Introduction to Data Analysis for Socioeconomics using STATA

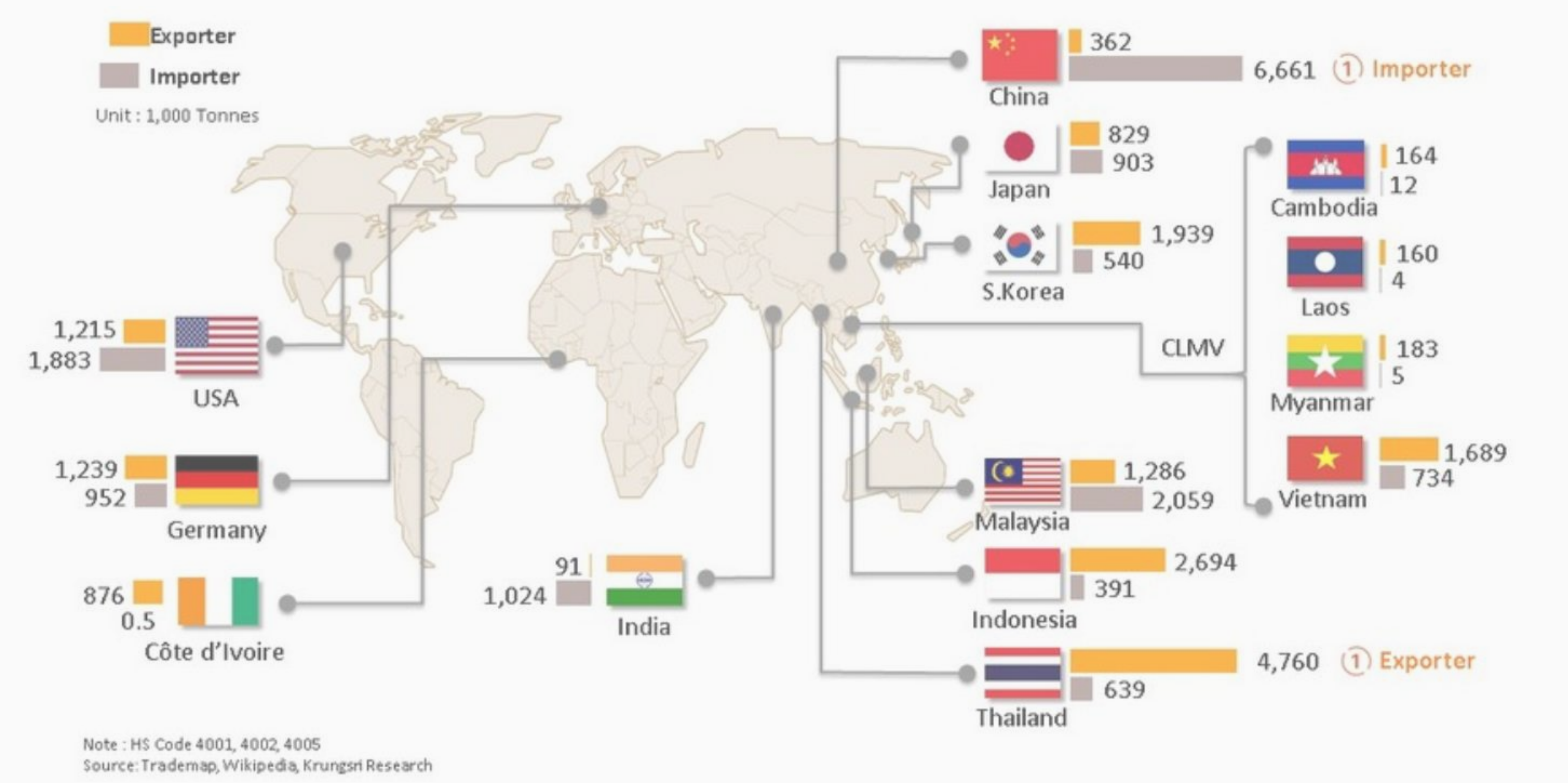
Dr.Arisara Romyen
The 10th AED 2021: 25-08-2021
Economic Faculty, Chiang Mai University

TOPICS

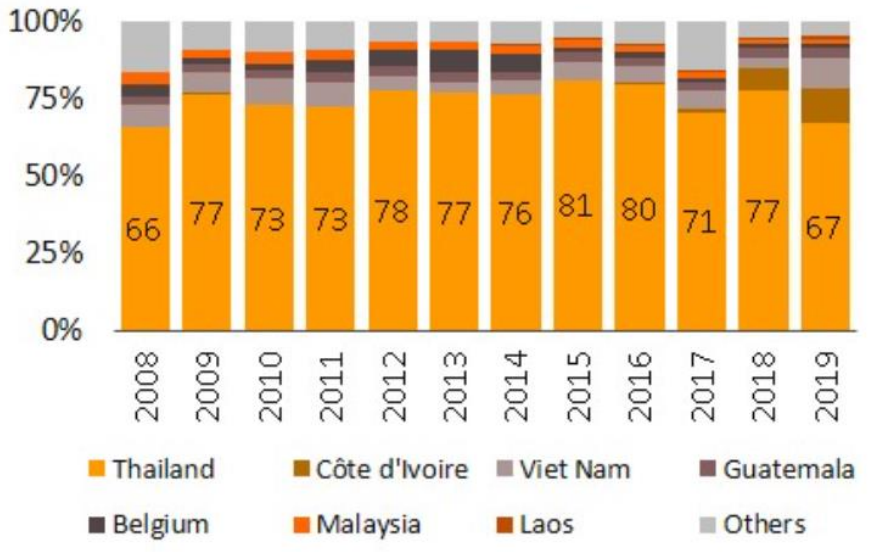
- ✓ Research Article related to a Logistic Model
- ✓ Introduction to a Logistic Model
- ✓ Conduct and Interpret a Logistic Regression
- ✓ STATA Tutorial in Binary and Multinomial Logistic models

Research Article related to a Logistic Model

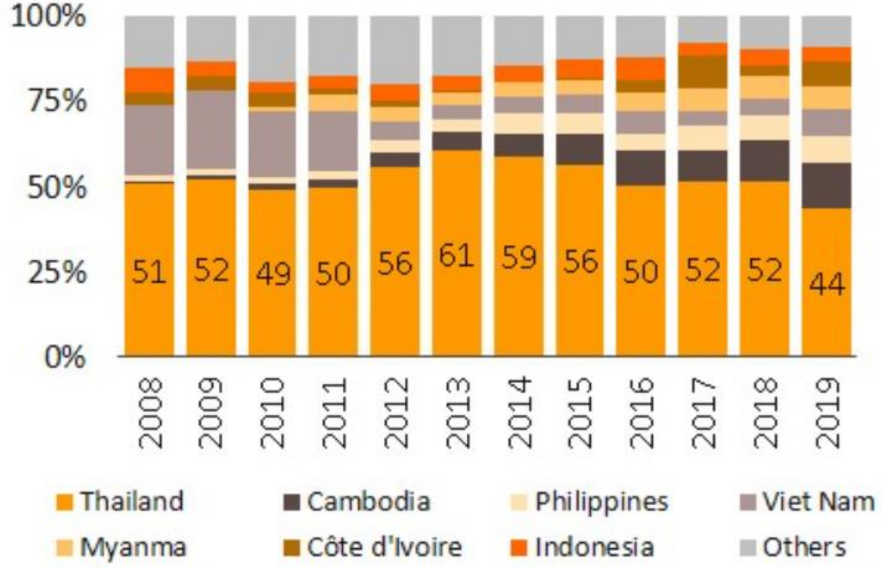
Major Natural Rubber Exporters and Importers (2019)



