# Modeling Agricultural Technology Adoption Using the Software STATA

CIMMYT-ALP
Training
Manual
No. 1/2005
(Part Two)

Augustine Langyintuo
Mulugetta Mekuria

Presented at a Training Course Organized by CIMMYT-ALP for its NARS Partners in Southern Africa on: "Econometric Application to Modeling the Adoption of Agricultural Technologies".
21 – 25 February, 2005, Harare, Zimbabwe

International Maize and Wheat Improvement Center
(CIMMYT)
PO Box MP 163, Mount Pleasant, Harare, Zimbabwe

**Table of Contents**

## 1. Introduction to the Training Course

### 1.1 Objectives of the course

As part of an effort to improve socio-economic capacity within the NARS, CIMMYT plans on organizing a one-week course on the "**Econometric Application to Modeling the Adoption of Agricultural Technologies"** scheduled for February 21 – 25, 2005 in Harare, Zimbabwe under the auspices of the Rockefeller Foundation Funded seed project "*Strengthening Seed Marketing Incentives in Southern Africa to Increase the Impact of Maize Breeding Research".* The main objective of the course is to expose participants to basic econometric methods required to build effective agricultural technology adoption models and modeling using the software package STATA. The training course in itself is divided into two complementary parts. Part One deals with the basic foundations of econometrics applied to technology adoption models. Part Two takes the concepts forward and applied to the setting up technology adoption models and implementing them using the software STATA.

### 1.2 Outline of the course in Part Two

The outline of the course in Part Two is as follows: (i) Importance of adoption studies: This section presents the relevant arguments why it is necessary to conduct adoption studies. (ii) Review of determinants of agricultural technology adoption: The main determinants of adoption are discussed under the following categorise (a) farmer's characteristics, (b) institutional factors and (c) technology specific attributes. (iii) Application of Non-Linear Probability Models to Specify Agricultural Technology Adoption Models. This section explores the use of non-linear models to estimate technology adoption. The choice of a Tobit model as an appropriate model when both adoption and intensity of adoption is to be estimated is discussed in section (iv): Other models for examining adoption: Tobit Model (v) Describing aggregate adoption process; This section makes an effort to present reasons why it is important to describe the whole adoption process and gives a method for carrying out such an examination. The final section (vii): Estimation of adoption models using STATA. Participants were taken through a hands-on-training on how to specify and analyse adoption models using the software STATA which was preceded by analysis of basic descriptive statistics.

This Part Two complements Part One in which the basic econometric foundations of modelling are discussed. Participants are encouraged to read PART One before Part Two.

### 1.3 Expectations of the course

It is expected that participants of the workshop would leave with some confidence in the conduct of adoption studies. They would also have some experience in the use of STATA. However, given the limited time devoted to the hands-on-training, participants were advised to continue to practice using the manuals to improve their proficiency.

## 2. Importance of adoption studies

### 2.1 Introduction

A number of agricultural economists and sociologists have attempted to understand why agricultural technologies developed by biophysical scientists have not been making the desired difference on farmers' production, income and livelihoods. Since the classic work of Rogers (1962) on the diffusion of agricultural new technologies in the US, the preoccupation of subsequent studies has been explaining the determinants of innovation diffusion and adoption.

**Adoption Defined**: Adoption can be defined in several ways but most important is to have the definition agreed upon so that the criteria for measurement are acceptable to all concerned. A simplistic definition of adoption is basically the use of a technology. This is further elaborated as the incidence/pattern and intensity of adoption. The incidence indicates whether a farmer has used a technology or not and the latter explains the degree of use of a technology. An example of pattern or incidence would be whether a farmer has used fertilizer or not (Yes or No). Percent of farmers that have applied fertilizers on their maize fields in 2004 in a region or district. The intensity would show the rate of fertilizer use per hectare that farmer applied.

### 2.2 Why Adoption studies are important

It is well understood that technology generation and development is an iterative process and the supply of technologies needs to be driven by demand from the users. Adoption studies are therefore important for the following reasons:

**i) To quantify the number of technology users over time to assess impacts or determine extension requirements:** An adoption study would help us in monitoring and feedback in technology generation. In a traditional technology generation/development and transfer continuum model, it is assumed that researchers would pass the technology onto extension agents to take it to farmers and the technology would work and be adopted by farmers. Many years of development efforts proved that such an approach has not worked. A participatory approach to technology development and transfer model is now very popular and contributed to better technology development and transfer. Adoption studies would provide further insights into the effectiveness of technology transfer.

**ii) To provide information for policy reform**: It is well documented that agricultural development efforts in sub-Saharan Africa are constrained by the lack or inadequate agricultural development polices that support development in general and agricultural research and development in particular. It is important that adoption studies emphasize and understand the policy bottlenecks to technology adoption. Such studies would enhance the development of effective policies for technology adoption.

**iii) To provide a basis for measuring impact:** A number of economists have estimated the high Rate of Return (RoR) to investments in agricultural research. Despite this, policy makers and donors are not convinced that their resource allocation to agricultural research brings the desired impacts and development. We are observing the downward trend in investments in agricultural research and transfer in most African countries. With the exception of South Africa and Ethiopia, most SSA National agricultural Research Services (NARS) are heavily dependent on donor support to implement even their day-to-day operations. Adoption studies if well conducted would hence enable us to measure the impact of technology generation and transfer and strengthen the case for more investment in agricultural research.

## 3. Review of determinants of agricultural technology adoption

Social scientists have been struggling over the years to explain the limited adoption of improved agricultural technologies by farmers. Although it is an important problem not much has been documented in southern Africa. To prepare participants for future work in technology adoption, this section is designed to discuss the factors that one would need to consider in studying adoption. Please do not consider the issues discussed here as an exhaustive but rather a motivation in the thinking process. The discussions are strategically grouped under (a) farmer characteristics, (b) institutional factors (including farm characteristics e.g., size), and (c) broad technological attributes as they relate to adoption of improved technologies.

### 3.1 Farmer Characteristics

#### 3.1.1 Gender

In southern Africa as in most parts of the developing world most extension workers are men and are usually biased towards men in their extension activities. Yet women play a significant role in agriculture especially widows. It is important to include gender in assessing the determinants of adoption of improved technologies. It is important to assess the relative benefits to be gained in redirecting extension efforts by gender.

#### 3.1.2 Age of Farmer

The role of a framer's age in explaining technology adoption is somewhat controversial in the literature. Older people are sometimes thought to be less amenable to change and hence reluctant to change their old ways of doing things. In this case, age will have a negative impact on adoption. On the other hand, older people may have higher accumulated capital, more contacts with extension, better preferred by credit institutions, larger family sizes, etc all of which may make them more prepared to adopt a technology than younger ones. Whatever the condition, it is important to include age as a factor that would help explain adoption decisions although years of experience in farming or sole decision making is sometimes preferred. Note, however, the two latter variables are usually correlated with age. The question of concern is what advantage is to be gained in targeting younger farmers versus older ones?

#### 3.1.3 Years of Experience as a Farmer or Sole Decision Maker on farm operations

The reason why researchers sometimes prefer to use years of experience in farming or main decision maker is that with increased farming experience, farmers are generally better able to assess the relevance of new technologies. This often comes from their interactions with their neighbors and the outside world. Because of their experience, they also tend to be better placed to acquire the needed skills to use the technologies compared with younger ones. Is that the case? One needs to investigate that.

### 3.1.4 Education

It is often assumed that educated farmers are better able to process information and search for appropriate technologies to alleviate their production constraints. The belief is that education gives farmers the ability to perceive, interpret and respond to new information much faster than their counterparts without education. But it is often the case in many countries that the majority of farmers are illiterate. Nevertheless it is significant to examine the role education plays in technology adoption decisions.

## 3.2 Institutional Factors

### 3.2.1 Farm Size

The size of the family farm is a factor that is often argued as important in affecting adoption decisions. It is frequently argued that farmers with larger farms are more likely to adopt an improved technology (especially modern varieties) compared with those with small farmers as they can afford to devote part of their fields (some times the less productive parts) to try out the improved technology. It is also known in the literature that lumpy technologies such as mechanized equipments that require economies of size for it to ensure profitability, there is often a minimum threshold farm size for adoption (e.g. animal traction). But in general, the directional effect of farm size on adoption is contradictory.

