ADB-WorldFish Technical Assistance to the Inland Fisheries Research and Development Institute (IFReDI), Royal Government of Cambodia
(TA Nº. T4025-CAM)

Training Workshop on Natural Resource Economics and Research Methods (Draft Manual)

23 February –3 March 2004
Penang, Malaysia

Mohammed A. Rab
Roehlano M. Briones
Andrew Tan Khee Guan
Ferdinand Paraguas

March 2004
Penang, Malaysia

Background

Rice and fish form the backbone of food security of Cambodia's 12 million people. The fisheries sector alone provides over 80% of animal protein supply, and employment and livelihoods to over 2 million people. Some recent findings suggest that per capita consumption of fresh and processed fish is nearly 75 kg per annum for the communities living in and around the waterways, river, lakes and flood-land. Given the low and highly volatile agricultural production, dependence of the rapidly growing population on fishery resources will remain high in the next decade. The country is in need of realizing economic gains from its vast inland water resources in sustainable ways within the shortest possible time and building capacity to monitor its natural resources.

Recognizing the high stakes involved in the fisheries sector, the Royal Government of Cambodia has taken specific steps to adopt and implement policies to promote sustainable management and conservation of natural resources that are critical to the livelihoods of millions. Given the importance, the Asian Development Bank (ADB) has been providing assistance to develop fisheries sector and to build capacity of the associated institutions as prioritized by the Royal Government of Cambodia. The newly created Inland Fisheries Research and Development Institute (IFReDI) is a step towards building institutional capacity of the Department of Fisheries (DoF) for research and development of the inland water resources that will help policy makers and stakeholders to take policy action for management and conservation of the resources that are under severe threat from increasing population pressure and over exploitation.

The ADB Technical Assistance (TA) in collaboration with the WorldFish Center was intended to build capacity of the IFReDI staff in the areas of research and development, institute management, technology transfer and policy development and dialogue. The WorldFish Center has implemented many of the activities of the TA including various research projects as part of "learning by doing" and a series of formal and on-the-job trainings in socioeconomic research methods, evaluation research design, statistics and valuation techniques of non-market resources. This training on "Natural Resources Economics and Research Methods" is carefully designed to address IFReDI's immediate needs for skill development that draws participants from across disciplines to make them understand the crucial importance of resource economics and impact analysis in policymaking. The most important strength of the training program is its comprehensiveness in contents and ability to mobilize resources from within and outside of WorldFish.

The training will focus on four broad areas such as basic economics, natural resources economics, research methods and statistics, and database management and statistical analytical techniques using software. To be precise, the theory part of the training will provide basic concepts of market economics and efficiency, causes and sources of market failure/inefficiency with special reference to natural/common property and environmental resources. The applied part of the training will provide basic theories in research methods and impact analysis, and techniques in data presentation that will include descriptive and inferential statistics with the exposure in computing techniques.

List of Participants & Observers

No.	Name	Position/Role	Organization
1	Mr. Srun Lim Song	Director/ National Project Director	IFReDI, DOF, Cambodia
2	Mr. Lieng Sopha	Deputy Director	IFReDI, DOF, Cambodia
3	Mrs. Hap Navy	Chief, Socioeconomics Division	IFReDI, DOF, Cambodia
4	Mr. Deap Loeung	Chief, Biology Unit/ Fishery Bioecologist	IFReDI, DOF, Cambodia
5	Mr. Seng Leang	Local Counterpart, Marketing	IFReDI, DOF, Cambodia
6	Mr. Chap Piseth	Vice Chief, Socioeconomics Division	IFReDI, DOF, Cambodia
7	Mr. Meas Vichit	Staff of Socioeconomics Division	IFReDI, DOF, Cambodia
8	Mr. Lieng Saroeun	Staff of Socioeconomics Division	IFReDI, DOF, Cambodia
9	Mr Hort Sitha	Deputy Chief, Admin. Unit	IFReDI, DOF, Cambodia
10	Mr Chheng Phen	Local Counterpart, Biology	IFReDI, DOF, Cambodia
11	Mr. Touch Bunthang	Local Counterpart, Technology Transfer	IFReDI, DOF, Cambodia
12	Mr. Ngor Peng Bun	Staff of Biology Division	IFReDI, DOF, Cambodia
13	Mr. Nom Sophiearith	Vice Chief, Biology Division	IFReDI, DOF, Cambodia
14	Mr. Hem Rady	Staff of Biology Division	IFReDI, DOF, Cambodia
15	Mrs. Chin Deth	Laboratory Technician	IFReDI, DOF, Cambodia
16	Mr. Chhun Haing Tong	Head of the Kandal Stung Field Research Station	IFReDI, DOF, Cambodia
17	Dr. Stephen J. Hall	Director General	WorldFish Center HQ, Malaysia
18	Dr. Paul S.Teng	Deputy DG-Research	WorldFish Center HQ, Malaysia
19	Dr. Mahfuzuddin Ahmed	Program Leader	WorldFish Center HQ, Malaysia
20	Dr. Mohammed A. Rab	Project Scientist	WorldFish Center HQ, Malaysia
21	Dr. Andrew Tan Khee Guan	Resource Personnel	Universiti Sains Malaysia, Penang.
22	Dr. Roehlano M. Briones	Post-Doctoral Fellow (Economics)	WorldFish Center HQ, Malaysia
23	Mr. Ferdinand J. Paraguas	Assistant Scientist	WorldFish Center HQ, Malaysia
24	Ms Jayamalar Francis	Project Management Assistant	WorldFish Center HQ, Malaysia
25	Ms Usha Kanagaratnam	Research Assistant	WorldFish Center HQ, Malaysia

Program

23 February 2004		
OPENING CEREMONY		
08:30	Registration	Secretariat
09:00	Welcome Address	Mahfuzuddin Ahmed, WorldFish Center
09:15	Opening Speech	Paul Teng, WorldFish Center
09:30	Overview of workshop	Mohammed A. Rab, WorldFish Center
09.45	**Tea break**	
10:00	Introduction to Economics	Roehlano M. Briones WorldFish Center
12:00	**Lunch Break**	
13:00	Supply and Demand –Basic Elements	Roehlano M. Briones WorldFish Center
14:30	Supply and Demand - Applications	Roehlano M. Briones WorldFish Center
16:00	**Tea break**	
16:15	Introduction to Research Methods	Mohammed A. Rab, WorldFish Center
24 February 2004		
08:30	Introduction and Economic Efficiency	Andrew Tan Khee Guan, Universiti Sains Malaysia, (USM)
10:30	**Tea break**	
10:45	Sources of Inefficiency and Market Failure	Andrew Tan Khee Guan, (USM)
12:45	**Lunch Break**	
14:00	Introduction to Research Methods *(Continued)*	Mohammed A. Rab, WorldFish Center
15:00	Research Design	Mohammed A. Rab, WorldFish Center
15:30	**Tea break**	
15:45	Why Worry about Data Management: An Outline to the Topic on Database Management and Data Analysis	Ferdinand Paraguas, WorldFish Center
25 February 2004		
08:45	Research Design *(Continued)*	Mohammed A. Rab, WorldFish Center
10:15	**Tea break**	
10:30	Policies to Address Efficiency Goals	Andrew Tan Khee Guan, (USM)
12:30	**Lunch Break**	
13:30	Reliability and Validity	Mohammed A. Rab, WorldFish Center
15:30	**Tea break**	

15:45	The Tonle Sap Economic Valuation Database: Introduction to Access	Ferdinand Paraguas, WorldFish Center
26 February 2004		
08:45	Measures of Supply and Demand Response	Roehlano M. Briones WorldFish Center
10:15	**Tea break**	
10:30	Valuing the Environment	Andrew Tan Khee Guan, (USM)
12:30	**Lunch Break**	
13:30	Analysis: Introduction to Descriptive Statistics	Mohammed A. Rab, WorldFish Center
15:30	**Tea break**	
15:45	The Tonle Sap Economic Valuation Database: Introduction to Access *(Continued)*	Ferdinand Paraguas, WorldFish Center
27 February 2004		
08:45	Production and Cost	Roehlano M. Briones WorldFish Center
10:15	**Tea break**	
10:30	Economics of Allocating Resources: Fisheries and Forests	Andrew Tan Khee Guan, (USM)
12:30	**Lunch break**	
13:30	Measures of Central Tendency	Mohammed A. Rab, WorldFish Center
15:30	**Tea break**	
15:45	The Tonle Sap Economic Valuation Database: Introduction to Access *(Continued)*	Ferdinand Paraguas, WorldFish Center
1 March 2004		
08:45	Profit and Supply	Roehlano M. Briones WorldFish Center
10:15	**Tea break**	
10:30	Economics of Allocating Resources: Agriculture and Water	Andrew Tan Khee Guan, (USM)
12:30	**Lunch break**	
13:30	Measures of Dispersion	Mohammed A. Rab, WorldFish Center
15:30	**Tea break**	
15:45	Using Microsoft Excel for Database Management and Initial Data Processing	Ferdinand Paraguas, WorldFish Center
2 March 2004		
08:45	Special Topic: International Trade	Roehlano M. Briones WorldFish Center
10:15	**Tea break**	
10:30	Correlation and Regression	Mohammed A. Rab, WorldFish Center
12:30	**Lunch break**	

13:30	Using Microsoft Excel for Database Management and Initial Data Processing *(Continued)*	Ferdinand Paraguas / Mohammed A. Rab, WorldFish Center
15:30	**Tea break**	
15:45	Data Management and Application	Ferdinand Paraguas / Mohammed A. Rab, WorldFish Center
3 March 2004		
08:45	Analysis of Household survey data using SPSS	Ferdinand Paraguas / Mohammed A. Rab, WorldFish Center
10:15	**Tea break**	
10:30	Analysis of Household survey data using SPSS *(Continued)*	Ferdinand Paraguas / Mohammed A. Rab, WorldFish Center
12:30	**Lunch break**	
13.30	Evaluation	
14:00	Analysis of Household survey data using SPSS *(Continued)*	Ferdinand Paraguas / Mohammed A. Rab, WorldFish Center
Concluding Session		
16:00	Tea break	All HQ staff- Auditorium Foyer
16:30	Chair	Mahfuzuddin Ahmed, Program Leader, PRIAP
16:45	Concluding Remarks	Mohammed A. Rab, WorldFish Center
17.00	Participants' Comments	Srun Lim Song, Director IFReDI
17.15	Concluding Speech and Certificate Distribution	Stephen Hall, Director General, WorldFish Center

1. Introduction

1. What is economics?

- Economics- the social science which studies how society allocates scarce resources to satisfy unlimited human wants.
    - resources → production→consumption→satisfaction of wants
    - wants are unlimited but resources are limited
    - reality of choice, benefits, and costs
- as science – a systematic inquiry, using empirical methods, oriented towards fundamental relationships underlying phenomena.
- social science – phenomena to be studied is society in its act of material production and consumption
    - hence observable, quantifiable – usually material objects in specific units
    - Ultimately claims must be tested against *data*.
- Major branches
    - Microeconomics –deals with the behavior of individual economic units (consumers, firms, workers, and investors), as well as the markets that these units comprise.
    - Macroeconomics –deals with aggregate economic variables (level and growth rate of national output, interest rates, and inflation).

1.2. Basic economic problems

- Economic Questions:
    - **How much** – of each good
    - **How** – combination of resources for producing each good (substitution possible)
    - **For whom** – allocation to each person
- Economic system – means by which a society addresses these problems. Types:
    - Custom – or traditional: for small economies, typically near subsistence, low levels of technology.
    - Market – use of free exchange. Typically use common denominator of value, or money, and exchanges quoted in terms of prices: hence, *price system*.
    - Planned – major allocation decisions made by government, through planning mechanism.

- Mixed – market economy with a mix of government intervention (a spectrum of forms). Most countries (including Philippines) are mixed.

- Normative and positive analysis

  - strictly speaking, science concerned only with a description and explanation of phenomena, which is *positive analysis* (what *is*)

  - application of economic science to improve material conditions may require judgments as to what constitutes an "improvement"

  - value judgments as to what is better or worse from a social or ethical or theological viewpoint (related to what *ought to be* or *should* be)

- Examples:

Normative: The government should repeal tariffs. Income distribution ought to be egalitarian.

Positive: Foreign prices of cement are lower than domestic prices. The inequality of income distribution in Cambodia is similar to that of Brazil.

Note: A positive statement may be either true or false.

- The problem of development – related to the basic questions.

  - how to make more goods available for all – **economic growth**

  - how to make it equitably distributed – **inequality reduction**

  - how to elevate everybody to a decent living standard – **poverty reduction**

  - how to avoid economic fluctuation – **stability**

  - how to sustain economic activity - **sustainability**

1.3. The method of economics

- Models – a simplification of reality

  - illustration: how useful is a map depends on what you want to do with it – do you want to get from point a to point b? do you want to know where there is traffic? do you want to know scenic spots?

  - models in economics – supply and demand. utility maximization. profit maximization.

- Model generates testable implications which are then compared to the data

2. Supply and demand: Introduction

2.1. Overview

Supply and demand: uses the concept of *equilibrium*

- *supply* denotes exchange offers of sellers
- *demand* denotes exchange offers of buyers
- when the exchange offers match, then there is equilibrium.

Equilibrium: Economics makes extensive use of the concept equilibrium.

- A system is in equilibrium when its component parts are in stable relationship with one another.

- equilibrium can be disturbed, and system will adjust from one static equilibrium to another → basis for framing hypothesis about directions of change

- In economics:
  - observed prices are equilibrium prices
  - changes in prices are changes in equilibrium prices.
  - equilibrium concept is main mechanism for generating testable predictions about cause-effect links to price and quantity traded in a market.

Set up

- we have to be careful to define the scope of a market, e.g. in defining geographic scope as well as range of products included

- Prices are given to the agents: the assumption of *price-taking* or *competitive* behavior

- Proper definition of a market plus competitive behavior leads to only one price in the market (competition between sellers eliminates differences in price)

Example: market for catfish in Seam Reap, quantities defined in terms of kg of catfish per day, price defined in terms of US cents per kg.

2.2. Demand

Demand schedule: a schedule that states the quantity of a good that buyers are ready to purchase, at various alternative prices, per unit time, within a market, all other factors constant.

Demand function: a function that relates the quantity of a good that buyers are ready to purchase, at various alternative prices, per unit time, all other factors constant.

Demand curve: a curve which plots the quantity of a good that buyers are ready to purchase, at various alternative prices, per unit time, all other factors constant.

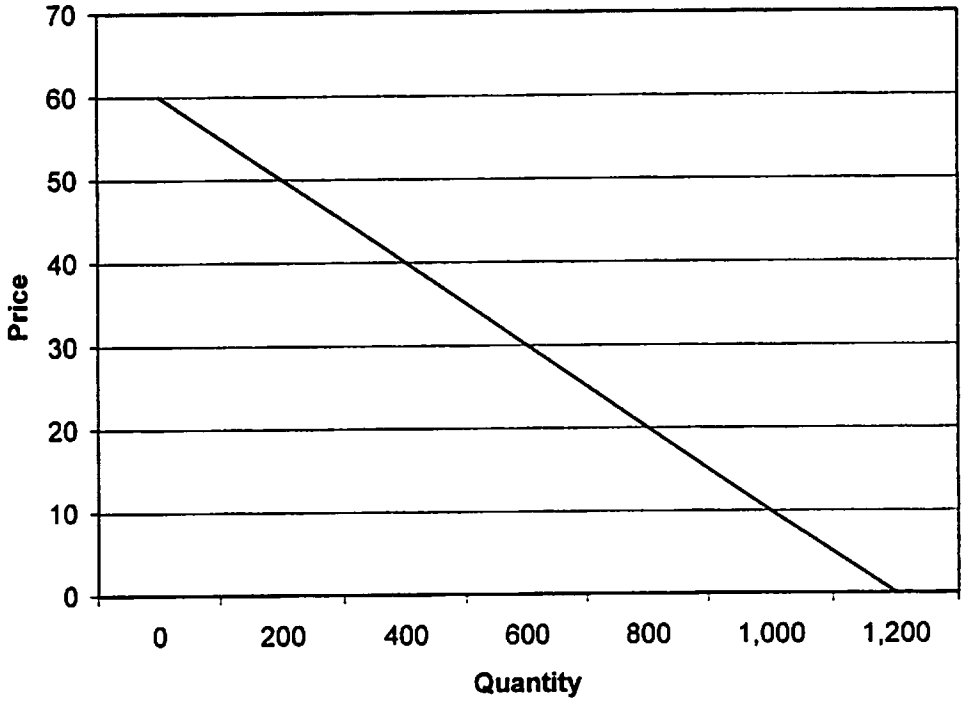
Example:

$Q_D = 1{,}200 - 20P$ where Q_D = quantity of rice demanded in the abovementioned market

Schedule:

Price	Quantity demanded (kg per day)
0	1,200
10	1,000
20	800
30	600
40	400
50	200
60	0

Figure 2.1. Demand Curve



Properties:

- price is negatively related to quantity, i.e. demand curve slopes downward – LAW OF DEMAND

- Note that these are purely hypothetical prices and quantities, which denote plans or intentions rather than actual purchases

Factors affecting demand:

1. Information on product characteristics
2. Tastes
3. Income
    a. if income has a positive effect: *normal good* (rice)
    b. if income has a negative effect: *inferior good*

4. Population – usually if population rises, demand increases
5. Prices of related goods
    a. If related good has a positive effect: *substitute* (rice and corn)
    b. If related good has a negative effect: *complement* (coffee and creamer)

2.3. Supply

Supply schedule: a schedule that states the quantity of a good that producers are ready to sell, at various alternative prices, per unit time, within a market, all other factors constant .

Supply function: a function that relates the quantity of a good that producers are ready to sell, at various alternative prices, per unit time, all other factors constant .

Supply curve: a curve ...

Example:

$Q_S = -150 + 25P$ where Q_S = quantity of rice supplied in the Philippines, in metric tons per day

- Note the negative intercept: simply means that price must be some positive number before any quantity can be supplied

Schedule:

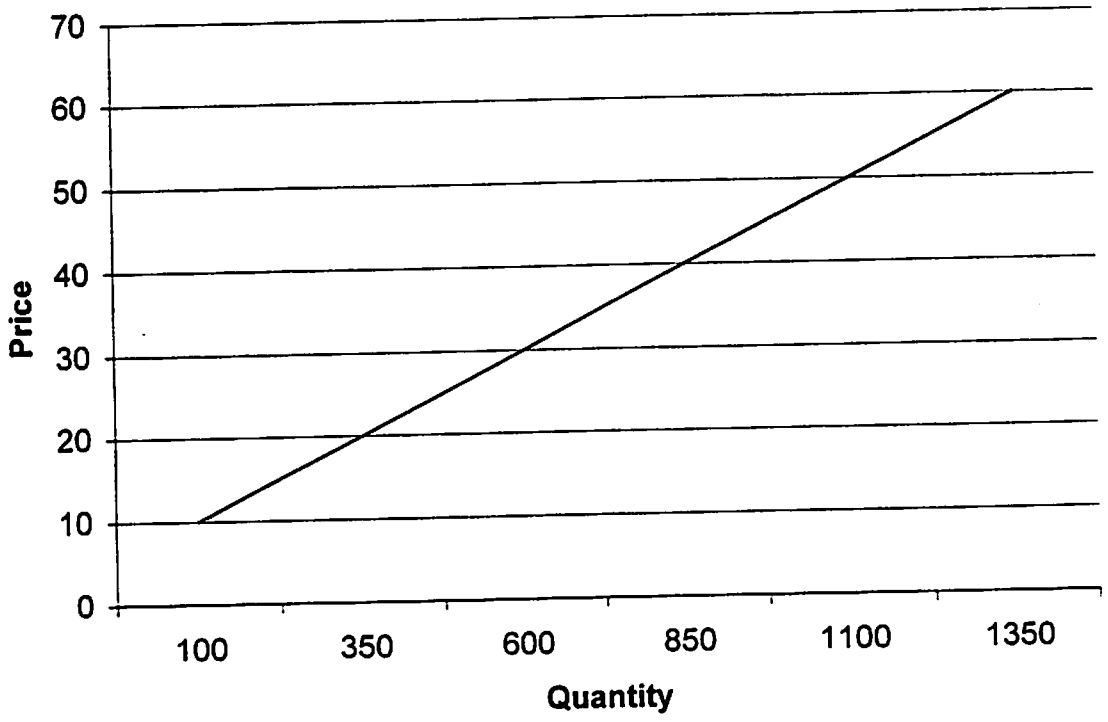
Price	Quantity supplied (metric tones per day)
10	100
20	350
30	600
40	850
50	1,100
60	1,350

- price is positively related to quantity, i.e. supply curve slopes upward– LAW OF SUPPLY (ultimately the basis of this is profit-seeking behavior of firm's owners)

Factors affecting supply:

1. Technology
2. Prices of inputs (availability of inputs)
3. Profitability in related product lines
4. Environmental shocks (especially in agriculture)

Figure 2.2. Supply curve



2.4. Price determination

Consider: supply and demand in one (schedule/graph)

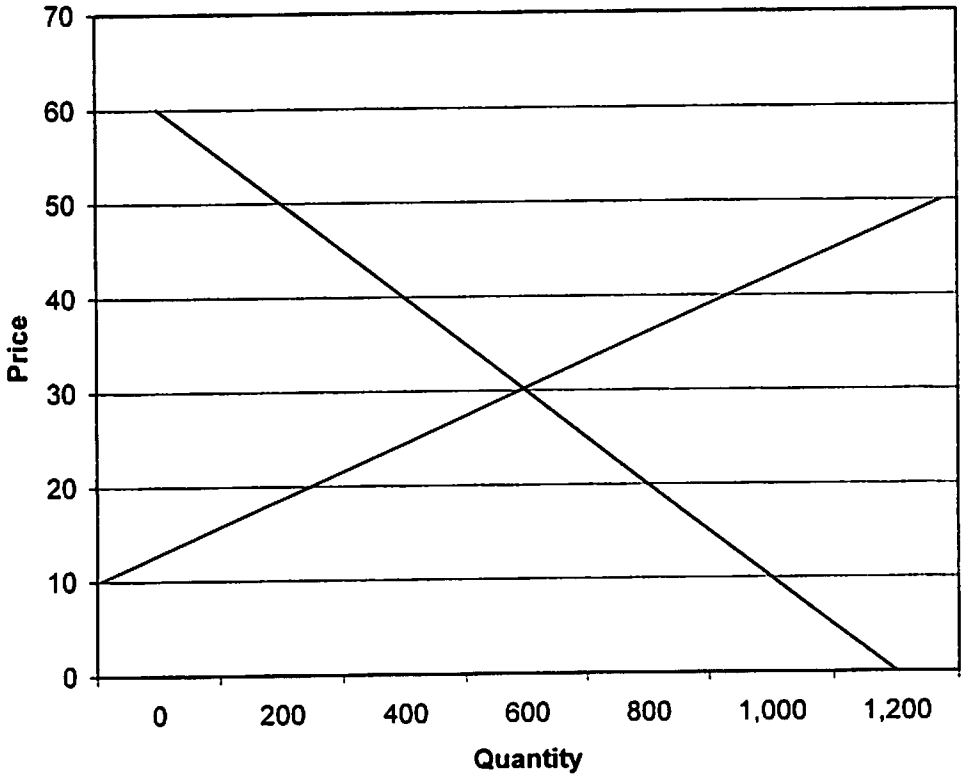
- At price = 4: quantity demanded exceeds quantity supplied; buyers as a whole are frustrated in their purchasing plans. SHORTAGE. Buyers are interested in changing the situation -> bid up the price.

- At price = 60: quantity demanded is less than quantity supplied; sellers as a whole are frustrated in their selling plans. SURPLUS. Sellers are interested in changing the condition -> bid down the price.

Price	Quantity supplied $Q_S = -150 + 25P$	Quantity demanded $Q_D = 1,200 - 20P$	Quantity traded	Excess supply/demand
10	100	1,000	100	
20	350	800	350	
30	600	600	600	
40	850	400	400	
50	1,100	200	200	
60	1,350	0	0	

- Suggests equilibrium condition: quantity demanded = quantity supplied. Everybody's plans are carried out – nobody is interested in changing the situation.

- adjustment in off-equilibrium situation carried by price adjustment; note that price is simply the terms of exchange. Hence, when buyers want more than are available, they persuade sellers to sell more by making the exchange terms more attractive to them; similarly for the sellers when they want buyers to buy more.



- adjustments are done by individuals, without any central direction. No negotiation needed between buyers and sellers – one can simply imagine only the price tags being altered

SOLUTION: Using supply and demand functions, note that equilibrium condition requires: $1200 - 20P = -150 + 25P$ or $P = 30$, $Q = 600$.

Problems

1. Which of the following is a possible demand function? Supply function? Why? (Q = quantity, P = price; constants a, b, c, and d are all positive)

   a. $Q = a + bP$ b. $Q = c + d/P$ c. $Q = a + b\sqrt{P}$

2. Supply is $Q = -440 + 125P$; Demand is $1000 - 163P$. Fill in the following table:

Condition	Price	Quantity traded	Shortage
Equilibrium			
Excess demand	4		

3. Supply is $Q = -100 + 40P$; Demand is $950 - 30P$. Fill in the following table:

Condition	Price	Quantity traded	Surplus
Equilibrium			
Excess supply	20		

3. Demand and Supply - Applications

3.1. Demand factors and the demand curve

- If demand factors (other than price) change, the entire demand schedule shifts.

  Example: From $Q_D = 1,200 - 20P$ to $Q_D = 1,000 - 20P$: shift left, or down; fall in demand.

Price	Quantity demanded $Q_D = 1,200 - 20P$	Quantity demanded $Q_D = 1,000 - 20P$
0	1,200	1,000
10	1,000	800
20	800	600
30	600	400
40	400	200
50	200	0
60	0	0

Factor	Demand change	Direction of shift
Information about favorable product characteristics	Increase	Rightward
Information about unfavorable product characteristics	Decrease	Leftward
Increased preference for the good	Increase	
Decreased preference	Decrease	
Increase in population	Increase	
Decrease in population	Decrease	
Normal goods:		
Higher income	Increase	
Lower income	Decrease	
Inferior goods:		
Higher income	Decrease	
Lower income	Increase	
Rise in the price of substitute goods	Increase	
Rise in the price of complementary goods	Decrease	
Fall in the price of substitutes	Decrease	
Fall in the price of complements	Increase	

- *Change in demand* versus *change in quantity demanded*: the latter is caused only by a change in price. The former is caused by a change in the factors affecting demand that were held constant in generating a demand function.

3.2. Supply factors and the supply curve

When other factors held constant in making the supply schedule are allowed to vary, the entire schedule shifts. Example: From $Q_S = -150 + 25P$ to $Q_S = -200 + 25P$: shift left, or down; fall in supply.

Price	Quantity supplied $Q_S = -150 + 25P$	Quantity supplied $Q_S = -200 + 25P$
0	-150	-200
10	100	50
20	350	300
30	600	550
40	850	800
50	1,100	1,050
60	1,350	1,300
	1,600	1,550

Factor	Supply change	Direction of shift
Technology improves	Increase Decrease	Rightward Leftward
Price of input falls Price of inputs rises	Increase Decrease	
Profitability or price of production alternatives rise Profitability or price of production alternatives fall	Decrease Increase	
Favorable environmental shock Unfavorable environmental shock	Increase Decrease	
Increase in population Decrease in population		

Change in supply versus *change in quantity supplied*: the latter is caused only by a change in price. The former is caused by a change in the factors affecting supply that were held constant in generating a supply function.

3.3. Explaining *changes* in price and quantity traded

Change	Price change	Quantity traded change
Increase in demand	Increase	Increase
Decrease in demand	Decrease	Decrease
Increase in supply	Decrease	Increase
Decrease in supply	Increase	Increase

- Price plays an allocative function: when a good is more desired by consumers, price rises to attract more resources into its production. When a good is less desired, price falls to allow resources to flow out to other uses. When a good is easier to produce, price falls to attract consumers to buy this good. When a good is more difficult to produce, price rises to restrain purchases of this good by consumers.

- This analysis highlights the empirically testable implications of the supply and demand model. After all, we cannot be test whether or not a price is an equilibrium price; but we can confirm whether or not predictions of how prices and quantities change are confirmed by the data.

Note: Factors that change demand do not change the supply schedule/curve/function. Demand factors can only affect quantity supplied, via changes in market price. Likewise factors that change supply do not change the demand schedule/curve/function. Supply factors can only affect quantity demanded, via changes in market price.

3.4. Application: technical change in agriculture

Quantity is in tons/year; price is in USD/ton. The annual demand curve for carp is given by:

$Q = 28{,}000 - 10P$

The annual supply curve is given by:

$Q = -2{,}000 + 15P$

Genetic improvement of carp leads to a new variety that grows 20% more per production cycle than the previous one.

New supply curve: $Q = -2{,}400 + 18P$

Condition	Price	Quantity
Without improvement	1,200	16,000
With improvement	1,085.7	17,142

Problems

1. Predict what happens to the following prices and quantities when the following events occur:

    a. Price and quantity of fish: an increase in the price of beef.

    b. Price and quantity of tires: an increase in automobile prices

    c. Price and quantity of fish: an increase in crude oil prices

    d. Price and quantity of rice: higher yet flood-free rainfall year round

    e. Price of digital cameras: improved technology allows ordinarily-priced video cameras to take identical quality pictures as digital cameras

2. Given baseline supply and demand curves: $Q_D = 1,200 - 20P$; and $Q_S = -150 + 25P$. Calculate equilibrium price and quantity. (See handout No. 2). Calculate what happens to equilibrium price and quantity, when:

    a. the demand curve becomes Q = 1,000 – 20P

    b. the supply curve becomes Q = -200 + 25P (in relation to the baseline)

    c. production expands by 25% due to technical change (in relation to the baseline)

4. Measures of Price Response

So far: have dealt only with directions of change, i.e."quantity demanded falls with an increase in price." To determine *how much* it falls, we need a measure of price response of demand. Likewise, to determine *how much* quantity supplied rises with price, we need a measure of price response of supply.

