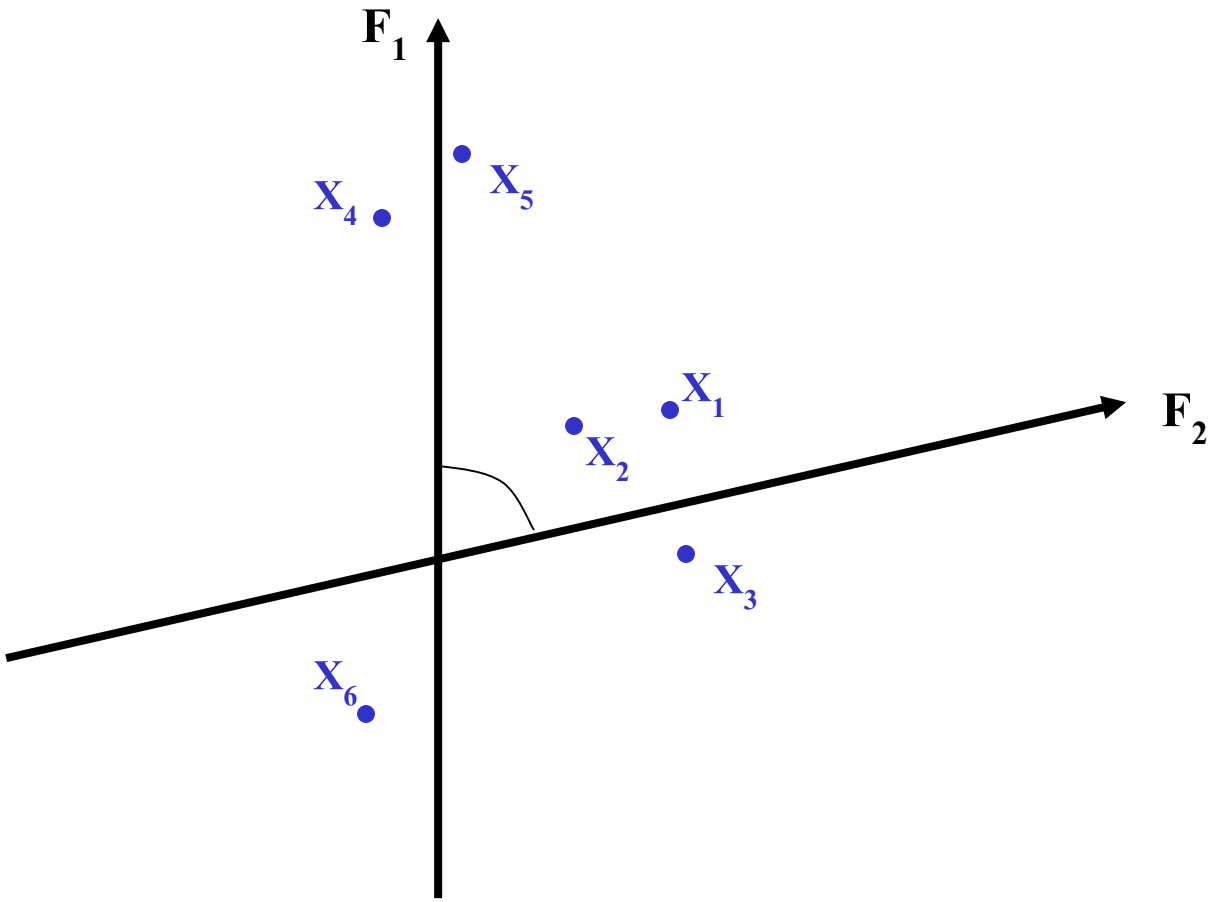
2.1 Oblimax

2.2 Quartimin

2.3 Covarimin

2.4 Bi - Quartimin

2.5 Oblimin


F1
F2
X1
X2
X3
X4
X5
X6

Fundamental Steps for EFA

Step 7

- **Interpret the Factor and Evaluate the Quality of the Solution**
  - Consider the meaningfulness and interpretability of the factors.
  - Eliminate poorly defined factors