It important to note, however, that farm size is often a variable that mirrors the effects of other interlinked variables and is very important to be able to separate these effects. For example, what may be showing up as the effect of farm size may be due to credit, labor availability etc. Understanding the relative roles of these variables is facilitated by adequate knowledge of the study area.

### 3.2.2 Membership of Farmers' Association

In most farming communities farmers form or join associations or cooperatives of various kinds for all sorts of reasons. Such associations or cooperatives sometimes afford farmers the opportunity to have better access to information, which is an important condition for adopting an improved technology. Some financial institutions are prepared to lend credit to farmers only when they are in an association or cooperative. Therefore belonging to an association or cooperative can influence farmer's decision to adopt an improved technology. The question is: how relevant are existing farmers' associations or cooperatives in a given community in influencing technology adoption? Should efforts be made in mobilizing farmers to form associations where they don't exist or to join any that are existing? These are some the questions one needs to answer in an adoption study.

### 3.2.3 Leadership Position in the Community

The tendency for an extension agent to link first to a leader in a given community is often very high. Coupled with the pressure of showing good leadership, community leaders are often known to adopt technologies faster than none leaders although their negative perceptions could negate thoughts on adoption and subsequently influencing their neighbors.

### 3.2.4 Access to Credit

In the literature it has been argued that the lack of credit is a constraint to adoption. This is often the case for lumpy technologies (e.g., mechanical technologies) in particular although improved adoption may require credit to procure complementary inputs to maximize their benefits. Farmers can invest in new technologies either from past accumulated capital or through borrowing from capital markets. The lack of sufficient accumulated savings by smallholder farmers prevents them from having the necessary capital for investing in new technologies. Also capital market failure exists in most developing countries due to the lack of information on interest rates and alternative credit sources.

### 3.2.5 Information

Exposure to information reduces subjective uncertainty and therefore increases the likelihood of adoption of new technologies. Various approaches have been used to capture the impact of information including:

1. Determining whether or not the farmer was visited by an extension agent in a given time,
2. Asking whether or not the farmer attended demonstration tests for the new technologies by extension agents, and
3. Observing the number of times the farmer has participated in on-farm tests, etc.

Sometimes information variables are not significant in explaining adoption because the proxy variable used do not measure what they are intended to measure. For instance based on past experience, farmers have lost confidence in the extension system. In such a case, farmers learn from the experiences of their neighbors and the relevant variable to include to measure “information” is “contact with other farmers using new innovations” and not contact with extension agent.

### 3.2.6 Output and Input Markets

Input and output markets are also known to influence the adoption of improved agricultural technologies. It is often useful to determine market accessibility of the villages being surveyed. One would ask the question: What are the distances of the villages to the nearest major input or output markets?

### 3.3 Technology Characteristics

Farmers make subjective inter-varietal comparisons of the attributes of new and local varieties and they adopt modern varieties only when they are perceived as having better characteristics than the locals. Adoption studies should pay special attention to farmers’ perceptions of technology characteristics (such as yield, taste, resistance to diseases and pests, etc) by using either dummy variables or farmers’ rating scores of inter-varietal comparisons of characteristics.

## 4. Methods for Analyzing Adoption Patterns:

It is important to understand how we measure and describe adoption and then see what is acceptable and useful to farmers, identify what is not, and suggest ways for improving the situation. We need to define the criteria for adoption or what constitutes adoption. The following considerations are critical when we want to measure adoption:

i) What kind of technology?
ii) How fast is the technology adopted?
iii) What are the implications of adopting a new technology on the farming system?
iv) How widely is the technology used on the farm?

Adoption studies require careful analysis and presentation of the results. However, the form and focus of the final report depends very much on the purpose of the study. Two main focal points can be considered:

1. Documenting the degree of adoption: This form focuses on describing adoption patterns, which requires having a clear understanding of the definition of adoption and presenting the degree of adoption, as well as a historical analysis of the adoption patterns.
2. Analysing adoption patterns to explain why a technology is accessed or not: This afford the presentation of possible reasons for differences in behaviour of farmers regarding adoption. Factors which affect adoption decisions are also presented. In many cases, further analysis may be required for a fuller understanding of why farmers accept or reject a new technology.

The task of the researcher undertaking adoption analysis is to use a wide range of tools that include the following:

1. Frequency tables,
2. Contingency tables,
3. Correlation analysis,
4. Linear regression, and
5. Binomial choice models

Frequencies are useful for individual farmer explanation but explanations do not provide much information on factors affecting other farmers. Besides, the factors are complex and diverse. Thus, use of statistical comparisons of the farmers' behavior with various characteristics helps to explain the adoption patterns. We know that farmers provide information but they do not articulate the causal linkage between the characteristics and adoption.

In general, the use of tables is a common approach in which it compares adopters and non-adopters. Differences are expected to offer insights into the rationale for adoption. The limitation is that differences are not proof of association. It is important to explore differences within the context of all the data available using for instance simple tables and statistical tests. Interpretation of data by the researcher must provide a consistent, coherent, and logical explanation for the adoption patterns that are observed. Statistical tests are simply a way of providing a quantitative estimate of the likelihood that the association observed between two variables could have occurred by chance, even if there were no relation between them. Statistical significance or insignificance does not always imply importance or unimportance of the relationship. Attention should always be paid into statistical significance and the theoretical importance of the variable.

Most socio economists involved in farm surveys are well acquainted with SPSS. SPSS has been a user-friendly program for analyzing survey data. It has its own limitation to undertake more rigorous analysis. Econometric Models that employ new and elaborated software programs to analyze both patterns and intensity of adoption are available. This training program will focus on this important subject.

## 5. Specifying Technology Adoption Models

A careful review of the adoption literature confirms the existence of three main paradigms used in analyzing adoption, which include (i) Innovation-diffusion, (ii) the economic constraint, and (iii) the adoption perception paradigms. These three paradigms are further explained below.

**1. The innovation diffusion model:** The innovation diffusion model follows Rogers' analysis of diffusion and concludes that access to information is a critical factor in the adoption and diffusion of technologies. Accordingly the problem of adoption is reduced to communicating information on the technology to the potential users. Adoption is also influenced by the farmers' characteristics i.e. age, gender, education levels, training etc. the model emphasizes extension as a means of increasing the adoption of new technologies.

**2. The economic constraint model:** The economic constraints model contends that economic constraints are major determinants. The approach also attempts to assert the superiority of the economic constraint model over the innovation-diffusion model but the conclusions have also been challenged.

**3. The adopter perception paradigm:** This is the least developed and used in the literature. It focuses on perceived attributes of the technology (e.g. taste, storability, etc. for cereal grains). Each of the models focuses on a few factors but all the three categories are important in explaining adoption patterns. Recent years, most studies try to include all the three categories of the factors in the analysis.

Current studies have demonstrated the significance of all three paradigms in technology adoption modelling. Consequently the specification below encompasses elements of the three paradigms presented above. The choice of linear probability model (LPM) in the current presentation has been justified in Part One of the current course.

### 5.1 Modeling agricultural technology adoption using the Tobit model approach

It is often the case that in adoption studies we do not only want to know probability that a farmer has adopted a technology but also the extent of use of the technology after adoption. To simultaneously explain probability of adoption, and intensity of use of the technology, the use of a Tobit model is appropriate. See Adesina and Zinnah (1993); Langyintuo et al, (2003) for details.