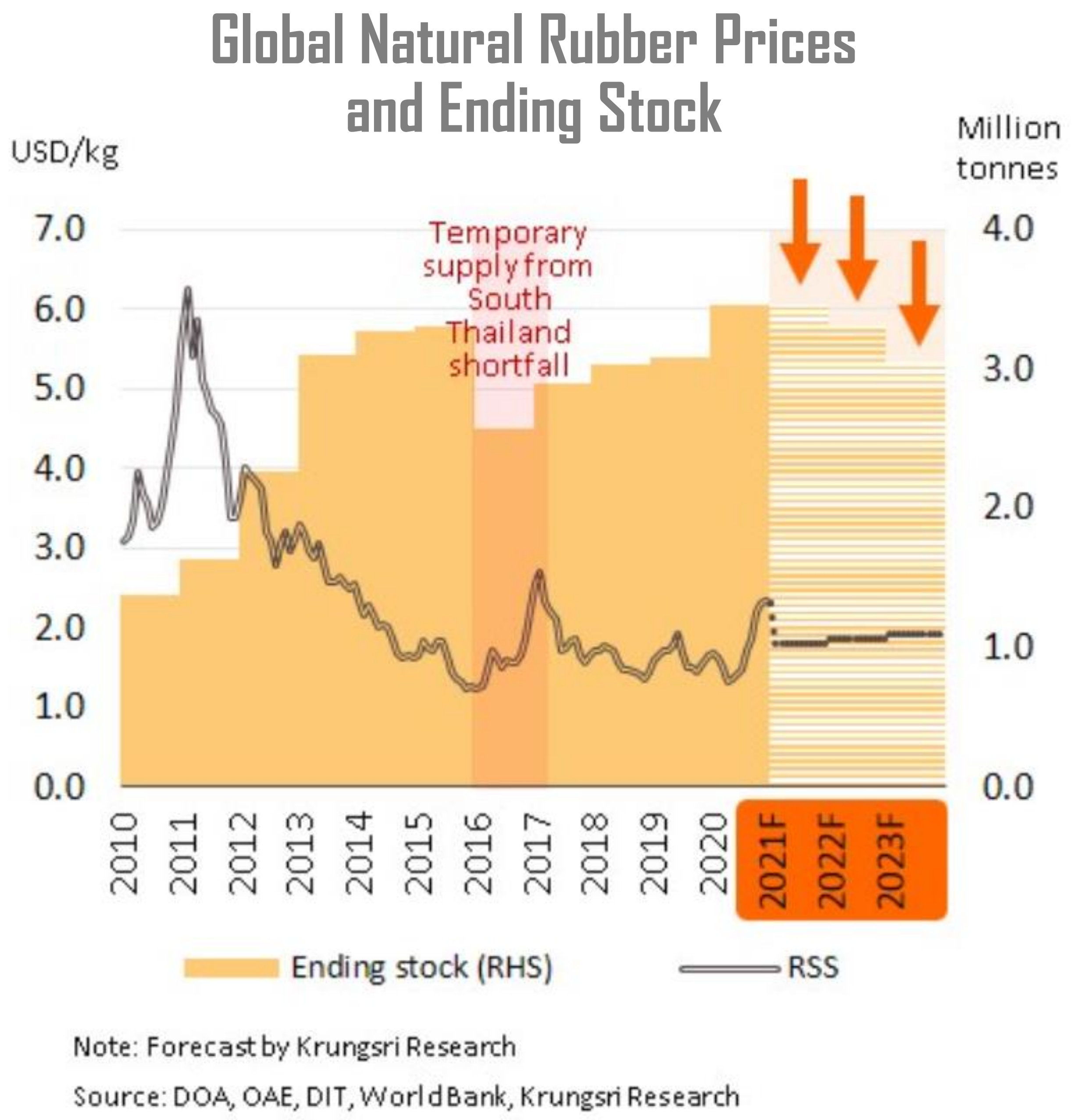
Share of World Concentrated Latex Exports



Share of World Ribbed Smoke Sheet Exports



Source: https://www.krungsri.com/en/research/industry/industry-outlook/Agriculture/Rubber/IO/io-rubber-21



Source: https://www.krungsri.com/en/research/industry/industry-outlook/Agriculture/Rubber/IO/io-rubber-21

In fact, **almost 90%** of rubber production in Thailand is produced by **the rubber smallholders.**

- Specifically, smallholder farmers growing rubber in **monocultures** often suffer income risk due to economic volatility.
- With a research question, Do they should diversify their farm incomes by adopting intercropping rubber or agroforestry?

Contents lists available at ScienceDirect

Kasetsart Journal of Social Sciences

journal homepage: http://www.elsevier.com/locate/kjss

Investigation of rubber-based intercropping system in Southern Thailand

Arisara Romyen*, Palakorn Sausue, Somboon Charenjiratragul [1]

Department of Agricultural Economics, Faculty of Economics, Prince of Songkla University, Songkhla 90110, Thailand

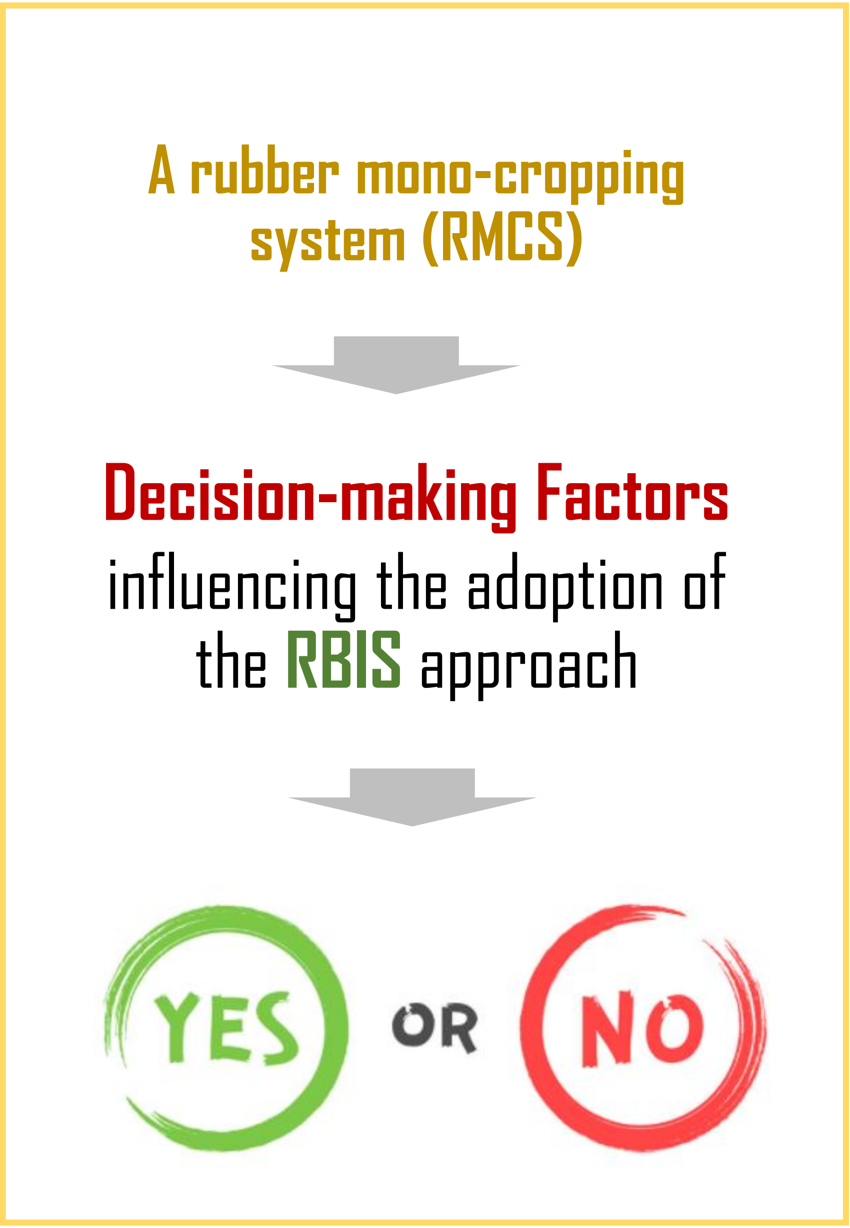
ARTICLE INFO

Article history:
Received 25 February 2016
Received in revised form 6 October 2016
Accepted 19 October 2016
Available online 22 December 2017

Keywords:
decision-making factors,
logistic regression analysis,
rubber monocropping system (RMCS),
rubber-based intercropping system (RBIS),
Southern Thailand

ABSTRACT

The objectives were: to investigate the differences in the socio-economic characteristics of rubber farmers who operate either a rubber mono-cropping system (RMCS) or a rubber-based intercropping system (RBIS), to identify RMCS farmer's attitudes toward RBIS, to determine the decision-making factors influencing the adoption of RBIS, and to examine the different types of intercrop available. The study areas were in Kaopra sub-district, Songkhla province and Tamod sub-district, Phattalung province, Thailand, since in these areas there is already some practice of RBIS. The findings revealed that the size of the rubber tapping area is a significant factor in the adoption of either RMCS or RBIS. The significant factors positively influencing RMCS farmers toward adopting RBIS were: members in the household, level of RBIS knowledge, attendance at an RBIS workshop, and rubber growing experience. The study's findings suggest that rubber intercropping tutorials are a driving force behind the adoption of RBIS. Whilst, it would seem to be a good idea to promote the expansion of the RBIS area in the future, this will be quite difficult to achieve in practice if left to happen naturally and there should be positive measures adopted to promote this expansion.