Re-Run and (Ideally) Replicate the Factor Analysis

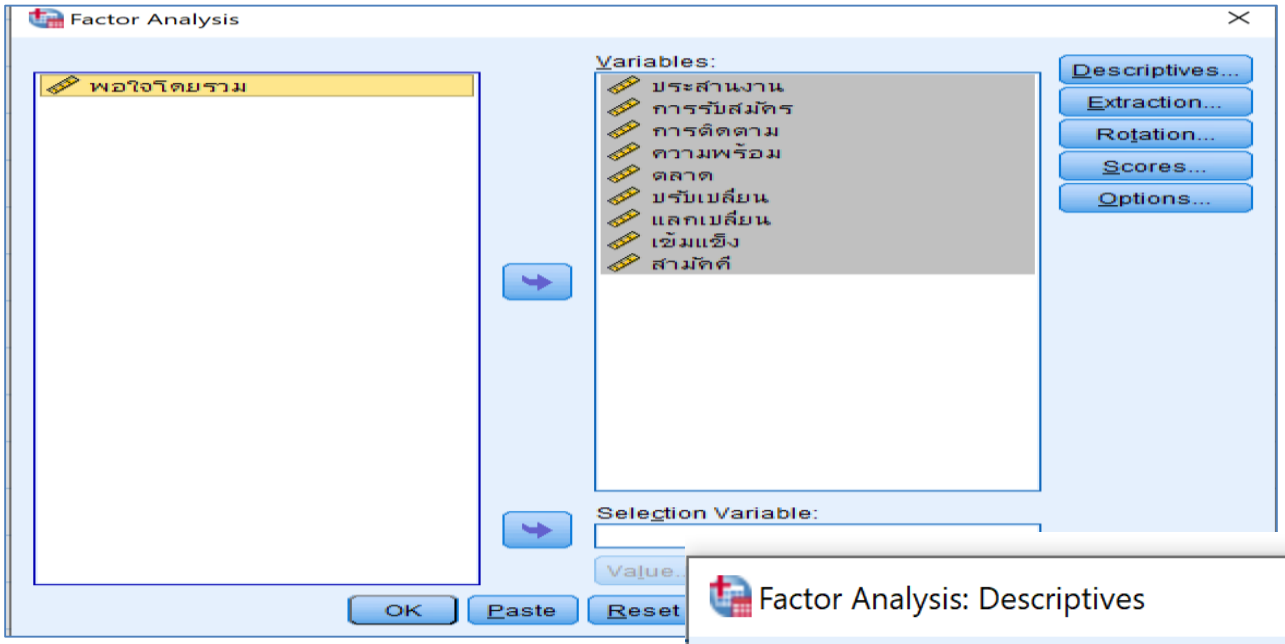
- If items or factors are dropped in preceding step, re-run the EFA

Limitation

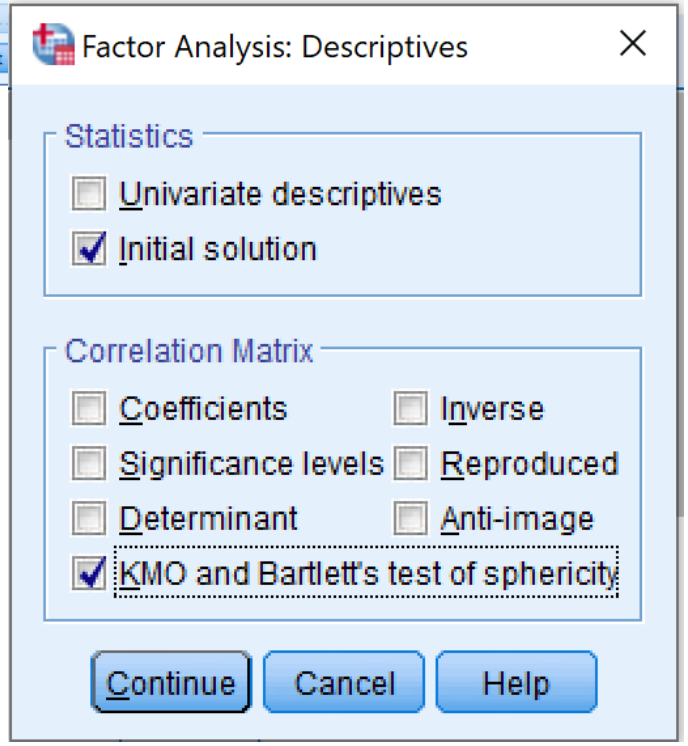
1. Highly subjective process
2. The determination of number of factor
3. Their interpretation
4. Rotation

Analyze – Dimension Reduction - Factor

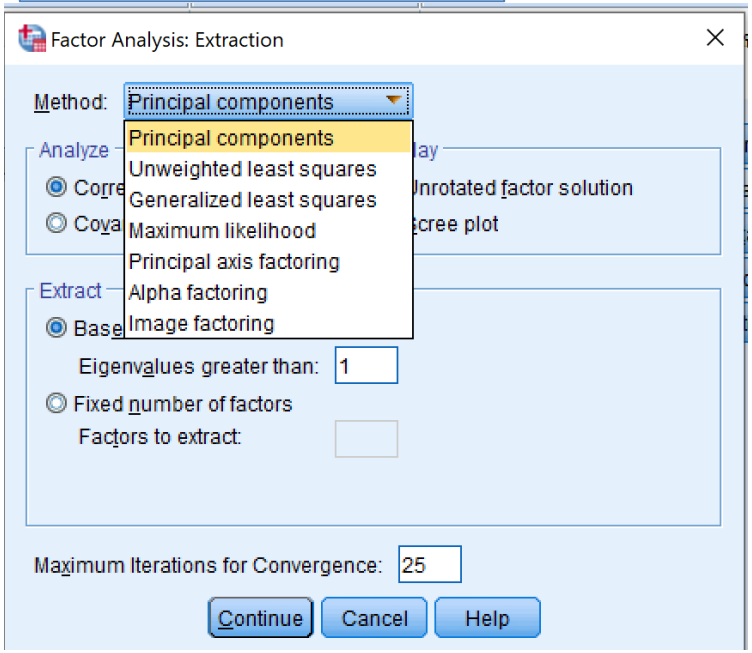
SPSS : Analyze ➔ Dimension Reduction ➔ Factor