4.1 Price response for demand

Given $Q_D = f(P)$, then a natural candidate for response is:

Slope = $\frac{Q_{D1} - Q_{D0}}{Q_{D0}}$; where base quantity is Q_{D0} and quantity after the price change is Q_{D1}.

Sope = $\frac{dQ_D}{dP}$

However we cannot compare demand response of different goods which are measured by different units (i.e. fish measured by tons; perfume measured by bottles)

Alternative: Elasticity

$$\varepsilon = \frac{\%\Delta Q_D}{\%\Delta P} = \frac{\frac{Q_1 - Q_0}{(Q_1 + Q_0)/2}}{\frac{P_1 - P_0}{(P_1 + P_0)/2}} \rightarrow \text{a negative number}$$

Example: $Q_D = 240 - 4P$; $P_0 = 40$, $P_1 = 45$; then
$Q_0 = 240 - 4*40 = 80$, $Q_1 = 240 - 4*45 = 60$

$$\varepsilon = \frac{\frac{-20}{70}}{\frac{5}{42.5}} = -2.43$$

point elasticity: $\varepsilon = \frac{\Delta Q_D / Q_D}{\Delta P / P}$

- If the demand curve is linear and given by $Q_D = a - bP$, then

$$\varepsilon = \frac{\Delta Q_D}{\Delta P} \cdot \frac{P}{Q} = -b \cdot \frac{P}{a - bP}$$

Example: $Q_D = 240 - 4P$

What is price elasticity at price equal to 40? Ans. –2.

Factors affecting elasticity:

- Degree of substitutability with other products (more substitutable, more elastic)
- Durability (more durable, more elastic) e.g. computer demand is elastic; even more so demand for semiconductors. Hence the distinction between short-run and long-run elasticity.
- Budget share of the commodity (the larger the share, the more elastic)

Other demand elasticities

- Income elasticity: response to income *Y*.

$$\varepsilon_Y = \frac{\%\Delta Q_D}{\%\Delta Y}$$

If positive: normal good
If negative: inferior good

- Cross-price elasticity: response of Q_{Di} to a price change of another good P_j.

$$\varepsilon_{ij} = \frac{\%\Delta Q_{Di}}{\%\Delta P_j}$$

If positive: substitute
If negative: complement

4.2 Price Response for Supply

How responsive is quantity supplied to change in price? Likewise apply an *elasticity* measure.

$\varepsilon = \frac{\%\Delta Q_S}{\%\Delta P}$ → positive number;

$$\varepsilon = \frac{\%\Delta Q_S}{\%\Delta P} = \frac{\frac{Q_{S1} - Q_{S0}}{0.5*(Q_{S1} + Q_{S0})}}{\frac{P_1 - P_0}{0.5(P_1 + P_0)}}$$ = (arc elasticity) → positive number

Point elasticity: $\frac{dQ}{Q} \Big/ \frac{dP}{P} = \frac{dQ}{dP} \cdot \frac{P}{Q}$

for linear supply curve $Q_S = -c + dP$, then $\varepsilon = \frac{dP}{-c + dP} > 0$

Factors affecting elasticity

- nature of the product (e.g. rice versus hamburgers)
- time horizon (elasticity of supply tends to be higher in the long-run compared to that of the short-run)

Problems

1. Supply is Q = -440 +125P; Demand is Q = 1000 – 163P. (See Handout No. 2). At equilibrium, point elasticity of demand is _____ while that of supply is _____.

2. The demand curve is given by: Q = 100 – 2P

    a. Compute the point elasticity of demand at P = 20.
    b. Compute the arc elasticity for P = 20 to P = 25.

3. At P = 20, quantity supplied is 40. If the supply curve is linear and supply elasticity is 0.8, what is the supply curve?

5. Production and cost

5.1. Outputs from inputs

The producer

Who is the "producer"? The entity that can make organize the process of utilizing resources to produce outputs. Example: farmer, fishpond operator, Microsoft, etc.

Production function

- the function that identifies the maximum output that a producer can achieve at alternative combinations of inputs.
- Example: $Q = f(L) = 4.5\sqrt{L}$; other input levels are fixed.
- The production function corresponds to a *production schedule*. In a single-input case, it can also be represented by a Total Product curve.
- Purely an engineering relation; depends on physical laws plus state of human ingenuity. Technology is given.
- Independent of preferences, or scarcity of resources, or prices of inputs or outputs.
- Time horizons
  - Short run: some inputs fixed.
  - Long run: all inputs variable. In which case:
- A short-run production schedule can be shifted upward by:
  - technological change
  - greater quantities of a fixed input

Productivity relations

Average product (AP) = output per unit input; AP=Q/L

Graphically: the slope of the line connecting TP and the origin

Marginal product = additional output for the last unit of input

$$MP = \frac{\Delta Q}{\Delta L}$$

Law of Diminishing Marginal Returns: When only one input is being varied, additional output per unit additional input eventually declines (i.e. MP must eventually fall towards zero)

Why? Each additional unit has fewer of the fixed inputs to work with.

4.2. Cost

The concept of cost

- Economic cost: the value of opportunities foregone with particular choices. Example: farmer owns property, uses a portion to dig a fishpond. Foregoes the space that could have gone to growing a crop. This foregone net harvest is part of the cost of the fishpond.

- Accounting cost: would only count actual market transactions as cost. Foregone opportunities that do not involve market transactions are not included (i.e. foregone use of land is often not included in inputs owned by the producer). Example: cost of family labor.

- Private versus social cost: private costs refer to opportunity costs from the viewpoint of a private individual. Some costs refer to foregone benefits from the viewpoint of society, which may differ from the individual's viewpoint.

    - Example: a hog farmer dumps waste into a stream → cost in terms of immediate nuisance is low; however downstream reduction in fish catch, loss in amenity value of stream, is ignored by the farmer.
    - Divergence between marginal and social cost is widespread, but social cost is difficult to measure (compared to private cost).

Fixed costs

- Fixed inputs → when valued, yields *fixed cost* (FC). Therefore holds only in the short-run.

Equals total cost even when no variable inputs are used (i.e. no output is produced).

Variable costs

Cost of the variable inputs → variable cost (VC).

Total cost

Total cost (C) is defined as: C = FC+VC.

Cost function

TC = C(Q) → function that relates C with alternative levels of Q.

Problems

1. Chair manufacturer using only labor, from 1 to 7 of labor. Calculate average and marginal products.

L	Q	AP	MP
0	0	-	-
1	10		
2	17		
3	22		
4	25		
5	26		
6	25		
7	23		

Does this exhibit diminishing returns?

2. Suppose land rent is 100 USD per ha. and labor cost is 0.20 USD per hour. Fill in the following table.

Land (ha)	**Labor** (person-hours per day)	**Rice** output (mt per cropping)	**Fixed cost** (USD per cropping)	**Variable cost** (USD per cropping)	**Total cost** (USD per cropping)
4	0	0	400	0	400
4	1	10	400	60	460
4	2	22	400	120	520
4	3	32	400	180	580
4	4	40	400	240	640
4	5	46	400	300	700
4	6	50	400	360	760
4	7	52	400	420	820
4	8	52	400	480	880
4	9	50	400	540	940

Handout No. 6
Basic Economics

6.Costs and Supply

6.1. Unit costs

Average costs

- Average fixed cost (AFC): $AFC = \frac{FC}{Q}$, or the fixed cost per unit output. Of course, this decreases as output increases.
- Average variable cost (AVC): AVC = VC/Q, i.e. variable cost per unit output.
- Average cost (AC): AC = C/Q =AFC+AVC, i.e. cost per unit output.

Marginal cost

Marginal cost (MC):

$$MC = \frac{\Delta C}{\Delta Q} = \frac{\Delta VC}{\Delta Q}$$

→ additional cost per unit change in output, (or how much cost will fall if output is reduced by one unit).

Example:

Output	Total cost	Fixed cost	Variable cost	Marginal cost	Average cost	Average variable cost	Average fixed cost
0	30	30	0	-	-	-	-
1	35	30	5	5	35	5	30
2	60	30	30	25	30	15	15
3	110	30	80	50	37	27	10
4	200	30	170	90	50	43	8
5	320	30	290	120	64	58	6
6	600	30	570	280	100	95	5

6.2. Marginal cost and production

Case of one variable input

- In this special case, the shape of the cost curves can be determined from the product curves
- We have: $MC = w\frac{\Delta L}{\Delta Q} = \frac{w}{MP}$
- When MP is rising, MC is falling, and where MP is falling, MC is rising.
- Hence, MC is rising → due to diminishing returns.

Case of more than one variable input

In the short run there may be more than one variable input. Marginal cost may still be upward sloping → use multiple variable version of diminishing returns (i.e. *concavity*)

6.3. Returns to production

Revenue

So far we have discussed costs and production. We now move to the revenue side. Impose a *fixed price assumption* (highly applicable to agricultural markets; questionable in many other cases).

$R = P^*Q = R(Q)$.

Hence, we can write profit as:

Profit = $R(Q) - C(Q)$

Marginal revenue

- Marginal revenue or MR: change in revenue per unit change in output sold.
- Under a fixed price assumption: each unit sold adds its own price to total revenue, hence price equals marginal revenue, that is:

  $P = MR$.

6.4. Profit-maximization

Consider the case in which the producer is in the market (i.e. producing). Question: *how much* does it produce?

The problem is to maximize profit (denoted by π) by choice of Q. Consider the *change in profit* as Q changes.

Change in profit can be split into: change in revenue and change in cost.

$$\frac{\Delta\pi}{\Delta Q} = \frac{\Delta R}{\Delta Q} - \frac{\Delta C}{\Delta Q}$$ → change in profit = $MR - MC$ = P – MC.

For an additional unit of Q:
If change in profit is *positive*: **keep on adding output** [P > MC]
If change in profit is *negative*: **cut back on output** [P < MC]

Example:

Q	P	TR	TC	MC
0	30	0	0	-
1	30	30	25	10
2	30	60	40	15
3	30	90	60	20
4	30	120	85	25
5	30	150	115	30
6	30	180	150	35

Marginal revenue is equal to price = 30; this equals marginal cost at Q = 5. Total profit is 35.

6.5. Supply

Individual producer's supply

Profit maximization: for every P, set Q such that $P = MC$.

Hence, $P = MC(Q) = f(Q)$. That is, f relates Q to P (through marginal cost).

For a producer that is in the market: The individual supply curve corresponds to its marginal cost curve (as long as it is producing).

Market supply

Obtained by horizontal addition of individual firms' supply curves.

Problem

Given the schedule of output and total cost given in Problem 2, Handout No. 5. Fill in the following tables (costs are in USD per cropping).

Output	Total cost	Fixed cost	Variable cost	Average cost	Average variable cost	Marginal cost
0						
10						
22						
32						
40						
46						
50						
52						
52						

What is the profit-maximizing supply when the price is 30?

Handout No. 7
Basic Economics

7. International Trade: Introduction

7.1. Absolute advantage

What do nations gain when they trade?

Assumptions

- Labor is the only resource
- Marginal product is constant (which implies: marginal product = average product)
- Earliest theory: *Absolute advantage*

  ⇒ A nation will sell goods in which its average productivity is superior to that of other countries.

  ⇒ Example: Given existing machinery, 100 person-days of labor can yield:

Table 1.

	Cellphones	Cameras
Japan	20	20
Sweden	25	15

Japan has absolute advantage in Cameras.
Sweden has absolute advantage in cellphones.
This suggests a "gain from trade."
Japan will sell cameras to and buy cellphones from Sweden.

7.2. Comparative advantage

Consider the following situation: 100 person-days of labor

Table 2.

	T-shirts	Spark plugs
Japan	200	100
Cambodia	100	1

Japan has absolute advantage in T-shirts and spark plugs.
There seems to be no gain from trade.
So there will be no trade between Japan and Cambodia.

- Over a hundred years ago, the economist David Ricardo showed that the above analysis is incorrect. In Table 2 there is a potential gain from trade.

- What matters is not absolute productivity, but *relative* productivity. In other words it is not absolute but comparative advantage which is a source of gains from trade.

- In Table 2: Cambodia gives up 1 spark plug to get 100 T-shirts. In Japan, the opportunity cost of 100 T-shirts is 200 spark plugs. Cambodia has a comparatively higher productivity in T-shirts (opportunity costs are lower).

- Meanwhile: Japan gives up 2 T-shirts to get 1 spark plug; Cambodia must give up 100 T-shirts to get 1 spark plug. Japan has comparatively higher productivity in spark plugs (opportunity costs are lower).

- Trading scenario: Japan will trade as long as relative price (exchange ratio) of spark plugs (for shirts) is greater than 2. Cambodia will trade as long as the relative price is below 100. At some in-between exchange ratio (e.g. 50 shirts per spark-plug), both countries gain from trade.

Country	Product *x*	Product *y*
A	4	5
B	2	4

A has absolute advantage in both products.

But relative cost of Product *y* is 4/5. In country *B* the relative cost of product *y* is ½. Then country *A* will gain by buying product *y* from country *B*.

Relative cost of Product *x* in country *A* is 1.2; in country *B* the relative cost is 2. Then country *B* will gain from buying product *x* from country *A*.

Possible exchange rate: 1.5 of product *x* per 1 unit of product *y*. Country *A* sells *x*, gets 0.3 more of *y* per unit of *x*; Country *B* buys *x*, and saves on 0.5 of *x* per unit of *y*.

Handout No. 7
Basic Economics

Problem

Country	Per 100 person-hours	
	Units of garment	Units of silicon chip
China	8	4
Philippines	6	2

China has the absolute advantage in _____.

Philippines has the absolute advantage in _____.

China has the comparative advantage in _____. This is because the opportunity cost of 1 of _____ is only _____ units of _____. Meanwhile for the Philippines the opportunity cost is _____ units.

Philippines has the comparative advantage in _____. This is because the opportunity cost of 1 of _____ is only _____ units of _____. Meanwhile for China the opportunity cost is _____ units.

At an exchange rate of _____ garments for 1 silicon chip, both countries stand to gain from trade. This is because Philippines will get an extra _____ units of _____, while China can get _____ units of _____.

Training Workshop on Natural Resource Economics & Research Methods
Conducted by the Policy Research & Impact Assessment Program
WorldFish Center
(23 Feb - 04 March 2004)

Natural Resource Economics
(Outline of Lectures)

Lecture	Topics
1	I. INTRODUCTION i. What is Environmental Economics? ii. What is Natural Resource Economics? II. THEORY & FUNDAMENTALS OF NATURAL RESOURCE ECONOMICS i. Economic Efficiency a. Static Efficiency *Demand & Willingness To Pay* *Total & Marginal Cost* *Net Benefits* b. Dynamic Efficiency *Choices Over Time* *Present Value and the Time Value of Money*
2	c. Sources of Inefficiency (Market Failure) *Externalities* *Public Goods* *Imperfect Competition* *Imperfect Information* *Inappropriate Gov't Intervention/Gov't Failure*
3	ii. Policies to Address Efficiency Goals a. Decentralized Policies *Liability Laws* *Property Rights* *Moral Suasion* b. Command-and-Control Strategies *Standards* *Enforcement* *Emergency Restrictions* c. Economic Incentive-Based Strategies *Emission Taxes* *Abatement Subsidies,* *Marketable/Transferable Discharge Permits* *Deposit-Refund Systems*

Lecture	Topics
4	iii. Valuing The Environment For Enviromental Decision Making a. Economic Impact Analysis b. Cost-Effectiveness Analysis c. Risk Analysis d. Measuring Non-Market Environmental Goods *The Travel Cost Method (TCM)* *The Hedonic Regression Method (HRM)* *The Contingent Valuation Method (CVM)*
5	III. THE ECONOMICS OF ALLOCATING NATURAL RESOURCES (ISSUES) i. Renewable Common-Property Resource (Fisheries) ii. Storable, Renewable Resource (Forests)
6	iii. Reproducible Private-Property Resource (Agriculture) iv. Replenishable but Depletable Resource (Water)

REFERENCES:

Baumol, W. J. and W. E. Oates. The Theory of Environmental Policy. Cambridge University Press, 2nd Edition, New York (1988).

Callan, S.J. & J.M. Thomas. Environmental Economics and Management: Theory, Policy, and Applications. The Dryden Press, Florida (2000).

Field, B. C. Environmental Economics. McGraw-Hill, Inc., New York (1994).

Hackett, S. Environmental and Natural Resources Economics: Theory, Policy, and the Sustainable Society. M.E. Sharpe Publishers (1998).

Kahn, J. R. The Economic Approach to Environmental and Natural Resources. The Dryden Press, 2nd. Edition, Florida (1998).

Lesser, J.A., D. E. Dodds, & R.O. Zerbe. Environmental Economics and Policy. Addison Wesley (1997).

Tietenberg, T. Environmental and Natural Resource Economics. Harper Collins Publishers, 3rd & 4th Edition, New York (1992 & 1996).

Training Workshop on Natural Resource Economics & Research Methods
Conducted by the Policy Research & Impact Assessment Program
WorldFish Center
(23 Feb - 03 Mar 2004)

Natural Resource Economics
(Lecture 1)

I. INTRODUCTION

i. What is Environmental Economics?

a. Economics – The study of how and why "people" (consumers, firms, non-profit organizations, government agencies) make decisions about the use of valuable and limited resources (labor, land, capital) to meet unlimited wants/desires.

b. Environmental Economics – Focuses primarily on how and why people make decisions that have environmental consequences. Also focuses on how we change economic institutions and policies to bring those environmental impacts more into balance with human desires and the needs of the ecosystem itself.

ii. What is Natural Resource Economics?

The study of how different natural resources or raw materials (input) are manipulated and transformed to produce consumer goods (output) by the production process. These resources include:

a. Renewable Resources – grows in tim and replenishes according to biological processes

Renewable Common-Property Resources (fisheries)
Reproducible Private-Property Resources (agriculture)
Storable, Renewable Resources (forests, sunlight/solar)

b. Non-renewable Resources – there are no processes of replenishment (once used, gone forever)

Depletable & Nonrecyclable Resources (petroleum, coal, uranium)
Recyclable Resources (minerals, paper, glass, plastics)
Replenishable But Depletable Resources (water)

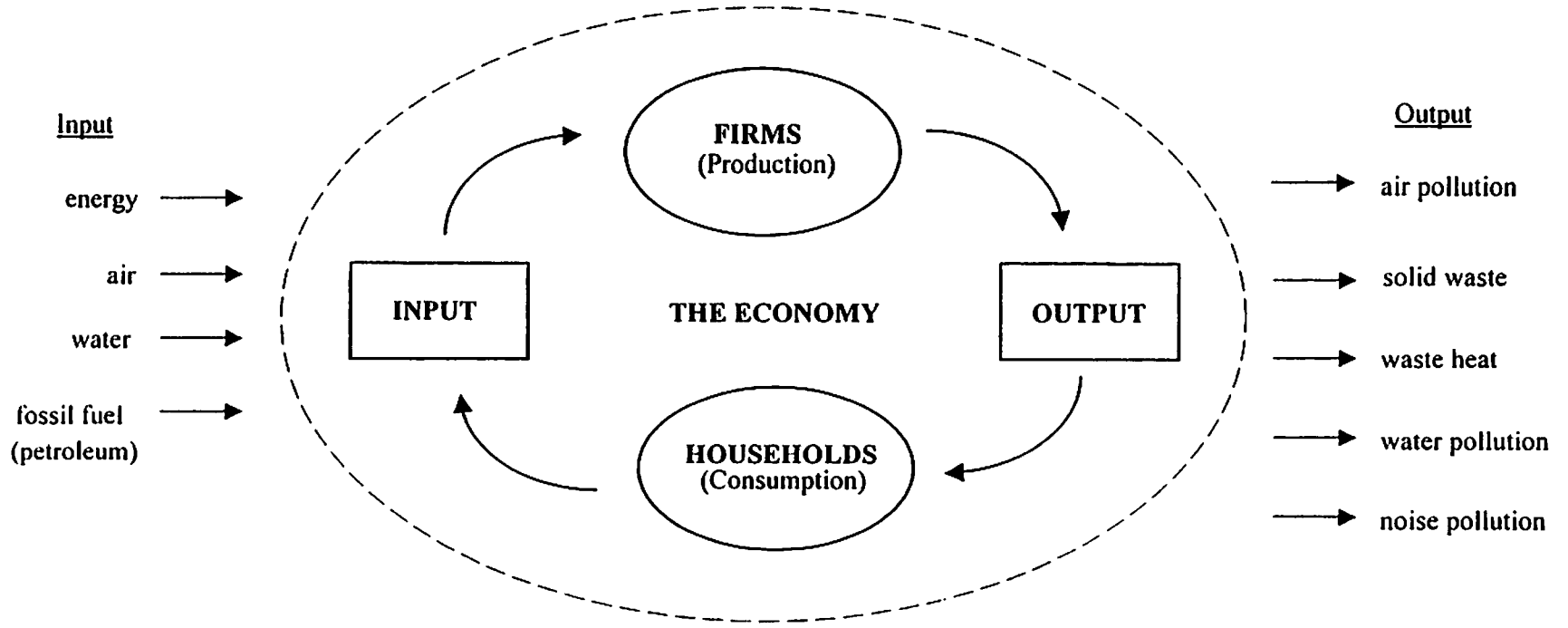


Figure 1: Economics & The Environment

c. How Does Environmental Economics Differ from Traditional Economics?

Nature of Market	Traditional Market	Environmental Market
<u>Goods Demanded</u>	Apples, Cars, TV, Books	Clean air, Clean water, Ocean fisheries, Healthy forests, Diverse Species (bio-diversity)
<u>Market Structure</u>	Well-defined	Not well-defined
<u>Differences</u> 1. Nature of Good (Private vs Public)	Private Good *-rivalry & excludability properties*	Public Good *- can be enjoyed ("consumed" by many individuals simultaneously without affecting individual consumption [non-rivalry & non-excludability]* [1]
2. Time (Static vs Intertemporal)	Static *- Most efficient way of producing goods from economic theory.* *e.g. To produce TVs, theory dictates the most efficient mix of electricity, labour, and plastic to produce TVs at the lowest cost. Producing an additional TV today does not prevent the production of a TV tomorrow*	Intertemporal *- Consumption decisions of some environmental goods may be irreversible and may have profound effects on general well-being of future generations* *- Burn fossil fuels today, create environmental problems for the future.* *- Extinction occurs (whales, turtles).* *- Generation of nuclear wastes (which retain radioactivity for hundreds of years)*

[1] rivalry – if I pay and consume, less for others to consume
excludability – if I pay, I have rights to the good

II. THEORY & FUNDAMENTALS OF NATURAL RESOURCE ECONOMICS[2]

To appreciate the process of how and why individuals make economic decisions that impact the environment and natural resources, it is necessary to place some sort of value on the various options (or actions) available. The use of criterions for judging the desirability of the various options is herewith discussed.

i. Economic Efficiency

a. Static Efficiency – an allocation of resources is said to satisfy the static efficiency criterion (or merely efficiency) if the net benefit from the use of the resource is maximized by that allocation. That is, an allocation is deemed to be efficient when:

$$\text{Max: Net Benefits} = \text{Total Benefits} - \text{Total Costs}$$

$$\text{Max: } NB(Q) = TB(Q) - TC(Q) \qquad (1)$$

a.1 Demand & Willingness To Pay (WTP)

Using the concept of Willingness to Pay (WTP), measurement of total benefits (TB) can be derived from the demand curve of a resource. Given an individual demand curve, dd_i:

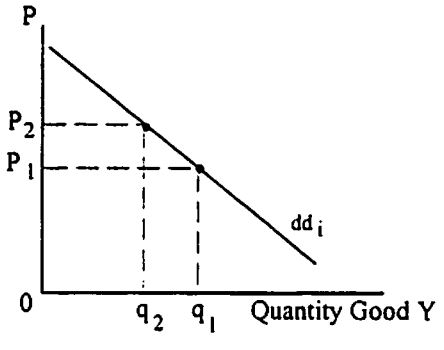


Figure 2: The Individual Demand Curve

In Figure 2, the individual demand curve is constructed by asking the question, "At a price of $P, how much of good Y would you buy?" The answer is then recorded as a point on the demand curve. By repeating the question many times for different prices, a locus of points can be traced. Connecting these points yields an *individual demand curve, dd_i.*

Adding up all the individual amounts demanded by all individuals by some stipulated price yields one point on the market demand curve. Connecting the points for various prices reveals the *market demand curve.*

---

[2] Refer readings: Tietenberg, T. Environmental and Natural Resource Economics. Harper Collins Publishers, 3rd Edition, New York. (1992): Chapter 2.

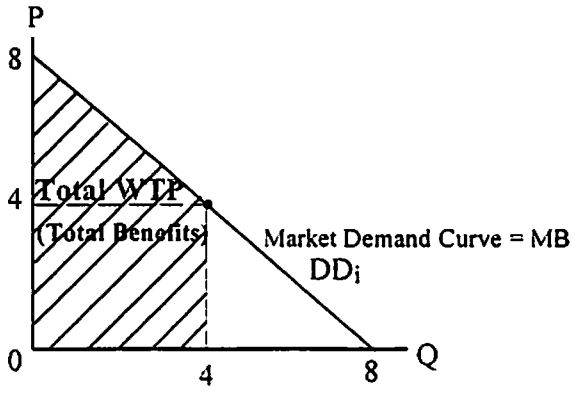


Figure 3: The Relationship of Demand to Willingness To Pay

In Figure 3, the corresponding point on the market demand curve represents the amount of money a consumer is willing to pay for the last unit of the good, or the marginal benefit (MB) that the consumer derives from each unit of good used. The total *willingness to pay* (WTP) for, say, 4 units is the sum of the willingness to pay for each of the 4 units -- measured by the sum of the WTP for the first, second, third, & fourth units, respectively.

The total WTP is the area under the continuous market demand curve to the left of the allocation in questioned. Given that total WTP is the concept used to define total benefits, it is, therefore, equal to the area under the market demand curve from the origin to the allocation of interest (shaded region).

a.2 Total & Marginal Cost

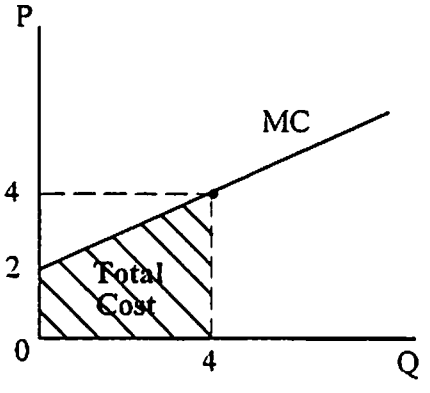


Figure 4: Measuring Total Costs

In Figure 4, total costs are measured as opportunity costs, i.e. the net benefit forgone because resources providing the service can no longer be used in their next most beneficial use. *The marginal opportunity cost* curve defines the additional cost of producing the last unit. In purely competitive markets, the marginal opportunity cost curve is identical to the supply curve.

Total (variable) cost is simply the sum of the marginal costs. Here, the total cost of producing 4 units is equal to the cost of producing the first plus second plus third plus the fourth unit. As with total WTP, the sum of the individual elements of a continuous marginal cost curve is the area under the marginal cost curve (shaded).

a.3 Net Benefits

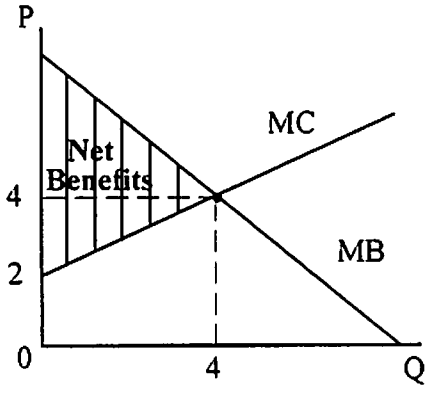


Figure 5: Measuring Net Benefits

In Figure 5, which combines the information in Figures 3 & 4, equilibrium is achieved when marginal benefit = marginal cost of production, $P^* = 4$ & $Q^* = 4$.

Since net benefit is defined as the excess of benefits over costs, it follows that net benefit is equal to the portion of the area under the demand curve which lies above the supply curve (shaded region)[3].

Mathematically,

From Eq. (1), Max: $NB(Q) = TB(Q) - TC(Q)$

To maximize NB, take the derivative ($\partial NB/\partial Q$) of net benefits and set equal to zero.

$$\partial NB/\partial Q = \partial TB/\partial Q - \partial TC/\partial Q = 0$$

Since $\partial TB/\partial Q - \partial TC/\partial Q = 0$,

then $\partial TB/\partial Q = \partial TC/\partial Q$.

Given that: $\partial TB/\partial Q$ = marginal benefits (MB) & $\partial TC/\partial Q$ = marginal costs (MC), thus net benefits are maximized when marginal benefits and marginal costs are equal.

Hence, $MB = MC$ (2)

When condition (2) is met, **net benefits are maximized and efficiency is achieved.**

[3] This criterion is derived from the **Pareto Optimality** concept. Under this concept, allocations are said to be Pareto optimal if no rearrangement of that allocation could benefit some people without affecting at least one other person. Efficient allocations are Pareto optimal. Since net benefits are maximized by an efficient allocation, it is not possible to increase the net benefit by rearranging the allocation. Subsequently, inefficient allocations are judged inferior because they do not maximize the net benefit. By failing to maximize net benefit, they are forgoing an opportunity to make some people better off without harming others.

b. Dynamic Efficiency

b.1 Choices Over Time

Since environmental decisions today can influence the quality of the environment and the stock of a natural resource far into the future, dynamic considerations have to be taken into account when valuing efficiency.