Introduction

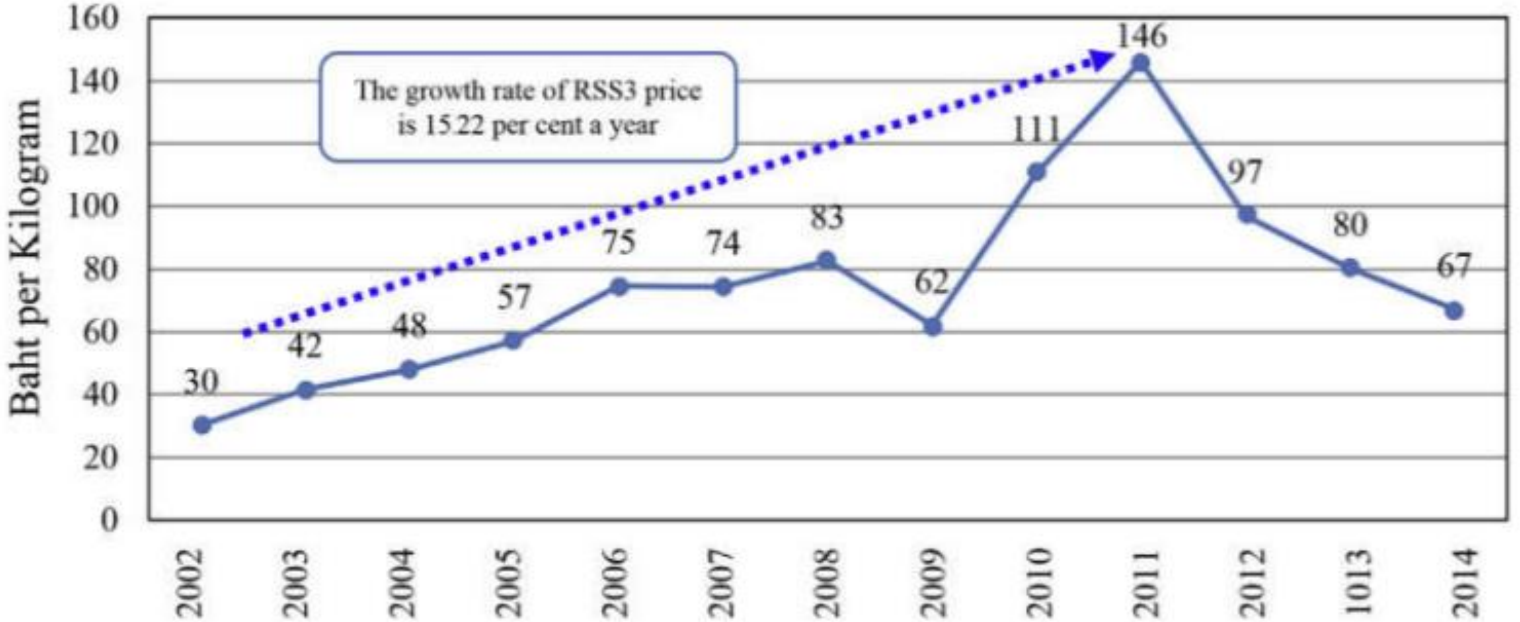
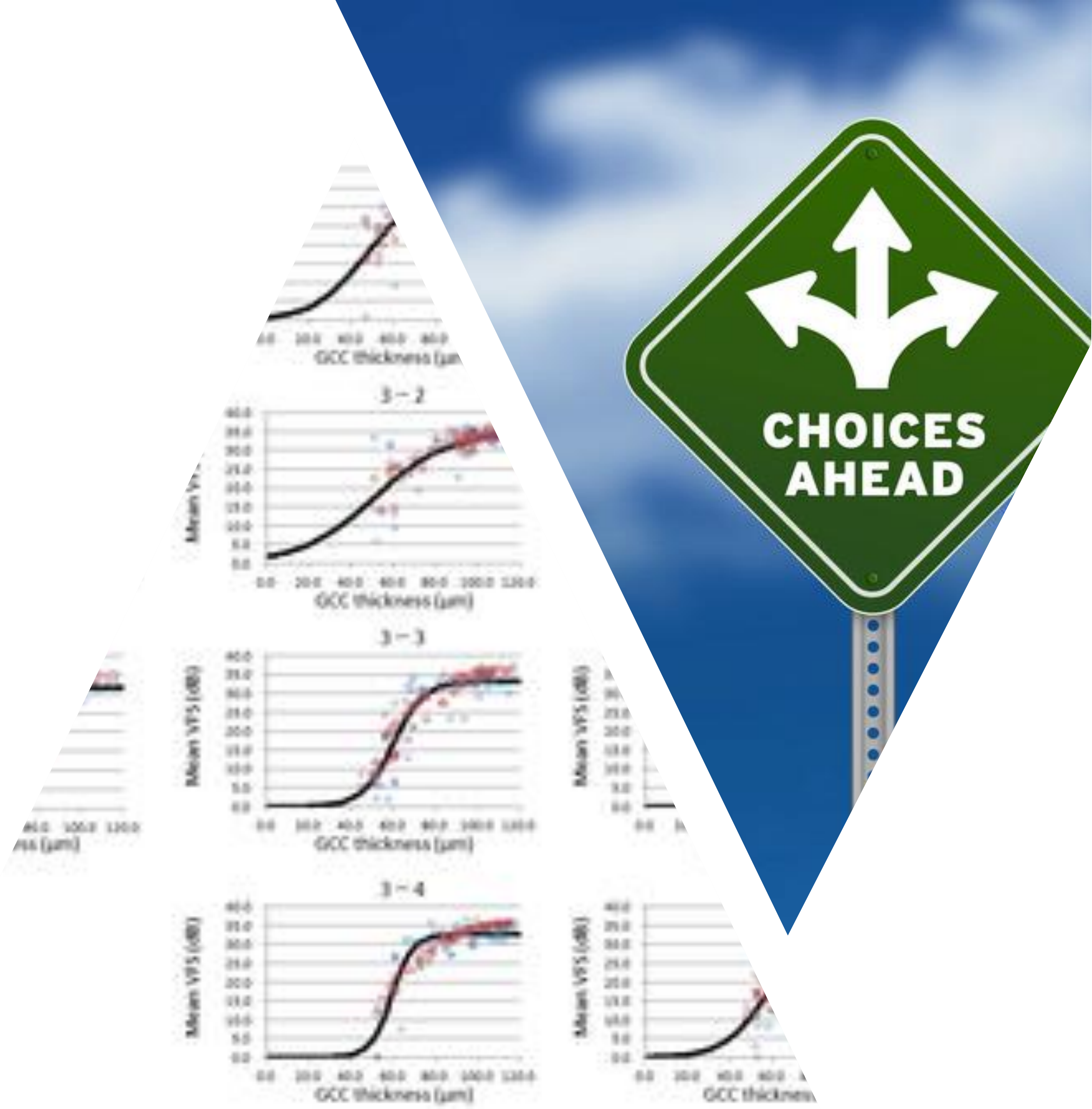


Figure 1 Average price of RSS3 at The Hat Yai Central Rubber Market between 2002 and 2014 (Rubber Research Institution of Thailand, 2014).
Note: The percentage growth rate has been calculated using the exponential curve, $Y = ab^T$.
Source: Charernjiratragul et al. (2015)

The rubber prices in domestic market in Thailand tend to fluctuated over time due to cyclical movements in the world market

Introduction to a Logistic Model



Categorical Response Variables

❑ Binary Response

- ✓ Whether or not a person buys: Y → BUY / NOT BUY
- ✓ Success of a medical treatment Y → SURVIVES / DIES