The fundamental assumption of the study is that farmers' adoption decisions on improved technology (e.g., improved maize variety) are based upon utility maximization (Rahm and Huffman, 1984). $U(M_{ji}, A_{ji})$ gives the non-observable underlying utility function, which ranks the preference of the *ith* farmer for the *jth* variety (*j = 1, 2: 1* = improved variety and *2* = traditional variety). Thus, the utility derivable from the new variety depends on *M*, which is a vector of farm and farmer specific attributes of the adopter and *A*, which is a vector of the attributes associated with the technology. Although the utility function is unobserved, the relation between the utility derivable from a *jth* variety is postulated to be a function of the vector of observed farm, farmer and variety specific characteristics (e.g. yield, taste, etc.) and a disturbance term having a zero mean:

$$U_{ji} = \alpha_j F_i(M_i, A_i) + e_{ji} \qquad j = 1, 2;\ i = 1, 2, \ldots, n \qquad \ldots (1)$$

As the utilities $U_{ji}$ are random, the *ith* farmer will select the alternative $j = 1$ if $U_{1i} > U_{2i}$ or if the non-observable (latent) random variable $y^* = U_{1i} - U_{2i} > 0$. The probability that $Y_i$ equal one (i.e., that the farmer adopts an improved technology) is a function of the independent variables:

$$\begin{aligned} P_i = P_r(Y_i = 1) &= P_r(U_{1i} > U_{2i}) \\ &= P_r[\alpha_i F_i(M_i, A_i) + e_{1i} > \alpha_2 F_i(M_i, A_i) + e_{2i}] \\ &= P_r[e_{1i} - e_{2i} > F_i(M_i, A_i)(\alpha_2 - \alpha_1)] \\ &= P_r(\mu_i > -F_i(M_i, A_i)\beta) \\ &= F_i(X_i\beta) \qquad \ldots (2) \end{aligned}$$

where $X$ is the $n$ x $k$ matrix of the explanatory variables and $\beta$ is a $k$ x 1 vector of parameters to be estimated, *Pr(.)* is a probability function, $\mu_i$ is a random error term, and $F_i(X_i\beta)$ is the cumulative distribution function for $\mu_i$ evaluated at $X_i\beta$. The probability that a farmer will adopt an improved technology is a function of the vector of explanatory variables and of the unknown parameters and error term. Equation 2 cannot be estimated directly without knowing the form of *F*. It is the distribution of $\mu_i$ that determines the distribution of *F*. The functional form of *F* is specified with a Tobit model, where $\mu_i$ is an independently, normally distributed error term with zero mean and constant variance $\sigma^2$:

$$\begin{aligned} Y_i &= X_i\beta \text{ if } i^* = X_i\beta + \mu_i > T \text{ (Adoption)} \\ &= \text{O if } i^* = X_i\beta + \mu_i \le T \text{ (Non-Adoption)} \qquad \ldots (3) \end{aligned}$$

Where:

$Y_i$ = probability of adoption (and intensity of use) of the improved technology

i* = non-observed latent variable

T = non-observed threshold level

The Tobit model therefore measures not only the probability that a farmer will adopt the improved technology but also the intensity of use of the technology once adopted. The empirical model can be used to draw economic implications for commodity improvement strategies for farmers, the effects of changes of given attributes and characteristics of farmers on adoption probabilities and use intensities can be obtained by decomposing the marginal effects Following a Tobit decomposition framework suggested by McDonald and Moffitt (1980).

**5.2 Decomposing marginal effects of tobit analysis**

Let E(P) be the expected value of the proportion of adoption across all observations conditional on the farmer being above the threshold limit. That is, we are concerned about use intensities of farmers who have already adopted an improved technology. Given the probability of adoption as F(z), where $z = XP/\sigma$, the relationship between these variables can be shown to be:

$$E(P) = F(z) * E(P) \quad \ldots (4)$$

For a given change in the level of a specific characteristic of interest, the effects on farmer adoption behavior can be broken down into two parts by differentiating (4) with respect to the specific characteristic change:

$$\partial E(P) / \partial X_i = F(z)\{\partial E(p) / \partial X_i\} + E(p)\{\partial F(z) / \partial X_i\} \quad \ldots(5.1)$$

Multiplying through by $X_i / E(P)$, the relation in (5.1) can be converted into elasticity forms as:

$$\{\partial E(P) / \partial X_i\} X_i / E(P) = F(z)\{\partial E(p) / \partial X_i\} X_i / E(P) + E(p)\{\partial F(z) / \partial X_i\} X_i / E(P) \quad \ldots(5.2)$$

Re-arranging (5.2) by using (4), the following decomposed elasticity equation can be obtained:

$$\{\partial E(P) / \partial X_i\} X_i / E(P) = \{\partial E(p) / \partial X_i\} X_i / E(P) + \{\partial F(z) / \partial X_i\} X_i / F(z) \quad \ldots(5.3)$$

Therefore, total elasticity of a change in the level of any given characteristic (which is assumed to be directly linked to adoption) consists of two effects: (a) the change in the elasticity of the use intensities of the improved technologies, for those farmers that are already adopters; and (b) the change in the elasticity of the probability of being an adopter.

### 5.3 Describing Aggregate Adoption Process

Aggregate adoption is the spread of a technology, measured by the aggregate level of use of the technology within a given geographical area or given population. It is usually important to assess aggregate adoption because it (i) gives temporal view of the historical adoption patterns of the technology in question, (ii) provides needed information for assessing future adoption patterns, (iii) is useful in projecting the future demands for inputs for the technology, and (iv) is useful for impact assessment studies.

As you may well know, technologies are not uniform; some are divisible (e.g., varieties, fertilizer, herbicides, etc) while others such as mechanization (animal traction, tractor use, etc) not (lumpy or indivisible). Divisible technologies are easily measured by the extent of the cropped area on which the technology is adopted (e.g., 80% of area is in modern varieties or quantity of herbicide used per ha). At the aggregate level, similar values can be developed to measure aggregate adoption of divisible technologies. On the other hand, indivisible technologies are necessarily dichotomous at the farm level (i.e., the dependent variable takes on the value of 1 if a farmer adopts and a value of zero if the farmer did not adopt the technology) but continuous at the aggregate level (e.g., the number of farmers that have adopted animal traction technology in a given zone).

In general the diffusion of a technology follows an S-shaped pattern or a "Sigmoid curve" (Figure 1), which largely reflects the way information spreads. Technologies are adopted at first by a few farmers (who are termed "early adopters") due to their early contact with information on the technology. Subsequently, several more farmers (termed "late adopters") begin to adopt the innovation as the information is diffused over time and space. This is called the "stage of rapid growth".

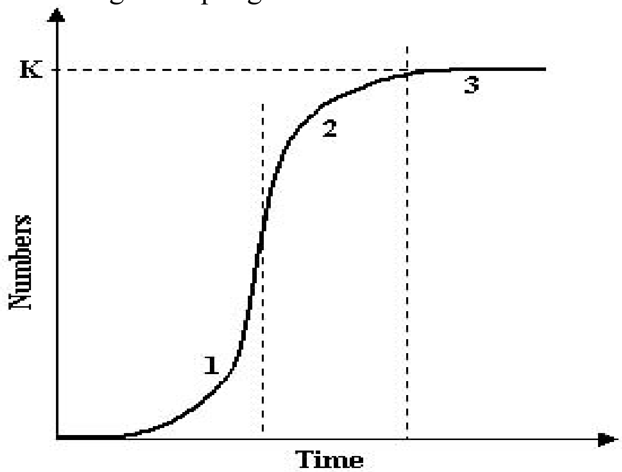


Figure 1: Sample sigmoid curve
Source: http://www.saburchill.com/IBbiology/chapters05/images/14060315.jpg

This pattern continues until the rate of adoption reaches a maximum level or "adoption ceiling". As many farmers adopt the technology, production may increase decreasing prices and profitability. This may cause the early adopters to drop out to look for new innovations given the higher opportunity cost of their resources.

### 5.4 Methodology for Examining Aggregate Adoption

The most appropriate methodology for analyzing the spread of a new technology (assume here an improved maize variety) is the logistic curve. The logistic function describing the shape of the logistic curve takes the general form:

$$\partial Y_t / \partial t = \beta Y_t / K(K - Y_t) \qquad \dots (6)$$

Where:
Yt = number of individuals who have adopted the improved varieties in time t (cumulative % of adopters or area under improved variety at time t).
K = the fixed population of potential adopters (upper bound on % adopters)
β = parameter reflecting the rate of adoption (probability of adopting upon learning about the technology)
Solving for Yt of the logistic differential equation generates the result:

$$Y_t = K/(1 + e(-\alpha - \beta_t)) \qquad \dots (7)$$

Where: Yt, and K are as defined above, α is a constant term that positions the curve on the time scale and β the rate or speed of the technology adoption process. Equation (7) can be manipulated to obtain a form that is linear in parameters, and readily estimated:

$$\ln(Y_t /(K - Y_t)) = \alpha + \beta_t \qquad \dots (8)$$

To determine the value of K (the maximum adoption ceiling), there are two possible approaches: First, one can run the above model for different selected values of K and choose the value of K that gives the model the highest $R^2$ value. Note, however, that the use of $R^2$ here does not mean anything as we are only trying to obtain the best fit for the function. Second, one may also plot the values of the adoption data and examine visually the maximum value of K (i.e, the upper bound on the adoption). Finally, one can also use a combination of the above methods. Note that in some applications it is possible for K to be endogenously determined; however, this will require the use of more complex non-linear regression algorithms.