For example, exhaustible energy resources, once used, are gone. Biological renewable resources (fisheries or forests) can be overharvested. So the question is, how can we make choices when the benefits and costs may occur at different points in time?

Thus, dynamic efficiency is a generalization of the static efficiency criterion to address choices when benefits and costs may occur at different points in time. This criterion provides a way to compare the net benefit received in one period with the net benefit received in another.

b.2 Present Value and the Time Value of Money

The concept that incorporates timing into comparing net benefits received in one period with the net benefits received in another period is called **Present Value.**

However, let us first discuss another concept referred to as **Future Value.** Amounts in the present will grow through some future dates. Given an initial amount of US$1.00 today invested at 10% interest yields US$1.10 a year from now – the return of the US$1.00 principal plus US$0.10 interest.

Taking compound interest into consideration, the Future Value (FV) is computed as:

$$FV(X) = X(1 + r)^n$$

where, X = value of initial investment, r = interest rate, and n = number of years.

Example 1: What could your US$1.00 earn in one year at 10% interest?

Using $FV(X) = X(1 + r)^n$

$FV(1) = 1.00(1 + 0.1)^1$

$FV(1) = 1.1$

Example 2: What could your US$100.00 earn in ten years at 10% interest?

Using $FV(X) = X(1 + r)^n$

$FV(100) = 100.00(1 + 0.10)^{10}$

$FV(100) = 259.00$

Alternatively, the **Present Value** of US$1.10 received one year from now is, therefore, US$1.00, because given US$1.00 now, you can turn it into US$1.10 a year from now by investing it at 10% interest.

The general formula for the present value of a one-time net benefit $\{NB_n\}$ to be received n years from today when the rate of interest is r is:

$$PV[NB_n] = \frac{NB_n}{(1+r)^n} \qquad (3)$$

while, the present value of a stream of net benefits $\{NB_0, \ldots, NB_n\}$ received over a period of n years is computed as:

$$PV[NB_0,\ldots,NB_n] = \sum_{i=0}^{n} \frac{NB_i}{(1+r)^i} \qquad (4)$$

where, r is the appropriate interest rate, and NB_0 is the amount of net benefits received immediately. This process is called *discounting*.

Example 3: What is the present or discounted value of US$1.40 received three years from now at 12% interest?

Using $PV(X) = B_n / (1+r)^n$

$PV(1.40) = 1.40 / (1+0.12)^3 = \underline{1.00}$

Example 4: What is the present value of US$1000 payable in ten years, if the discount rate is 8%?

Using $PV(X) = B_n / (1+r)^n$

$PV(1000) = 1000 / (1+0.08)^{10} = \underline{463}$

Example 5: What is the present value of the following pattern of net benefits on the last day of each of the next five years: US$3,000, US$5,000, US$6,000, US$10,000, and US$12,000, given that interest rate is 6% for those five years?

Using $PV[NB_0,\ldots,NB_n] = \sum_{i=0}^{n} \frac{NB_i}{(1+r)^i}$

$$PV = \frac{3{,}000}{(1+0.06)^1} + \frac{5{,}000}{(1+0.06)^2} + \frac{6{,}000}{(1+0.06)^3} + \frac{10{,}000}{(1+0.06)^4} + \frac{12{,}000}{(1+0.06)^5}$$

$PV = 2830.2 + 4449.9 + 5037.7 + 7920.9 + 8967.1 = \underline{US\$29{,}205.8}$

This means that if US$29,205.8 is put into a savings account earning 6% interest and US$3,000, US$5,000, US$6,000, US$10,000, and US$12,000 is withdrawn from that account on the last day of each of the next five years, the balance in the account at the last day would be zero. Thus, one should be indifferent to receiving US$29,205.8 now, or in the specific five-year stream of benefits totalling US$36,000; given one, you can get the other. This method translates everything back to its current worth.

Exercise 1:

In the context of environmental and resource economics, determine the desirability of a project to upgrade a sewage treatment plant in Malaysia to protect area beaches and other marine resources. Assume that in each year of the project (lifetime), the rate of interest to be used is 10% while the benefits and costs are as follows:

Year	Benefits	Costs
2004	.....(Present).....	
2005	US$200	US$550
2006	US$250	US$200
2007	US$400	US$100

Hint: In order to determine if the project is a good idea, one must take the PV of the benefits and costs in each year and then determine if the PV of the benefits minus the present value of the costs (present value of the whole time stream of net benefits) is positive or negative.

Training Workshop on Natural Resource Economics & Research Methods
Conducted by the Policy Research & Impact Assessment Program
WorldFish Center
(23 Feb - 03 Mar 2004)

Natural Resource Economics
(Lecture 2)

c. Sources of Inefficiency (Market Failure)

When markets fail to allocate resources in an efficient manner, the resulting inefficiency is referred to as **market failure**. In general, five categories of market failure exists:

(1) externalities,
(2) public goods,
(3) imperfect competition,
(4) imperfect information, and
(5) inappropriate government intervention / government failure.

c.1 Externalities

• *Externalities* exist when the price system fail to allocate resources efficiently. This occurs when " ... the welfare of some agent, either a firm or household, depends on his or her activities and on activities under the control of some other agent as well"[1].

• A negative externality or external diseconomy arises when another person's actions affect me adversely. [Example: I am concentrating in class when two other students start to talk loudly. Their actions reduces my concentration in class].

• A positive externality or external economy arises when other people's actions affect me positively. [Example: My neighbour's attractive garden help me sell my house at a premium].

• Both the above types of externalities involve no exchange through a market. In the case of my neighbour, he/she is not paid anything for the benefits others receive from the garden. In the class example, those students do not help me understand the materials better while distracting me.

• Examples in environmental natural resources include air and water pollution – cases where individuals and firms discard waste products into the environment without acknowledging the damages these products cause to others.

[1] Callan & Thomas, Chapter 3.

• Under basic economic competitive market assumptions, supply represents the marginal costs of production while demand represents the marginal benefits of consumption, both based on private or internal decision making. Thus:

Supply (S): $P = 10.0 + 0.075Q$ (Marginal Costs: MC)
Demand (D): $P = 42.0 - 0.125Q$ (Marginal Benefits: MB)

• In the absence of externalities, private costs are equal to social costs (since zero social costs are present). Thus, we can formally refer to the supply function as the marginal private cost (MPC) of production while the demand relationship is referred to as the marginal private benefit (MPB) function. Thus:

MPC: $P = 10.0 + 0.075Q$
MPB: $P = 42.0 - 0.125Q$

• In Figure 1, the competitive equilibrium is achieved at P_c & Q_c.

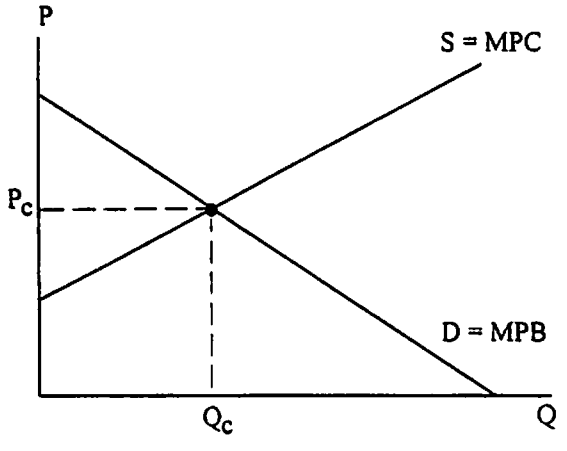


Figure 1: Competitive Equilibrium: Supply (Marginal Private Costs) and Demand (Marginal Private Benefits) Curves

• However, when externalities are present, a polluting firm incurs **private costs** (consisting of land, labor, and capital costs for the producers, which are costs to both the producer and society as a whole), as well as **social costs** (which include the above mentioned private costs plus additional external costs consisting of the damage generated by industrial pollution to society) through its normal operation of the price system.

• Hence, if we were to model the hypothetical marginal external cost (MEC) function as MEC = 0.05Q, the MEC must be added to the firm's MPC to derive the marginal social cost (MSC) equation as shown below:

$$\begin{aligned} MSC &= (MPC) + (MEC) \\ &= (10 + 0.075Q) + (0.05Q) \\ MSC &= 10 + 0.125Q \end{aligned}$$

c.2 Public Goods

• *Public goods* are defined as those that exhibit both consumption indivisibilities (non-rivalry) and non-excludability.

• **Non-rivalry** refers to a circumstance where one person's consumption of the resource does not diminish the amount available for others (zero marginal cost).

• **Non-excludability** refers to a circumstance where once the resource is provided, even those who cannot afford to pay for it cannot be excluded from enjoying the benefits that it provides.

• The following spectrum in Figure 3 describes the level of "pureness" of public and private goods. Goods exhibiting the highest degree of non-rivalry and non-excludability properties are described as "**pure public goods**", while goods that exhibit the lowest degree of non-rivalry and non-excludability properties are described as "**pure private goods**".

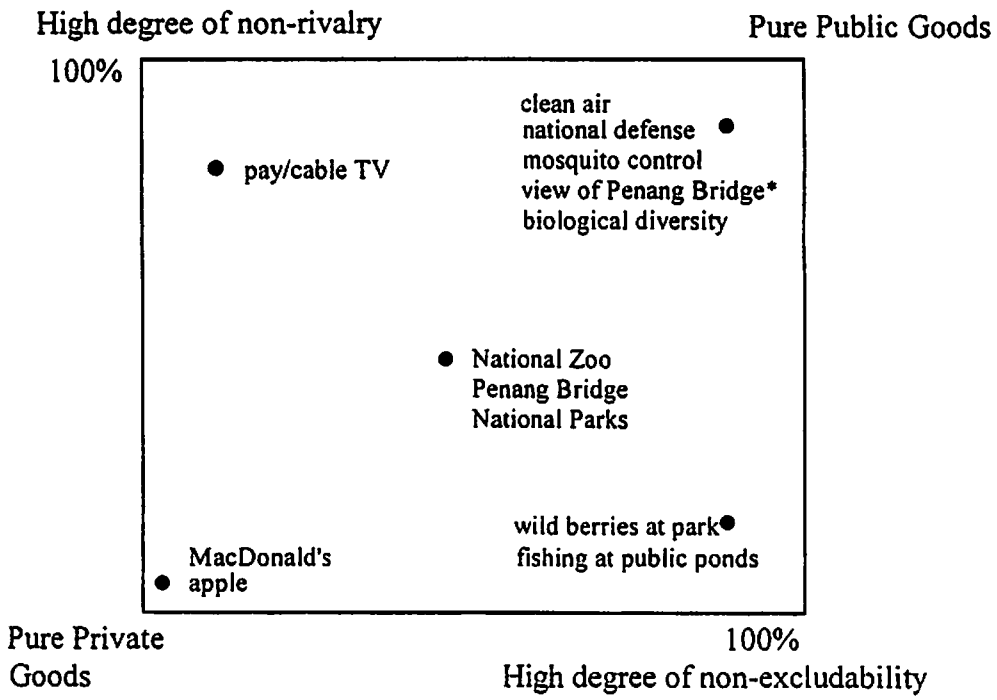


Figure 3: The Spectrum of Public & Private Goods[3]

[3] Adapted from Kahn, J. R. Chapter 2.

• Market failure exists because the market fails to provide an efficient allocation of resources due to the existence of (public) goods that can be provided to users at zero marginal cost and must be provided on a non-exclusive basis to everyone.

• Given the non-excludability nature of a public good (where once the resource is provided, even those who cannot pay for it cannot be excluded from enjoying the benefits that it confers), consumers may hide their true willingness to pay (or even underpay).

• Unlike private goods whereby the market demand curve is found by summing up individuals' demands horizontally, individual demand curves for public goods are added *vertically*. This is because once such a commodity if provided, it is available at the same quantity to all consumers (non-rivalry characteristic). Each consumer should thus express a unique "willingness-to-pay" for the public good based on the benefits each expects to derive from consumption.

Quant. Demanded	Consumer 1's WTP $p_1 = 10 - 0.1Q_d$	Consumer 2's WTP $p_2 = 15 - 0.2Q_d$	**Combined Demand / WTP for Consumer 1 & 2** $p_1 + p_2 = 25 - 0.3Q_d$
0	10.00	15.00	25.00
5	9.50	14.00	23.50
10	9.00	13.00	22.00
15	8.50	12.00	20.50
20	8.00	11.00	19.00
25	7.50	10.00	17.50
30	7.00	9.00	16.00

• Figure 4 considers a situation with two individuals (A & B) and their demand curves for a public good (D_A & D_B). Because individual A values the public good more, his demand curve is always above that for individual B, i.e. individual A is willing to pay more for any given quantity than individual B.

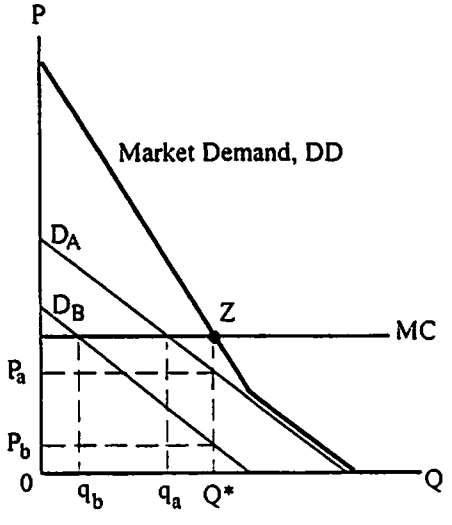


Figure 4: Public Goods & Inefficiency

• The market demand curve (DD) is represented by the vertical summation of the individual demand curves because everyone can simultaneously consume the same amount of public good (non-rivalry characteristic) and also producing one more unit of the public good would benefit both people (non-excludability characteristic).

• Assuming a constant marginal cost (MC) or supply (SS) curve, market equilibrium is met at point Z [where MC (SS) = DD]. Since net benefits for the public good are represented by the portion of the area under the market demand curve which lies above the MC curve, Q* would be the socially optimal level of output that maximizes net benefits.

• However, given the inherent characteristics of public goods (non-rivalry & non-excludability), the free-rider problem – where consumers receive the benefits of any public good purchased by others – prevents the provision of any optimal level of output (Q*) to be realized.

• Consider the demand curves below: Individual B would be willing to purchase q_b amount of the public good since this is the amount that would maximize his individual net benefits. When individual A comes along, he notices that individual B has already purchased q_b, and would therefore only need to purchase $(q_a - q_b)$ to maximize his own individual net benefits (given that q_b had already been purchased). As such, the efficient level of output (Q*) can never be realized since total collection from individuals A & B would be sufficient to defray the cost of q_a units only.

• The importance of public goods to the study of natural and environmental resources stems from numerous types of environmental pollution (air, water), destruction of bio-diversity, management of fish and wildlife populations, and heritage conservation.

c.3 Imperfect Competition

• *Imperfect competition* occurs when either the buyer (monopsony) or seller (monopoly) in the market is able to exercise an inordinate amount of power over the price or output.

• Consider a market where only one seller exists (monopoly). In Figure 5, the monopolist maximizes its profits by equating marginal revenue and marginal costs of production (MR = MC at point Z). Equilibrium output and price levels are Q_M & P_M respectively.

• In contrast, output and price in a competitive market are Q_C & P_C respectively.

• Thus, the monopolist output (Q_M) is lower than the efficient output level under perfect competition (Q_C); the monopolist also charges a higher price (P_M) compared to the competitive price (P_C).

• For output between Q_M and Q_C, the benefits of the good are more than the costs of the good. This excess benefit (depicted by area QYZ) represents the costs to society of having this lower than optimal level of output due to actions taken by the monopolist.

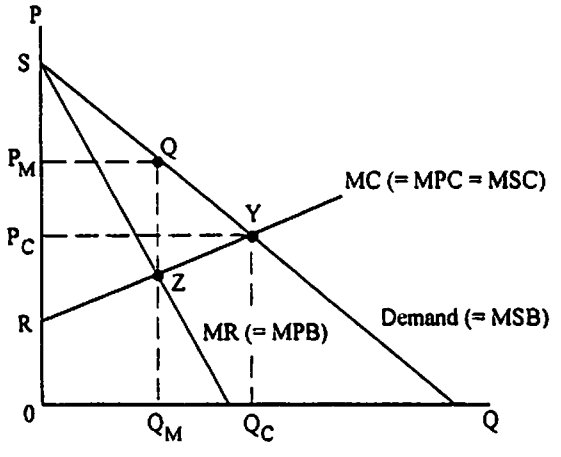


Figure 5: Imperfect Competition (Monopoly) & Inefficiency

• Under competitive market conditions, demand represents the marginal social benefit (MSB), while supply represents the marginal private and social costs of production (MPC = MSC) since no externality is present in this case.

• However, the monopolist's marginal revenue function (MR) represents its marginal private benefits to the firm. Hence, its MR = MPB.

• Thus, the monopolist do not maximize net social benefits by equating marginal social benefits with marginal social costs (MSB = MSC at point Y) but equates marginal private cost to marginal private benefit (MPB = MPC at point Z) at output Q_M instead. This output level is lower than the socially optimal level of Q_C.

• The importance of imperfect competition to the study of natural and environmental resources is related to many extractive industries such as OPEC, which forms a cartel resulting in higher than normal prices and lower than normal production. Other examples include regulated monopolies such as electric power and natural gas industries.

c.4 Imperfect Information (Uncertainty)[4]

• *Imperfect information* means that some segment of the market – consumers or producers or both – does not know the true costs or benefits associated with the good or activity. Under this scenario, one would not expect the forces of supply and demand to equate marginal social benefits with marginal social costs (i.e. MSB ≠ MSC).

• Instances in environmental and natural resource economics where imperfect information occurs include destruction of rainforests (bio-diversity), and the hazards of using chemicals in homes (pesticides, solvents etc.).

• However, one must be careful to distinguish between imperfect information involving a public good and an externality, and imperfect information involving a private good. Destruction of rainforests and bio-diversity involve public goods and externalities. There is an inherent market failure in the indiscriminate destruction of rainforests based on comparisons of private costs and benefits, not social costs and benefits. There is also imperfect information in that the exact science of bio-diversity is not yet fully understood. In such cases, imperfect information does not cause market failure per se. However, it makes it more difficult to develop public policy dealing with the particular market failure.

• On the other hand, the hazards and dangers of using chemicals in homes (pesticides, solvents, styrofoam) are not an externality; rather, it is a by-product of mankind's activities. Market failure occurs when people do not understand the true health consequences of using these chemicals and do not take proper mitigative measures (abstainment, reduction in use, using alternative products, recycle). If people understood the dangerous health consequences of these products, there would be an adjustment of prices and corrective measures to generate the optimal level of use on these products in the long run.

c.5 Inappropriate Government Intervention / Government Failure

• Market failure occurs when governments interfere with the free market and causes a disparity between private and social values. In such cases, the government intervenes in the economy not to correct a divergence between private and social costs, but for some other purpose, thereby causing a divergence between private & social costs.

• *Government failure* exists when political rather than economic institutions causes environmental problems and creates inefficiency. Special interest groups (e.g. logging, tobacco, petroleum companies and even landowners) use the political process to engage in *rent seeking* to secure protective legislation. Successful rent seeking might increase the net benefits going to that special interest group, but it will also frequently decrease net benefits to society[5], thus contributing to an inefficient outcome.

---

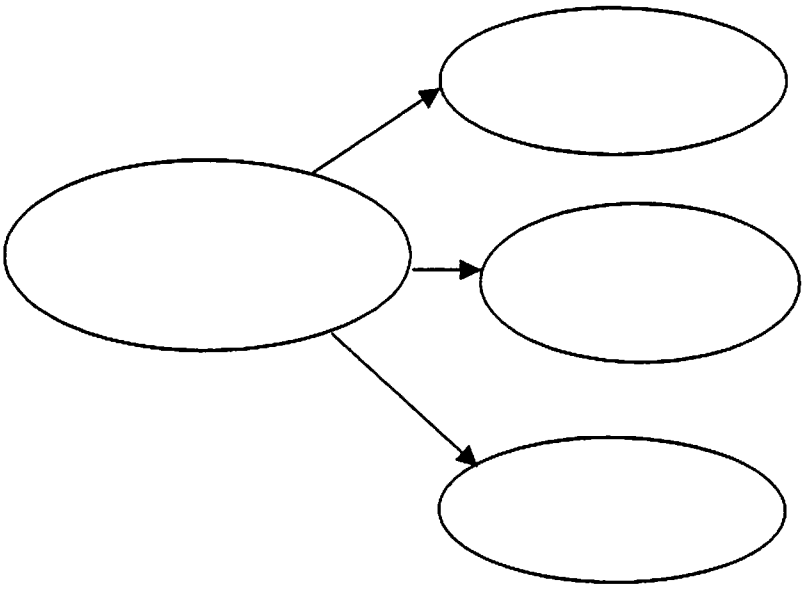
[4] Refer readings to Kahn, J. R. Chapter 2.
[5] For additional readings, refer Tietenburg, Chapter 3.

Training Workshop on Natural Resource Economics & Research Methods
Conducted by the Policy Research & Impact Assessment Program
WorldFish Center
(23 Feb - 03 Mar 2004)

Natural Resource Economics
(Lecture 3)

ii. Policies to Address Efficiency Goals

Environmental problems surface when resource allocations are either inefficient or expected to leave future generations worse off than the current generation. Various forms of policies are seen as corrective measures to improve social welfare[1]. These measures include: (a) decentralized policies, (b) command-and-control strategies, and (c) economic incentive-based strategies.



[1] While government intervention is discussed in this section, it is by no means the ultimate form of corrective measure in combatting inefficient resource allocations. Government intervention could sometimes be the cause of the problem (via distorting taxes, subsidies, regulations, restraint of trade or even special interests) and less rather than more government might be the most direct solution.

a. Decentralized Policies[2]

Decentralized policies refer to those that allow individuals/parties involved to work it out themselves. Decentralized approaches have the common advantage that parties directly involved are the ones who may know the most about damages and abatement costs, and therefore can presumably find the right balance between them.

These decentralized policies include: (1) liability laws, (2) property rights, and (3) moral suasion.

a.1 Liability Laws – involves the court of law with a defendant (accused polluter) against a plaintiff (party claiming damage). By law, if an individual agent inflicts a damage to another party, liability rules allow the damage costs to be attributed to the agent who caused the damage.

These laws will tend to bring private and social costs into line, while also introducing an incentive to force would-be polluters to make careful decisions. Knowing that they will be held liable for environment damages in effect internalizes what would otherwise be ignored external costs or effects.

Advantages

- Efficient level of pollution could be reached as both parties (polluters & those damaged) weigh the relative compensation and abatement costs via laws.
- Useful when polluter behavior is extremely difficult to monitor (e.g. oil spills).
- Useful when relatively few people are involved, causal linkages are clear, and damages are easy to measure.

Disadvantages

- Burden of proof need to be met and it is extremely difficult to proof causality given current scientific knowledge (e.g. cigarettes cause cancer?)
- Unlikely to work in large-scale (when large numbers of people are involved), technically complicated environmental problems of contemporary societies.
- Transactions costs may be high (e.g. legal costs, gathering evidence, presenting case, challenging opponents etc.)
- Long time lags

[2] Refer readings to Field, B. C. Environmental Economics. McGraw-Hill, Inc. New York, (1994): Chapter 10.

a.2 Property Rights – refers to "a bundle of entitlements defining the owner's rights, privileges, and limitations for use of the resource"[3]. An efficient (well-defined) property rights structure has these main characteristics:

- **Universality**: All resources are privately owned, and all entitlements are completely specified.
- **Exclusivity**: All benefits and costs accrued as a result of owning and using the resources should accrue to the owner, and only to the owner, either directly or indirectly by sale to others.
- **Transferability**: All property rights should be transferable from one owner to another in a voluntary exchange.
- **Enforceability**: Property rights should be secure from involuntary seizure or encroachment from others.

Advantages

- When property rights are established, the public-good environmental quality can be transformed into a private good and efficient allocations are possible since owners and others who would like to use environmental assets can negotiate agreements that balance the relative costs of different alternatives.

- Individuals involved in decentralized bargaining may know more about the relative values involved – abatement costs and damages – than anyone else, so efficiency outcomes may be expected.

- There is no need for central bureaucratic organization to make decisions, which may turn out to be counter productive or political in nature.

Disadvantages

- The main characteristics of well-defined property rights structure may be difficult to adhere to, thus breakdowns may occur.

- High transaction costs may occur given large and complex cases of environmental degradation where free-rider problems abound.

[3] Tietenburg, Chapter 3.

a.3 Moral Suasion – refers to government programs of persuasion that appeal to a person's sense of moral values or civic duty, in order for him/her to voluntarily refrain from taking part in environmental degrading activities.

Advantages

- May be effective if people (household, firm or organization) who are being asked to change their behavior believe it is in their individual or collective interests to do so (e.g. Say NO to Smoking!).

- Preferred method when violaters are normally scattered throughout a population in a way that makes it impractical to monitor them and detect violations as they occur (e.g. Singapore's cleanliness programs).

- May have widespread positive spillover effects as citizens are brought to feel a greater sense of civic virtue (e.g. recycling programs).

Disadvantages

- Not all people are equally responsible from an ethical standpoint as some people will respond to moral arguments, while others will not. The burden of this policy falls on the part of the population that is morally more sensitive; those who respond less to moral arguments will be free riding on the others, enjoying the benefits of others' moral restraint but escaping their rightful share of the burden.

- In the long run, moral free riding may tend to erode the general level of civic and moral responsibility. Although perhaps effective in the short run, it could actually have the opposite effect in the long run.

b. Command-and-Control Strategies[4]

Command-and-control strategies or direct controls refer to enforcement machineries (e.g. courts, police, fines) used on individuals to bring about socially desirable behavior.

This approach relies on various types of standards in order to improve environmental quality. In general, a standard is simply a mandated level of performance enforced by law. The spirit of a standard is, if you want society not to do something (e.g. pollution), pass a law prohibiting it, and the authorities would be entrusted to enforce the law.

There are three main types of standards: (1) emission, (2) technology, and (3) emergency restrictions. Others include ambient, performance and health standards.

b.1 Emission (Performance-Based) Standards – refer to never-exceed levels applied directly to the quantities of emissions coming from pollution sources. These standards limit the discharges of specified pollutants.

Emission standards are a type of performance standard because they refer to end results meant to be achieved by polluters who are regulated. However, these standards do not stipulate the technology to be used to achieve the desired pollution limit, thus allowing polluters to choose themselves each individual method to meet emissions limit.

Examples are requirements that farmers reduce their use of a particular pesticide below some level; upper limits on the quantity of residuals flow per minute; and, emission concentration of parts per million of biochemical oxygen demand (BOD) in wastewater.

b.2 Technology Standards – also termed as "design standards" or "engineering standards", these types of standards require potential polluters to adopt certain specified equipment/technology/techniques/practices for pollution abatement.

It actually dictates to polluters certain operating practices or particular equipment to be used in the precise same way.

Examples would include a requirement that all cars use unleaded gasoline; all motorcycles are equipped with catalytic converters; and the prohibition of all CFC related equipment within a specified period.

---

[4] Field, Chapter 11.

b.3 Emergency Restrictions – these direct controls are used as temporary bans on specific polluting activities to reduce emission levels during emergency conditions.

These direct controls, used during events that occur in random or unpredictable fashion, would be favorable over taxes, which may not provide enough flexibility to adequately deal with the new set of circumstances.

For example, automobiles with either odd or even numbers are not allowed to travel during high pollution days in Mexico & China. In Malaysia, when the air pollution index reaches a certain level, open burning is banned.

Advantages

- Standards are popular because its targets are clearly specified, simple, and direct. Easy to implement, commonly used, and highly effective. Authorities enact specific standards, injunctions or other immediate prohibitive treatment once a problem is identified. Failure to comply would result in punitive reparation. So a certain degree of certainty that pollution will be reduced is guaranteed when monitoring is difficult.

- Standards also agree with our ethical sense that pollution is bad and ought to be declared illegal. No-nonsense instrument that deny polluters a license to pollute.

- Appropriate when externality imposes a high-localized level of social cost, such as hazardous waste and that these costs can only be reduced by a reduction in the externality generating activities.

- Standards do not require tax collections and may therefore have lower administrative costs as compared to emission taxes.

- Appropriate during presence of unpredictable emergency events.

Disadvantages

- Standards assume perfect information and zero transaction costs exists. This is not always the case in the real world.

- Standards get set through some sort of political/administrative process that may be affected by all kinds of regulatory distortions (e.g. vested interest, lag time).

- Standards require polluters to adopt the same technological solutions, regardless of whether such technologies are the most efficient available, regardless of whether polluters have equal costs of reducing pollution, and regardless of whether costs of reductions exceed the benefits.

- Uniform emission restrictions may not allocate production between plants optimally as it fails to equate marginal social costs across different plants.

- A technology standard places the burden on the government to make the correct technology decisions. As polluters are not charged for their emissions (polluters do not pay the full marginal social cost of their output), there exists no incentives or rewards to finding better approaches to reduce emissions.

- There is also an incentive not to look for new control technologies because firms fear that they will be forced to adopt any new technology that works, regardless of cost.

- Enforcement or monitoring may be a problem since public enforcement agencies always work under limited budget.

c. Economic Incentive-Based Strategies

As opposed to defining certain behaviors as legal or illegal and specifying penalties for engaging in illegal behavior, economic incentive-based strategies rely on making individual self-interest coincide with the social interest.