❑ Ordinal Response

- ✓ Opinion poll responses Y → AGREE / Neutral / DISAGREE

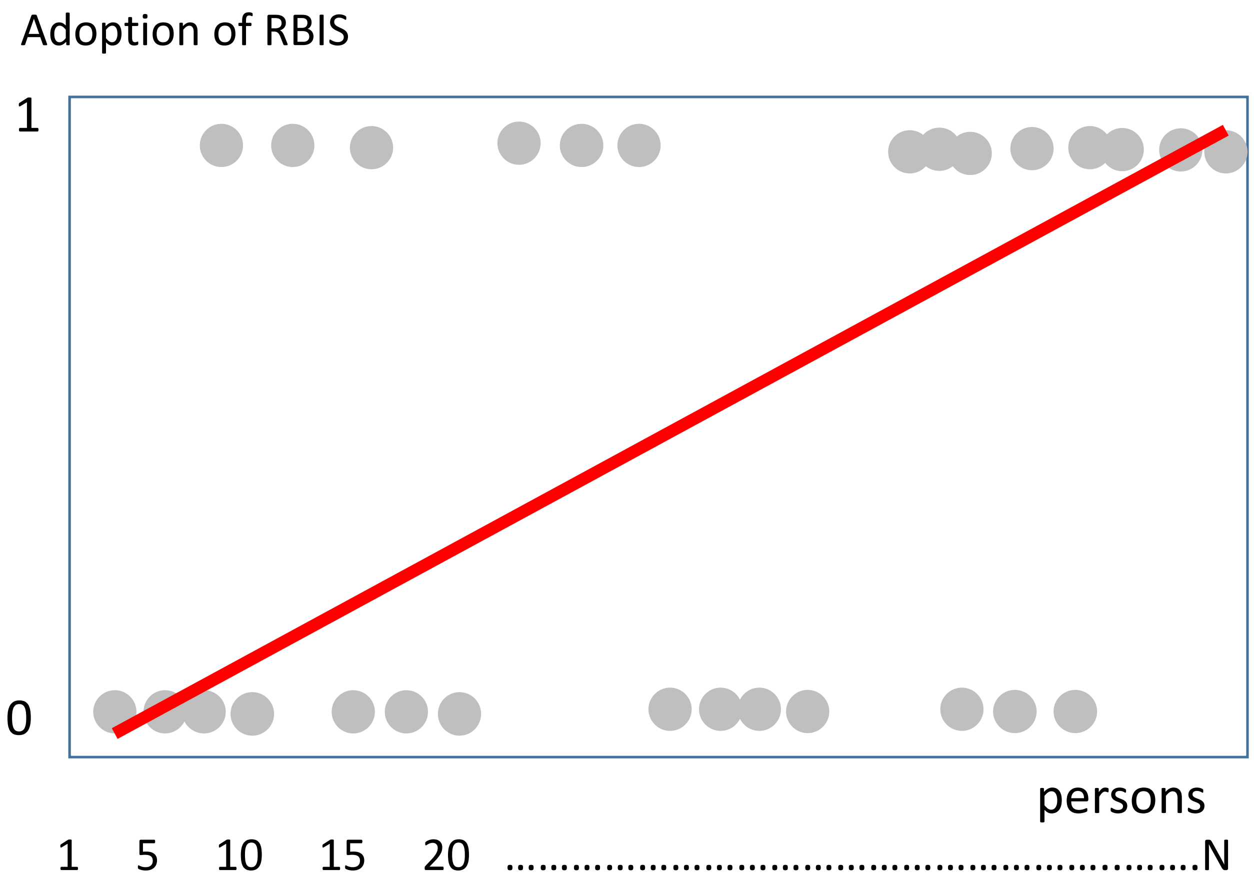
Categorical Response Variables

- In this class of models, we consider the case where the Dependent variable can take the value of 0 or 1. They are often termed Dichotomous variables
- The literature on this type of model is extensive, it can include cases where there are more than 2 possible outcomes, however we are only covering an introductory section of this area or econometrics.
- These types of model tend to be associated with the Cross-Sectional Econometrics rather than time series.

Decision-making Factors
influencing the adoption of the **RBIS** approach

Y=1 Y=0

FAILURE of the Linear Regression



The Dependent variable has two values,

- the value 1 has a probability of *p* and
- the value 0 has a probability of (1-*p*)

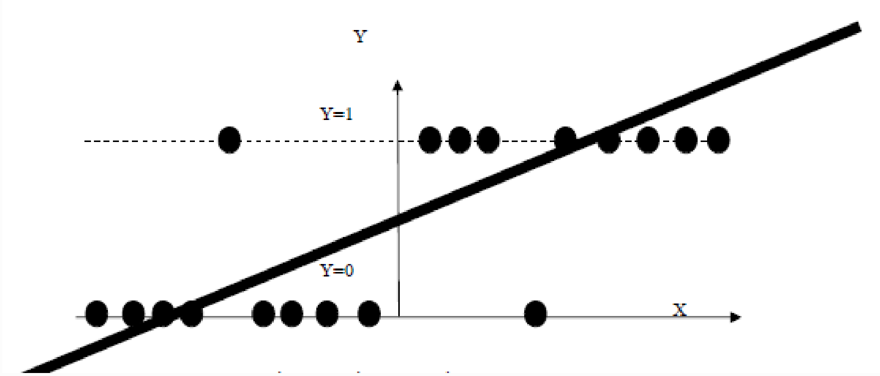
This is known as the Bernoulli Probability Distribution. In this case the Expected Value of a random variable following
a Bernoulli distribution is the probability the variable equals 1. Since the probability of *p* must lie between 0 and 1, then the expected value of the dependent variable must also lie between 0 and 1.
The Error term is not normally distributed, it also follows the Bernoulli distribution.
The variance of the error term is heteroskedastistic.
The variance for the Bernoulli distribution is p(1-p), where p is the probability of a success.
The value of the R-squared statistic is limited, given the distribution of the LPMs.

Possibly the most problematic aspect of the LPM is the non-fulfillment of the requirement that the estimated value of the dependent variable y lies between 0 and 1.

Which function is preferred to classify the dataset (Y) as shown ??

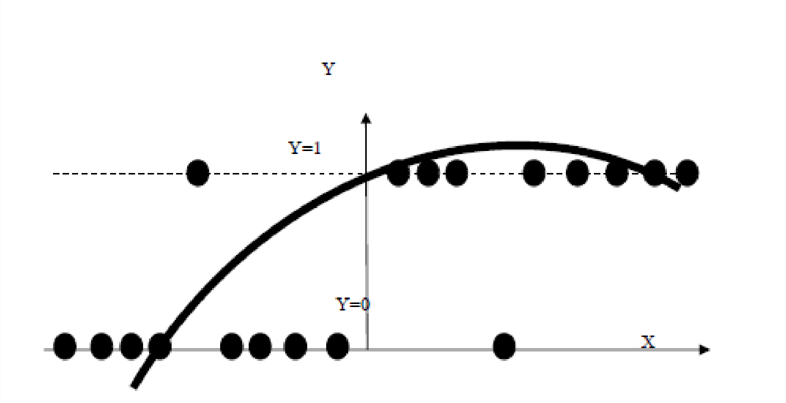
Function 1: Linear Model

It's not really to interpret Y.



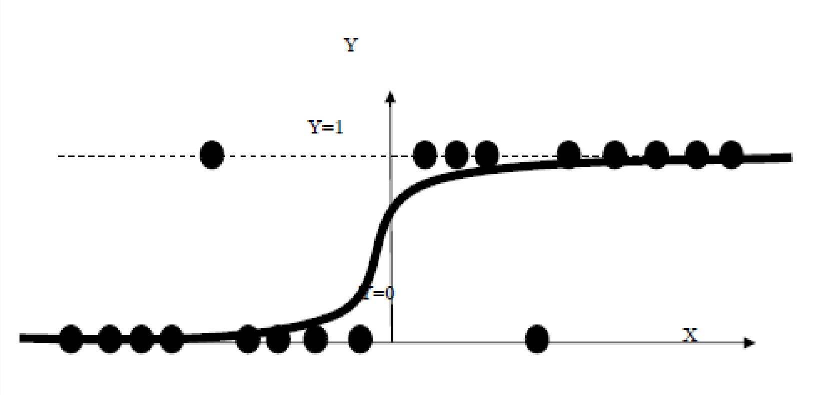
Function 2: Polynomial Model

It's not really to interpret Y.