Limitations of the adoption curves include:

1. The parameters estimated for the model are constant over time periods, which assumes that relative p ratios, institutional factors or infrastructure, or the technology itself, remain constant over the period in which the analysis is done,
2. K is assumed to be a constant and exogenous. However, it is possible for the value of K to be endogenous (e.g., it could depend on other institutional factors),
3. The logistic curve assumes cumulative adoption of the technology by farmers over time, which may not be correct as some farmers might have discontinued the use of the technology during the period of investigation. Therefore, it is important for researchers to find out if some farmers have discontinued use of the technology during the period of the analysis and to investigate the reasons why they discontinued technology use.

## 6. Estimation of Adoption Models Using STATA

This section limits itself to the key relevant aspects of STATA usage to give participants an opportunity to start and operate STATA meaningful. The operational procedures demonstrated here are extracted from the Stata manuals and hence inadequate to replace the manuals. Participants are therefore urged to practice using the manuals provided to improve their proficiency. Areas covered are (1) Starting and stopping a STATA session, (2) Data management (limited to data entry using the editor, pasting data from a spreadsheet and importing SPSS or Excel files), and (3) basic procedure in analyzing (adoption) data using Tobit model, which is preceded by estimating various descriptive statistics of actual field data.

### 5.1 Starting and Stopping a STATA Session

When Stata is launched from the Windows Start menu or by clicking on the icon on the desktop, the first screen you will see is as below in Figure 2. If Stata could not be loaded because the license file could not be found or for any other reason, you may have to go to Troubleshooting starting and stopping Stata in the reference manual to diagnose the problem and solve it. The tool bar menus have been labelled. If you forget what the button means or does, hold the mouse pointer over the button for a moment and a box will appear with the description of that particular button. The main window components of Stata are displayed in Figure 3.

It is assumed that Stata is installed and running properly. To verify that Stata installation was performed correctly, type *verinst* in the Command window (See Figure3) and press Enter. You should see an output in the Results window informing you that Stata is correctly installed. You can type *exit* and press return to exit Stata or choose exit from the File menu.

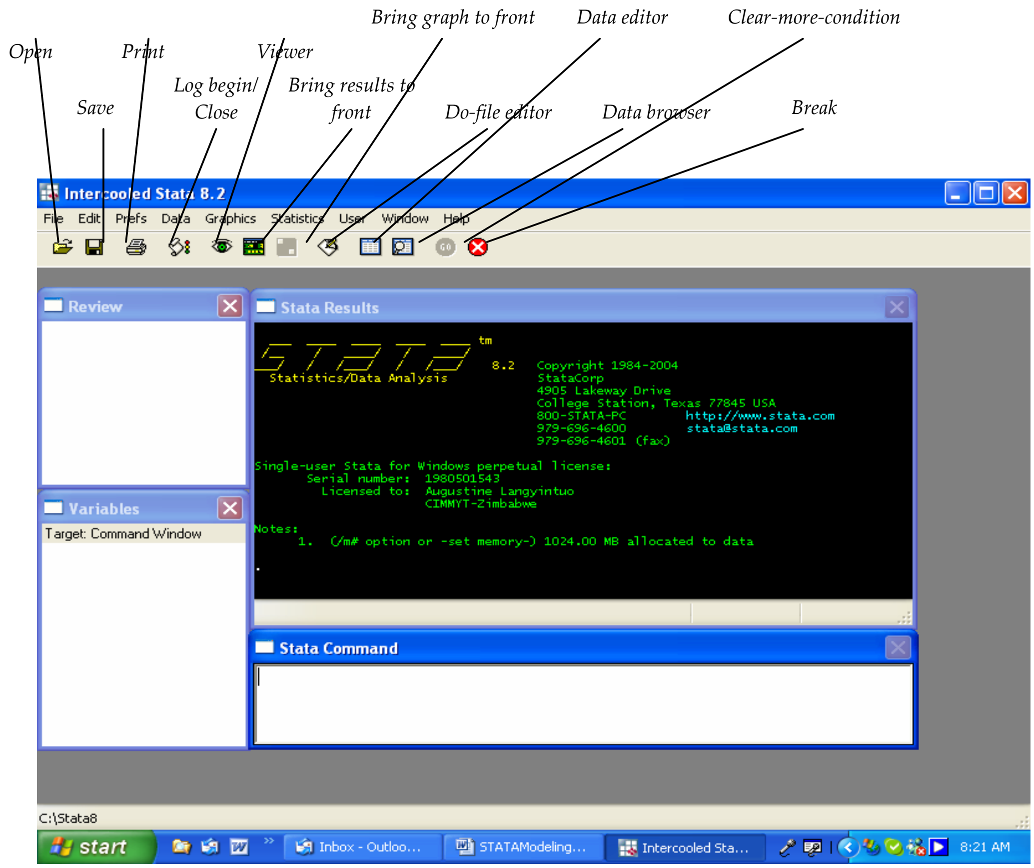


**Figure 2: Stata screen when logging on**

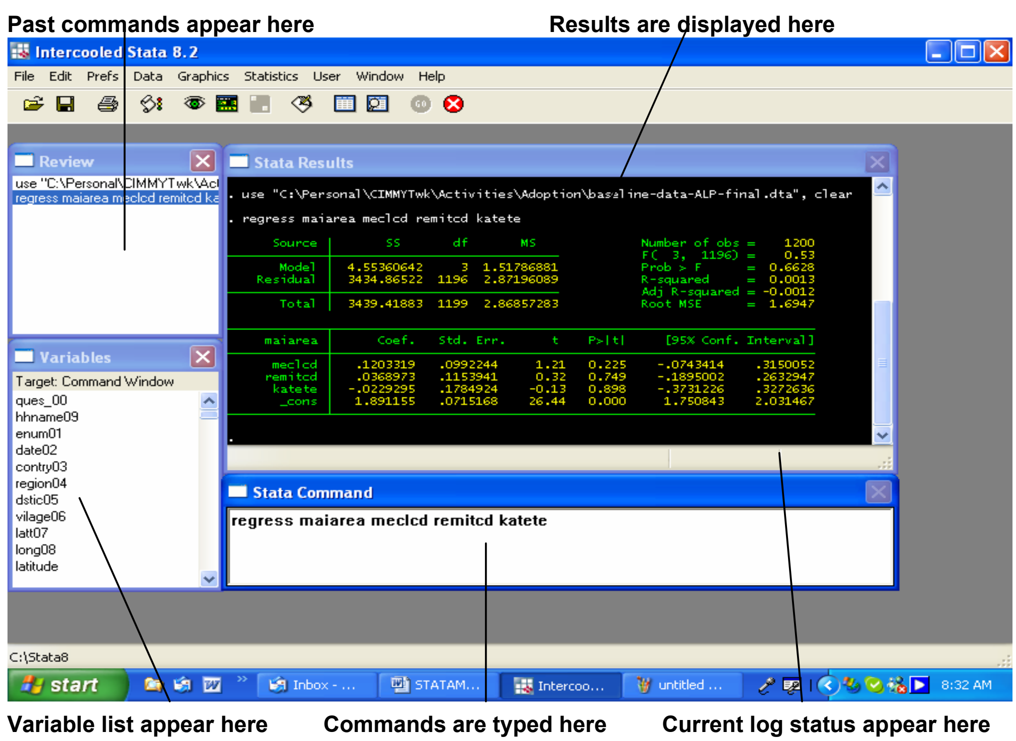


**Figure 3: The main components of the Stata window**

If you see an error message, *verinst unrecognized command r(199),* complaining that the command was not found,*;* then not all of Stata is installed and you will need to re-install Stata and carry out the verification test once more to be sure.

### 6.1.1 The Working Directory

If you look at the screenshot in Figure 1, you will notice a status bar at the bottom of the screen that contains C: \DATA. Stata is telling you that C: \DATA is the current working directory. The working directory is where graphs and datasets will be saved unless you specify another directory. Once you have started Stata, you can change the current working directory with the *cd* command.