There are four main types of economic incentive-based strategies: (1) emission taxes, (2) subsidies, (3) marketable/transferable discharge permits, and (4) deposit-refund systems.

c.1 Emission Taxes – polluters are required to pay a pollution charge (or tax) according to the level of their emissions[5]. Since externality-generating polluters do not produce at the level where marginal private social costs are equal to marginal social costs, a pollution charge is necessary to induce the necessary optimality conditions to hold.

The increased cost due to the pollution charge causes an optimal reduction in total industry emissions. The essence of the tax approach is to provide an incentive for polluters to find the best way to reduce emissions, rather than having a central authority to determine how it should be done.

Advantages

- The imposition of pollution charges induces polluters to invest in R&D for pollution abatement alternatives since the polluter still has to pay charges for the optimal amount of pollution (as long as the costs of emissions controls are less than the tax).

- By leaving polluters free to determine how best to reduce emissions, we call into play their own energy and creativity and their desire to minimize costs, in order to find the least cost way of reducing emissions.

- Provides a source of revenue for the government while correcting the distorted pattern of resource use that results when environmental resources are used as free inputs.

---

[5] This is in contrast to the effects on an output tax.

Disadvantages

- Cost-minimizing industries view such charges (taxes) as an added burden to their production plans, and hence, will typically resist the enactment of such programs.

- Pollution charges and taxes suffer from administrative and bureaucratic inefficiencies. There exists a lag time necessary for these policies to take effect. These reforms are also often very difficult to implement in a short period of time and not practical during periods of emergency.

- Will not have as much certainty of achieving the target level of pollution when compared to command and control policies since policy makers seldom have perfect information on the responses of polluters to the taxes.

- Errors in tax rate might have carry-over effects on other parts of the economy.

c.2 Subsidies – given the reasoning that if a system of taxes can reduce pollution, a mirror-image set of subsidy payments can also achieve the same results. The subsidy acts as a reward for reducing emissions; while also acting as an opportunity cost, because when a polluter chooses to emit a unit of effluent, it is in effect forgoing the subsidy payment it could have had if it had chosen not to produce that unit of effluent instead.

Subsidies for environmental cleanup fall into two categories: (i) those that subsidize the cost to firms of installing necessary pollution control equipment, and (ii) those that pay firms directly for the amount of pollution they produce or don't produce.

Advantages

- Theoretically should work given the same incentives as an emissions tax. Essentially subsidizing a negative externality.

Disadvantages

- Subsidies transfer resources toward polluters since they are receiving a payback rather than making a payment. This might not be politically acceptable as politicians would be reluctant to support a system that pay polluters.

- Subsidies for pollution control rarely produce desired results and can, in fact, cause more environmental problems by reducing the costs of production of goods having external costs.

- Emissions per firm may go down but the number of firms in the industry, and therefore total emissions, may increase.

- Strategic behavior might result as firms increase their initial amount of pollution before the system goes into place in order to obtain a larger subsidy.

c.3 Marketable/Transferable Discharge Permits[6] – a system where property rights are defined for environmental resources and then offered for sale to the highest bidder.

Under this system, the environmental authority specifies the acceptable level of pollutant emissions (ceiling) and issues a specific number of permits for this amount. These initial permits are allocated freely, sold at a price, or auctioned by environmental authorities to perspective polluters.

Polluters may then trade these permits among themselves. Market forces provide all necessary trade information and dictate a market-clearing price for those permits. This results in cost-minimizing.

Advantages

- Politically more acceptable for people to agree on a pollution-control policy that begins by distributing valuable new property rights than by notifying people they will be subject to a new emissions tax.
- Because the marketable/transferable discharge permits system responds to market demand and supply forces, the system is independent of bureaucratic decision-makings once the permits are issued. The administrative costs are not excessive because trades are made in observed markets.
- The environmental authority can directly limit pollution to their target level by restricting the quantity of permits distributed.
- Similar to pollution tax, this system induces polluters to invest in R&D for pollution abatement alternatives because by not reducing their emissions, firms are forgoing the increased revenues they could have obtained by selling some of their permits. However, compared to a tax, this system is politically more acceptable.
- There is an incentive for sources to police/monitor themselves, at least informally, since firms who are cheating are essentially lowering the market price of permits.

Disadvantages

- When market power exists among polluting plants, certain plants may form a cartel and attempt to manipulate the market-clearing price of the permits. This diminishes the cost-effectiveness of the system.
- This system takes the burden of pollution control out of the hands of engineers and places it under the operation of a market.
- Problems such as who get the permits at the beginning, the strength of their incentives to minimize costs, the transaction rules set by the administering public agency, and the ability to monitor and enforce compliance may surface.

---

[6] For an indepth discussion on this topic, refer to Baumol, W. J. and Oates, W. E. The Theory of Environmental Policy. Cambridge University Press, New York, (1993): Chapter12.

c.4 Deposit-Refund Systems (DRS) – this strategy combines the characteristics of a tax and a subsidy in order to promote actions by individuals or firms that otherwise would not have been undertaken.

By imposing a deposit (tax) in the amount of the potential damage during the initial stage of the transaction, individuals or firms are charged for the rights to inflict negative effects on the environment.

A subsidy (refund) is then offered when certain conditions are met or when those rights are not exercised. This system is well suited to situations where a product is widely dispersed when purchased and used, and where disposal is difficult or impossible for authorities to monitor.

For example, in the USA, a number of states have enacted DRS for beverage containers, both to reduce litter and to encourage recycling. Others include waste lubricating oil, batteries, and automobile tires.

Advantages

- DRS allow the polluter to look for the least-cost manner to dispose hazardous waste, as the deposit becomes a fee if the polluter does not comply satisfactorily.
- The onus is on polluters to prove their compliance to the authorities, thus saving on monitoring, record keeping, and other informational gathering costs.
- Even if illegal or improper disposal of hazardous waste were to persist, the government may clean up the environment themselves since the actual or expected damages is already covered by actual payments in the form of the deposit (tax).
- Few regulatory distortions would exist among the various government agencies as bureaucratic fragmentation is minimized.

Disadvantages

- If the deposit part of the DRS is badly designed or subject to low-cost abuses, incentives may arise for "pirate-producers" to produce the refundable items specifically for the purposes of being turned in to collect the refund.

Training Workshop on Natural Resource Economics & Research Methods
Conducted by the Policy Research & Impact Assessment Program
WorldFish Center
(23 Feb - 03 Mar 2004)

Natural Resource Economics
(Lecture 4)

iii. Valuing the Environment for Environmental Decision Making

a. Economic Inpact Analysis[1]

- Economic Impact Analysis requires a basic understanding of how economies function, and how their various parts fit together.

- Basically, it centers on how some action – a new law, a new technological breakthrough, a new source of import – will affect an economic system. Some level of econometrics may be required at some point.

- Selected examples include: Studying ...

  * the impact of environmental regulations on growth rates and unemployment,

  * the effects of an environmental law on the growth of the pollution-control industry,

  * the response of the beef industry to new packaging regulations,

  * the impact of a wetlands law on the rate of population growth and tax base in the community,

  * how efforts to control CO_2 emissions might impact the relative growth rates of rich and poor countries,

  * the impact of water quality on citrus production,

  * the impact of Turtle Excluder Devices (TEDs) on the shrimp production,

  * season closures on shrimp,

  * how a new residential development on a filled wetland affects local and regional economy.

[1] Field, pg. 110-111.

b. Cost Effectiveness Analysis: Fundamental Approach[2]

- The basic type of cost-effectiveness analysis takes an objective as given, then costs out various alternative ways of attaining that alternative.

- For example, suppose the government is planning to supply fresh water to a remote village in East Malaysia. Assuming several possibilities exists:

(1) It could drill new wells into the ground,

(2) It could build a connector to the water supply system of a neighboring town,

(3) It could build a small reservoir.

- A cost-effectiveness analysis would estimate the costs of these different alternatives by showing how they compared in terms of, say, the costs per million gallons of delivered water into the village system.

- This method can be considered as half of a benefit-cost analysis where costs, but not benefits, are estimated in monetary terms.

- It may make good sense to do a fundamental cost-effectiveness analysis even before there is a strong commitment to the objective being costed. In many cases, the value of a given objective may not exactly be known.

- Once a cost-effectiveness analysis is done, they may be able to tell, at least in relative terms, whether any of the different alternatives would be desirable. They may be able to say something like:

> *"We don't know exactly how much benefits are in monetary terms, but we feel that they are more than the costs of several of the alternatives that have been costed out, so we will go ahead with one or both of them."*

[2] Field, pg. 111-113.

c. Risk Analysis[3]

- Risk analysis involves a detailed examination of (i) risk assessment, (ii) risk evaluation, and (iii) risk management alternatives.

- This type of analysis is performed to understand the nature of unwanted, negative consequences to human life, health, property, or the environment.

- This analytical process provides information regarding undesirable events by quantifying the probabilities and expected consequences for identified risks.

- For example, a risk assessment study can be carried out to determine the acceptable levels of risk to the residents living in a community near a hazardous waste dump. This involves calculating the risks that the hazardous waste might end up contaminating the surrounding area and studying the likely impacts on the health of community residents if exposed to the waste material[4].

- Risk management then involves considering different policy options for managing the waste dump after the levels and significance of the risk have been established. This could call upon comparative risk analysis where the investigation of different policy options and the levels of risk each would produce.

- Other forms of risk analysis include calculating the value of a statistical life[5]. For example, suppose statistical analysis indicates that on the average, workers accept an annual wage premium of US$500 in return for working in jobs where the annual likelihood of a fatal accident is an additional 1 person for every 10,000 full time equivalent workers (all other work characteristics are held constant).

- If we assume that those workers are fully informed and the labor market is competitive, then we can expect the following equations to hold:

wage premium = (value of statistical life) x (increased probability of death)

value of statistical life = (wage premium) ÷ (increased probability of death)

500 1/10,000

which, in our example above, yields a value of US$5 million for the value of a statistical life.

---

[3] Field, pg. 129-139.
[4] This process might involve engineers, hydrologists, biologists and other health experts.
[5] Hackett, pg. 109.

d. Measuring Non-Market Environmental Goods

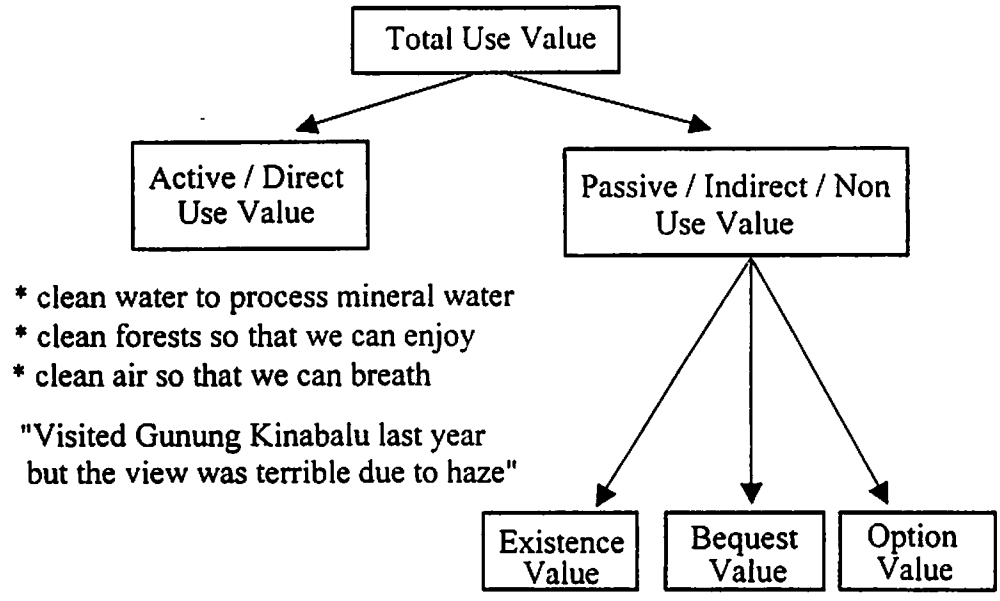


Figure 3: The Various Components of Total Use Value

Existence Value - "Never intend to visit Gunung Kinabalu but glad that the views there are majestic"
(just knowing it is there, utility increases; though might be scared of heights)

Bequest Value - "I want my children and grandchildren to be able to enjoy and see Gunung Kinabalu in all its splendour"

Option Value - "May visit the Gunung Kinabalu next year. Want a guarantee that the view will be great"
(might exercise the option to visit, might not)

Total Use Value: defined as the total of use value and non-use value of resources. It is what a good or resource is worth, as reflected in market transactions or, indirectly, through its use.

Direct Use Value: this direct value affects the individual him/herself. E.g. We would directly enjoy the benefits of better recreational facilities or better health effects through clean air.

Indirect Use Value: the value that will affect the individual indirectly. E.g. If we have storm protection, we will indirectly benefit when the storm comes.

Existence Value: the value of knowing that the resource is there to be enjoyed for personal fulfillment. E.g. Knowing that the Malayan Tiger is there and would not be extinct.

Bequest Value: the value of preserving the resource for future (next-of-kin) use. Analogous to options value but this is for the use by future generations or next-of-kin. E.g. If we preserve the habitats of the Malayan Tiger, our children can enjoy/benefit by our safeguarding the tiger.

Options Value: the value of preserving the resource for (personal) future use. E.g. Preserve the forest so that we may be able to use it for future research.

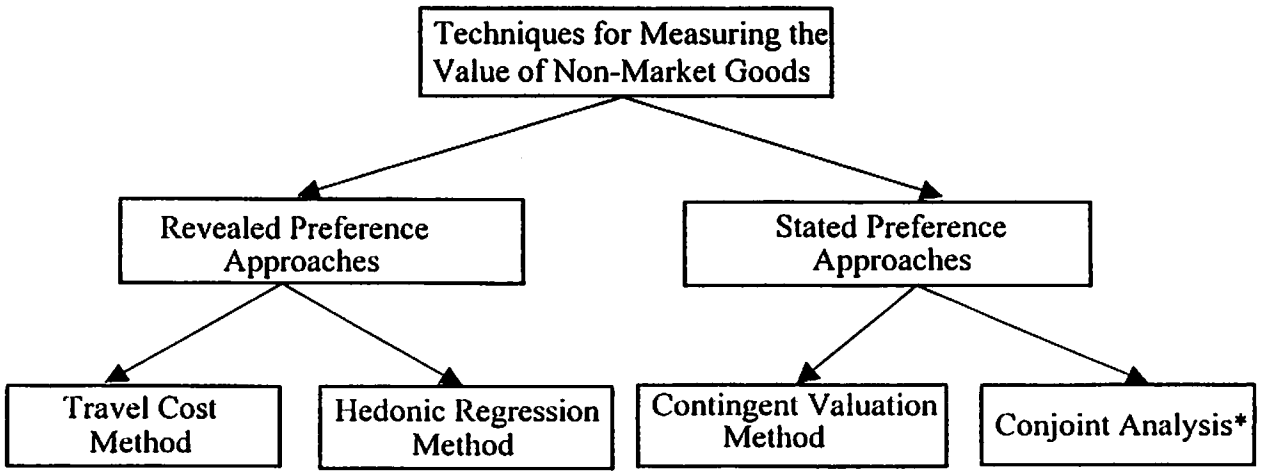


Figure 4: Valuation Techniques to Value Non-Market Goods

d.1 The Travel Cost Method (TCM)

• The TCM is used to value environmental resources associated with recreational activity. Based on this model, travel costs to a recreational site is viewed as a proxy for market prices when estimating the demand curve for a recreational site.

• When recreationists are surveyed about the number of trips they make and their travel cost to the recreational site, their preference can be revealed to estimate a demand curve based on people's willingness to incur travel costs to make the visit and the frequency of visits.

• Travel expenditures give an idea of the minimum cost that people are willing to pay for access to environmental amenities. For higher costs, it is expected that the number of trips made would be low. Conversely, if costs are low, it is expected that the number of trips made would be high.

• TCM offers a way of acquiring information on the lower-bound active-use value that people assign to an area. Data are based on actual expenditures rather than hypothetical surveys but cannot measure passive uses where people do not travel to an area.

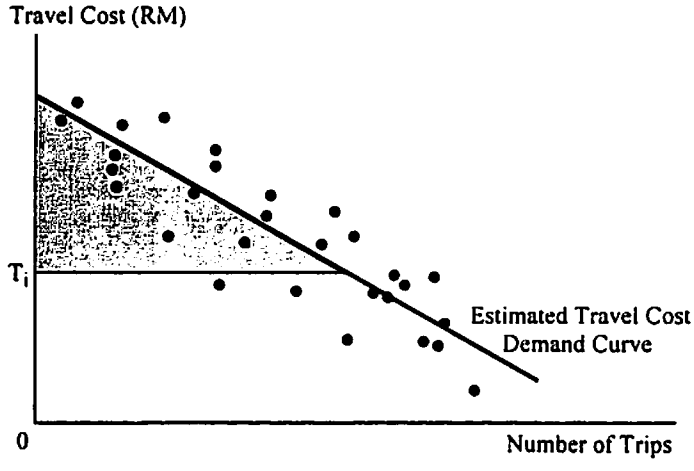


Figure 5: Estimation of Travel Cost Demand Curve

• Figure 5 contains the results of such a survey where observations on travel cost and number of trips are collected and each point represent the combination of travel costs and number of trips reported by an individual surveyed. The fitted line through those points represents an individual demand curve.

• The value of the site to an individual can then be estimated by: (i) computing the consumer surplus[6] for each individual in the survey, (ii) averaging the consumers' surplus, and then (iii) multiplying by the estimated number of recreationists.

[6] Consumer Surplus is the area under the demand curve and above the price (the travel cost) that the individual actually incurred. For example, if TC_i was the travel cost paid by the individual, the individual's consumer surplus would be equal to the shaded area in Figure 3.

• The travel cost demand curve is often expanded to include other explanatory variables such as: frequency of visit, income, family size, educational level, and other relevant socio-economic variables.

Travel Cost = f(No. of Trips, Income, Family Size, Educ. Level, Distance Travelled, ...)

• Using this method, Bell and Leeworthy[7] surveyed 826 tourists who visited the beaches in Florida. They gathered data on days spent on the beach, expenses incurred while visiting, the cost of their travel to Florida, as well as other factors that influence visiting the beach such as age, children, income, and perceived quality of such an experience. Using multiple regression analysis, they were able to isolate the impact of travel cost and number of beach visits. The average tourist spent nearly 5 days on the beach, and spent on average US$85 per day. From this information, the researchers estimated that consumer surplus for 4.7 days was on average US$179, for a daily average of US$38. With 70 million tourists annually visiting the Florida beaches, these beaches were found to yield a lower-bound estimate of US$2.7 billion in active-use value.

[7] Bell, F., and V. Leeworthy. "Recreational Demand by Tourists for Saltwater Beach Days." *Journal of Environmental Economics and Management*. 18(3), 1990: 189-205.

d.2 The Hedonic Regression Method[8] *(HRM)*

- The HRM is a method for revealing willingness-to-pay and is based on the theory of consumer behavior. This theory suggests that individuals value a good because they value the characteristics of the good rather than the good itself.

- For example, an individual would value a car not only because the car directly gives him/her utility, but because of the indirect value of the characteristics of the car such as safety, operating cost per mile, luxury, comfort, and status as well. Thus, an examination of how the price of a car varies with changes in the levels of these characteristics can reveal the prices of these characteristics.

- Another example wpuld be to examine the value of a wetland area derived from expenditures on wetland-related activities such as wildlife hunting or wildlife observation.

- Yet another typical example of HRM can be found in housing studies. Assume two housing areas with the exact same characteristics (price, size of land, quality of finishing, distance to schools etc.) but located at different ends of the city (north & south). Suppose a polluting steel factory is built at the north of the city. Now, people will prefer to live in the cleaner southern part of the city. As people move from the northern part of the city to the southern part, house prices will fall in the north (as reflected in the decrease in demand) and increase in the south (as reflected in the increase in demand). Hence, there will always be an upward pressure on southern prices and downward pressure on northern prices as long as people prefer living in the south compared to the north. This price movement will eventually cease when the price differential is large enough to make people indifferent between living in the clean south and the dirty north. Since the only difference between the two areas is the higher air pollution in the north, the price differential between the two areas reveals people's WTP to avoid the air pollution[9].

---

[8] Kahn, pg. 93-98.

[9] Other examples where observable behavior can reveal an individual's willingness-to-pay would be: (i) recreationists who are willing to travel further into the wilderness to enjoy higher quality outdoor recreational sites (cleaner beaches, cleaner air, better fishing etc.), and (ii) individuals who consider the wage that can be earned in a particular city, as well as a host of other factors including negative characteristics (such as crime, pollution, high cost of living, extreme climate, congestion etc.) and positive characteristics (such as schools, recreational opportunity, social life, sports, mild climate etc).

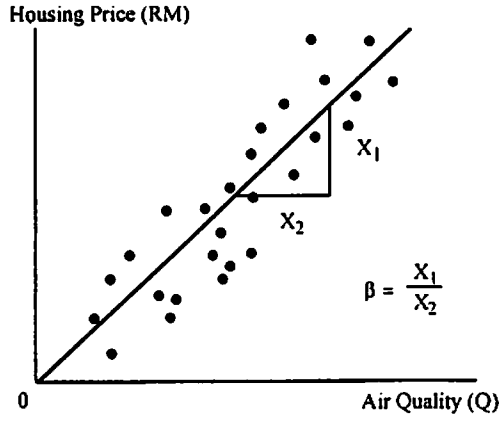


Figure 6: Estimation of Hedonic Regression Demand Curve

• In Figure 6, assume that all characteristics of houses in a neighbourhood are the same throughout a city, except air pollution (which varies with location). While each dot represents the housing price and air quality levels associated with each individual house in the city, houses in areas with higher air quality will have higher prices.

• This is evident in the general positive relationship of dots, which can be formalized into a straight line through the cluster of points using ordinary least-squares regression analysis. In this case, the line describes the fashion in which improvements in air quality lead to increases in housing prices.

• The simple regression yields: $H = \alpha + \beta Q$

where, H = housing price, Q = air quality, and α & β = intercept and slope respectively. Here, β represents how many units of H will increase for each unit increase in air quality.

• More complex (multiple) regression analysis yields the following equation to explain variations in housing prices when looking at the spectrum of housing prices in a city:

$$H = \alpha + \beta_1 Q + \beta_2 S + \beta_3 D + \ldots$$

where, H = housing price, Q = air quality, S = house size, D = Distance to Schools, and β_i = regression coefficients. Here, the estimates of β represents how many units of H will increase for each unit of increase in each of the independent variables.

• Using this method, several researchers have estimated the willingness-to-pay for reductions in air pollutants. Bloomquist et. al.[10] estimated the household willingness-to-pay per unit of air quality change against particulates to be US$360 per mg/m^3, while Brookshire et. al.'s[11] estimate of willingness-to-pay per unit of air quality change against NO_2 to be US$45.92 for poor to fair and US$59.02 for fair to good.

---

[10] Bloomquist, G.C., M.C. Berger, and J.P. Hoehn. "New Estimates of Quality of Life in Urban Areas." *American Economic Review* 78(1988): 89-107.

[11] Brookshire, D.S., M.A. Thayer, W.W. Shultze, and R.L. d'Arge. "Valuing Public Goods: A Comparison of Survey and Hedonic Approaches." *American Economic Review* 72 (1982): 165-177.

d.3 The Contingent Valuation Method (CVM)

- The CVM is the only one of the three discussed above that is capable of measuring passive-use values, such as existence and option values.

- In essence, the CVM constructs a "hypothetical market" and ascertains value through surveys by asking individual's willingness-to-pay to bring about a change in environmental quality.

- Such environmental change could be in the form of improved visibility from lessened air pollution, the protection of endangered species, or the preservation of a wilderness area.

- Respondents are asked in a variety of ways, using both open-ended questions (respondents are asked to state their maximum willingness-to-pay) and close-ended questions (respondents are asked to say whether or not they would be willing-to-pay a particular amount to preserve the environmental resource.

- Hackett[12] noted that the typical CVM works as follows:

1. Choose a particular environmental amenity that involves a public policy decision, like salmon habitat restoration in a particular river, or expansion of a wildlife refuge.
2. Perform a pretest in which you survey a small focus group and determine the "likely" range of values that people would be willing to pay (WTP) for the amenity.
3. Use these preliminary values to make up a survey instrument. The survey instrument (a) informs people of the precise nature of the proposed activity or protection; (b) asks a predetermined number of randomly selected people whether they would be willing to pay US$X in additional tax revenues (or maybe as a higher product price or as a voluntary contribution) if as a result the environmental amenity is protected in the way described; (c) repeats the question in (b) for different US$X values, where a different randomly selected respondent group is asked about each different US$X value; and (d) ask follow-up questions on why respondents answered the way they did and on explanatory demographic characteristics such as income, age, gender, education, concern for the environment, and degree of active use.
4. Analyze the data using relevant statistical techniques to estimate a demand curve (WTP function): WTP = *f*(income, age, sex, education, concern for environment, degree of active use)

- Using this method, Walsh[13] determined how much people value allocating an additional 2.6 million acres as federal wilderness in Colorado, USA. The survey was designed to examine key value areas – use, option, and existence. On average, recreation was worth US$18.50/visitor day – yielding a total of US$28 million/year while passive-use values (existence, option) totalled US$135 million/year. This translates into billions when one calculates the present value of this stream of benefits into the future. Other researchers like Loomis[14] used CVM to quantify non-marketed environmental benefits The results indicated that economic benefit to California residents of preserving Mono Lake could conservatively be estimated to be US$1.5 billion annually.

---

[12] Hackett, Steve. Environmental and Natural Resources Economics: Theory, Policy, and the Sustainable Society. M.E. Sharpe, Publishers. (1998).

[13] Walsh, R.G. et.al. *Wilderness Resource Economics: Recreational Use and Preservation Values*. Denver: American Wilderness Alliance (1982).

[14] Loomis, J. "Measuring the Economic Benefits of Removing Dams and Restoring the Elwha River: Results of a Contingent Valuation Survey." *Water Resources Research* 32 (Feb 1996): 441-447.

Training Workshop on Natural Resource Economics & Research Methods
Conducted by the Policy Research & Impact Assessment Program
WorldFish Center
(23 Feb - 03 Mar 2004)

Natural Resource Economics
(Lecture 5)

III. THE ECONOMICS OF ALLOCATING RENEWABLE & NON-RENEWABLE RESOURCES (ISSUES)

i. Renewable Common-Property Resources (Fisheries)

• Renewable resources are those which stock can be continually replenished. However, renewable resources are not perpetual as some – living populations, such as plants and animals – are also exhaustible if not managed effectively.

• The vast majority of fisheries are renewable common-property and open-access resources, with almost no private ownership (although private aquaculture/fish farming is a rapidly growing industry).

• Open-access (or free-access) resources refer to those that do not have property rights, or are accessible to any party (Table 1).

• Common-property resources, on the other hand, is an institution whereby a resource is owned in common by a well-defined group and each member of the group has well-defined rights to use the resource (hence, restriction or access can be limited to a certain group).

• For example, countries bordering the sea have declared that their ownership rights extend some 200 miles out to sea. Within the 200 miles, the area is common-property for the countries' fishermen as foreign fishing vessels are not allowed into the waters. Beyond the 200 miles, the area is open-access to any international fishing/whaling or trawling activities.

Type of Limitation	Sole Owner	Corporate Ownership	Common Property	Open Access Limited User	Open Access Unlimited User
Access limitation	One person	Corporation only	Group members only	Group members only	Anyone
Use limitation	Use limited by owner's decision	Use limited by mgmt's decision	Use limited by rules	Use unlimited	Use unlimited

Source: Lesser, J.A., D.E. Dodds, & R.O. Zerbe, Jr. Environmental Economics & Policy, Addison Wesley Longman, Inc. Massachusetts (1997): Chapter 6.

Table 1: Comparison of Ownership Structures

a. Efficient Allocation of the Catch from a Fishery[1]

The Biological Dimension

For Population Sizes	Characteristics
S_1 to S^*	population growth ↑ as population ↑
S^* to S_2	initial ↑ in population lead to ↓ in growth

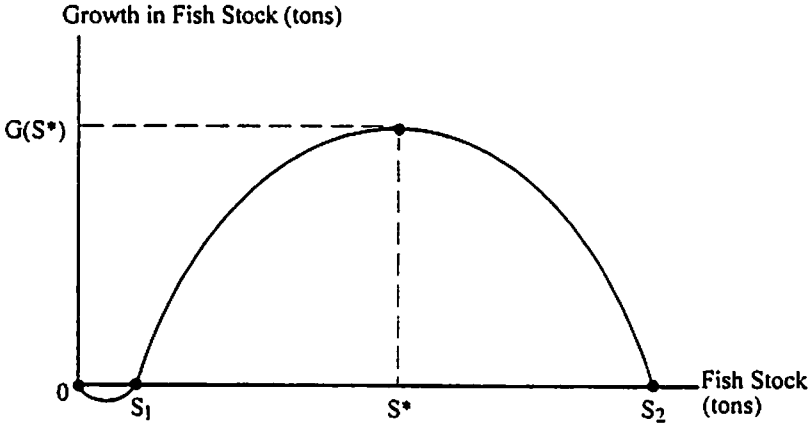


Figure 1: Relationship Between the Fish Population & Growth:
Growth Function with Critical Depensation

The Biological Dimension: Defined

S_1 is known as the *minimum viable population*. It represents the level of fish population below which growth in population is negative (deaths and out-migration exceeds births and in-migration). In contrast to S_2, this equilibrium is unstable as population to the right of S_1 lead to positive growth. When the population moves to the left of S_1, the population declines until it eventually becomes extinct. In this region, no forces act to return the fish population to a viable yield.