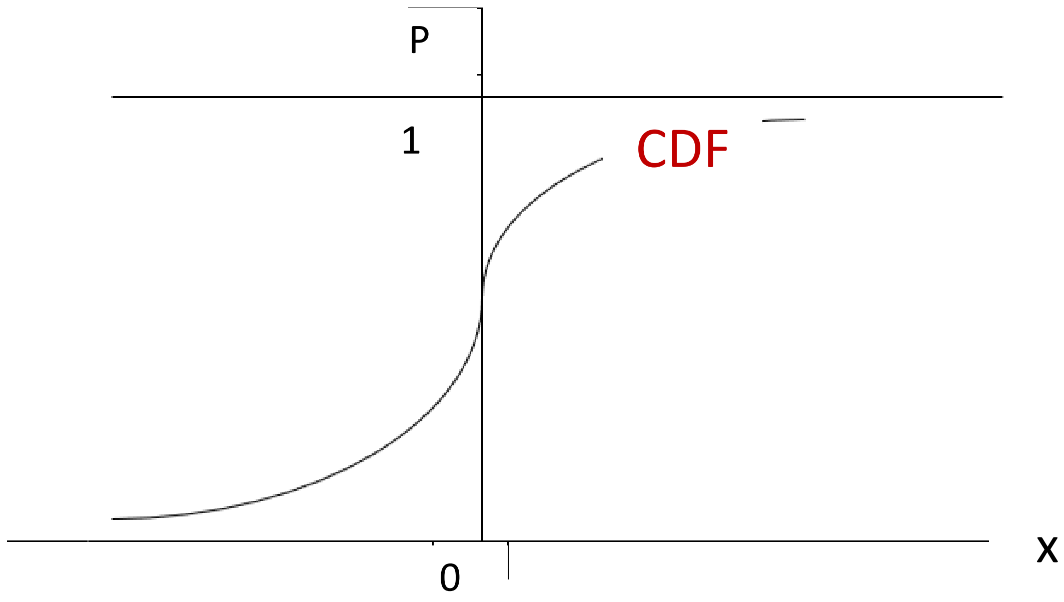
Function 3: Logistic Model

It's precisely to interpret Y.



The Logit Model

- The core criteria around the problems referred to the Bernoulli distribution.
- This needs the apply of a 'S' Shaped Curve, which is similar to the cumulative distribution function (CDF) of a random variable.
- The CDFs used to represent a Discrete variable are the logistic (Logit model) and normal (Probit model).

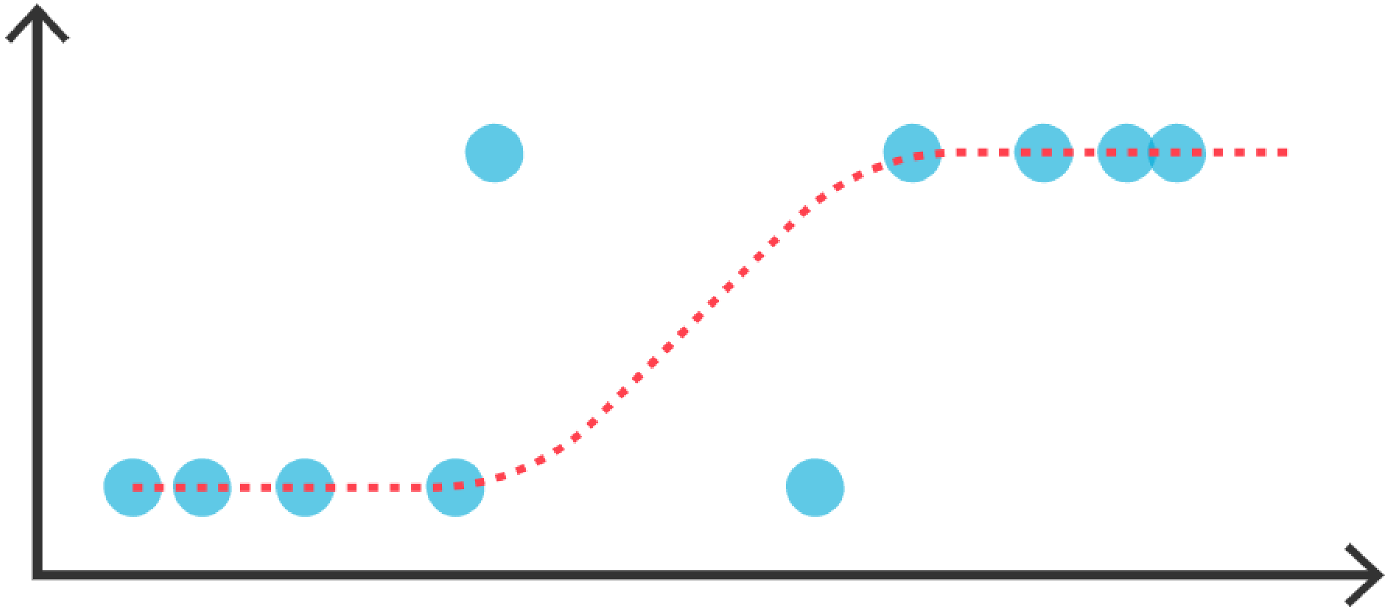


The Logit Model

- In a case of the probability $y = 1$, **the cumulative logistic distribution function** is expressed as:

$$y_i = \alpha_0 + \alpha_1 x_i + u_i$$

$$p_i = E(y = 1 / x_i) = \alpha_0 + \alpha_1 x_i$$

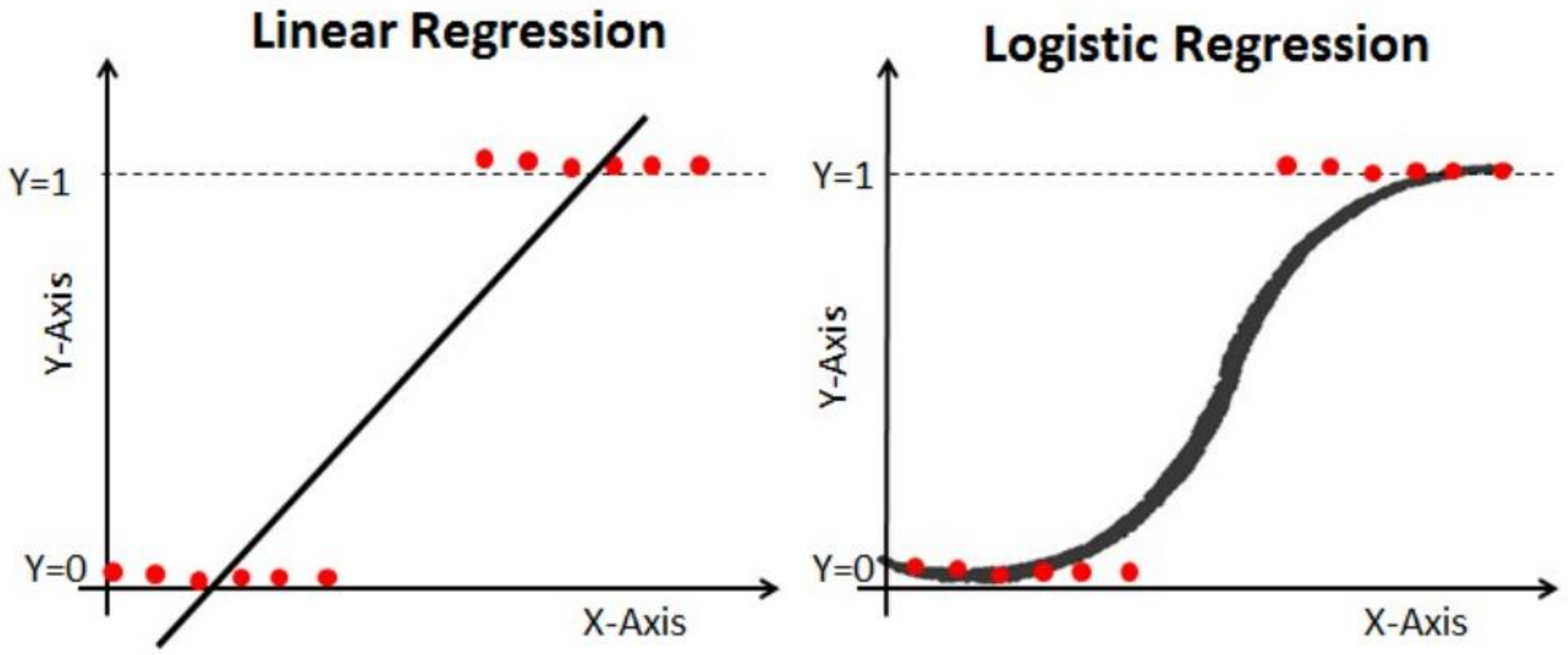


- The Cumulative Logistic Distributive Function can then be written as:

$$p_i = \frac{1}{1 + e^{-z_i}}$$

$$\text{Where}: z_i = \alpha_0 + \alpha_1 x_i$$

The Logit Model



source : www.datacamp.com

- **Probability Function of the Logistic Model**

Probability to Event: Pr(Y=1): $\Pr(Y=1) = \frac{1}{1+e^{-\beta x'}}$

Probability to Non-Event: 1– [Pr(Y=1)]: $\frac{1+e^{-\beta x'}}{1+e^{-\beta x'}} - \frac{1}{1+e^{-\beta x'}} = \frac{e^{-\beta x'}}{1+e^{-\beta x'}}$