### 6.1.2 Stata's Interface

Stata is a command-driven application with great flexibility. That is, commands are issued to Stata to perform an action such as load a file, do a regression, etc. Whereas advanced users may type commands by hand, starters would prefer to access the commands in a point-and-click manner by pulling down Stata's menus and selecting items that invoke dialog boxes to build Stata commands. Stata's Data, Graphics, and Statistics menus provide point-and-click access to almost every command in Stata.

When you open a dataset in Stata the Variables window in Figure 3 shows you a list of variable names in that dataset, which can help you fill out the dialog boxes for Stata commands. Notice the "Target" area at the top of the Variables window. When you click on a variable name in the Variables window, that name is typed into the specified target. By default, the Command window is the target, so if you click on a variable name, it will be typed into the Command window.

### 6.1.3 Increasing the Amount of Memory Stata Uses

By default, Intercooled Stata uses 1 megabyte of memory for data. Depending on the size of your datasets, you may wish to adjust the amount of memory Stata allocates for data temporarily or permanently.

To temporarily reset the amount of memory Intercooled Stata is using for data, start Stata and type *set memory 1g* for example. The 1g means Stata is to allocate 1 gigabyte to its data area. If you were to type 5m, Stata would allocate 5 megabytes. To make such a change permanent, you need to type *set memory1g, permanently*. If you change the memory beyond the capacity of your machine, Stata will prompt you that your operating system is unable to provide the amount of memory that you requested. Note that allocating too much memory to Stata may leave little memory available for other applications.

## 6.2 Data Management in STATA

There are three main ways of managing data in STATA: (1) direct entry using the data editor, (2) copying and pasting data from other spread sheets, and (3) importing data from a different data management source.

### 6.2.1 Direct Entering Using The Data Editor:

**To enter the editor:**

1. Click on the *Data Editor* button on the icons pane on top, or
2. Type *edit* and press Enter in the Command window at the bottom of the screen.

To input data variable-by-variable:

1. Click on the top cell in the first empty column
2. Enter the value
3. Press Enter to go down

To input data observation-by-observation:

1. Click on the first cell in the first empty row
2. Type the value
3. Press *Tab* to go right
4. After the first observation's values have been entered, click on the second cell in the first column
5. Type the values for the second observation; press *Tab* to go right
6. At the end of the second observation (and all subsequent ones),
7. *Tab will* automatically take you back to the first column

Numeric data and string data (i.e., data consisting of characters) are entered the same way. Quotes (“”) around strings are unnecessary. Note that for missing values, simply press Enter or Tab or type in “.” (i.e., period) and press Enter or Tab. [STATA supports different types of missing values: See manual for details.]

The editor initially names variables varl, var2, etc. You can rename them using the following procedure:

1. Double-click anywhere in the variable's column to bring up the Variable Information dialog.
2. Enter the new name of the variable of between 1 to 32 characters long. The characters can be letters: A-Z, a-z or digits: 0 – 9, or underscores but no spaces or special characters (#, $, !, etc). The first character must be a letter or an underscore, but using an underscore to begin your names is not recommended since Stata's built-in variables begin with an underscore. STATA is case-sensitive so that *gender, Gender,* and *GENDER* are different names.

Suppose that we have the following dataset from a survey:

| farmer | age | hhhyrs | Farmsize | hhsize |
|---|---|---|---|---|
| phiri | 30 | 11 | 3 | 7 |
| banda | 25 | 8 | 3 | 7 |
| ellena | 40 | 20 | 3 | 7 |
| moffat | 26 | . | 9 | 7 |
| mary | 46 | 30 | 3 | 6 |
| pristone | 63 | 40 | 5 | 7 |
| stanley | 44 | 21 | 3 | 6 |
| tryson | 65 | 46 | 5 | 9 |
| jere | 31 | 12 | 4 | 6 |
| paulo | 26 | 7 | 5 | 3 |

Note that we do not know *hhhyrs* for Moffat. Entering this data will be demonstrated in the work-session using the data editor as follows:

1. Click on the Data Editor button or type edit in the Command window.
2. Enter the data variable-by-variable (by pressing the Enter key after each value) or observation-by-observation (by pressing the tab key after each value). Remember that Columns correspond to variables and rows to observations. The Tab key is smart; after the first observation has been entered, Stata knows how many variables you have. So, at the end of the second observation (and all subsequent observations), Tab will automatically take you back to the first column.

Note: STATA will not allow empty columns or rows in the *middle of* your dataset. If you skip over some columns or rows, STATA will fill in the intervening columns or rows with missing value, which is undesirable. Therefore, whenever you enter new variables or observations, always begin in the first empty column or row.

***[You may refer to Chapter 8 of Getting Started with Stata for Windows on how to label your data].***

To save the new dataset for use with STATA:

1. Choose Save As... from the File menu and select Stata Data (*.dta) from the Save as type list
2. Or type save old *filename* in the Command window

To resave a dataset that has been changed (overwriting the original data file):

1. Pull down the File menu and choose Save

Or click on the Save button
Or type save *filename,* replace
Or simply type save, replace

To use the file you have previously saved in STATA *.dta format, you proceed as follows:

1. Pull down the File menu
2. Choose Open...

Or type use *filename*
Or click on the Open button

When loading a dataset, the dataset currently in memory is discarded. You either save the current dataset first or allow it to be discarded. When you choose File-Open..., will be ask ed for your OK before discarding it. Or type use *filename,* clear, Or type clear and then use *filename.*

### 6.2.2 Copying and Pasting Data From Other Spread Sheets:

The editor is like a spreadsheet; Columns correspond to variables and rows to observations. You navigate by clicking on a cell or by using the arrow keys as you would do in Excel or lotus. You can copy and paste data between STATA's editor and other spreadsheets. To copy and paste data from other spread sheets:

1. Highlight the data that you wish to copy in the other spreadsheet and then pull down Edit and choose Copy (you can also copy STATA data and paste in other spread sheets).

Or click once on a variable name to select the entire column.

Or click once on an observation number to select the entire row.
Or click and drag the mouse to select a range of cells.

2. After copying the data, you can paste it into STATA's editor by selecting the top left cell of the area to which you wish to paste, pulling down Edit, and choosing Paste (or right clicking you mouse and choosing paste from the pull down menu).

After pasting the data you can modify the data by choosing the cell, typing the value, and pressing *Enter* or *Tab*. Remember that using *Enter* moves you down to the next cell in the column while the *Tab* moves you to the right to the next cell in the row (until you get to the last filled-in column, then it takes you back to the first column). To save the file created by pasting the data from another spreadsheet, follow the steps outlined under "To save the new dataset for use with STATA" above. [*This will be demonstrated in the practice session*].

### 5.2.3 Importing Data from a Different Data Management Source:

**Importing data from a previously saved SPSS file:** If you have previously entered data in say SPSS you can import it into STATA. Note that if the data was entered in SPSS, you may save the file in as a "SAS Transport (*.xpt)" file before it can be imported into STATA. Assuming you want to import the file "malawi-data.xpt" previously saved, you may proceed as follows:

1. Click File on the STATA menu
2. Select Import
3. Select FDA data (SAS XPORT) to open the "fdause Import FDA data (SAS XPORT)" window.
4. Click on Browse and navigate to the appropriate directory to select "malawi-data.xpt".
5. Click Ok and the file name and the complete directory extension would appear next to the Browse button.
6. Accept the file by clicking on the OK button.

The Review window would show the file and its directory extensions while the Variable window will display all the variables in the file. Make sure that the directory names and your file name do not contain spaces. If for example you saved your file as: "*malawi data.xpt*"and you attempt to import it into STATA, you will receive the following error message written in red: ***invalid 'data.xpt'***. Simply re-save the file as "*Malawi-data.xpt*" or "*malawidata.xpt*" without the space. Similarly, if your directory name(s) contain spaces, e.g., c:\survey data\malawi-data.xpt, you will encounter an error loading Malawi-data.xpt. The imported file can then be saved as *.dta (a STATA format) for subsequent usage.