S_2 is known as the *natural equilibrium*. Since this is the population size which would persist in the absence of external forces. Reductions in the fish stock (due to mortality or out-migration) would be exactly offset by increases in the stock (due to births, growth of the fish in the remaining stock, and in-migration). This natural equilibrium persists because it is stable. A stable equilibrium is one in which movements away from this population level set forces in motion to restore it. For example, if the stock temporarily exceeded S_2, it would exceed the capacity of its habitat (called carrying capacity). This tendency for the population size to return to S_2 works in the other direction too. Suppose the population is temporarily reduced below S_2 (since the stock is now smaller), growth would be positive and the size of the stock would increase. Over time, the fishery would move to the right until S_2 is reached again.

S^* is known as the *maximum sustainable yield* (MSY) population or the population size which yields the maximum growth. Hence, the MSY is equal to this maximum growth and it represents the largest catch that can be perpetually sustained. If the catch is equal to the growth, the sustainable yield for any population size between S_1 and S_2 can be determined by the corresponding vertical axis.

[1] Tietenburg (Chapter 12).

b. Static Efficient Sustained Yield

• The MSY is, however, not the efficient allocation. This is because efficiency is associated with maximizing the net benefit from the use of the resource. To define the efficient allocation, we must include the costs & benefits of harvesting.

Assumptions:

1. The price of fish is constant and does not depend on the amount sold,
2. The marginal cost of a unit of fishing effort is constant,
3. The amount of fish caught per unit of effort expended is proportional to the size of fish population (the smaller the population, the fewer fish caught per unit of effort),
4. In any sustainable yield, catches, population, effort levels, and net benefits remain constant over time. As such static efficient sustainable yield allocation maximizes the constant net benefit.
5. This model represents the long run equilibrium between fishing effort and stocks.

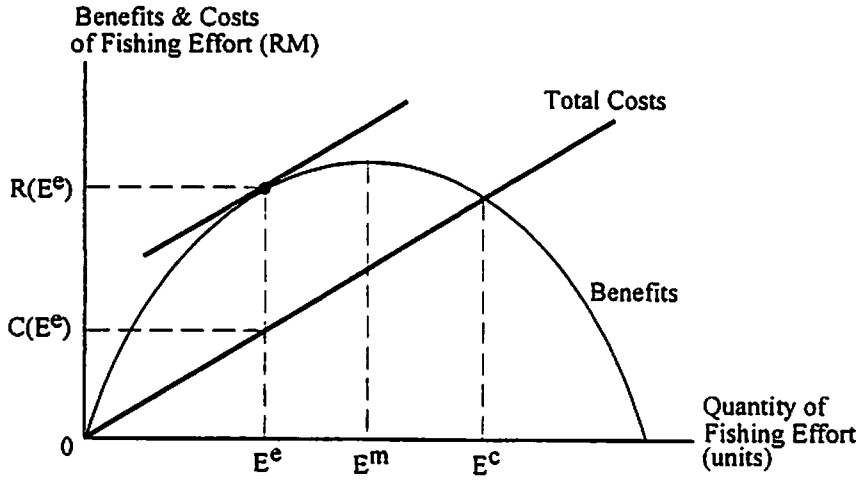


Figure 2: Efficient Sustainable Yield for a Fishery

• In Figure 2, the benefits (revenues) and costs are portrayed as a function of fishing effort (measured in vessel years or hours of fishing). The shape of the revenue function is dictated by the shape in Figure 1 since price of fish is assumed constant.

• As levels of effort are increased, eventually a point is reached (E^m) where further effort reduces the sustainable catch and revenue for all years. That point corresponds to the MSY in Figure 1 which involves identical population and growth levels.

• The net benefit is presented in the diagram as the difference (vertical distance) between benefits (price x quantity caught) and costs (the constant marginal cost of effort x units of effort expended). The efficient level of effort is E^c, the point in Figure 2 where the vertical distance between benefits and costs is maximized.

- E^e is the efficient level of effort because it is where marginal benefit (which graphically is the slope of the total benefit curve) is equal to marginal cost (the constant slope of the total cost curve).

- Levels of effort higher than E^e are inefficient because the additional cost associated with them exceeds the value of the fish obtained.

- To determine whether or not the MSY is efficient, we see that the MSY would be efficient only if the marginal cost of additional effort were zero. (Why? – What is the marginal benefit at the MSY?).

- Since this is not the case, the efficient level of effort is less than that necessary to harvest the MSY. Thus, the static efficient level of effort leads to a larger fish population than the MSY level of effort.

- To recap, suppose a technological change were to lower the marginal cost of fishing (e.g. sonar detection). This results in a rotation of the total cost curve to the right (less steep). With this new cost structure, the old level of effort is no longer efficient. The marginal cost of fishing (slope of the total cost curve) is now lower than the marginal benefit (slope of the total benefit curve). Since the marginal cost is constant, the equality of marginal cost and marginal benefit can only result from a decline in marginal benefits.

- This implies an increase in effort. The new static efficient sustainable yield equilibrium implies more effort, a lower population level, a larger catch, and a higher net benefit for the fishery.

c. How Market Allocates this Common-Property Resource

• A competitive sole owner (with well-defined property rights to the fish) would maximize profits. The owner can thus increase profits by increasing fishing effort until marginal revenue (MR) equals marginal cost (MC) at point E in Figure 3.

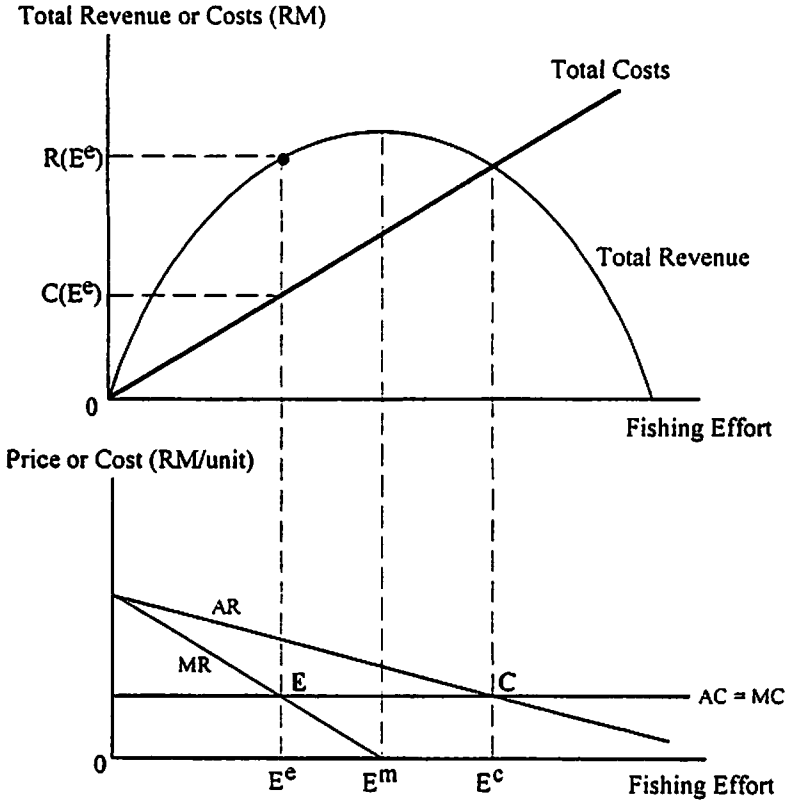


Figure 3: Market Allocation in a Common-Property Fishery: Maximizing (TR-TC)

• In Figure 3, the effort level E^c is also the static efficient sustainable produces that yields positive profits equal to the difference between $R(E^c)$ and $C(E^c)$.

• The sole owner would choose not to expend more effort than E^c because to do so would reduce the profits of the fishery, resulting in a personal loss to him. When access to the fishery is restricted, a decision to expend effort beyond E^c reduces profits to the fishery as a whole but not to that individual fisherman. Most of the decline in profits falls on other fishermen.

• In ocean fisheries, however, sole owners seldom exist as ocean fisheries are typically common-property resources – no one exercises complete control over them. Since property rights to the fishery are not conveyed to any single owner, no single fisherman can exclude others from exploiting the fishery (case of public good).

• In Figure 3, once too many fishermen have unlimited access to the same common-property resource, the property rights to the fish are no longer efficiently defined.

• As such, in a free-access resource, the individual fisherman has an incentive to expend further effort until profits are zero. In Figure 3, that point is at effort level E^c, where average benefit and average cost are equal (at point C). It is now easy to see the contemporaneous externality – too much effort is being expended to catch too few fish, and the cost is substantially higher than it would be in an efficient allocation.

• For example, many fisheries are currently plagued by these problems. In the Bering Sea and Aleutian Islands, the efficient number of motherships (used to take on and process the catch at sea, so the catch boats do not have to return to port as often) was estimated to be 9 but the current level was 140. As a result, a significant amount of net benefits were lost (US$124 million/year). Had the fishery been harvested more slowly, the same catch could have been achieved with fewer boats used closer to their capacity.

d. Public Policies in the Quest for Efficient & Sustainable Management in Fisheries

d.1 Aquaculture

• Since inefficient management of fisheries result from treating it as common, rather than private property, one obvious solution is to allow some fisheries to be privately, rather than commonly, owned. This approach works when the fish is not very mobile, confined to artificial barriers, or when they instinctively return to their place of birth to spawn.

• In such cases, the owner is encouraged to invest in the resource and undertake measures that will increase the productivity (yield) of the fishery. This movement toward controlled raising and harvesting of fish is called *aquaculture*.

d.2 Raising the Real Costs of Fishing

• If harvesting costs are increased, the Total Cost (TC) line in Figures 2 & 3 will rotate upward (anti-clockwise). Total fishing effort will be reduced, as will total fisheries harvest. One way to increase costs is to restrict the available technologies that can be used for harvesting.

• For example, banning huge drift nets increases the costs of fishing. However, such policies might be inefficient as efficiency implies not only that the catch must be at the efficient level, but it must also be extracted at the lowest possible cost. The latter condition was violated by these policies.

d.3 Individual Transferable Quotas (ITQs)

• The economic viability of ITQs are similar to the marketable pollution permits discussed earlier. ITQs operate with the following characteristics[2]:
  1. The quotas entitle the holder to catch a specified weight of a specified type of fish,
  2. The total amount of fish authorized by the quotas held by all fishermen should be equal to the efficient catch of the fishery, and
  3. The quotas should be freely transferable among fishermen.

---

[2] For a more thorough discussion, refer to Tietenburg.

d.4 Open-Access Regulations

• Regulations modify fishing behavior of participants in the fishery without directly affecting participation in the fishery. These techniques are designed to maintain fish stocks at some target level. These restrictions generally take the form of restrictions on:

1. how fish may be caught (restrict sonar & spotter planes, net size),
2. which fish may be caught (restrict the minimum size fish legal to be harvest),
3. when fish may be caught (restrict certain period during spawning season),
4. where fish may be caught (restrict to 200-mile limit),
5. amount of fish that may be caught.

ii. Storable, Renewable Resources (Forests)

• Forests represent an example of a storable, renewable resource. Typically, tree stands[3] go through three distinct growth phases – slow volume growth in the early stage, rapid growth in the middle years, and slower growth as the stand reaches full maturity.

• The owner who harvests the timber receives income from its sale but the owner who delays harvest will receive additional growth. This amount of growth depends on the part of the growth cycle the stand is in.

• The efficient time to harvest a stand of timber is when the present value of net benefits is maximized.

• This occurs when the marginal benefit from delaying the harvest one more year is equal to the marginal cost of the delay.

- *For longer-than-efficient delays, the additional costs outweigh the increased benefits (in terms of opportunity costs foregone);*

- *For earlier-than-efficient harvests, more benefits (in terms of the increased value of timber) are given up than costs saved.*

a. Maximizing the Physical Quantities of Harvested Wood[4]

• To maximize the physical amount of wood to be derived from the forest, one can use the option of letting the forest grow until it reaches its peak volume of wood and then cut it, the forest is replanted, and the process is allowed to repeat itself.

• Or the other option would be to choose the length of the harvest-replant-harvest cycle (rotation of the forest) to maximize total harvests of wood that can be achieved over time.

• Table 2 and Figures 4 & 5 describe the growth of a one-acre stand of a hypothetical tree species as a function of the age of the trees in the stand (all the trees in the stand are the same age).

[3] It is important to look at the growth of a stand of trees, rather than an individual tree. This is because a tree in isolation grows more quickly compared to one competing with other trees for sunlight, nutrients, & water.

[4] Kahn, J.R. The Economic Approach to Environmental and Natural Resources. The Dryden Press, 2nd. Edition. Florida (1997): Chapter 11.

Table 2: Total Volume of Wood, Annual Increment, & Mean Annual Increment (cubic feet of wood) on a One Acre Stand of a Hypothetical Tree Species

Age	Total Volume	Annual Increment[5]	Mean Annual Increment[6]
10	0	0.0	0.00
20	75	7.5	3.75
30	200	12.5	6.66
40	750	55.0	18.75
50	1400	65.0	28.00
60	2100	70.0	32.50
70	2550	45.0	34.30
80	2900	35.0	36.25
90	3200	30.0	35.55
100	3450	25.0	34.50
110	3650	20.0	33.18
120	3800	15.0	31.67
130	3900	10.0	30.00
140	3950	5.0	28.20
150	3975	2.5	26.60
160	3975	0.0	24.84

Source: Kahn, J.R.

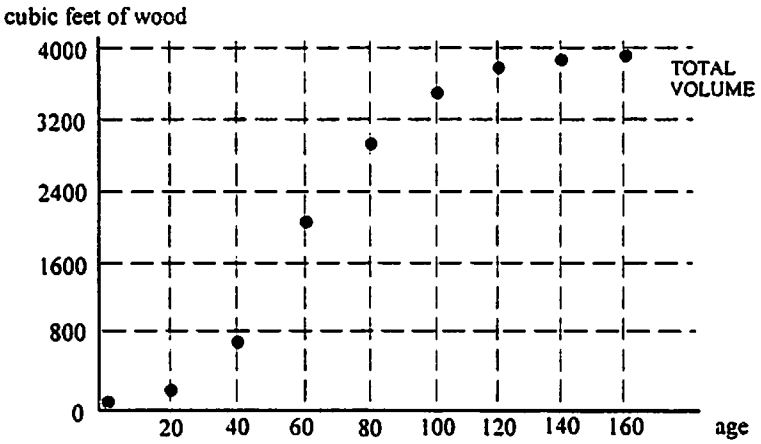


Figure 4: Total Volume of Marketable Wood of a Hypothetical Tree Species – One Acre Stand

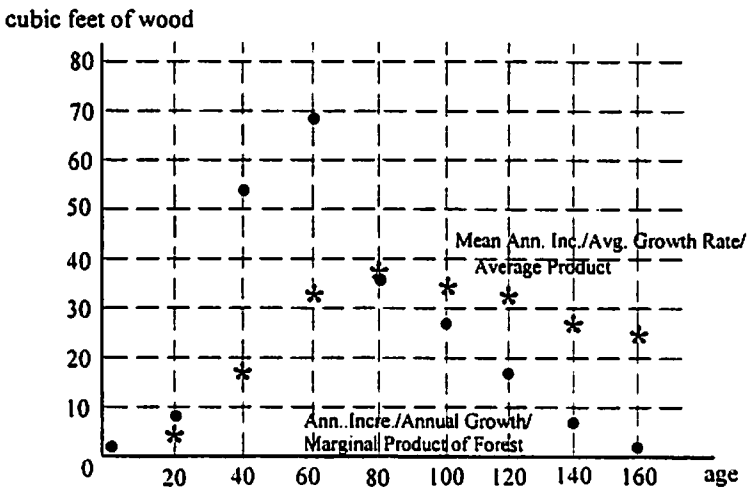


Figure 5: Annual Increment and Mean Annual Increment for a Hypothetical Tree Species - One Acre Stand

[5] Annual Growth/Marginal Product of the forest (marginal with respect to time).

[6] Annual Amount of Growth/Average Growth Rate/Average Product of the forest (the total product divided by the number of years of the rotation).

• From Figure 4, the volume of wood first increases at an increasing rate and then at a decreasing rate. Note that when the forest grows to its peak size, the growth (annual increment) declines to zero. This occurs when the stand of trees is 160 years old.

• In order to maximize the total amount of wood that is harvested over time, two trade-off alternatives exists:

Alternative 1: One way to increase the flow of wood is to harvest more frequently, but the more frequently the forest is harvested & replanted, the younger & smaller the trees will result.

Alternative 2: The other way to increase the flow of wood is to harvest less frequently and have bigger harvests.

• The optimal compromise between these two conflicting strategies is to harvest at the forest age that maximizes the average growth of the tree over its lifetime. If average growth is maximized over a sequence of multiple rotations, then total growth (and the total flow of wood) will be maximized as well.

• Notice in Figure 5 that the average growth rate (often called mean annual increment) is maximized when it is equal to the marginal growth rate (annual increment), which occurs at approximately 80 years for this hypothetical tree species. This time period is substantially shorter than the rotation that maximizes individual harvest (160 years).

• Although maximizing the physical quantities of wood has been used as a management strategy, it is an economically inefficient policy even when managers are only concerned about the benefits arising from harvesting wood.

• The reason for the inefficiency is that the costs and benefits associated with different quantity levels have not been incorporated into the management strategy. In the case of forests, one must consider the costs and benefits associated with making the rotation longer or shorter.

b. Optimal Rotation

• The choice of the optimal length of rotation is a conceptually simple problem. The forest manager simply asks,

> *"Are the benefits of making the rotation a year longer (or a year shorter) greater than the costs?"*

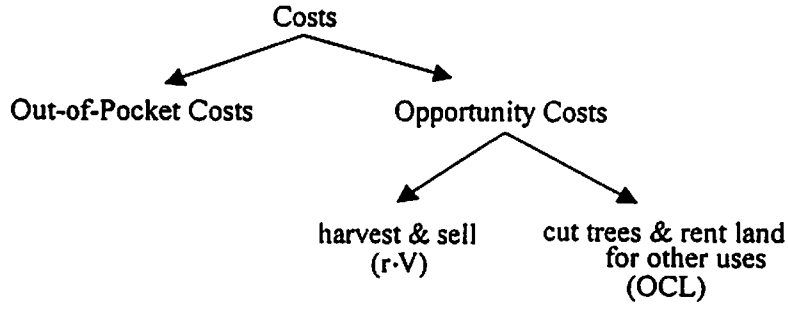
• Thus, the forest manager's should maximize the present value of the costs and benefits by deciding the optimal rotation length. The optimal rotation length is determined by comparing the costs and benefits of increasing/decreasing the rotation length.

b.1 Benefits:

• The benefits of allowing the trees to grow come from the fact that, if you wait, you have more wood to sell in the future. As one might expect, the benefits will be critically dependent upon the shape of the marginal growth (annual increment) function. This additional revenue or change in stumpage value is represented by the function $\Delta V / \Delta t$ in Figure 6, where $\Delta V / \Delta t$ = rate of change of stumpage value [V(t)] and represents the change in stumpage value as rotation length is changed.

b.2 Costs:

• The costs of letting the trees grow for another year (or a shorter period of time) can be summarized as follows:



Out-of-pocket Costs[7] – include expenses for disease prevention, thinning, pruning, fire prevention, and control of pests.

Opportunity Costs – based on foregone income from two sources: (i) the income that could be derived from earning interest on the revenues that would have been obtained had the trees been harvested instead of allowed to grow for another year (rV), and (ii) the rent that could have been obtained had the forest been cut and the land rented for other uses (OCL).

[7] To simplify analysis, the out-of-pocket costs are usually assumed to be zero, and only the opportunity costs are considered when determining the optimal rotation. An additional assumption made is that the real price of a cubic foot of wood does not change over time because if the price were to change, then there would be additional benefits (or costs) associated with increasing or decreasing the rotation.

(i) The foregone interest opportunity cost is represented by *rV(t)*, where r is the interest rate and V(t) is the stumpage value. This function will be based on the total growth function and it reaches its maximum (point S) at the time when lengthening the rotation has no impact on stumpage value ($\Delta V / \Delta t = 0$). This point is labeled R_M in Figure 6.

(ii) The next cost to be considered is the opportunity cost of the land (OCL). Since we have limited, by assumption, the use of the land to forestry, it may at first seem as if there is no opportunity cost associated with letting the trees grow longer. However, by letting the trees grow longer, not only is the interest income that could be realized from the sale of the harvest deferred, but the interest that could be earned on the proceeds to be derived from the sale of the land is deferred as well. This opportunity cost also equals the annual rental value of the land, since equilibrium in land markets implies that the rent is equal to the product of the market value of the land and the rate of interest. Notice that the maximum of this opportunity cost of the land (OCL) will occur when the rotation is at its optimal length (point Q in Figure 6). This is because the forest will be most valuable (have the highest rent) when it is optimally managed.

• Reiterating, the sum of the two opportunity costs is represented by the curve labeled rV + OCL. When these marginal opportunity costs are equal to the marginal benefits of changing rotation length ($\Delta V / \Delta t$) at point P in Figure 6:

Marginal Benefits	=	Marginal Opportunity Costs
$(\Delta V / \Delta t)$	=	rV + OCL

the present value of the whole future stream of harvests is maximized, and the rotation can be said to be optimal. This optimal rotation length is equal to R*.

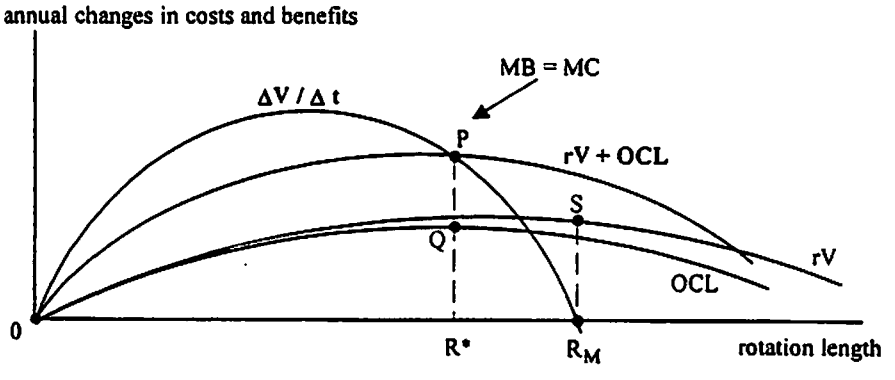


Figure 6: Determination of the Length of the Optimal Rotation

Training Workshop on Natural Resource Economics & Research Methods
Conducted by the Policy Research & Impact Assessment Program
WorldFish Center
(23 Feb - 03 Mar 2004)

Natural Resource Economics
(Lecture 6)

III. THE ECONOMICS OF ALLOCATING RENEWABLE & NON-RENEWABLE RESOURCES (ISSUES)

iii. Reproducible Private-Property Resources (Agriculture)[1]

While agriculture is impacted by environmental degradation, agriculture is also a significant source of environmental degradation as land, water, and air are typically common property resources.

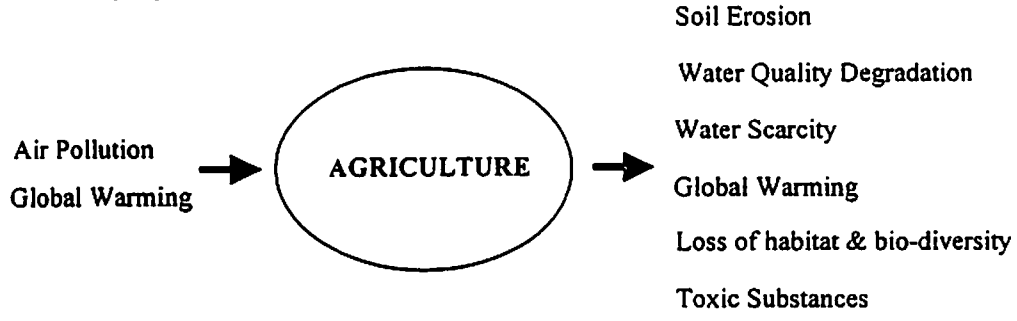


Figure 1: Relationship Between the Agricultural Sector & the Environment

a. Impacts of the Agricultural Industry on Environmental Quality[2]

- Soil Erosion

Runoff from agricultural fields → Topsoil lost → Nutrients lost → Less productive soil → Less productive crop
→ Transportation of fertilizer, pesticides, herbicides into ground & surface water
→ Landslides

- Loss of Habitat & Bio-Diversity

Agricultural land → Converted from other uses (wildlife habitat, bio-diversity, recreation, existence values)

Farmer considers private costs and benefits before converting and does not consider social costs; hence market failure (externality)

- Toxic Gases/Substances

CO_2 – fossil fuel in farm machinery; burning of forests for agriculture

Methane – agricultural sources such as manure piles; wet rice cultivation; digestive processes of ruminants such as sheep, goat, and cattle

Nitriuos Oxide Emissions - chemical & organic fertilizers

---

[1] Kahn, J.R. The Economic Approach to Environmental and Natural Resources. The Dryden Press, 2nd. Edition. Florida (1997): Chapter 16.
[2] Refer Figure 4 in above section.

b. Agriculture & Public Policy

b.1 Non-Point Source Pollution

• Agricultural pollution of surface and groundwater generally originates from non-point sources rather than from point sources[3].

• As such, it is much more difficult to monitor and measure the release of pollution by a particular polluter (the farmer). Due to this problem, it is more difficult to implement economic incentive mechanisms such as a per-unit pollution tax or marketable permits to control such pollution.

• Thus, policy incentives should focus on command-and-control techniques or economic incentives oriented toward inputs (e.g. fertilizer or pesticide taxes) rather than waste outputs.

b.2 Price-Supports

• Agriculture is also a price-supported industry. These price supports serve to raise prices above marginal social cost. Price supports can be of two forms: (i) Purchased-based, and (ii) Quantity-based price supports.

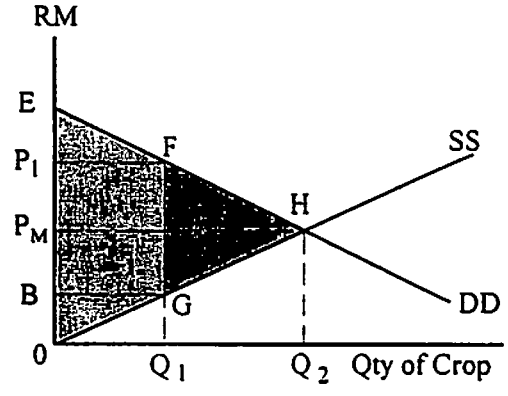


Figure 2: Welfare Losses from Quantity or Purchased-Based Price Supports
(No Pollution)

• In Figure 2, if the government wants to increase price from the market price (P_M) to P_I, it could either buy Q_2 - Q_I units of the crop (purchase-based price support), or it could pay farmers not to grow Q_2 – Q_I units of the crop (quantity-based price support).

• In the absence of any pollution or market failures, the social losses associated with the price support that raises prices from P_M to P_I are equal to the darker shaded triangle. The lighter shaded trapezoid shows the net benefits of the Q_I units of the crop that consumers purchase at the supported price P_I. In this context, there are no differences between the two types of price supports in terms of inefficiency (darker shaded triangle).

---

[3] *Point sources* refer to sources of pollution where pollution is released into the environment at a distinct location such as at the end of an effluent pipe. Example: Paper manufacturing plant polluting a river. *Non-point sources* of pollution occur when agricultural pollution flows into the environment over a large area rather than the pollutant entering at a specific location. Example: Soil nutrients, pesticides, and herbicides are carried by rainwater runoff into lakes and rivers. The runoff enters the lakes and rivers along the entire length of the interface between land and water.

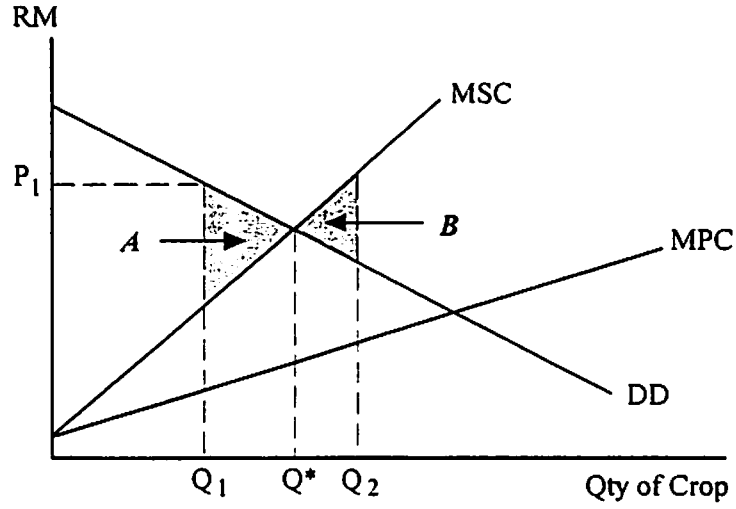


Figure 3: Welfare Losses With Pollution and Purchased-Based Price Supports

- However, the presence of pollution causes the marginal private cost (MPC) and the marginal social cost (MSC) curves to diverge (Figure 3 & Figure 4). In Figure 3, the price is supported at P_1 with a purchase-based price support system where consumers buy Q_1 units while Q_2 units are produced. The government purchases this difference.

- In this case, neither Q_1 nor Q_2 are optimal as the optimal level of output occurs at Q* [where MSC = MSB (DD)]. Since Q_2 units are produced, there are excess social costs by the amount of shaded triangle B (since MSC > MSB in this region).

- In addition, since consumers consume only Q_1 units, they lose consumer's surplus equal to the upper half of shaded triangle A while producers lose producer's surplus equal to the lower half of triangle A.