To consider the odd ratio:

The likelihood that an event will occur over the likelihood that the event will not occur.

$$\frac{\Pr(Y=1)}{1-[\Pr(Y=1)]} = \frac{1}{1+e^{-\beta x'}} / \frac{e^{-\beta x'}}{1+e^{-\beta x'}}$$

$$= \frac{1}{1+e^{-\beta x'}} * \frac{1+e^{-\beta x'}}{e^{-\beta x'}}$$

$$= \frac{1}{e^{-\beta x'}}$$

$$= e^{\beta x'}$$

$$\frac{\Pr(Y=1)}{1-[\Pr(Y=1)]} = e^{\beta x'}$$

Taking a logarithm;

$$\log\left[\frac{\Pr(Y=1)}{1-[\Pr(Y=1)]}\right] = \exp\ e^{\beta x'}$$

$$\log\left[\frac{\Pr(Y=1)}{1-[\Pr(Y=1)]}\right] = \beta x'$$

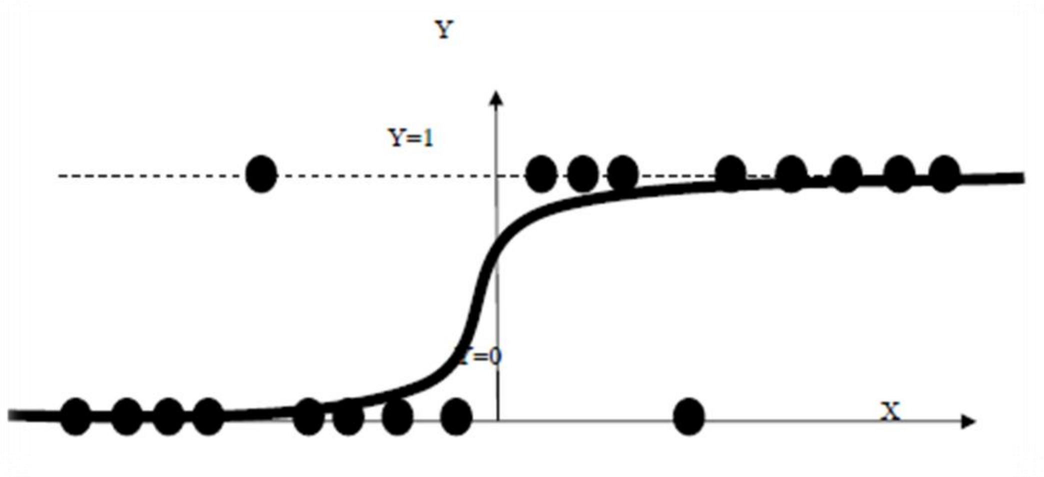
Hence, the odd ratio can be expressed by relevant determinant factors

จาก $\log\left[\frac{\Pr(Y=1)}{1-[\Pr(Y=1)]}\right] = \beta x'$

$$\log\left[\frac{\Pr(Y=1)}{1-[\Pr(Y=1)]}\right] = \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_n x_n$$

the Odd ratio $= \beta_1 x_1 + \beta_2 x_2 + \ldots\ldots + \beta_n x_n$

- ✓ OR >1: Predictor ↑ ,Probability of event occurring ↑
- ✓ OR <1: Predictor ↓ ,Probability of event occurring ↓



The Multinomial Logit Model

- The target variable has three or more nominal categories without ordering, such as predicting that which superstar is more preferred, which food is more preferred, predicting the type of tree. The Multinomial Logit Model replies on characteristics as:

$$P\left[choice = j \middle| z_{it,i,t}\right] = \frac{\exp(\alpha_j + \gamma'_j, z_{it})}{\sum_{j=1}^{J(i,t)} \exp(\alpha_j + \gamma'_j, z_{it})}$$

The Multinomial Logit Model

- A Relative Risk Ratio (RRR) is used to express the probability to the event occurring instead of the OR
- The RRR of a coefficient indicates how the risk of the outcome falling in the comparison group compared to the risk of the outcome falling in the referent group changes with the variable in question. An RRR > 1 indicates that the risk of the outcome falling in the comparison group relative to the risk of the outcome falling in the referent group increases as the variable increases.

Dependent Variables

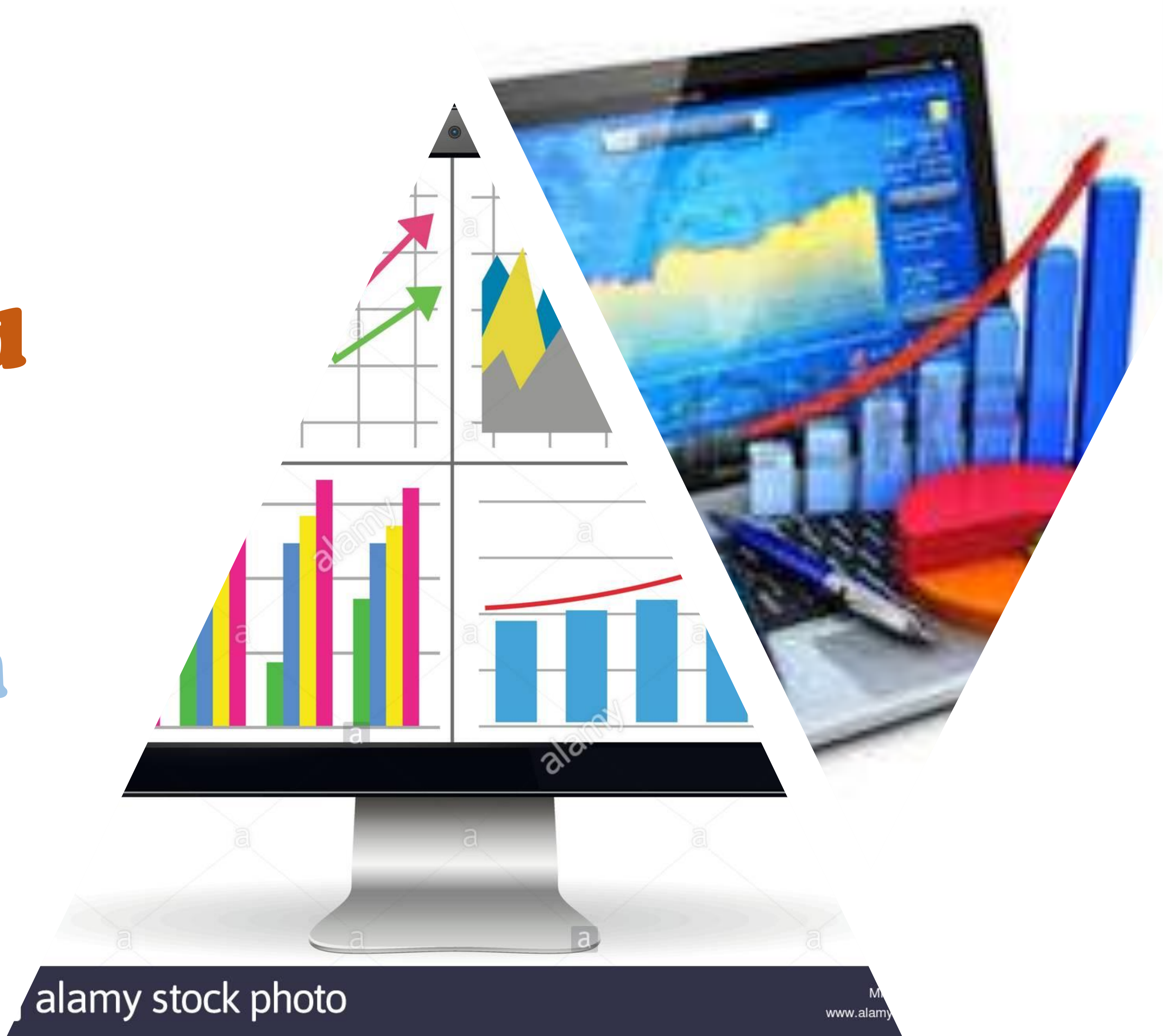
Yi
- 11: Perennial plant
- 22: Fruit
- 33: Vegetable

reference

11 VS 22 (reference=33)
11 VS 33 (reference=22)
22VS 33 (reference=11)

https://stats.idre.ucla.edu/stata/output/multinomial-logistic-regression/

Conduct and Interpret a Logistic Regression



Methodology

I

To explore the decision-making factors influencing the adoption of the RBIS

The dependent variables is defined as Dichotomous variables

$$Y_i \begin{cases} 1, \text{ if RMCS planters decide to operate RBIS in the next 5 years} \\ 0, \text{ otherwise} \end{cases}$$