**Importing data from a previously saved Excel spreadsheet:** If you intend to use a previously saved Excel data in STATA, you will need to save the file as: Text (Tab delimited) with an extension .txt. But be sure that the number of variables does not exceed 256. Variables in excess of 256 will be discarded. To import a previously saved file called "Mozam-data.txt" proceed as follows:

1. Click File on the STATA menu
2. Select Import
3. Select "ASCII data created by a spreadsheet" to open the "Insheet – Import ASCII data" window.
4. Click on Browse and navigate to the appropriate directory to select "Mozam-

data.txt".
5. Click Ok and the file name and the complete directory extension would appear next to the Browse button.
6. Accept the file by clicking on the OK button.

The Review window would show the file and its directory extensions while the Variable window will display all the variables in the file. The imported file can then be saved as *.dta (a STATA format) for subsequent usage.

### 6.2.4 Exiting the Data Editor

To exit the editor, click on the editor's close box (the box with an X at the right of the editor's title bar). Note that changes made in the editor are not saved until you tell Stata to save them. The data that you have entered only exist in the computer's memory. Save your dataset by pulling down File and choosing Save As....(enter file name). Beware that you cannot save your dataset to disk until you exit the editor. If you pull down the File menu while in the editor, Save and Save As... will be grayed out.

### 6.2.5 Listing Data

The *list* command is used to display data in the Results window by typing *list* in the command window. When a ***-more-*** appears in the Results window especially when listing a long data, press the Enter key or click on ***–more-*** with your mouse to continue listing or Break key (ctrl-Break) to stop. Simply typing list will display all the variables in the data set. To display a single variable, type *list* and the name of the variable. For example if you want to list farm sizes of farmers, type *list farmsize*. For multiple variables, type *list* followed by the names of all the variables you wish to list. You may abbreviate *list* by *l*. Variable names can also be abbreviated. For example *farmsize* can be abbreviated to *farm*. So you can type *l farm* to list farm sizes. To list all variables starting with for example *fa*, type *l fa**.

To list the fourth observation, type *list in 4*; the last but one observation, *list in -2*; the last observation, *list in -1* or *list in l* (i.e., the letter L not one). To list observations 2 through 8 type *list 2/8*.

You can conditionally list observations using the *list if* command. For example, to list farmsizes greater than 3, type *list if farmsize>3*. Stata supports the following logical operators:

| | |
|---|---|
| < | less than |
| <= | less than or equal |
| == | equal |
| > | greater than |
| >= | greater than or equal |
| ! = | not equal (~= can also be used) |
| & | and |
| \| | or |
| ! | not (logical negation; ~ can also be used) |
| ( ) | parentheses specify order of evaluation |

[***Please refer to Chapter 11 of Getting Started with Stata for Windows for more on listing***].

### 6.2.6 The *Generate* and *Replace* Commands for Creating New Variables

You can create a new variable that is an algebraic expression of other variables with the **generate** command. For example to generate a variable called *manland*, which is the ratio of household size to by farm size we proceed as follows:
*generate manland=hhsize/farmsize.* The *generate* command can be abbreviated to *g, ge, or gen.*

If you wanted to change the contents of another variable *mlratio* with the hhsize/farmsize then you use the *replace* command as follows:
*Replace mlratio= hhsize/farmsize.* **No abbreviation for *replace* command allowed.**

Mathematical operators necessary for the implementation of the **generate** and **replace** commands are:

| Arithmetic | | Logical | | Relational (numeric and string) | |
|---|---|---|---|---|---|
| + | Addition | ! | Not | > | Greater than |
| - | Subtraction | \| | Or | < | Less than |
| * | Multiplication | & | and | >= | Greater than or equal |
| / | Division | | | <= | Less than or equal |
| ^ | power | | | == | equal |
| | | | | != | Not equal |

### 6.2.7 Deleting Variables and Observations

You can delete data from the active fill using the drop command. To drop all data in the memory type *drop –all*. To drop all the data and clear the memory, simply type *clear*. If you want to drop a single variable, type drop followed by variable name and to drop multiple variables, type drop followed by the names of all the variables you wish to drop. The drop functions follow those of the list functions. For all other operations replace *list* with *drop*.

## 6.3 Analyzing Adoption Data

The analysis section would concentrate on presenting simple descriptive statistics of data (means, frequency distributions, cross-tabulations, correlations, etc), and estimating, Probit/logit and Tobit models.

### 6.3.1 Simple Descriptive Statistics

After collecting the data one would like to see the distribution of key variables in the data set. For example, it will be important to examine the central tendency of the ages of farmers, farm sizes households, etc. as well as the distribution of household heads in terms of their literacy status, adoption or non-adoption of improved varieties, etc. In some cases one may want to see if there has been any change in the number of farmers currently adopting a given technology compared with a reference point. To demonstrate these kinds of analysis and others, we will use sample data from Malawi collected during the recent CIMMYT region-wide survey on seed markets. These are contained in the file "maldata.xpt".

#### 6.3.1.1 Estimating the Mean of a Variable

In Stata, you can obtain the mean, standard deviation, minimum and maximum of a given variable by entering the *summarize* command. For example, to find the means of the variables crpland, totfarm, agehh in maldata.xpt. Proceed as demonstrated below using either the command prompt or pull down menu.

From command prompt type:
summarize totfarm cropland agehh

From pull-down Menu select:

1. Statistics
2. Summaries, tables & tests
3. Summary statistics
4. Summary statistics to move to the *Summarize – Summary statistic*s window.
5. Select *totfarm, cropland*, and *agehh* from the list of variables you see in the variables window to the left screen by clicking on them with your mouse after positioning your in the *Variables: (leave empty for all variables) space*
6. Check the box next to Standard option
7. Click *OK* to continue

Output:

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| totfarm | 300 | 4.001833 | 2.677184 | .5 | 15.5 |
| cropland | 300 | 3.5385 | 2.308218 | .3 | 13 |
| agehh | 300 | 44.56667 | 16.70576 | 18 | 89 |

If you were to type summarize without specifying any variables, Stata will produce a summary table for all the variables in the file. If you want to have the percentiles printed as well then you need to include detail option as follows:

From command prompt type:
summarize agehh, detail

From pull-down Menu select:

1. Statistics
2. Summaries, tables & tests
3. Summary statistics
4. Summary statistics to move to the *Summarize – Summary statistic*s window.
5. Select *agehh* from the list of variables you see in the variables window to the left screen by clicking on it with your mouse after positioning your in the *Variables: (leave empty for all variables) space*
6. Check the box next to *Display additional statistics*
7. Click *OK* to continue

Output:

Age of household head

| Percentiles | | Smallest | | |
|---|---|---|---|---|
| 1% | 19 | 18 | | |
| 5% | 22 | 19 | | |
| 10% | 24.5 | 19 | Obs | 300 |
| 25% | 30 | 19 | Sum of Wgt. | 300 |
| 50% | 42 | | Mean | 44.56667 |
| | | Largest | Std. Dev. | 16.70576 |
| 75% | 56 | 82 | | |
| 90% | 70 | 85 | Variance | 279.0825 |
| 95% | 75.5 | 85 | Skewness | .4898273 |
| 99% | 83.5 | 89 | Kurtosis | 2.368574 |

You may want to summarize household sizes (or any variable) for adopters separated from non-adopters using the variable *impvnow* (which takes a value of 1 for adopters and zero otherwise) as shown below:

From command prompt type:
sort impva03
by impva03: summarize agehh

From pull-down Menu select:
1. Statistics
2. Summaries, tables & tests
3. Summary statistics
4. Summary statistics - to move to the *Summarize – Summary statistics* window.
5. With the main button highlighted, select *agehh* from the list of variables you see in the variables window to the left screen by clicking on it with your mouse after positioning your in the *Variables; (leave empty for all variables) space*
6. Click on *by/if/in* button
7. Check the box next to *Repeat command for groups defined by* and select impva03 from the list of variables in the variables window to the left of the screen
8. Check the box next to Display additional statistics
9. Click *OK* to continue

Output:
-> impva03 = 0

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| agehh | 116 | 45.11207 | 18.13609 | 18 | 85 |

-> impva03 = 1

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| agehh | 184 | 44.22283 | 15.77919 | 20 | 89 |

### 6.3.1.2 Hypothesis Test of the Differences in Means of Two Variables

You may want to know if the difference in means between adopters and non-adopters is significant or not. The *ttest* command can be used as demonstrated below:

From command prompt type:
ttest agehh, by(impva03)

From pull-down Menu select:
1. Statistics
2. Summaries, tables & tests
3. Classical mean comparison test
4. With *Main* button highlighted, click the down arrow next to the Variable name to select *agehh* from the list of variables. For the Group variable name, click the down arrow to select *impva03* from the list of variables.
5. Click *OK* to continue

Two-sample t test with equal variances

| Group | Obs | Mean | Std. Err. | Std. Dev. | [95% Conf. Interval] | |
|---|---|---|---|---|---|---|
| 0 | 116 | 45.11207 | 1.683894 | 18.13609 | 41.7766 | 48.44754 |
| 1 | 184 | 44.22283 | 1.163257 | 15.77919 | 41.92771 | 46.51795 |
| combined | 300 | 44.56667 | .9645076 | 16.70576 | 42.66858 | 46.46475 |
| diff | | 892429 | 1.983216 | | -3.01364 | 4.792126 |

Degrees of freedom: 298
Ho: mean(0) - mean(1) = diff = 0

| Ha: diff < 0 | Ha: diff != 0 | Ha: diff > 0 |
|---|---|---|
| t = 0.4484 | t = 0.4484 | t = 0.4484 |
| P < t = 0.6729 | P > t = 0.6542 | P > t = 0.3271 |

### 6.3.1.3 Frequency distribution of variables

We can take a count of a given variable, e.g., *educhh* (i.e., educational level of household head from our sample data set using the *tabulate* command as follows:

From command prompt type:
```
tabulate educhh
```

From pull-down Menu select:
1. Statistics
2. Summaries, tables & tests
3. Tables
4. One-way tables - to move to the *Tabulate1 – One-way tables* window.
5. With *Main* button highlighted, click the down arrow next to the Categorical variable to select *educhh* from the list of variables.
6. Click *OK* to continue

Output:

| Education level of Household head | Freq. | Percent | Cum. |
|---|---|---|---|
| 1 | 69 | 23.00 | 23.00 |
| 2 | 211 | 70.33 | 93.33 |
| 3 | 18 | 6.00 | 99.33 |
| 4 | 2 | 0.67 | 100.00 |
| Total | 300 | 100.00 | |

We can compare the educational status between males (1) and females (0).

From command prompt type:
```
tabulate educhh gender
```

From pull-down Menu select:
1. Statistics
2. Summaries, tables & tests
3. Tables
4. One-way tables - to move to the *Tabulate1 – One-way tables* window.
5. With *Main* button highlighted, click the down arrow next to the Categorical variable to select *educhh* from the list of variables.
6. Click on *by/if/in* button
7. Check the box next to *Repeat command for groups defined by* and select *impva03* from the list of variables in the variables window to the left of the screen
8. Click *OK* to continue

Output:

tabulate educhh gender

| Education level of household head | Gender of household head | | |
|---|---|---|---|
| | 0 | 1 | Total |
| 1 | 37 | 32 | 69 |
| 2 | 68 | 143 | 211 |
| 3 | 2 | 16 | 18 |
| 4 | 1 | 1 | 2 |
| Total | 108 | 192 | 300 |

**5.3.1.4 Hypothesis Test of the Differences in proportions**

The statistical differences in proportion of males and female educational levels can be examined using chi squared statistic as shown below. Note that the actual values are printed above the proportions for each category and chi squared statistic printed suggest a significant difference between males and females in their educational standards.

From command prompt type:
tabulate educhh gender, chi2 col

From pull-down Menu select:

1. Statistics
2. Summaries, tables & tests
3. Classical proportion comparison test
4. With *Main* button highlighted, click the down arrow next to the Variable name to select *educhh* from the list of variables. For the Group variable name, click the down arrow to select *gender* from the list of variables.
5. Click *OK* to continue

Output:

```
+-------------------+
Key
-------------------
frequency column percentage
+-------------------+
```

| Education level of household head | Gender of household head 0 | 1 | Total |
|---|---|---|---|
| 1 | 37 | 32 | 69 |
| 34.26 | 16.67 | 23.00 | |
| 2 | 68 | 143 | 211 |
| 62.96 | 74.48 | 70.33 | |
| 3 | 2 | 16 | 18 |
| 1.85 | 8.33 | 6.00 | |
| 4 | 1 | 1 | 2 |
| 0.93 | 0.52 | 0.67 | |
| Total | 108 | 192 | 300 |
| | 100.00 | 100.00 | 100.00 |
| Pearson chi2(3) = | 15.6141 | Pr = 0.001 | |

Value
Percent

You may also want to use cross-tabulations to see the dynamics of technology adoption in a given locality. For example if you have a reference date when farmers adopted the technology and you want to see at another date if the numbers have changed or not you can use the *tabulation* command and test whether the differences are significant with chi squared test as above. The output you expect is:

From command prompt type:
tabulate impvab4 impva03, cell chi2

From pull-down Menu select:

1. Statistics
2. Summaries, tables & tests
3. Tables
4. Two-way tables - to move to the *Tabulate2 – Two-way tables* window.
5. With *Main* button highlighted, click the down arrow next to the Row variable to select *impvab4* from the list of variables and in the Column variable select impva03.
6. Check the boxes next to the Pearson's Chi-squared and Within-column relative frequencies
7. Click *OK* to continue

Output:

```
+-----------------+
Key
```

```
-----------------
frequency   cell percentage
+-----------------+
```

| Have you ever planted any improved maize | Planted improved variety in 2003 0 | 1 | Total |
|---|---|---|---|
| 0 | 50<br>16.67 | 1<br>0.33 | 51<br>17.00 |
| 1 | 66<br>22.00 | 183<br>61.00 | 249<br>83.00 |
| Total | 116<br>38.67 | 184<br>61.33 | 300<br>100.00 |

Value
Relative %

Pearson chi2(1) = 91.3335 Pr = 0.000

### 6.3.1.5 Correlation Matrices

In order that you may estimate efficient models, it might be important to examine the correlation between variables. For instance, will it make sense to include age of farmer (agehh) and years of farming (farmyrs) in an adoption model? Correlation matrices can be generated and examined before answering such a question. The *correlate* command can be used to examine the correlation between *agehh* and *farmyrs*, as below.

**From command prompt type:**
correlate agehh farmyrs

**From pull-down Menu select:**
1. Statistics
2. Summaries, tables & tests
3. Summary statistics
4. Correlation and covariances.
5. With *Main* button highlighted, click the down arrow next to Variables: (leave empty for all) and select *agehh* and *farmyrs*
6. Click *OK* to continue

**Output:**
(obs=300)

| | agehh | farmyrs |
|---|---|---|
| agehh | 1.0000 | |
| farmyrs | 0.9376 | 1.0000 |

A similar procedure can be used to generate correlation matrices for many variables as shown below.

**From command prompt type:**
correlate agehh farmyrs educhh croplan gender hhsize

**From pull-down Menu select:**
1. Statistics
2. Summaries, tables & tests
3. Summary statistics
4. Correlation and covariances.

5. With *Main* button highlighted, click the down arrow next to Variables: (leave empty for all) and
6. select *agehh, farmyrs, educhh, cropland, gender* and hhsize
7. Click *OK* to continue

**Output:**
(obs=300)

| | agehh | farmyrs | educhh | cropland | gender | hhsize |
|---|---|---|---|---|---|---|
| agehh | 1.0000 | | | | | |
| farmyrs | 0.9376 | 1.0000 | | | | |
| educhh | -0.2350 | -0.2362 | 1.0000 | | | |
| cropland | 0.1162 | 0.0995 | -0.0346 | 1.0000 | | |
| gender | -0.0341 | -0.0372 | 0.2066 | 0.1946 | 1.0000 | |
| hhsize | 0.3142 | 0.2784 | 0.0138 | 0.1621 | 0.1328 | 1.0000 |

### 6.3.3 Estimating a Tobit Model

In attempting to describe the factors influencing the adoption and use intensity of an improved variety, for example, the Tobit command can be used in Stata. To run a probit or logit model, simply replace "tobit" with "probit" or "logit". Remember be make sure that the dependent variable is dichotomous. Tobit, censored-normal, fits a model of depvar on indepvars where the censoring values are fixed. The syntax is:

tobit depvar [indepvars], ll, ul

11(#) and ul(#) indicate the censoring points in tobit. You may specify one or both. 11( ) indicates the lower limit for left censoring. Observations with *depvar* < ll ( ) are left-censored; observations with depvar > ul ( ) are right-censored; and remaining observations are not censored. You do not have to specify the censoring values at all. It is enough to type 11, ul, or both. When you do not specify a censoring value, tobit assumes that the lower limit is the minimum observed in the data (if 11 is specified) and the upper limit is the maximum (if ul is specified).