- Thus, the losses from pollution in combination with a purchased-based price support system are equal to the areas of triangles A & B.

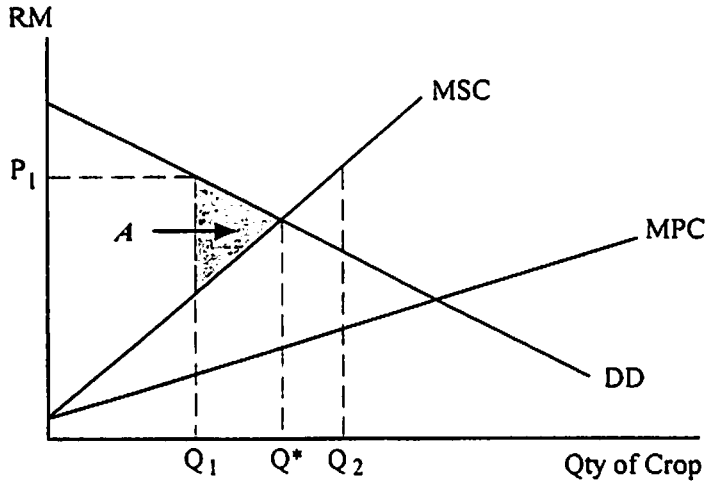


Figure 4: Welfare Losses With Pollution and Quantity-Based Price Supports

• Figure 4 shows the social losses where pollution is present and the price is supported at P_1 with a quantity-based price support system. Since the area between $Q_2 - Q_1$ is not produced (farmers are paid to take this amount out of production), there is no pollution and no extra costs associated with that non-existent production. The social loss is equal only to the area of the shaded triangle, which is the lost consumer's surplus and producer's surplus associated with consuming below the optimal level.

• The important point to this discussion is that the type of price support system employed has an effect on pollution and other environmental effects associated with agriculture. If the price supports are generated by taking land out of production, then the environmental externalities will be less severe than a price support system that relies on government purchases of crops.

c. Current Policies & New Policies Suggested

• Agricultural policies should attempt to: (i) discourage erosion, and on-site & off-site impacts of erosion, (ii) discourage excessive use of fertilizers and pesticides, (iii) increase food safety, and (iv) restore marginal farmland to natural habitat and protect existing habitat.

• Since fertilizer and pesticide runoff associated with agriculture is non-point source in nature, a per-unit tax or a system of marketable pollution permits would not be feasible as the amount of pollution per farm per unit time cannot be easily measured.

• Command-and-control (CAC) regulations can also be used to require or prohibit certain agricultural methods. For example, cultivating within a certain distance of a water body can be prohibited. CAC regulations or economic incentives can also be used to protect habitat and bio-diversity. For example, conversion of certain types of habitat can be banned, or government agencies (or NGOs) could buy critical habitat from farmers.

iv. Replenishable but Depletable Resource (Water)

a. The Efficient Allocation of Scarce Water Resource

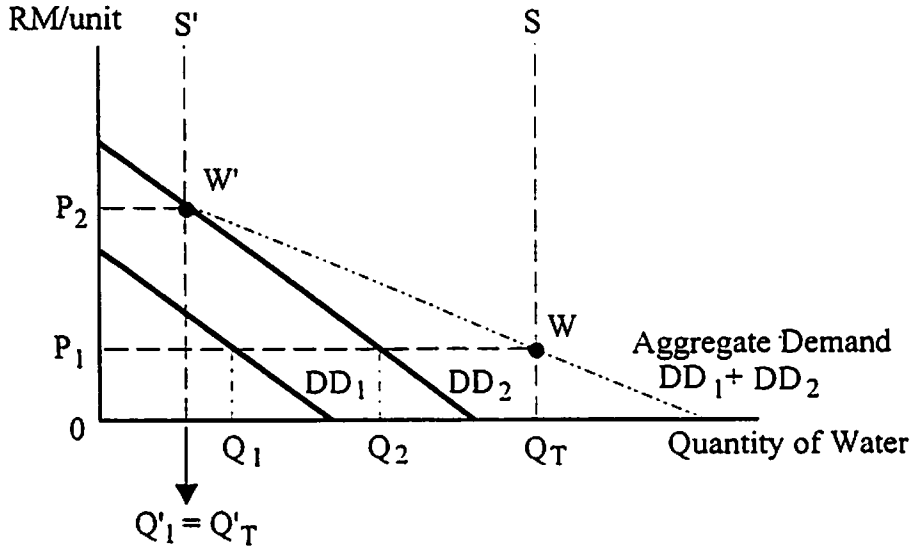


Figure 5: The Efficient Allocation of Water Supplies

• In Figure 5, two competing users of water are depicted. Each user has a demand curve for water (DD_1 & DD_2). In this two-user case, the total market demand curve or aggregate demand curve ($DD_1 + DD_2$) for water equals the horizontal sum of the individual demand curves (depicted by dashed line).

• Given the initial supply of water as S, the total quantity of water demanded is given by Q_T (point W). The efficient supplies of water to each user are Q_1 and Q_2 respectively.

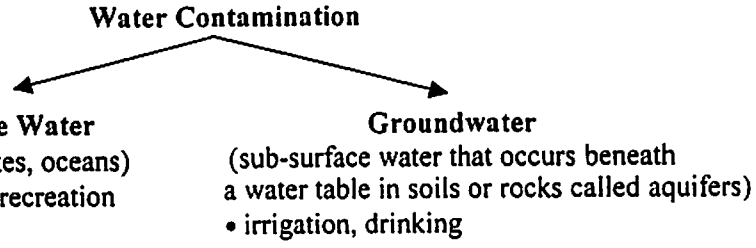
• However one of the problems with surface water supplies is that it varies from time to time, primarily due to climatic changes. If water supplies are restricted severely, it is possible for the allocation of the existing supply to be limited to one party.

• In Figure 5, suppose that water supplies are now restricted to S' gallons. S' intersects the market demand curve for water at W'. At W', the marginal benefit of water equals P_2. This is greater than user one's price for even 1 gallon. Thus the entire supply of water will be efficiently allocated to user 2.

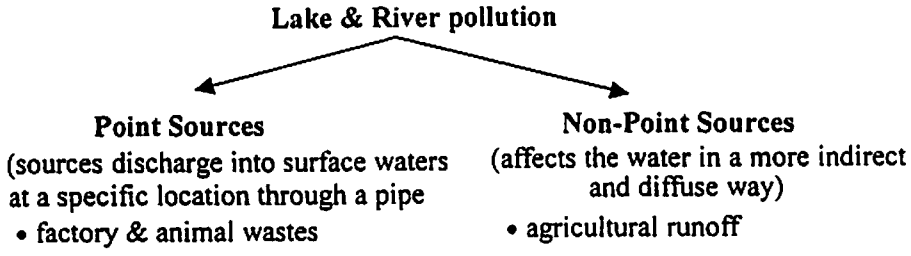
• Unfortunately, allocation of water based on economic efficiency is rare. During dry seasons, for example, user 1 would object to all the water being allocated to user 2, even though the latter group valued that water more highly. These sorts of objections occur because water is viewed as an entitlement rather than a tradable resource.

b. Nature of Water Polluting Problems

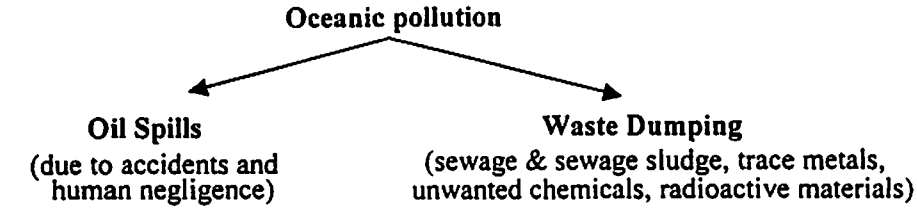
• Besides inefficiency problems associated with water allocation, other problems that arise include degrading water usage:



• For lake and river pollution, two sources of contamination exists:



• For oceanic pollution, two sources of contamination exists:



• Groundwater contamination occurs when polluting substances leach into a water-saturated region. While many potential contaminants are removed by filtration as water moves slowly through the layers of rock and soil, toxic organic chemicals may not be filtered out in the process.

• Unlike surface water, which is replenishable, once these chemicals pollute groundwater, further cleansing is no longer possible. In addition, though some groundwater is renewed by percolation of rain or melted snow, most was accumulated over geologic time and because of its location, cannot be recharged once it's depleted.

• As such, groundwater allocation must be addressed with different policies than surface water due to different replenishment rates. In general, surface water supplies are replenished far more quickly than groundwater supplies. Thus surface water can be thought of as a renewable resource, but efficient allocation of groundwater must be considered over time like other depletable resources[4].

---

[4] First, there is the opportunity cost of not having the water available for another current use, which is termed the *contemporaneous opportunity cost*. Second, because current use depletes the stock and makes it unavailable for future use, there is the opportunity cost of not having the water available for future use. This opportunity cost is termed the *intertemporal opportunity cost*. Both these contemporaneous & intertemporal opportunity costs are components of user cost and must be incorporated into price if water is to be efficiently allocated among all present & future alternative uses.

c. Water Pollution Control Policies

c.1 Command-and-Control: Technology Based Emission Standards

• A technology-based effluent standard (TBES) is an effluent standard set at the level of emissions that a source would produce if it were employing a particular type of abatement technology.

• To establish a TBES, the environmental authorities would study the effluent abatement technologies and procedures available to a particular type of industrial operation (e.g. particular types of equipment, raw materials, recycling machinery, pre-treatment processes, or effluent removing techniques). After having selected one technology from among the many available, it sets the emissions standard at the level of emissions that are produced when that technology is used by firms in the industry.

• However, problems arising from equal treatment-type programs like TBESs are:

(1) inefficiency since the same standards for one industry will be applied to all industries regardless of damages and control costs,
(2) cost-ineffectiveness since this procedure subjects the same standards to all firms within each industry. While these standards will be cost-effective if all individual plants in each industry have exactly the same marginal abatement costs, this is highly unlikely to be the case,
(3) weak enforcement and time consuming due to the enormous effort to establish consistent effluent standards for each and every individual source.

c.2 Economic Incentives System: Marketable Pollution Permits

• Emissions permits could have the following characteristics:

(1) Initial distribution of emissions rights are based on historic levels of emissions by major polluters,
(2) The rights are distributed free and they are valid only for five years; after that they must be reissued by the administrative agency,
(3) Traded emission permits also have a maximum life of five years,
(4) Permits are allowed to be traded so that firms with relatively high abatement costs could buy permits from those with low marginal abatement costs.

• The basic problem with emissions trading approaches is that they require relatively larger numbers of participants if competitive permit markets are to develop. But most water pollution problems are local, or at most regional, in nature. Thus, the potential number of traders is likely to be too small for effective trading.

• Other problems with emissions trading approaches include calculating the initial efficient amount of permits to be distributed, waivers of emission limitations on municipal discharges, bureaucratic delays to justify permit trading, and the fact that permits being reissued after a certain time period creates uncertainty over whether purchased permits will be counted in a firm's new base when reauthorization occurs.

• This perhaps implies that more emphasis should be put on the use of emissions taxes as a means of attaining cost-effective water pollution control programs.

d. Non-Point Pollution

• In contrast to the control of point sources, current laws do not address non-point pollution in a rigorous manner although in many areas, this type of pollution is a significant part of the total problem. This is due to the nature of non-point emissions being diffuse and not concentrated into specific outfalls that make them very difficult to control.

• Traditional approaches like emissions standards and emissions taxes are problematic because it is difficult to measure emissions accurately. This means that the locus of control will normally have to be pushed back directly onto the practices and technologies that typically lead to substantial non-point source runoff.

• Another method of controlling non-point source emissions is to tax those activities or materials that lead to the emissions, rather than the emissions themselves, if possible. Taxes might be out on fertilizer used by farmers, for example, or on lawn chemicals used by suburban dwellers and golf courses. The objective is to induce a reduction in the use of materials that may ultimately end up in rivers, lakes, or groundwater aquifers.

e. Oil Spills[5]

• Current water pollution policies toward oil spills relies heavily on the legal system to internalize the costs of a spill though liability law. By forcing owners of a vessel to pay for the costs of cleaning up the spill, including compensation for natural resource damages, a powerful incentive to exercise care is created.

• However, this outcome is not necessarily efficient in practice due to two problems: high administrative costs & limitations of financial liability. Since problems with high administrative costs are self-explanatory, we shall just proceed with a brief discussion on problems associated with the limitations of financial liability during oil spills.

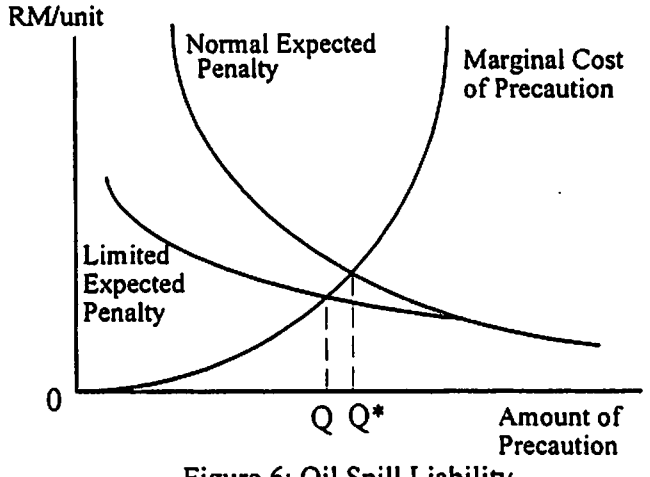


Figure 6: Oil Spill Liability

[5] Tietenburg, Chapter 17.

• In Figure 6, the amount of precaution taken by vessel owners is depicted in the horizontal axis while the costs of precaution and penalty costs are drawn as the vertical axis. As shown above, the marginal cost of precaution increases as the amount of precaution increases. The inverse relationship between penalties (normal and limited) suggests that larger amounts of precaution are presumed to yield smaller marginal reductions in the likelihood and magnitude of resulting accidents.

• Vessel owners are expected to minimize costs by choosing the level of precaution, which equates marginal cost of additional precaution with the resulting reduction in the marginal expected penalty.

• With unlimited liability, the vessel owner's cost-minimizing choice of precaution is depicted as Q* because as long as the imposed penalty equaled the actual damage and the probability of having to pay the damage, once an accident occurred, was 1, this outcome would normally be efficient. Taking precautions would simply be cheaper than paying for the cleanup.

• However, with limited liability, a different outcome arises, Under this scenario, the expected penalty function rotates downward for levels of precaution lower than that level which would produce an accident resulting in damages exactly equal to the limit.

• Lower levels of precaution imply damage which exceeds the limit, but the vessel owner would not have to pay anything above the limit[6].

• Thus, the effect of limited liability on the vessel owner's choice of precaution levels is that as long as the liability limit is binding (which appears to routinely be the case with recent spills), owners will choose too little precaution (at Q instead of Q*).

[6] The deviation in magnitude of the limited expected penalty function is greatest at lower levels of precaution; it declines to zero at that precaution level where the expected magnitude of an accident is equal to the liability limit.

Techniques of Writing Proposal and Report

Developing Research Idea
- The research journal
- The proposal

Writing the introduction of proposal: Basic elements
- Establishing the importance of your study
  - Demonstrate the concept's prevalence
  - Demonstrate the concept's relevance to real life
  - Demonstrate historical precedence
  - Stating your hypothesis

Justification of the study
- The exploratory study
  - New is not enough
  - Spell out your reasoning
  - Defend your procedures
- The direct replication
  - Document the original study's importance
  - Explain why the results might not replicate
- The systematic replication
  - Improved power
  - Improved construct validity
  - Improved external validity
- The conceptual replication
  - Using a different measure
  - Using a different manipulation
  - Using a different design

Writing the method section
- Participants
- The design section
- Apparatus
- Procedure
- Putting the method section together

Writing the results section

Writing the discussion section

Putting the front and back on
- The title and the title page
- Abstract
- References

RESEARCH DESIGN

I What is Research Design?

Research design can be thought of as the *structure* of research. We often describe a design using a concise notation that enables us to summarize a complex design structure efficiently. What are the "elements" that a design includes? They are:

Observations or Measures. These are symbolized by 'O' in design notation. An **O** can refer to a single measure (e.g., a measure of body weight). If we need to distinguish among specific measures, wecan use subscripts with the **O**, as in $\mathbf{O_1}$, $\mathbf{O_2}$, and so on.

Treatments or Programs. These are symbolized with an 'X' in design notations. The **X** can refer to an intervention or program. We can use subscripts to distinguish different programs or program variations.

Groups. Each group in a design is given its own line in the design structure. If the design notation has three lines, there are three groups in the design.

Assignment to Group. It is designated by a letter at the beginning of each line (i.e., group) that describes how the group was assigned. The major types of assignment are:

R = random assignment, **N** = nonequivalent groups, **C** = assignment by cutoff

Time. Time moves from left to right. Elements that are listed on the left occur before elements that are listed on the right.

1.1 Design Notation Examples

R O x O

R O O

Notes: Os indicate different waves of assignment, Vertical alignments of Os show that pretest and posttest are measured at the same time, R indicates the groups are randomly assigned, X is the treatment, There are two lines, one for each group

R O_1 x $O_{1,2}$

R O_1 $O_{1,2}$

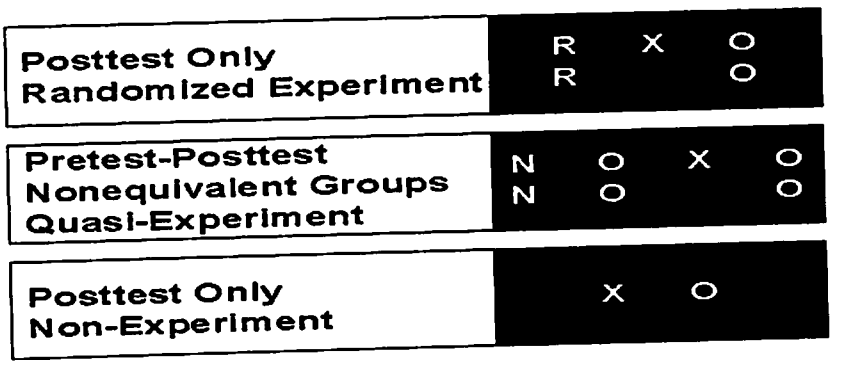
II Types of Research Design

We can classify designs into thee categories:

1. **Randomized Experiment** uses random experiment.
2. **Quasi-experimental** lacks randomized assignment, but use multiple groups or multiple waves of measurement.
3. If none of the above two are followed then we cal it as **non-experimental.**

This classification of design is useful to describe internal validity. A randomized experiment generally is the strongest of the three designs when research interest is to establish cause-effect relationship. A non-experiment is generally the weakest in this respect. In fact, the simplest form of non-experiment is a one-shot survey design that consists of nothing but a single observation O. This is probably one of the most common forms of research and, for some research questions -- especially descriptive ones -- is clearly a strong design.

To illustrate the different types of designs, consider one of each in design notation. The first design is a posttest-only randomized experiment. You can tell it's a randomized experiment because it has an R at the beginning of each line, indicating random assignment. The second design is a pre-post nonequivalent groups quasi-experiment. We know it's not a randomized experiment because random assignment wasn't used. And we know it's not a non-experiment because there are both multiple groups and multiple waves of measurement. That means it must be a quasi-experiment. We add the label "nonequivalent" because in this design we do not explicitly control the assignment and the groups may be nonequivalent or not similar to each other. Finally, we show a posttest-only non-experimental design. You might use this design if you want to study the effects of a natural disaster like a flood or tornado and you want to do so by interviewing survivors.



Before analyzing different types of research design let us discuss two crucial elements in research - validity and reliability.

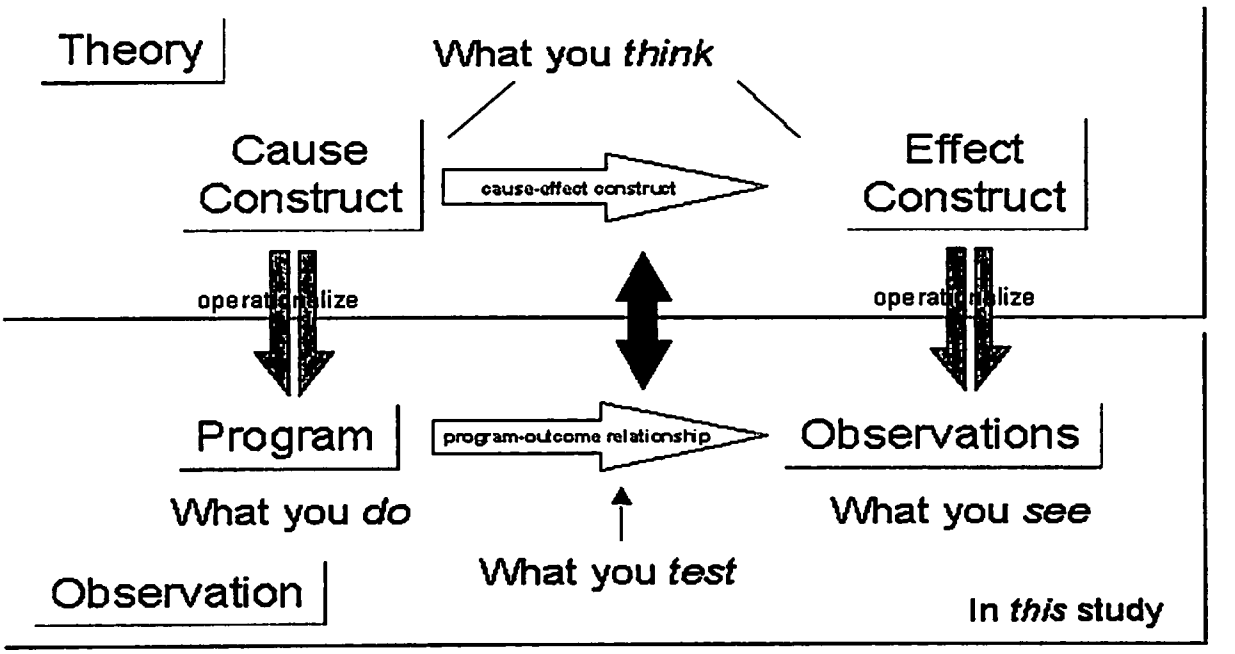
III Validity and Reliability

Validity of any conclusion based on either qualitative or quantitative research and reliability of data are critically important. Any research that is based on unreliable data cannot provide any valid conclusion. Therefore, these two concepts are interrelated.

Validity

Validity is the best available approximation to the truth of a given proposition, inference, or conclusion, which is based on reliable measures, samples and design. Technically, we should say that a measure leads to valid conclusions or that a sample enables valid inferences, and so on.

Almost every social research involves measurement or observation. And, whenever we measure or observe we are concerned with whether we are measuring what we intend to measure or with how our observations are influenced by the circumstances in which they are made. We subdivide validity into four types. Each type addresses a specific methodological question. In order to understand the types of validity, we have to know how we investigate a research question.



The figure shows that there are really two realms involved in research. The first, on the top, is the theory "what we think". The second, on the bottom, is observations. It is the real world into which we translate our ideas -- our programs, treatments, measures and observations. When we conduct research, we are continually flitting back and forth

between what we think about the world and what is happening. When we are investigating a cause-effect relationship, we have a theory of what the cause is (***the cause construct***). For instance, if we are testing a new extension program, we have an idea of what it would look like ideally. Similarly, on the effect side, we have an idea of what we are ideally trying to affect and measure (***the effect construct***). But each of these, the cause and the effect, has to be translated into real things, into a program or treatment and a measure or observational method. We use the term ***operationalization*** to describe the act of translating a construct into its manifestation. In effect, we take our idea and describe it as a series of operations or procedures. Now, instead of only being an idea in our minds, it becomes a public entity that anyone can look at and examine it.

The four validity types are built on one another. The bottom two (conclusion and internal) refer to observation on the bottom of the figure, the top one (external) being primarily concerned about the range of our theory and the last one (construct) emphasize the linkages between the bottom and the top. In short the four validity types and the question each address are:

1. Conclusion Validity: In this study, is there a ***relationship*** between the two variables?
2. Internal Validity: *Assuming that there is a relationship in this study,* is the relationship a ***causal*** one?
3. Construct Validity: *Assuming that there is a causal relationship in this study,* can we claim that the program reflected well our ***construct*** of the program and that our measure reflected well our idea of the ***construct*** of the measure
4. External Validity: *Assuming that there is a causal relationship in this study between the constructs of the cause and the effect,* can we ***generalize*** this effect to other persons, places or times? There are different kinds of validity.

Threats to Validity

For any inference or conclusion, there are always possible ***threats to validity*** – for which the conclusion or inference might be wrong. Therefore, the aim should be to reduce the most likely threats to validity. For instance, imagine a study examining whether there is a relationship between the extension of a specific technology and subsequent rates of adoption of that technology. Assume that the study is completed and no significant correlation between extension and adoption rates is found.

Now, how could this conclusion be wrong, or what are the threats to validity? It may be that there isn't sufficient statistical power to detect a relationship even if it exists due to small sample size or the measure of amount of extension is unreliable. It may also be that assumptions of the correlation/association tests are violated given the variables used. Random heterogeneity in the respondents may increase the variability in the data and made it harder to see the relationship of interest. The inference that there is no relationship will be stronger (greater conclusion validity) if one can show that these alternative explanations are not credible. The theory of validity, and the many lists of specific threats, provides a useful scheme for assessing the quality of research conclusions.

Minimizing Threats to Validity

Good research designs minimize the plausible alternative explanations for the hypothesized cause-effect relationship. But such explanations may be ruled out or minimized in a number of ways other than by design. There are five ways to minimize threats to validity, one of which is by research design:

1. **By Argument.** The most straightforward way to rule out a potential threat to validity is to simply argue that the threat in question is not a reasonable one. Such an argument may be made either ***a priori*** **or** ***a posteriori***, although the former will usually be more convincing than the latter. For example, depending on the situation, one might argue that an instrumentation threat is not likely because the same test is used for pre and post test measurements and did not involve observers who might improve, or other such factors. In most cases, ruling out a potential threat to validity by argument alone will be weaker than the other approaches listed below. As a result, the most plausible threats in a study should not, except in unusual cases, be ruled out by argument only.
2. **By Measurement or Observation.** In some cases it will be possible to rule out a threat by measuring it and demonstrating that either it does not occur at all or occurs so minimally as to not be a strong alternative explanation for the cause-effect relationship. Consider, for example, a study of the effects of an advertising campaign on subsequent sales of a particular product. In such a study, history would be a plausible alternative explanation. For example, a change in the local economy, the removal of a competing product from the market, or similar events could cause an increase in product sales. One might attempt to minimize such threats by measuring local economic indicators and the availability and sales of competing products. If there were no change in these measures coincident with the onset of the advertising campaign, these threats would be considerably minimized. Similarly, if one is studying the effects of special mathematics training on math achievement scores of children, it might be useful to observe everyday classroom behavior in order to verify that students were not receiving any additional math training to that provided in the study.
3. **By Design.** Here, the major emphasis is on ruling out alternative explanations by adding treatment or control groups, waves of measurement, and the like. This topic will be discussed in more detail below.
4. **By Analysis.** There are a number of ways to rule out alternative explanations using statistical analysis. One interesting example is provided by Jurs and Glass (1971). They suggest that one could study the plausibility of an attrition or mortality threat by conducting a two-way analysis of variance. One factor in this study would be the original treatment group designations (i.e., program vs. comparison group), while the other factor would be attrition (i.e., dropout vs. non-dropout group). The dependent measure could be the pretest or other available pre-program measures. A main effect on the attrition factor would be indicative of a threat to external validity or generalizability, while an interaction between group and attrition factors would

point to a possible threat to internal validity. Where both effects occur, it is reasonable to infer that there is a threat to both internal and external validity.

5. **By Preventive Action.** When potential threats are anticipated they can often be ruled out by some type of preventive action. For example, if the program is a desirable one, it is likely that the comparison group would feel jealous or demoralized. Several actions can be taken to minimize the effects of these attitudes including offering the program to the comparison group upon completion of the study or using program and comparison groups which have little opportunity for contact and communication. In addition, auditing methods and quality control can be used to track potential experimental dropouts or to insure the standardization of measurement.

The five categories listed above should not be considered mutually exclusive. A good research plan should, where possible make use of multiple methods for reducing threats. In general, reducing a particular threat by design or preventive action will probably be stronger than by using one of the other three approaches.