The multiple Logistic regression:

$$logit\,Y = \beta_0 + \beta_1(RUB\_MEM) + \beta_2(SECOC) + \beta_3(RUBLAND) + \beta_3(RUBLAND) + \beta_4 INTERC + \beta_5(RUB\_EXP) + \beta_6(RRAF) + \beta_7(KNOWLEG) + \beta_8(ATTEND)$$

The explanatory variables

Variable	Label
logit Y	Dependent variable (takes the value 1 if the respondent intends to operate RBIS in next 5 years, 0 if the respondent decides against intercropping)
RUB_MEM	Household members (number of people)
SECOC	Extra occupation (dummy variable where 1 refers to having an alternative occupation and 0 is no alternative occupation)
RUBLAND	Rubber land area (rai)
INTERC	Practice intercropping (dummy variable where 1 refers to conducting intercropping and 0 is not conducting intercropping)
RUB_EXP	Years' experience of growing rubber
RRAF	Member of Rubber Replanting Aid Fund (RRAF) (Dummy variable where 1 refers to being a member of RRAF and 0 is not being a member.
KNOWLEG	Rubber intercropping knowledge (scale variable: 1—5, from low to high knowledge)
ATTEND	Attendance at a rubber intercropping tutorial (Dummy variable where 1 refers to having attended a tutorial and 0 is not having attended one.)

Methodology

II

To investigate the variety of inter-crop grown on rubber lands.

The crops were divided into three categories:

$$\mathbf{Y_i} \begin{cases} \text{1. Perennial Plants} \\ \text{2. Fruit trees} \\ \text{3. Home-grown} \end{cases}$$

The Multinomial Logistic regression:

$$logit\ A\left[\frac{P(group1)}{P(group2)}\right] = \beta_0 + \beta_1(RUB\_MEM) + \beta_2(SECOC) + \beta_3(RUBLAND) + \beta_3(RUBLAND)$$

$$+\beta_4 INTERC + \beta_5(RUB\_EXP) + \beta_6(RRAF) + \beta_7(KNOWLEG) + \beta_8(ATTEND)$$

Results

Table 4
Decision-making factors influencing the adoption of the RBIS approach within the next 5 years

Explanatory variable	Coef.	Robust S.E.	z	*p*	Marginal effect
RUB_MEM	0.221	0.136	2.02	.043**	0.041
SECOC	0.266	0.312	1.12	.265 NS	0.049
RUBLAND	0.006	0.011	0.59	.554 NS	0.002
INTERC	0.024	0.028	−0.83	.407 NS	0.004
RUB_EXP	0.018	0.010	−1.79	.074*	0.004
RRAF	0.229	0.341	0.85	.397 NS	0.042
KNOWLEG	0.332	0.202	2.29	.022**	0.061
ATTEND	0.619	0.558	2.07	.039**	0.114
Constant	0.464	0.352	−0.83	.407 NS	

Note: * significant at the .10 level, ** significant at the .05 level, NS not significant

The rubber experience variable (RUB_EXP) correlates with an increasing tendency to adopt the RBIS approach (ME = 0.004) since experienced RMCS farmers have accumulated a lot of knowledge regarding agriculture. This longer experience might cause them to adopt RBIS. In addition, they have inevitably suffered fluctuations in the rubber price over the decades, and therefore they understand well the market mechanisms of the global rubber market. Operating an RBIS approach might improve their household wealth.

The rubber intercropping knowledge variable (KNOWLEG) is also a factor which increases the likelihood of the adoption of the RBIS approach (ME = 0.061). RMCS farmers are able to obtain training in rubber intercropping which is organized by local civic groups, the Rubber Replanting Aid Fund, or other government agencies. Generally, rubber farmers (based on their perception or habit) have tended to view RBIS as an obstacle to the management of their rubber plantations. However, acquiring more knowledge about rubber intercropping can alter such attitudes. In fact, RBIS is not a new agricultural concept in the Kaopra sub-district, of Songkhla province or in the Tamod sub-district of Phattalung province; nevertheless, this concept has not hitherto been a common practice.

Similarly, variable for attendance at a rubber intercropping tutorial (ATTEND) is positively correlated with the intention to adopt RBIS and shows the largest magnitude effect (ME = 0.114). It is evident that attendance at a rubber intercropping tutorial directly provides new information and promotes this ides to those who attend.

Results

Table 5
Crop types selected by intended adopters of RBIS

Type of crop	n = 345	Percentage
Perennial plants	199	57.7
Fruit trees	106	30.7
Homegrown vegetables	40	11.6

To calculate the probabilities of the adoption of crop different types for rubber intercropping, a multinomial logistic regression model was applied to predict preferences between different types of crops based on three crop comparisons: (1) perennial plants versus homegrown vegetables, (2) fruit trees versus homegrown vegetables, and (3) perennial plants versus fruit trees. Table 5 shows the number of farmers preferring each crop together with their respective percentages. In summary, the cultivation of perennial plants such as Iron Wood, Eagle Wood, and Champak was the most popular model (57.7%), followed by fruit trees and homegrown vegetables at 30.7 percent and 11.6 percent, respectively.

Table 6
Factors affecting the decision to choose crop type for rubber intercropping

Explanatory variable	Model[1]: perennial plants *vs.* homegrown vegetables			Model[2]: fruit trees *vs.* homegrown vegetables			Model[3]: perennial plants *vs.*fruit trees		
	Coef.	Sig	Exp(B)	Coef.	Sig	Exp(B)	Coef.	Sig	Exp(B)
Rub_MEM	0.170	.381	1.185	0.865	.409	1.181	0.003	.978	1.003
SECOC	−0.336	.368	0.715	0.167	.649	0.836	−0.157	.541	0.855
RUBLAND	0.015	.425	1.015	−0.179	.309	1.020	−0.005	.671	0.995
INTERC	−0.010	.810	0.990	0.020	.892	1.006	−0.016	.590	0.984
RUB_EXP	0.015	.401	1.015	−0.002	.928	0.998	0.017	.166	1.017
RRAF	1.050	.005***	2.858	0.631	.109	1.879	0.420	.143	1.521
KNOWLEG	0.180	.392	1.197	−0.278	.221	0.757	0.458	.002***	1.581
ATTEND	0.238	.565	1.269	−0.280	.538	0.755	0.519	.082*	1.680
Constant	−0.355	.666		0.865	.309		0.034	.034**	

Note: Model[1, 2]: homegrown vegetables is the reference category and in Model[3] fruit trees is the reference category. * significant at the .10 level, ** significant at the .05 level, *** significant at the .01 level, [NS] not significant

Results

Model 1: perennial plants versus homegrown vegetables

Being a member of the RRAF was significant in choosing perennial plants as against homegrown vegetables (exp = 2.858). In 2013, the RRAF promoted corporate social responsibility (CSR) for the expansion of green areas in rubber plantations projects with a budget of THB 1.5 million over the six southern provinces of Thailand. Many perennial plants such as Iron Wood, Mahogany, *Shorea roxburghii*, and *Tectona grandis* were distributed. Consequently, rubber farmers who are members of the RRAF, tend to grow perennial plants rather than homegrown vegetables. Those farmers who grow perennial plants are also likely to earn higher incomes from the variety of plants which they grow.