**From command prompt type:**
tobit imvaprop gender agehh hhsize assocn totfarm hhexpt ngocon3 rnkpalat rnkcost rnkyield rnkspest mango nzimba, ll ul

**From pull-down Menu select:**
1. Statistics
2. Linear regression and related
3. Censored regression
4. Tobit regression
5. With *Main* button highlighted, click the down arrow next to Dependent variable and select *imvaprop. C*lick the down arrow next to Independent variables and select *gender agehh hhsize assocn totfarm hhexpt ngocon3 rnkpalat rnkcost rnkyield rnkspest mango nzimba*
6. Click *OK* to continue

**Output:**
Tobit estimates Number of obs = 300
LR chi2(13) = 64.27
Prob > chi2 = 0.0000
Log likelihood = 8.0593015 Pseudo R2 = 1.3347

| imvaprop | Coef. | Std. Err. | t | P>t | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| gender | 0.020080 | 0.021298 | 0.940000 | 0.347000 | -0.021839 | 0.062000 |
| agehh | -0.000489 | 0.000631 | -0.770000 | 0.439000 | -0.001731 | 0.000754 |

| | | | | | | |
|---|---|---|---|---|---|---|
| hhsize | 0.004086 | 0.003874 | 1.050000 | 0.293000 | -0.003540 | 0.011711 |
| assocn | 0.046004 | 0.025064 | 1.840000 | 0.067000 | -0.003330 | 0.095337 |
| totfarm | -0.015102 | 0.004085 | -3.700000 | 0.000000 | -0.023142 | -0.007061 |
| hhexpt | 0.000001 | 0.000001 | 2.090000 | 0.038000 | 0.000000 | 0.000003 |
| ngocon3 | 0.057112 | 0.022671 | 2.520000 | 0.012000 | 0.012489 | 0.101734 |
| rnkpalat | 0.026590 | 0.029177 | 0.910000 | 0.363000 | -0.030838 | 0.084017 |
| rnkcost | -0.003960 | 0.024846 | -0.160000 | 0.873000 | -0.052863 | 0.044943 |
| rnkyield | 0.147316 | 0.033170 | 4.440000 | 0.000000 | 0.082029 | 0.212602 |
| rnkspest | -0.003919 | 0.034136 | -0.110000 | 0.909000 | -0.071109 | 0.063270 |
| mango | 0.061858 | 0.027445 | 2.250000 | 0.025000 | 0.007840 | 0.115876 |
| nzimba | 0.019978 | 0.028224 | 0.710000 | 0.480000 | -0.035574 | 0.075529 |
| _cons | -0.120896 | 0.057092 | -2.120000 | 0.035000 | -0.233268 | -0.008524 |

```
_se            .1508347     .084144 (Ancillary parameter)
  Obs. summary:        116  left-censored observations at imvaprop<=0
  183      uncensored observations
  1 right-censored observation  at imvaprop>=.6956522
```

Interpretation of the results will be carried out during the hands-on-training session.

### 6.3.4 Generating Marginal Effects or Elasticities

You can obtain the marginal effects by the following procedure soon after running the tobit regression as follows:

**From command prompt type:**

```
mfx compute, dydx at (mean)
```

**From pull-down Menu select:**

1. Statistics
2. General post-estimation
3. Obtain marginal effects or elasticities after estimation
4. In the mfx compute – marginal effects or elasticities window, check the box next to d(y)/d(x) and means
5. OK

**Output:**

Marginal effects after tobit
y = Fitted values (predict)
= .02599984

| variable | dy/dx | Std. Err. | z | P>\|z\| | [ 95% C.I. ] | | X |
|---|---|---|---|---|---|---|---|
| gender* | .0203105 | .02095 | 0.97 | 0.332 | -.020742 | .061363 | .64 |
| agehh | -.0005002 | .00062 | -0.81 | 0.420 | -.001717 | .000716 | 44.5667 |
| hhsize | .0036951 | .00382 | 0.97 | 0.333 | -.003783 | .011173 | 6.11 |
| assocn* | .0502776 | .02463 | 2.04 | 0.041 | .001995 | .09856 | .183333 |
| totfarm | -.015009 | .00402 | -3.74 | 0.000 | -.022882 | -.007136 | 4.00183 |
| hhexpt | 1.61e-06 | .00000 | 2.30 | 0.022 | 2.4e-07 | 3.0e-06 | 11384.3 |
| ngocon3* | .0595012 | .02228 | 2.67 | 0.008 | .015834 | .103168 | .266667 |
| rnkcost* | .0075658 | .02465 | 0.31 | 0.759 | -.040738 | .05587 | .783333 |
| rnkavail* | .0572662 | .02046 | 2.80 | 0.005 | .01716 | .097373 | .4 |
| rnkyield* | .1382094 | .03285 | 4.21 | 0.000 | .073818 | .202601 | .836667 |
| rnkspest* | -.0016273 | .03347 | -0.05 | 0.961 | -.067218 | .063963 | .153333 |
| rnkpalat* | .0050824 | .02958 | 0.17 | 0.864 | -.052887 | .063052 | .15 |
| mango* | .0654961 | .02702 | 2.42 | 0.015 | .012543 | .118449 | .333333 |
| nzimba* | .0146838 | .02779 | 0.53 | 0.597 | -.039777 | .069144 | .333333 |

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Bibliography/Further Reading

Adesina, A.A., Zinnah, M. 1993. "Technology Characteristics, Farmer Perceptions, and Adoption Decisions: A Tobit Model Application in Sierra Leone". *Agricultural Economics* Vol. 9: 297-311.

Feder, G., Slade, R. (1984). The acquisition of information and the adoption of new technology. *Amer. J. of Agric. Econ.* 66(2): 312-320.

Greene, W., (2000). *Econometric Analysis*. 4th Edition. Upper Saddle River, New Jersey.

**Augustine S. Langyintuo**, Timothy J. Dalton and Thomas Randolph, 2002. The role of information asymmetries, asset fixity and farmer perceptions in the adoption of improved rice varieties in Northern Ghana. Paper submitted as a selected paper to the 25th International Conference of Agricultural Economists, Durban, South Africa, 16-22 August, 2003.

McDonald, J.F., Moffitt, R.A., (1980). "The Uses of Tobit Analysis", *Review of Economics and Statistics*, Vol. 62, pp. 318-321.

Norris, P.E., Batie, S.S. (1987). Virginia farmers' soil conservation decisions: An application of Tobit analysis. Southern Journal of Agricultural Economics 19(1): 79-89.

## STATA Exercise

Participants are encouraged to accomplish the following exercise. You may refer to your notes but it is important to go through the whole exercise.

1. **Importing data from a previously saved SPSS file5.2.5 Listing Data:** You have been provided with a data set saved as zamdata.xpt file. Import the data into Stata

2. **Use of the *Generate* Commands:** Create a new variable called *imvprop* which is obtained by dividing *impava* (improved variety area) by *croplan* (cropped land).

3. **Estimating Means:** Calculate the means of *agehh* (age of household head), *hhsize* (house hold size), *crplan* (size of cropped land), *totfarm* (size of total farm land)

4. **Hypothesis Test of the Differences in Means:**
    a. Compare the means calculated in (3) between males and females.
    b. Compare the means calculated in (3) between adopters and non-adopters.

5. **Frequency distribution:** How many adopters and non adopters are in the sample by
    a. Whole sample?
    b. District?

6. **Hypothesis Test of the Differences in proportions:** How different are the relative proportions of males versus females attending various levels of education in the dataset by
    a. The whole sample?
    b. District?

7. **Correlation Matrices:** How related are *agehh* and *decyrs* (years household head assumed full responsibility to making farming decisions).

8. **Running a Tobit Model:** Run a Tobit model or *imvprop* on selected independent variables of your own choice and interpret the results.