IV Reliability

Reliability has to do with the quality of measurement. In its everyday sense, reliability is the "consistency" of measures. Therefore, we need to understand different types of measurement error. The foundation of the theory of measurement is the **true score theory**. True score theory maintains that every measurement is an additive composite of two components: true score and random error.

$$X = T + e_X$$

The true score theory is a simple and good model for measurement, but it may not always be an accurate reflection of reality. In particular, it assumes that any observation is composed of the true value plus some random error value. But, what if all error is not random? One way to deal with this notion is to revise the simple true score model by dividing the error component into two subcomponents, **random error** and **systematic error**.

$$X = T + e_r + e_s$$

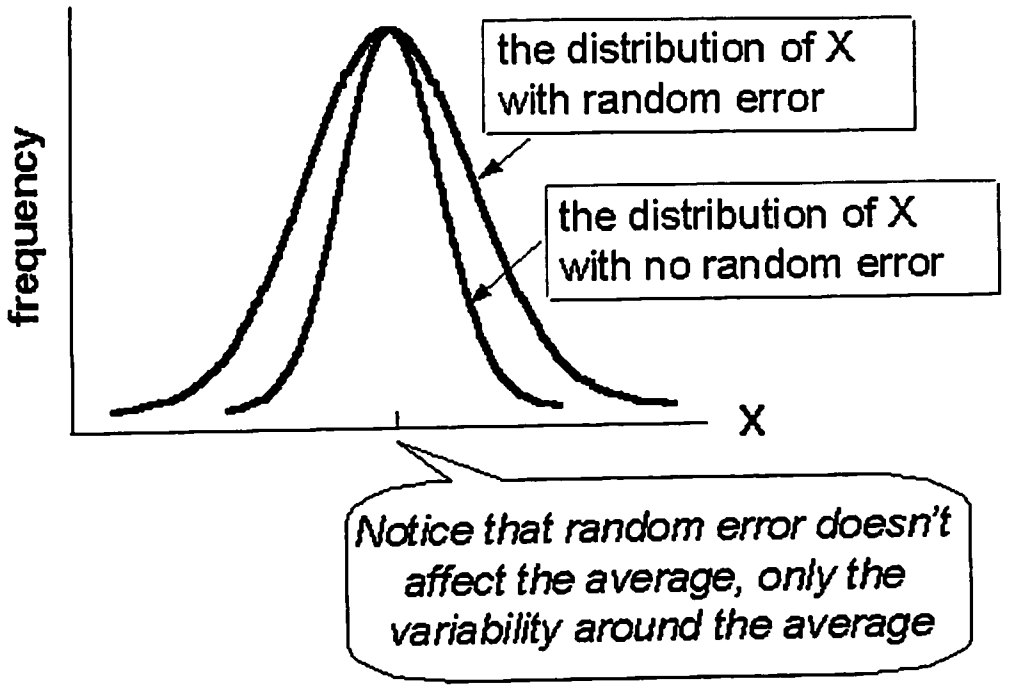
Random Error

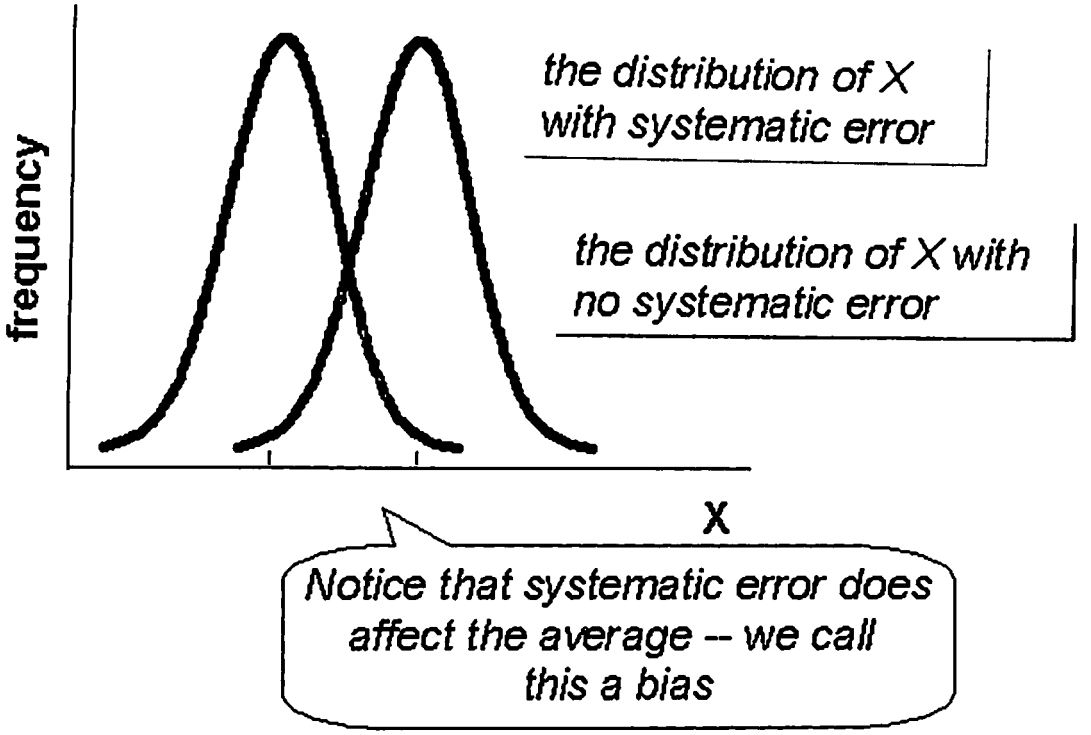
Random error is caused by any factors that randomly affect measurement of the variable across the sample. The important thing about random error is that it does not have any consistent effects across the entire sample. Instead, it pushes observed scores up or down randomly. This means that if we could see all of the random errors in a distribution they would have to sum to 0 -- there would be as many negative errors as positive ones. The

important property of random error is that it adds variability to the data but does not affect average performance for the group. Random error is sometimes considered ***noise***.

Systematic Error

Systematic error is caused by any factors that systematically affect measurement of the variable across the sample. For instance, if there is loud traffic going by just outside of a classroom where students are taking a test, this noise is liable to affect all of the children's scores -- in this case, systematically lowering them. Unlike random error, systematic errors tend to be consistently either positive or negative -- because of this; systematic error is sometimes considered to be ***bias*** in measurement.





Reducing Measurement Error

Pre-testing, pilot testing of instruments
Train data collectors
Check and recheck data
All data entry for computer should "double-punched" and verified
Use statistical procedure to adjust for measurement error
Multiple measure of the same construct

V Experimental designs

Experimental designs are considered the most "rigorous" of all research designs as it provides internal validity. Internal validity is at the center of all causal or cause-effect inferences. When we want to determine whether some program or treatment causes some outcome or outcomes to occur, then we are interested in having strong internal validity. Essentially, we want to assess the proposition:

If the program is given, then the outcome occurs

Unfortunately, it's not enough just to show that when the program or treatment occurs the expected outcome also happens. That's because there may be lots of reasons, other than the

program, for why we observed the outcome. To really show that there is a causal relationship, we have to simultaneously address the two propositions:

If the program is given, then the outcome occurs

and

If the program is *not* given, then the outcome does *not* occur

If we can provide evidence for both of these propositions then we can say that we have isolated the program from all other potential causes of the outcome, which points to the causal effectiveness of the program.

For example, we can select two groups that are as similar as we can possibly make them. If we are confident that the two situations are comparable, then we can administer our program to one group and the other group will remain without the program. That's exactly what an experimental design tries to achieve. In the simplest type of experiment, we create two groups that are "equivalent" to each other. One group (the program or treatment group) gets the program and the other group (the comparison or control group) does not. In all other respects, the groups are treated the same. They have similar people, live in similar contexts, have similar backgrounds, and so on. Now, if we observe differences in outcomes between these two groups are must be due to the program.

Now, how can we create two groups that are "equivalent"? In experimental design this is done by random assignment, which form the basis for obtaining two similar groups. If we randomly assign people to two groups, and we have enough people in our study to achieve the desired probabilistic equivalence, then we may consider the experiment to be strong in internal validity and will be able to have a good assessment whether the program causes the outcome(s).

However, it is practically very difficult to have experimental design. Many things may go wrong such as we may not have a large enough sample, we may have people who refuse to participate in our study or who drop out halfway through, we might be resisted by our staff in a study who would like some of their "favorite" people to get the program.

In some instances some artificial situation may be created to assess causal relationship with high internal validity. In that case we will be compromising with external validity or generalization. That is, we have reduced external validity in order to achieve greater internal validity.

VI Quasi-Experimental Design

Quasi-experimental design includes all the elements of experimental design but the random assignment. With respect to internal validity, it is inferior to randomized experiments. However, most compelling about this design is that it is frequently implemented than their randomized ones.

Most commonly used quasi-experimental design is the nonequivalence group design. In its simplest form it requires a pretest and posttest for a treatment and comparison group. Lack of random assignment, and the potential nonequivalence between the groups, complicates the statistical analysis of the nonequivalent groups design.

The second design is the regression discontinuity design. It is an important and often misunderstood as alternative to randomized experiments because its distinguishing characteristic -- assignment to treatment using a cutoff score on a pretreatment variable -- allows us to assign the treatment that need it most.

Other quasi-experiments with specific applicability are Proxy Pretest Design, Double Pretest Design, Nonequivalent Dependent Variables Design, Pattern Matching Design, and the Regression Point Displacement design. There is one major class of quasi-experimental designs that are not included above -- the interrupted time series designs.

Much contemporary social research is devoted to examining whether a program, treatment, or manipulation causes some outcome or result. For example, we might wish to know whether a new educational program causes subsequent achievement score gains, whether a novel drug causes a reduction in symptoms, and so on. Cook and Campbell (1979) argue that three conditions must be met before we can infer that such a cause-effect relation exists:

1. **Covariation.** Changes in the presumed cause must be related to changes in the presumed effect. Thus, if we introduce, remove, or change the level of a treatment or program, we should observe some change in the outcome measures.
2. **Temporal Precedence.** The presumed cause must occur prior to the presumed effect.
3. **No Plausible Alternative Explanations.** The presumed cause must be the only reasonable explanation for changes in the outcome measures. If there are other factors which could be responsible for changes in the outcome measures we cannot be confident that the presumed cause-effect relationship is correct.

In most social research the third condition is the most difficult to meet. Any number of factors other than the treatment or program could cause changes in outcome measures. Campbell and Stanley (1966) and later, Cook and Campbell (1979) list a number of common plausible alternative explanations (or, threats to internal validity). For example, it may be that some historical event which occurs at the same time that the program or treatment is instituted was responsible for the change in the outcome measures; or, changes in record keeping or measurement systems which occur at the same time as the program might be falsely attributed to the program.

VII Design Construction

Basic Design Elements. Most research designs can be constructed from four basic elements:

1. **Time.** A causal relationship, by its very nature, implies that some time has elapsed between the occurrence of the cause and the consequent effect. While for some phenomena the elapsed time might be measured in microseconds and therefore might be unnoticeable to a casual observer, we normally assume that the cause and effect in social science arenas do not occur simultaneously, In design notation we indicate this temporal element horizontally - whatever symbol is used to indicate the presumed cause would be placed to the left of the symbol indicating measurement of the effect. Thus, as we read from left to right in design notation we are reading across time. Complex designs might involve a lengthy sequence of observations and programs or treatments across time.
2. **Program(s) or Treatment(s).** The presumed cause may be a program or treatment under the explicit control of the researcher or the occurrence of some natural event or program not explicitly controlled. In design notation we usually depict a presumed cause with the symbol "X". When multiple programs or treatments are being studied using the same design, we can keep the programs distinct by using subscripts such as "X_1" or "X_2". For a comparison group (i.e., one which does not receive the program under study) no "X" is used.
3. **Observation(s) or Measure(s).** Measurements are typically depicted in design notation with the symbol "O". If the same measurement or observation is taken at every point in time in a design, then this "O" will be sufficient. Similarly, if the same set of measures is given at every point in time in this study, the "O" can be used to depict the entire set of measures. However, if different measures are given at different times it is useful to subscript the "O" to indicate which measurement is being given at which point in time.
4. **Groups or Individuals.** The final design element consists of the intact groups or the individuals who participate in various conditions. Typically, there will be one or more program and comparison groups. In design notation, each group is indicated on a separate line. Furthermore, the manner in which groups are assigned to the conditions can be indicated by an appropriate symbol at the beginning of each line. Here, "R" will represent a group which was randomly assigned, "N" will depict a group which was nonrandomly assigned (i.e., a nonequivalent group or cohort) and a "C" will indicate that the group was assigned using a cutoff score on a measurement.

VIII The Nature of Good Design

Throughout the design construction task, it is important to have some criteria that should be achieved before finally accepting a design strategy. Criteria discussed below are suggestive that are found in good research design. All these criteria point to the need to tailor research

designs based on specific research questions and the research environment rather than blindly following textbooks.

1. **Theory-Grounded.** Research strategies should reflect the theories, which are being investigated. Hypothesis regarding the theoretical expectations can be incorporated into the design. For instance, when theory predicts a specific treatment effect on one measure but not on another, the inclusion of both in the design improves discriminant validity and demonstrates the predictive power of the theory.
2. **Situational.** Good research designs reflect the settings of the investigation. Intergroup rivalry, demoralization, and competition might be assessed through the use of additional comparison groups who are not in direct contact with the original group.
3. **Feasible.** Good designs should be such that they are easy to implement. The sequence and timing of events are carefully thought out. Potential problems in measurement, adherence to assignment, database construction and the like, are anticipated. Additional groups or measurements are included in the design to explicitly correct for such problems.
4. **Redundant.** Good research designs have some flexibility built into them. Often, this flexibility results from duplication of essential design features. For example, multiple replications of a treatment help to insure that failure to implement the treatment in one setting will not invalidate the entire study.
5. **Efficient.** Good designs strike a balance between redundancy and the tendency to overdesign. Where it is reasonable, other, less costly, strategies for ruling out potential threats to validity are utilized.

This is by no means an exhaustive list of the criteria by which we can judge good research design nevertheless, goals of this sort help to guide the researcher toward a final design choice and emphasize important components, which should be included.

SAMPLING THEORY

1. Introduction

The principal objective of any research is to generalize about a proposition across time and space. Scientific sampling helps to achieve the approximate truth of propositions, inferences, or conclusions that involve generalization. There are two major approaches to provide generalization about any population. The first approach is the **Sampling Model.** In the sampling model, we first identify the population to be generalized. Then, we draw a representative sample from that population and conduct research with the sample. However, there are several problems with the sampling model with respect to generalization. First, timing of the study; second, it is not always easy to draw a fair or representative sample, third, it's impossible to sample across time and space.

The second approach to generalizing is the **Proximal Similarity Model.** Under this model, we begin by thinking about different contexts of generalization and developing a theory about which contexts are more like our study and which are less so.

A threat to external validity is the explanation of how you might be wrong in making a generalization. There are three major threats to external validity -- people, places or times because these are the ways one could be wrong. Critics may argue that the results of a study are due to the unusual type of people selected or because of the unusual place or, they might suggest that the study was done in a peculiar time. For instance, if someone did smoking cessation study week after the Surgeon General issues the well-publicized results of the latest smoking and cancer studies, he might get different results than if it were done the week before.How can we improve external validity? Based on the sampling model we can draw sample (random or nonrandom) from a population. Once sample is selected, assure that the respondents participate in the study - keep dropout rates low. A second approach would be to use the theory of proximal similarity more effectively.

2. Sampling Terminology

In most social research we are interested to generalize about specific groups. The group we wish to generalize to is often called the **population** in our study. Let's imagine that we wish to generalize about urban homeless males between the ages of 30 and 50 in the United States. For this population of interest it is really hard to get a reasonable sampling plan. It is time consuming and expensive to have listing of population in all the cities of USA. Therefore, we should make a distinction between the population to be generalized, and the population that will be accessible to us. We'll call the former the **theoretical population** and the latter the **accessible population.** In this example, the accessible population might be homeless males between the ages of 30 and 50 in six selected urban areas across the U.S.

Now, we have to get a list of the members of the accessible population. The listing of the accessible population from which we draw sample is called the **sampling frame**. If we conduct phone survey and select names from the telephone book, the book would be your sampling frame. However, this is not the great way to sample because significant proportions of the population either don't have a phone or have moved in or out of the area since the last book was printed. In this case, we might identify the area code and all three-digit prefixes within that area code and draw a sample simply by randomly dialing numbers (known as *random-digit-dialing*). In this case, the sampling frame is not a list *per se*, but is rather a procedure that we follow as the actual basis for sampling. The **sample** is the group of people who we select to be in our study. We may not be able to contact all the people we sampled, or some could drop out over the course of the study. The group that actually completes the study is a sub-sample of the sample -- it doesn't include non-respondents or dropouts.

At this point, you should appreciate that sampling is a difficult multi-step process and that there are lots of places you can go wrong. In fact, as we move from each step to the next in identifying a sample, there is the possibility of introducing systematic error or **bias**. For instance, even if we are able to identify perfectly the population of interest, we may not have access to all of them. And even if we do, we may not have a complete and accurate enumeration or sampling frame.

3. Statistical Sampling Terms

When we sample, the units that we sample -- usually people -- supply us with one or more responses. In this sense, a **response** is a specific measurement value that a sampling unit supplies. When we look across the responses that we get for our entire sample, we use a **statistic**. There are a wide variety of statistics we can use -- mean, median, mode, standard deviation, variance, etc. Our objective of sampling is to get an estimate for the population we sampled. If we measure the entire population and calculate a value like a mean or average, we call it a **parameter** of the population.

3.1 The Sampling Distribution

How do we get from our sample statistic to an estimate of the population parameter? A crucial concept is the **sampling distribution**. To understand it let us do a thought experiment --instead of just taking a single sample we take three independent samples of the same population. For each sample we might have different statistics, say for example we get mean. That is we have now three means of three different sample of the same population. It is obvious, even though these means are based on the sample from the same population, these will vary slightly. Now if we estimate the mean of the sample means, we will get the estimate of the population parameter. If we take infinite samples from the same population and compute the average for each one and plot them on a histogram or bar graph we should find that most of them converge on the same central value, and fewer samples have averages farther away up or down from that central value. The bar graph will look like a *bell shape curve,* which is an indication of a "normal" distribution of statistics. The distribution of an infinite number of samples of the same

size is known as the **sampling distribution**. If we take the average of the sampling distribution -- the average of the averages of an infinite number of samples -- we would be much closer to the true population average -- the parameter of interest. The standard deviation of the sampling distribution tells us something about how different samples would be distributed. In statistics it is referred to as the **standard error.** A standard deviation is the spread of the scores around the average in a single sample. The standard error is the spread of the averages around the average of averages in a sampling distribution.

3.2 Sampling Error

The standard error is called **sampling error**. Sampling error gives us some idea of the precision of our statistical estimate. A low sampling error means that we had relatively less variability or range in the sampling distribution. The greater the sample standard deviation, the greater is the standard error. The standard error is also related to the sample size. The greater the sample size, the smaller is the standard error. Why? Because the greater the sample size, the closer your sample is to the actual population itself. If you take a sample that consists of the entire population you actually have no sampling error because you don't have a sample, you have the entire population. In that case, the mean you estimate is the parameter.

4. Probability Sampling Method

A **probability sampling** method is any method of sampling that utilizes some form of *random selection*. In order to have a random selection method, you must set up some process or procedure that assures that the different units in your population have equal probabilities of being chosen. Humans have long practiced various forms of random selection, such as picking a name out of a hat, or choosing the short straw. These days, we tend to use computers as the mechanism for generating random numbers as the basis for random selection.

4.1 Simple Random Sampling

The simplest form of random sampling is called **simple random sampling**. Here's the quick description of simple random sampling:

Objective: To select n units out of N such that each ${}_NC_n$ has an equal chance of being selected.

Procedure: Use a table of random numbers, a computer random number generator, or a mechanical device to select the sample

Simple random sampling is simple to accomplish and is easy to explain to others. Because simple random sampling is a fair way to select a sample, it is reasonable to generalize the results from the sample back to the population. Simple random sampling is not the most statistically efficient method of sampling because of the chance in the draw,

hard to get good representation of subgroups in a population. To deal with these issues, we have to turn to other sampling methods.

4.2 Stratified Random Sampling

Stratified Random Sampling, also sometimes called *proportional* or *quota* random sampling, involves dividing your population into homogeneous subgroups and then taking a simple random sample in each subgroup. There are several major reasons why you might prefer stratified sampling over simple random sampling. First, it assures representation of subgroups as well as the overall population. Second, stratified random sampling will generally have more statistical precision than simple random sampling. This will only be true if the strata or groups are homogeneous.

4.3 Systematic Random Sampling

Here are the steps you need to follow in order to achieve a **systematic random sample**:

Number the units in the population from 1 to N

Decide on the n (sample size

k = N/n = the interval size

Randomly select an integer between 1 to k then take every kth unit

Let's assume that we have a population of size N=100. We want to take a sample of n=20. To use systematic sampling, the population must be listed in a random order. The sampling fraction would be f = 20/100 = 20%, the interval size, k, is equal to N/n = 100/20 = 5. Now, select a random integer from 1 to 5, say 4. Now, to select the sample, start with the 4th unit in the list and take every k-th unit (every 5th, because k=5). That is we would be sampling units 4, 9, 14, 19, and so on to 100.

4.4 Cluster (Area) Random Sampling

When we have to sample a population that is disbursed across a wide geographic region then we may use cluster sampling. In cluster sampling, we follow these steps:

divide population into clusters

then take every kth unit

randomly sample clusters

measure <u>all</u> units within sampled clusters

Cluster or area sampling, then, is useful in situations like this, and is done primarily for efficiency of administration

4.5 Multi-Stage Sampling

When we combine sampling methods, we call this **multi-stage sampling**.

5. Non-probability Sampling

The difference between non-probability and probability sampling is that the former does not involve random selection and probability sampling does. Does that mean that non-probability samples aren't representative of the population? Not necessarily. But it does mean that non-probability samples cannot depend upon the rationale of probability theory. We can divide non-probability sampling methods into two broad types: *accidental* or *purposive*. Most sampling methods are purposive in nature because we usually approach the sampling problem with a specific plan in mind. The most important distinctions among these types of sampling methods are the ones between the different types of purposive sampling approaches.

5.1 Accidental, Haphazard or Convenience Sampling

One of the most common methods of sampling goes under the various titles listed here. Interviews conducted frequently by television news programs to get a quick reading of public opinion, use of college students in much psychological research are primarily matters of convenience. In clinical practice, we might use clients who are available to us as our sample. In many research contexts, we sample simply by asking for volunteers. Clearly, the problem with all of these types of samples is that we have no evidence that they are representative of the populations we're interested in generalizing.

5.2 Purposive Sampling

In purposive sampling, we sample with a purpose in mind. We usually would have one or more specific predefined groups we are seeking. Purposive sampling can be very useful for situations where one needs to reach a targeted sample quickly. With a purposive sample, we are likely to get the opinions of our target population, but we are also likely to overweight subgroups in your population that are more readily accessible.

5.3 Modal Instance Sampling

In statistics, **mode** is the most frequently occurring value in a distribution. In sampling, when we do a modal instance sample, we are sampling the most frequent case, or the "typical" case. In a lot of informal public opinion polls, for instance, they interview a "typical" voter. There are a number of problems with this sampling approach. First, how do we know what the "typical" or "modal" case is? We could say that the modal voter is a person who is of average age, educational level, and income in the population. But, it's not clear that using the averages of these is the fairest. And, how do you know that those

three variables -- age, education, income -- are the only or event the most relevant for classifying the typical voter? What if religion or ethnicity is an important discriminator? Clearly, modal instance sampling is only sensible for informal sampling contexts.

5.4 Expert Sampling

Expert sampling involves the assembling of a sample of persons with known or demonstrable experience and expertise in some area. Often, we convene such a sample under the auspices of a "panel of experts." There are actually two reasons of doing expert sampling. First, it would be the best way to elicit the views of persons who have specific expertise. But the other reason you might use expert sampling is to provide evidence for the validity of another sampling approach you've chosen.

5.5 Quota Sampling

In quota sampling, we select non-randomly according to some fixed quota. There are two types of quota sampling: *proportional* and *non proportional*. In **proportional quota sampling** we want to represent the major characteristics of the population by sampling a proportional amount of each.

Nonproportional quota sampling is a bit less restrictive. In this method, minimum number of sampled units in each category is selected. This method is the non-probabilistic analogue of stratified random sampling in that it is typically used to assure that smaller groups are adequately represented.

5.6 Heterogeneity Sampling

We sample for heterogeneity when we want to include all opinions or views, and we aren't concerned about representing these views proportionately. Another term for this is sampling for *diversity*. In many brainstorming or nominal group processes (including concept mapping), we would use some form of heterogeneity sampling because our primary interest is in getting broad spectrum of ideas, not identifying the "average" or "modal instance" ones.

5.7 Snowball Sampling

In snowball sampling, you begin by identifying someone who meets the criteria for inclusion in your study. You then ask them to recommend others who they may know who also meet the criteria. Although this method would hardly lead to representative samples, there are times when it may be the best method available. Snowball sampling is especially useful when you are trying to reach populations that are inaccessible or hard to find.

EVALUATION

I Introduction to Evaluation

Evaluation is a methodological area that is closely related to, but distinguishable from more traditional social research. Evaluation utilizes many of the same methodologies used in traditional social research, but because evaluation takes place within a political and organizational context, it requires group skills, management ability, political dexterity, sensitivity to multiple stakeholders and other skills that social research in general does not rely on as much. Here we introduce the idea of evaluation and some of the major terms and issues in the field.

1.1 Definitions of Evaluation

Evaluation is the systematic acquisition and assessment of information to provide useful feedback about some object

Evaluation is a *systematic* endeavor and the term 'object' refers to a program, policy, technology, person, need, activity, and so on. The definition emphasizes **acquiring and assessing information rather than assessing worth or merit** because all evaluation work involves collection and sorting data to make judgments about validity of the information and inferences we derive from it.

1.2 The Goals of Evaluation

The goal of most evaluations is to provide "useful feedback" to a variety of audiences including sponsors, donors, client-groups, administrators, staff, and other relevant constituencies. Most often, feedback is perceived as "useful" if it helps in decision-making. But the relationship between an evaluation and its impact is not a simple one -- studies that seem critical sometimes fail to influence short-term decisions, and studies that initially seem to have no influence can have a delayed impact when more congenial conditions arise. Despite this, there is broad consensus that the major goal of evaluation should be to influence decision-making or policy formulation through the provision of empirically driven feedback.

1.3 Evaluation Strategies

'Evaluation strategies' means broad and overarching perspectives on evaluation. Four major groups of evaluation strategies are:

Scientific-experimental models, management-oriented systems models, qualitative/anthropological models, participant-oriented models

Scientific-experimental models are the most historically dominant evaluation strategies. Borrowing values and methods from science, social sciences derive accuracy, objectivity and the validity of the information it generate. Scientific-experimental models include the tradition of experimental and quasi-experimental designs; objectives-based research that comes from education; econometrically oriented perspectives including cost-effectiveness and cost-benefit analysis; and the recent articulation of theory-driven evaluation.

The second class of strategies is **management-oriented systems models** such as Program Evaluation and Review Technique (**PERT**), and Critical Path Method (**CPM**). Both have been widely used in business and government in the developed countries. Another strategy currently used by the USAID and other donor agencies is the Logical Framework or "Logframe" model. The management-oriented systems models emphasize comprehensiveness in evaluation, placing evaluation within a larger framework of organizational activities.

The third class of strategies is the **qualitative/anthropological models**. They emphasize the importance of observation, the need to retain the phenomenal quality of the evaluation context, and the value of subjective human interpretation in the evaluation process. Included in this category are the approaches known in evaluation as naturalistic or 'Fourth Generation' evaluation; the various qualitative schools; critical theory and art criticism approaches; and, the 'grounded theory' approach of Glaser and Strauss among others.

Finally, a fourth class of strategies is termed **participant-oriented models**. As the term suggests, they emphasize the central importance of the evaluation participants, especially clients and users of the program or technology. Client-centered and stakeholder approaches are examples of participant-oriented models, as are consumer-oriented evaluation systems.

Among these strategies which one is the best? There is no inherent incompatibility between these broad strategies -- each of them brings something valuable. In fact, in recent years attention has increasingly turned as to how to integrate results from evaluations that use different strategies and different methods.

1.4 Types of Evaluation

There are many types of evaluations that depend on the object being evaluated and the purpose of the evaluation. The basic distinction in evaluation types is **formative** and **summative**. Formative evaluations strengthen or improve the object being evaluated by examining the delivery of the program or technology, the quality of its implementation, and the assessment of the organizational context, personnel, procedures, inputs, etc. Summative evaluations, in contrast, examine the effects or outcomes of some object by describing what happens after the delivery of the program or technology; assessing whether the object caused the outcome; determining the overall impact of the causal factor beyond only the immediate target outcomes; and, estimating the relative costs associated with the object.

1.4.1 **Formative evaluation** includes several evaluation types:

Needs assessment determines who needs the program, how great the need is, and what might work to meet the need

Evaluability assessment determines whether an evaluation is feasible and how stakeholders can help shape its usefulness

Structured conceptualization helps stakeholders define the program or technology, the target population, and the possible outcomes.

Implementation evaluation monitors the fidelity of the program or technology delivery

Process evaluation investigates the process of delivering the program or technology, including alternative delivery procedures

1.4.2 **Summative evaluation** can also be subdivided:

Outcome evaluations investigate whether the program or technology caused demonstrable effects on specifically defined target outcomes

Impact evaluation is broader and assesses the overall or net effects -- intended or unintended -- of the program or technology as a whole

Cost-effectiveness and cost-benefit analysis address questions of efficiency by standardizing outcomes in terms of their dollar costs and values

Secondary analysis reexamines existing data to address new questions or use methods not previously employed

Meta-analysis integrates the outcome estimates from multiple studies to arrive at an overall or summary judgment on an evaluation question

1.5 Evaluation Questions and Methods

In formative research the major questions and methodologies are:

What is the definition and scope of the problem or issue?

Formulating and conceptualizing may follow different methods that may include brainstorming, focus groups, nominal group techniques, Delphi methods, brain writing, stakeholder analysis, lateral thinking, input-output analysis, and concept mapping.

Where is the problem and how big or serious is it?

The most common method used in this case is "needs assessment" that may include: analysis of existing data sources, and the use of sample surveys, interviews of constituent populations, qualitative research, expert testimony, and focus groups.

How should the program or technology be delivered to address the problem?

Some of the methods listed above apply here. Other methods like simulation techniques, or multivariate methods like multi-attribute utility theory or exploratory causal modeling; decision-making methods; and project planning and implementation methods like flow charting, PERT/CPM, and project scheduling.

How well is the program or technology delivered?