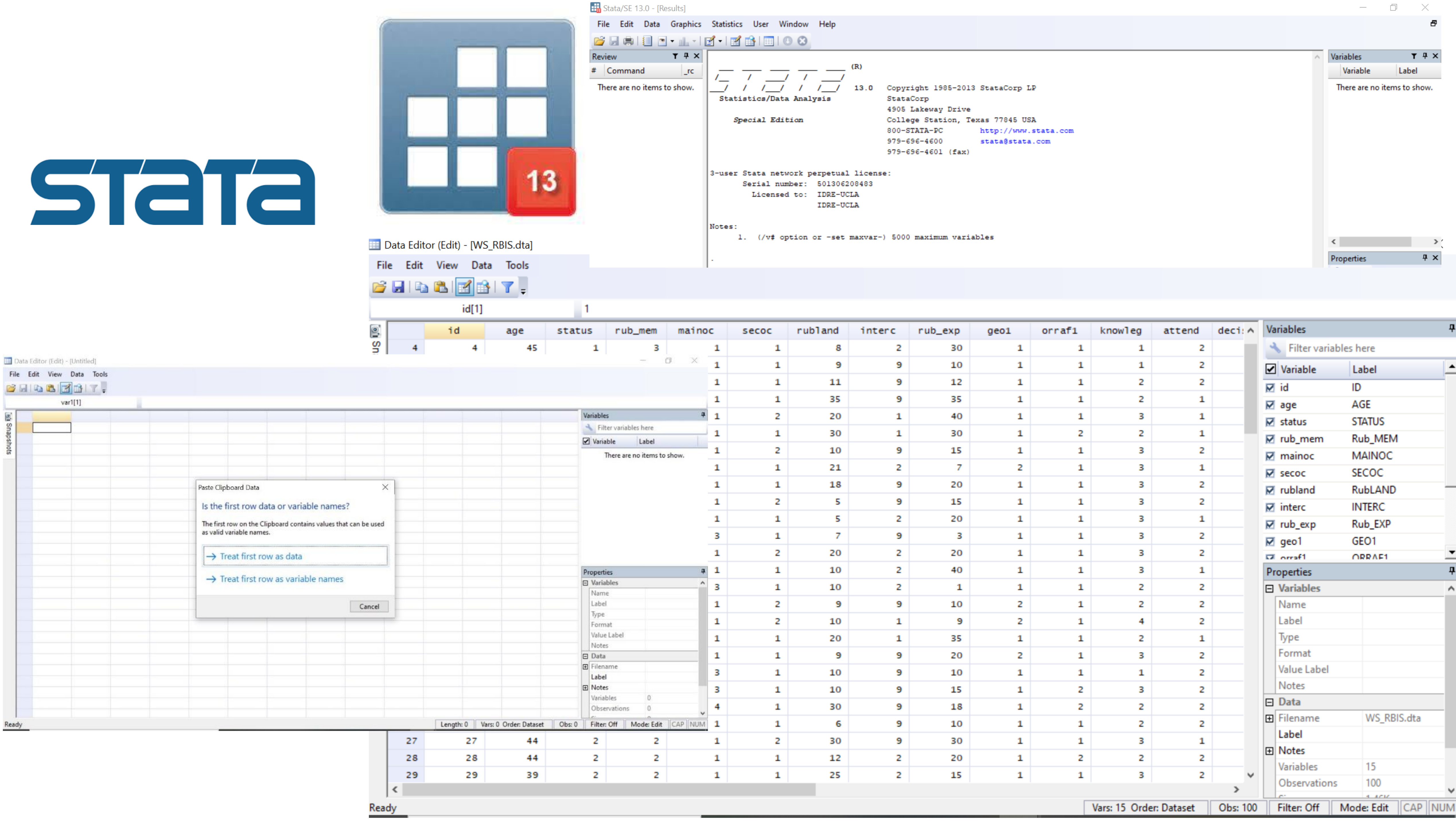
Model 2: fruit trees versus. homegrown vegetables

Interestingly, there were no significant differences between fruit trees and homegrown vegetables because the cultivation of either type of crop is often subject to market failure. The problems associated with the production and marketing of these types of crops are therefore similar.

Model 3: perennial plants 2 versus fruit trees

The rubber intercropping knowledge variable (KNOWLEG) and attendance at a rubber intercropping tutorial variable (ATTEND) appear to be important factors in the adoption of the RBIS approach and in this comparison, the exp (B) values of these two variables were 1.581 and 1.680, respectively, indicating that rubber farmers with more knowledge and those who have attended an RBIS tutorial are more likely to select perennial plants for the reasons mentioned above under Model 1.

STATA Tutorial in Binary and Multinomial Logistic models


Stata/SE 13.0 - [Results]
File Edit Data Graphics Statistics User Window Help
Review
Command
_rc
There are no items to show.
13.0 Copyright 1985-2013 StataCorp LP
Statistics/Data Analysis
StataCorp
4905 Lakeway Drive
Special Edition
College Station, Texas 77845 USA
800-STATA-PC http://www.stata.com
979-696-4600 stata@stata.com
979-696-4601 (fax)
3-user Stata network perpetual license:
Serial number: 501306208483
Licensed to: IDRE-UCLA
IDRE-UCLA
Notes:
1. (/v# option or -set maxvar-) 5000 maximum variables
Variables
Variable
Label
There are no items to show.
Properties
Data Editor (Edit) - [WS_RBIS.dta]
File Edit View Data Tools
id[1]
1
id age status rub_mem mainoc secoc rubland interc rub_exp geo1 orraf1 knowleg attend
Variables
Filter variables here
Variable Label
id ID
age AGE
status STATUS
rub_mem Rub_MEM
mainoc MAINOC
secoc SECOC
rubland RubLAND
interc INTERC
rub_exp Rub_EXP
geo1 GEO1
Properties
Variables
Name
Label
Type
Format
Value Label
Notes
Data
Filename WS_RBIS.dta
Label
Notes
Variables 15
Observations 100
Ready
Vars: 15 Order: Dataset
Obs: 100
Filter: Off
Mode: Edit
CAP
NUM
Data Editor (Edit) - [Untitled]
File Edit View Data Tools
var1[1]
Snapshots
Paste Clipboard Data
Is the first row data or variable names?
The first row on the Clipboard contains values that can be used as valid variable names.
Treat first row as data
Treat first row as variable names
Cancel
Variables
Filter variables here
Variable Label
There are no items to show.
Properties
Variables
Name
Label
Type
Format
Value Label
Notes
Data
Filename
Label
Notes
Variables 0
Observations 0
Ready
Length: 0
Vars: 0 Order: Dataset
Obs: 0
Filter: Off
Mode: Edit
CAP
NUM

WS_RBIS_command.do | Untitled.do

```
##Preparing dataset

describe

##summary##
tabulate decisionl
tab orrafl

##Binary Logistic Model##

logit decisionl age rub_mem rubland rub_exp orrafl knowleg attend
logit decisionl age rub_mem rubland rub_exp orrafl knowleg attend, or

logistic decisionl age rub_mem rubland rub_exp orrafl knowleg attend
logistic decisionl age rub_mem rubland rub_exp orrafl i.knowleg attend


##Marginal Effect##
margins, dydx(age rub_mem rubland rub_exp orrafl i.knowleg attend)


==============================================================================
==============================================================================
```

```
####Multinomail Logit Model##


##change string to numberical figure
encode type, generate(type2)
describe
tab type2
tab type2, nolabel
recode type2 (1=12) (2=10) (3=11) (4=13)

mlogit type2 age rub_mem rubland rub_exp orrafl knowleg attend, baseoutcome(11)
mlogit type2 age rub_mem rubland rub_exp orrafl i.knowleg attend


##RRR##
mlogit type2 age rub_mem rubland rub_exp orrafl knowleg attend, baseoutcome(11) rrr



##Marginal Effect
margins, dydx (status rub_mem secoc rub_exp orrafl knowleg)
```

```
. logit decision1 age rub_mem rubland rub_exp orraf1 knowleg attend , or

Iteration 0:   log likelihood = -232.30256
Iteration 1:   log likelihood = -220.45559
Iteration 2:   log likelihood = -220.17445
Iteration 3:   log likelihood = -220.17383
Iteration 4:   log likelihood = -220.17383

Logistic regression                             Number of obs   =        400
                                                LR chi2(7)      =      24.26
                                                Prob > chi2     =     0.0010
Log likelihood = -220.17383                     Pseudo R2       =     0.0522
```

decision1	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.9879021	.0121495	-0.99	0.322	.9643743	1.012004
rub_mem	1.214668	.1522158	1.55	0.121	.9501438	1.552836
rubland	1.010131	.0112898	0.90	0.367	.9882441	1.032503
rub_exp	.9910399	.013934	-0.64	0.522	.9641027	1.01873
orraf1	1.256952	.3364203	0.85	0.393	.7438688	2.123933
knowleg	1.388776	.1858904	2.45	0.014	1.06831	1.805374
attend	.5262457	.1550159	-2.18	0.029	.2954273	.9374033
_cons	3.519532	3.182052	1.39	0.164	.5982783	20.70459

Thank You

Questions please