Click Descriptive



Click Extraction



KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.806
Bartlett's Test of Sphericity	Approx. Chi-Square	2372.596
	df	36
	Sig.	.000

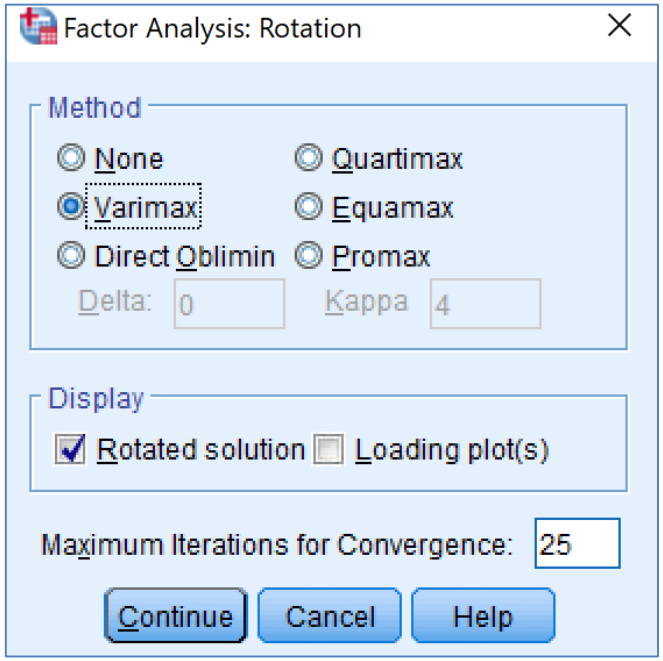
Communalities		
	Initial	Extraction
พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	1.000	.861
พอใจการรับสมัคร/การคัดเลือกเกษตรกร	1.000	.890
พอใจด้านการติดตามกิจกรรมการเกษตรของเจ้าหน้าที่ในพื้นที่	1.000	.785
พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	1.000	.944
พอใจการเชื่อมโยงตลาด	1.000	.943
ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	1.000	.840
พอใจการพบปะแลกเปลี่นความคืบหน้า อุปสรรค	1.000	.869
พอใจความเข้มแข็งของผู้นำ	1.000	.804
พอใจความสามัคคีของสมาชิกในกลุ่ม	1.000	.615
Extraction Method: Principal Component Analysis.		

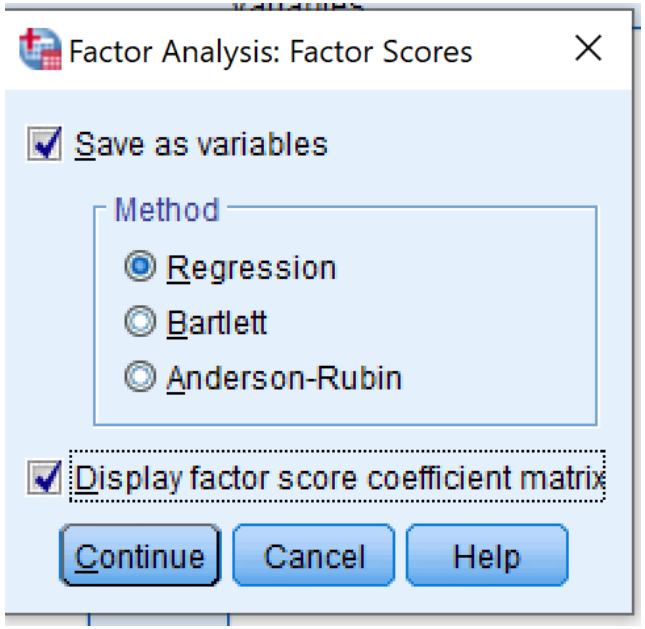
Total Variance Explained						
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.019	55.764	55.764	5.019	55.764	55.764
2	1.509	16.764	72.528	1.509	16.764	72.528
3	1.024	11.374	83.903	1.024	11.374	83.903
4	.561	6.230	90.133			
5	.327	3.635	93.768			
6	.269	2.986	96.754			
7	.137	1.525	98.279			
8	.099	1.105	99.384			
9	.055	.616	100.000			
Extraction Method: Principal Component Analysis.						

Component Matrix[a]			
	Component		
	1	2	3
พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.761	.485	.217
พอใจการรับสมัคร/การคัดเลือกเกษตรกร	.751	.524	.227
พอใจด้านการติดตามกิจกรรมการเกษตรของเจ้าหน้าที่ในพื้นที่	.669	.487	.316
พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.752	.148	-.596
พอใจการเชื่อมโยงตลาด	.726	.126	-.633
ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.828	-.383	.085
พอใจการพบปะแลกเปลี่นความคืบหน้า อุปสรรค	.820	-.439	.070
พอใจความเข้มแข็งของผู้นำ	.757	-.454	.158
พอใจความสามัคคีของสมาชิกในกลุ่ม	.636	-.423	.179
Extraction Method: Principal Component Analysis.			
a. 3 components extracted.			

Step 5 : Rotation

save





Rotated Component Matrix[a]			
	Component		
	1	2	3
พอใจการประสานงานของเจ้าหน้าที่ในการดำเนินโครงการ	.226	.865	.247
พอใจการรับสมัคร/การคัดเลือกเกษตรกร	.194	.891	.239
พอใจด้านการติดตามกิจกรรมการเกษตรของเจ้าหน้าที่ในพื้นที่	.187	.858	.118
พอใจความพร้อมของแหล่งจำหน่ายผลผลิต	.253	.272	.897
พอใจการเชื่อมโยงตลาด	.242	.226	.913
ความพร้อมของเกษตรกรในการปรับเปลี่ยนกิจกรรมสำหรับการทำเกษตรอินทรีย์	.845	.239	.264
พอใจการพบปะแลกเปลี่ยนความคืบหน้า อุปสรรค	.874	.189	.265
พอใจความเข้มแข็งของผู้นำ	.864	.181	.157
พอใจความสามัคคีของสมาชิกในกลุ่ม	.767	.143	.085
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.			
a. Rotation converged in 4 iterations.			

Total Variance Explained									
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.019	55.764	55.764	5.019	55.764	55.764	3.058	33.981	33.981
2	1.509	16.764	72.528	1.509	16.764	72.528	2.550	28.337	62.318
3	1.024	11.374	83.903	1.024	11.374	83.903	1.943	21.584	83.903
4	.561	6.230	90.133						
5	.327	3.635	93.768						
6	.269	2.986	96.754						
7	.137	1.525	98.279						
8	.099	1.105	99.384						
9	.055	.616	100.000						
Extraction Method: Principal Component Analysis.									

Descriptive Statistics			
	Mean	Std. Deviation	N
พอใจโดยรวม	3.99	1.961	296
เกษตรกร	.0000000	1.00000000	296
เจ้าหน้าที่	.0000000	1.00000000	296
แหล่งจำหน่าย	.0000000	1.00000000	296

Variables Entered/Removed[a]			
Model	Variables Entered	Variables Removed	Method
1	แหล่งจำหน่าย, เจ้าหน้าที่, เกษตรกร[b]	.	Enter
a. Dependent Variable: พอใจโดยรวม			
b. All requested variables entered.			

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.855[a]	.730	.728	1.023
a. Predictors: (Constant), แหล่งจำหน่าย, เจ้าหน้าที่, เกษตรกร				

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	828.328	3	276.10	263.7	.000[b]
	Residual	305.658	292	1.047		
	Total	1133.98	295			
a. Dependent Variable: พอใจโดยรวม						
b. Predictors: (Constant), แหล่งจำหน่าย, เจ้าหน้าที่, เกษตรกร						

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	3.993	.059		67.150	.000		
	เกษตรกร	1.045	.060	.533	17.548	.000	1.000	1.000
	เจ้าหน้าที่	.914	.060	.466	15.350	.000	1.000	1.000
	แหล่งจำหน่าย	.938	.060	.478	15.741	.000	1.000	1.000
a. Dependent Variable: พอใจโดยรวม								

บรรณานุกรม

กัลยา วานิชย์บัญชา "การวิเคราะห์ข้อมูลหลายตัวแปร"

กัลยา วานิชย์บัญชา"การวิเคราะห์สถิติขั้นสูงด้วย SPSS "

กัลยา วานิชย์บัญชา"สถิติสำหรับงานวิจัย "

กัลยา วานิชย์บัญชา การวิเคราะห์สมการโครงสร้าง (SEM) ด้วย AMOS

กัลยา วานิชย์บัญชา และ ฐิตา วานิชย์บัญชา : การใช้ SPSS For Windows ในการวิเคราะห์ข้อมูล