Qualitative and quantitative monitoring techniques, the use of management information systems, and implementation assessment would be appropriate methodologies here.

The questions and methods addressed under summative evaluation include:

What type of evaluation is feasible?

Assessment of the problem to decide what type of evaluation is needed. Standard approaches should follow to select appropriate evaluation design.

What was the effectiveness of the program or technology?

Choose from observational and co relational methods for demonstrating whether desired effects occurred, and quasi-experimental and experimental designs for determining whether observed effects could reasonably be attributed to the intervention.

What is the net impact of the program?

One can apply econometric methods for assessing impact and cost/benefit analysis for cost effectiveness along with qualitative methods that enable to summarize the full range of intended and unintended impacts.

Analysis of Data

I Introduction

Analysis of data is fairly straightforward if collection of data follows all the scientific steps we discussed earlier. That is steps such as defining the research problem; development and implementation of a sampling plan; conceptualizing, operationalizing and testing of measures; and developing a design structure were followed. In most social research the data analysis involves three major steps:

Data Preparation involves checking for accuracy; entering the data into computer; transforming data; and developing and documenting a database structure that integrates the various measures.

Descriptive Statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data. With descriptive statistics we are simply describing what is, what the data shows.

Inferential Statistics investigate questions, models and hypotheses. In many cases, the conclusions from inferential statistics extend beyond the immediate data alone. For instance, we use inferential statistics to try to infer from the sample data what the population thinks. Or, we use inferential statistics to make judgments of the probability that an observed difference between groups is a dependable one or one that might have happened by chance in this study. Thus, we use inferential statistics to make inferences from our data to more general conditions; we use descriptive statistics simply to describe what's going on in our data.

II. Conclusion Validity

In many ways, conclusion validity is the most important of the four validity types because it is relevant whenever we are trying to decide if there is a relationship in our observations. It can be defined as:

> **Conclusion validity is the degree to which we can conclude about the conformity of the relationships.**

For instance, if we're doing a study that looks at the relationship between socioeconomic status (SES) and fishing, we eventually want to reach some conclusion. Based on our data, we may conclude that there is a positive relationship that persons with higher SES tend to have fishing as an occupation. Conclusion validity is the degree to which the conclusion we reach is credible or believable. Whenever we investigate a relationship, we

essentially have two possible conclusions -- either there is a relationship in our data or there isn't. In either case, however, we could be wrong in our conclusion. We might conclude that there is a relationship when in fact there is not, or we might infer that there isn't a relationship when in fact there is. So, we have to consider all of these possibilities when we talk about conclusion validity.

Threats to Conclusion Validity

Threats to conclusion validity are those that can lead to an incorrect conclusion about a relationship we try to establish. One may make two kinds of errors about relationships - conclude that there is no relationship when in fact there is or conclude that there is a relationship when in fact there is no.

Statistical Power. The rule of thumb in social research is that we want statistical power to be greater than 0.9 in values. That is, at least 90 chances out of 100 of finding a relationship when there is one. There are several factors that interact to affect power- increase sample size, increase alpha, and increase effect size.

Good Reliability. Reliability is related to the idea of noise or "error" that obscures wer ability to see a relationship. In general, we can improve reliability by doing a better job of constructing measurement instruments, by increasing the number of questions on a scale or by reducing situational distractions in the measurement context.

Good Implementation. When we are studying the effects of interventions, treatments or programs, we can improve conclusion validity by assuring good implementation. This can be accomplished by training program operators and standardizing the protocols for administering the program relationship.

Data Preparation

Data Preparation involves checking or logging the data in; checking the data for accuracy; entering the data into the computer; transforming the data; and developing and documenting a database structure that integrates the various measures.

Logging the Data

In any research project we may have data coming from a number of different sources at different times such as mail survey data, coded interview data, pretest or posttest data or observational data

Checking the Data For Accuracy

As soon as data is received we should screen it for accuracy. In some circumstances doing this right away will allow we to go back to the sample to clarify any problems or errors. There are several questions we should ask as part of this initial data screening:

Are the responses legible/readable?

Are all-important questions answered?

Are the responses complete?

Developing a Database Structure

The database structure is the manner in which we intend to store the data for the study so that it can be accessed in subsequent data analyses. There are generally two options for storing data on computer -- database programs and statistical programs. Usually database programs are the more complex of the two to learn and operate, but they allow the analyst greater flexibility in manipulating the data.

In every research project, one should generate a printed **codebook** that describes the data and indicates where and how it can be accessed. Minimally the codebook should include the following items for each variable:

variable name
variable description
variable format (number, data, text)
instrument/method of collection
date collected
respondent or group
variable location (in database)
notes

The codebook is an indispensable tool for the analysis team. Together with the database, it should provide comprehensive documentation that enables other researchers who might subsequently want to analyze the data to do so without any additional information.

Entering the Data into the Computer

There are a wide variety of ways to enter the data into the computer for analysis. Probably the easiest is to just type the data in directly. In order to assure a high level of data accuracy, the analyst should use a procedure called **double entry**.

Descriptive Statistics

Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data.

Univariate Analysis

Univariate analysis involves the examination across cases of one variable at a time. There are three major characteristics of a single variable that we tend to look at:

- the distribution
- the central tendency
- the dispersion

In most situations, we would describe all three of these characteristics for each of the variables in our study.

The Distribution. The distribution is a summary of the frequency of individual values or ranges of values for a variable. Or, we might group income into four or five ranges of income values.

Category	Percent
Under 35	9%
36-45	21
46-55	45
56-65	19
66+	6

Table 1. Frequency distribution table.

One of the most common ways to describe a single variable is with a ***frequency distribution***. Frequency distributions can be shown in two ways, as a table or as a graph. Table 1 shows an age frequency distribution with five categories of age ranges defined. The same frequency distribution can be depicted in a graph as shown in Figure 2. This type of graph is often referred to as a histogram or bar chart.

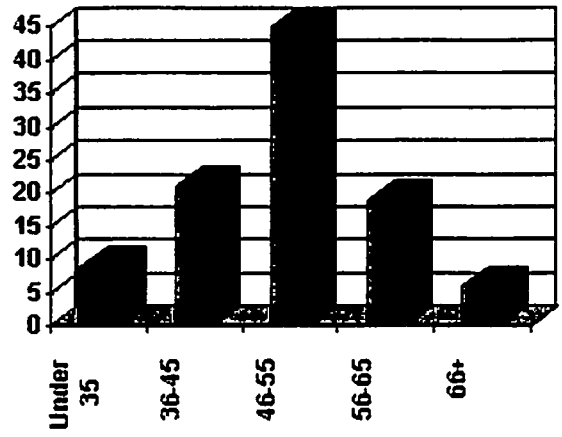


Figure 1. Frequency distribution bar chart.

Discrete example: Suppose we take a sample of 50 Cambodian families, and record the number of children in each family as X. We call X a discrete random variable. Suppose number of children in the 50 families is distributed as below:

Table 2. Frequency Distribution of the Number of Children in 50 Cambodian families

Number of Children	Frequency	Cumulative Frequency	Relative Frequency
0	9	9	.18
1	7	16	.14
2	12	28	.24
3	9	37	.18
4	5	42	.10
5	6	48	.12
6	0	48	0
7	2	50	.04
	N=50	N=50	1.00

Continuous example: Suppose we take a sample of 200 Cambodian men and record each height in inches. We call height X a continuous variable, since its possible values may not be whole number, it may vary continuously (X could be 64, 65, 66,..... or anywhere between, such as 64.325....inches.

Table 3. Frequency and relative frequency of the heights of 200 men

Cell Boundaries	Cell Midpoint	Frequency *f*	Cumulative frequency	Relative Frequency *(f/n)*
58.5-61.5	60	04	004	.02
61.5-64.5	63	12	014	.06
64.5-67.5	66	44	060	.22
67.5-70.5	69	64	124	.32
70.5-73.5	72	56	180	.28
73.5-76.5	75	16	196	.08
76.5-79.5	78	04	200	.02
		n=200		1.00

Distributions may also be displayed using percentages

Percentiles: In some instances data can be presented in terms of percentiles. For example, in table 3, what can we say about a man 64.5 inches tall? Obviously he is relatively small, with only a few are smaller than him. Specifically, we can say that only 8% are smaller than him. His height is said to be in the 8^{th} percentile. At the other end, a height of 73.5 inches would be the 90^{th} percentile.

Quartiles: The percentiles that cut data into four quarters have special names: the 25^{th} and 75^{th} percentiles are called the lower and upper quartiles. The 50^{th} percentile is called median, because it is the middle value that cuts the data in half. We can calculate the 50^{th} percentile or median for the height data as follows:

$$\textbf{Median=X=67.5 + (20/30)3 = 69.4}$$

Central Tendency. The central tendency of a distribution is an estimate of the "center" of a distribution of values. There are three major types of estimates of central tendency are Mean, Median, Mode.

The **Mean** or average is probably the most commonly used method of describing central tendency. To compute the mean all we do is add up all the values and divide by the number of values. For example, consider the test score values:

15, 20, 21, 20, 36, 15, 25, 15

The sum of these 8 values is 167, so the mean is 167/8 = 20.875.

The Mean: It is the most frequent measure of central tendency as it represents all the data points although sometimes over or under estimated by extreme data points. Suppose we

surveyed 5 households on daily catch of fish as 8 kg, 10 kg, 6kg, 7 kg and 9 kg. The average (mean) catch will be (8+10+6+7+9)/5= 6 kg. To generalize this, suppose a sample of n observations is denoted by X_1, X_2, X_3, ..., X_n, Then the average is found by summing and dividing by the sample size n:

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$$

$$\overline{X} = \frac{\sum X_{i=1}^{n}}{n}$$

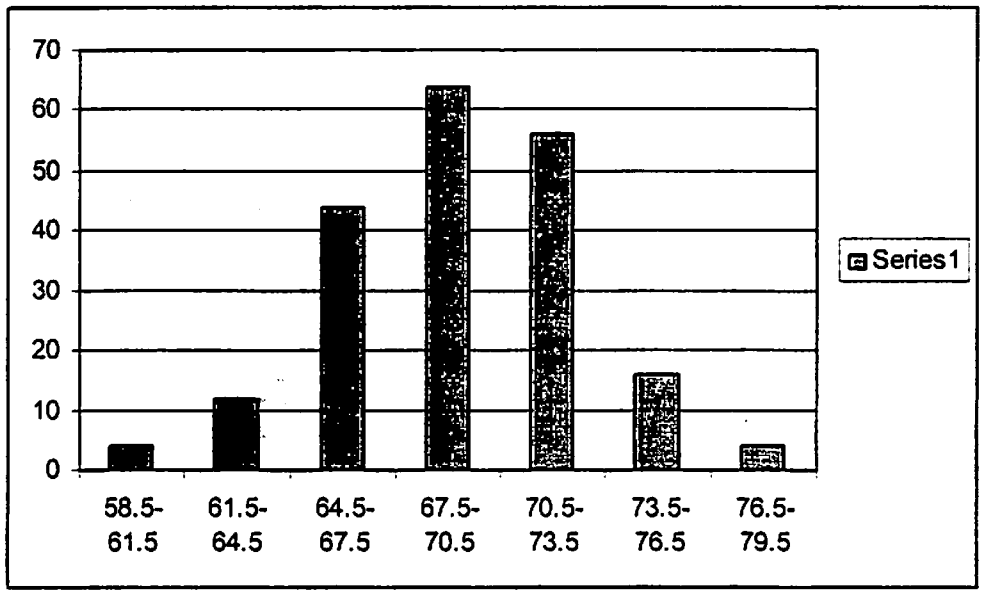
Grouped data example: Table 3 provides example of grouped data. More than one individual lie in each class of heights, we call it frequency (*f*). In this case we find average by multiplying class frequencies by the mid point value of each class and dividing by *n*. We can generalize this as below:

$$\overline{X} = \frac{\sum fx}{n}$$ where *x* represents mid value of each group.

Table 3. Calculation of Mean for grouped data

Cell Boundaries	Cell Midpoint (*x*)	Frequency *f*	*fx*
58.5-61.5	60	04	240
61.5-64.5	63	12	750
64.5-67.5	66	44	2904
67.5-70.5	69	64	4416
70.5-73.5	72	56	4032
73.5-76.5	75	16	1200
76.5-79.5	78	04	312
		n=200	$\overline{X} = \frac{\sum fx}{n}$ =13,860/200 =69.3

Figure 2. Histogram of the grouped data



The **Median** is the score found at the exact middle of the set of values. One way to compute the median is to list all scores in numerical order, and then locate the score in the center of the sample. For example, if there are 500 scores in the list, score #250 would be the median. If we order the 8 scores shown above, we would get:

15,15,15,20,20,21,25,36

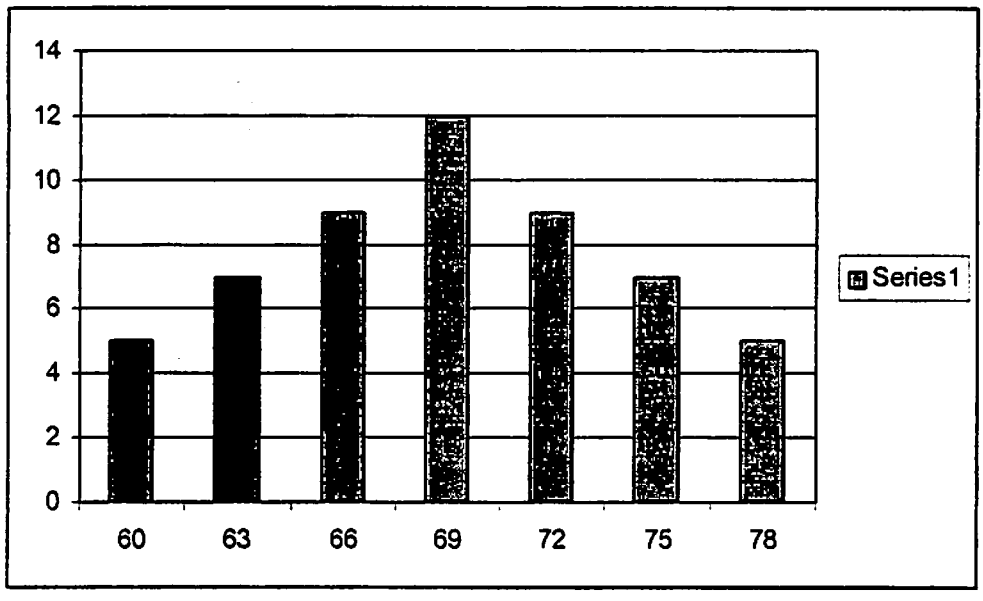
There are 8 scores and score #4 and #5 represent the halfway point. Since both of these scores are 20, the median is 20. If the two middle scores had different values, we would have to interpolate to determine the median.

The median is less sensitive to extreme scores than the mean and this makes it a better measure than the mean for highly skewed distributions. The median income is usually more informative than the mean income, for example.

The sum of the absolute deviations of each number from the median is lower than is the sum of absolute deviations from mean. The mean, median, and mode are equal in symmetric distributions. The mean is higher than the median in positively skewed distributions and lower than the median in negatively skewed distributions.

The **mode** is the most frequently occurring value in the set of scores. To determine the mode, we might again order the scores as shown above, and then count each one. The most frequently occurring value is the mode. In our example, the value 15 occurs three times and is the model.

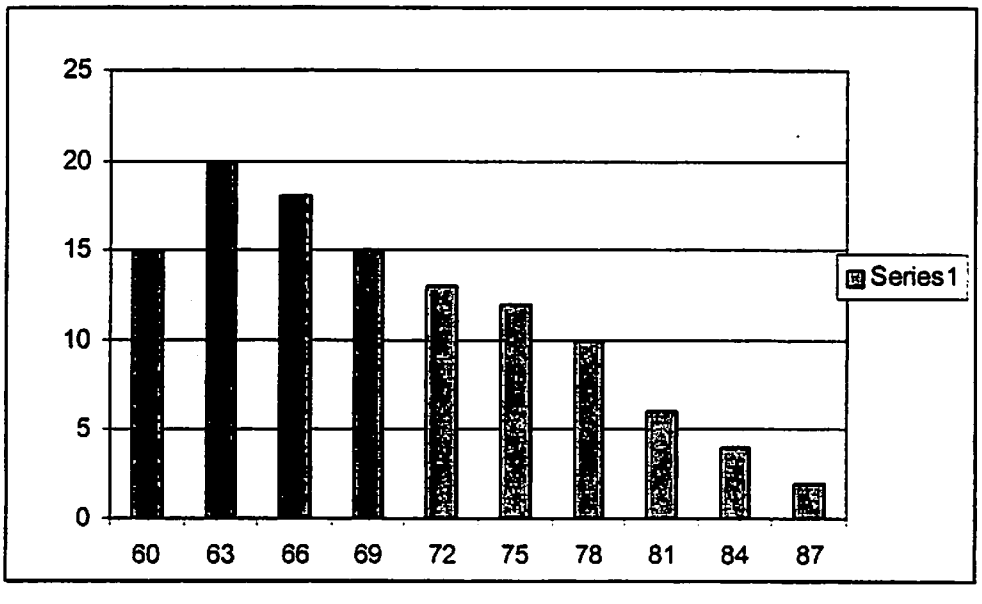
Figure 3. Histogram showing Mode



The advantage of the mode as a measure of central tendency is that it's meaning is obvious. Further, it is the only measure of central tendency that can be used with nominal data.

The mode is greatly subject to sample fluctuations and is therefore not recommended to be used as the only measure of central tendency. A further disadvantage of the mode is that many distributions have more than one mode. These distributions are called "multimodal."

Figure 4. Negatively Skewed Distribution



Measures of Dispersion

Dispersion refers to the spread of the values around the central tendency. There are different measures of dispersion:

The Range
The Inter-Quartile Range (IQR)
The Mean Absolute Deviation (MAD)
The Mean Squared Deviation (MSD)
The Standard Deviation (SD)and
The Variance.

The **Range** is simply the highest value minus the lowest value. In our example distribution, the high value is 36 and the low is 15, so the range is 36 - 15 = 21.

The **IQR** is the difference between the upper quartile and the lower quartile:

IQR = Upper Quartile – Lower Quartile

Limitation of the IQR: It does not base on all the observation in the sample.

The **Mean Absolute Deviation (MAD)** is the average of the absolute deviations from each observation from the mean. Again lets take the set of scores:

15,20,21,20,36,15,25,15

to compute the MAD, we first find the distance between each value and the mean. We know from above that the mean is 20.875. So, the differences from the mean are:

15 - 20.875 = -5.875
20 - 20.875 = -0.875
21 - 20.875 = +0.125
20 - 20.875 = -0.875
36 - 20.875 = 15.125
15 - 20.875 = -5.875
25 - 20.875 = +4.125
15 - 20.875 = -5.875

Sum = 0; $\mathbf{MD=\sum(X-\overline{X})/n=0/n=0;}$

$\mathbf{AD=\sum|X-\overline{X}|=38.725;}$

$\mathbf{MAD=\frac{1}{n}\sum|X-\overline{X}|=38.725/n=4.84}$

Notice that values that are below the mean have negative discrepancies and values above it have positive ones.

Although MAD intuitively is a good measure of spread, in many ways a better solution to the problem of "canceling signs" is to square each deviation. The **Mean Squared Deviation** and the **Standard Deviation** are more accurate and detailed estimate of dispersion because an outlier can greatly exaggerate the range.

Next, we square each discrepancy:

-5.875 * -5.875 = 34.515625
-0.875 * -0.875 = 0.765625
+0.125 * +0.125 = 0.015625
-0.875 * -0.875 = 0.765625
15.125 * 15.125 = 228.765625
-5.875 * -5.875 = 34.515625
+4.125 * +4.125 = 17.015625
-5.875 * -5.875 = 34.515625

Now, we take these "squares" and sum them to get the Sum of Squares (SS) value. Here, the sum is 350.875.

$=\frac{1}{n}\sum(X-\overline{X})^2$ =350.875/8=43.86

Next, we divide this sum by the number of scores minus 1. Here, the result is 350.875 / 7 = 50.125. This value is known as the **variance**. To get the **Standard Deviation (SD)**, we take the square root of the variance (remember that we squared the deviations earlier). This would be SQRT(50.125) = 7.079901129253. We can express this exercise in terms of formulae:

$$SD = \sqrt{\frac{\sum(X-\overline{X})^2}{n-1}}$$

Definition of Standard Deviation: Square root of the sum of the squared deviations from the mean divided by the number of scores minus one.

Although we can calculate these univariate statistics by hand, it gets quite tedious when we have more than a few values and variables. Every statistics program is capable of calculating them easily for we. For instance, I put the eight scores into SPSS and got the following table as a result:

N	8
Mean	20.8750
Median	20.000
Mode	15.0
Std. Deviation	7.079
Variance	50.125
Range	21.0

The standard deviation allows us to reach some conclusions about specific scores in our distribution. Assuming that the distribution of scores is normal or bell-shaped (or close to it!), the following conclusions can be reached:

approximately 69% of the scores in the sample fall within one standard deviation of the mean
approximately 95% of the scores in the sample fall within two standard deviations of the mean
approximately 99% of the scores in the sample fall within three standard deviations of the mean

For instance, since the mean in our example is 20.875 and the standard deviation is 7.0799, we can from the above statement estimate that approximately 95% of the scores will fall in the range of 20.875-(2*7.0799) to 20.875+(2*7.0799) or between 6.7152 and 35.0348.

Mean and Standard Deviation from the Relative Frequency Distribution

Sometimes data are available in summary form in relative frequency distribution. WE can derive appropriate formulae for $\overline{X}$ is below:

$$\overline{X} = \frac{1}{n}(x_1 f_1 + x_2 f_1 + \ldots\ldots + x_n f_n)$$

$$= x_1\left(\frac{f_1}{n}\right) + x_2\left(\frac{f_2}{n}\right) + \ldots\ldots + x_n\left(\frac{f_n}{n}\right)$$

$$= \overline{X} = \sum x\left(\frac{f}{n}\right)$$

That is, $\overline{X}$ is just the sum of the x values weighted according to their relative frequencies *f/n*. In the same way the MSD can also be calculated from the relative frequencies:

$$\text{MSD} = \sum (x - \overline{X})^2 \left(\frac{f}{n}\right)$$

Finally, we can easily calculate s^2 from the MSD as

$$\text{Variance} = s^2 = \left(\frac{n}{n-1}\right)\text{MSD}$$

Table 4. Calculation of Mean and Standard Deviation from the Relative Frequency Distribution

x	f/n	$x(f/n)$	$x-\overline{X}$	$(x-\overline{X})^2$	$(x-\overline{X})^2\left(\frac{f}{n}\right)$
60	.02	1.20	-9	81	1.62
63	.06	3.78	-6	36	2.16
66	.22	14.52	-3	9	1.98
69	.32	22.08	0	0	0
72	.28	20.16	3	9	2.52
75	.08	6.00	6	36	2.88
78	.02	1.56	9	81	1.62
		$\overline{X}=69.30$ $\overline{X}=69$			MSD=12.78 s^2=(200/199)*12.78 =12.84

Correlations

The correlation is one of the most common and most useful statistics. A correlation is a single number that describes the degree of relationship between two variables. Let's work through an example to show we how this statistic is computed.

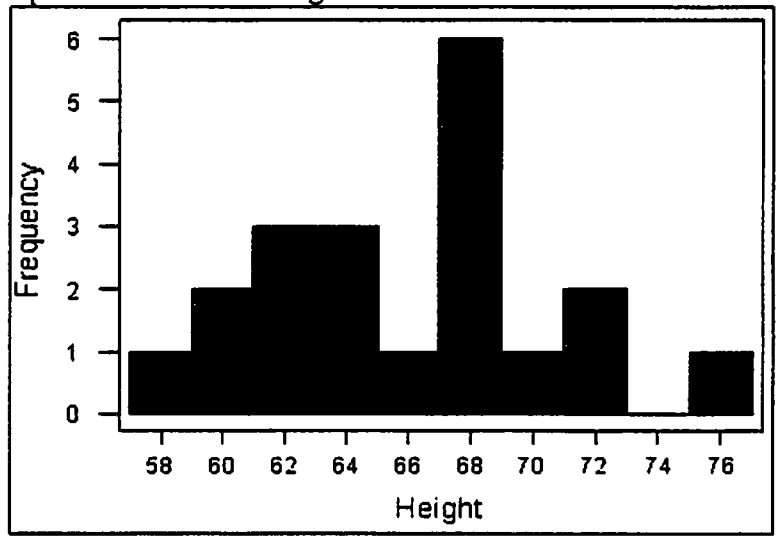
Correlation Example

Let's assume that we want to look at the relationship between two variables, height (in inches) and self-esteem. Perhaps we have a hypothesis that there is a relationship between height and self esteem. Let's say we collect some information on twenty individuals (all male). Height is measured in inches. Self esteem is measured based on the average of 10 1-to-5 rating items (where higher scores mean higher self esteem). Here's the data for the 20 cases:

Table 6. Relationship between height and self esteem

Person	Height	Self Esteem
1	68	4.1
2	71	4.6
3	62	3.8
4	75	4.4
5	58	3.2
6	60	3.1
7	67	3.8
8	68	4.1
9	71	4.3
10	69	3.7
11	68	3.5
12	67	3.2
13	63	3.7
14	62	3.3
15	60	3.4
16	63	4.0
17	65	4.1
18	67	3.8
19	63	3.4
20	61	3.6

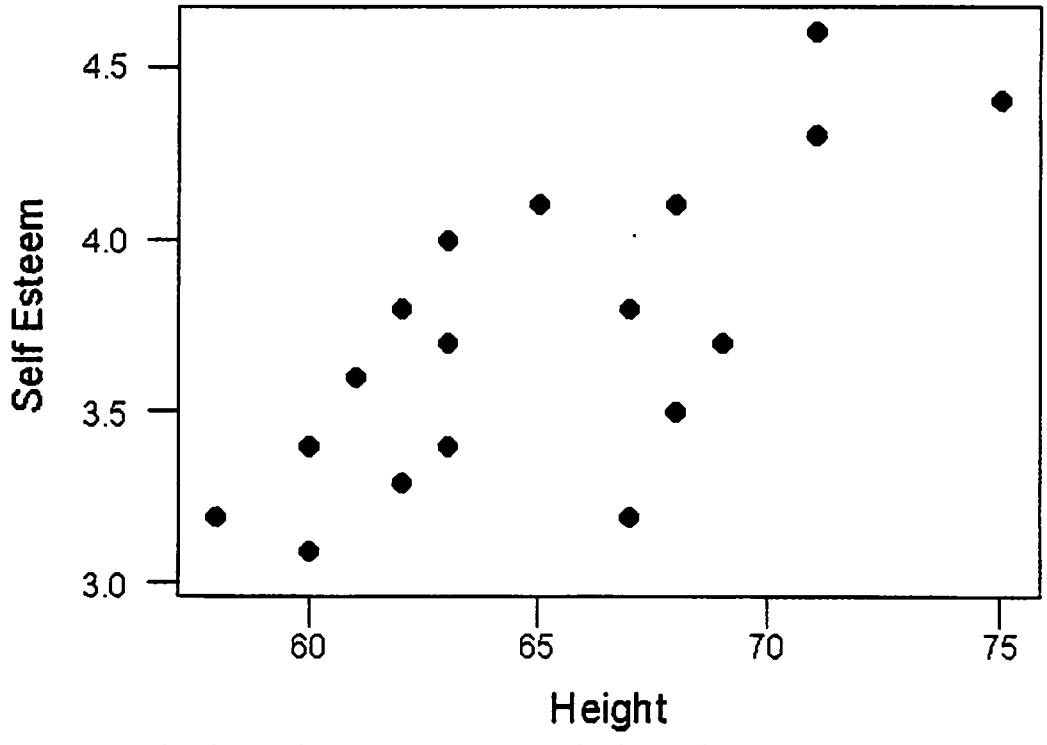
Now, let's take a quick look at the histogram for each variable:



And, here are the descriptive statistics:

Mean	SD	Variance	Sum	Min	Max
65.4	4.406	19.41	1308	58	75
3.755	0.426	0.1815	75.1	3.1	4.6

Finally, we'll look at the simple bivariate (i.e., two-variable) plot:



Since the correlation is nothing more than a quantitative estimate of the relationship, we would expect a positive/negative correlation.

Calculating the Correlation:

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\left[N\sum X^2 - (\sum X)^2\right]\left[N\sum Y^2 - (\sum Y)^2\right]}}$$

N = Number of pairs of scores

We use the symbol r to stand for the correlation. Let's look at the data we need for the formula. Here's the original data with the other necessary columns:

Person	Height (x)	Self Esteem (y)	x*y	x*x	y*y
1	68	4.1	278.8	4624	16.81
2	71	4.6	326.6	5041	21.16
3	62	3.8	235.6	3844	14.44
4	75	4.4	330	5625	19.36
5	58	3.2	185.6	3364	10.24
6	60	3.1	186	3600	9.61
7	67	3.8	254.6	4489	14.44
8	68	4.1	278.8	4624	16.81
9	71	4.3	305.3	5041	18.49
10	69	3.7	255.3	4761	13.69
11	68	3.5	238	4624	12.25
12	67	3.2	214.4	4489	10.24
13	63	3.7	233.1	3969	13.69
14	62	3.3	204.6	3844	10.89
15	60	3.4	204	3600	11.56
16	63	4	252	3969	16
17	65	4.1	266.5	4225	16.81
18	67	3.8	254.6	4489	14.44
19	63	3.4	214.2	3969	11.56
20	61	3.6	219.6	3721	12.96
Sum =	**1308**	**75.1**	**4937.6**	**85912**	**285.45**

The first three columns are the same as in the table above. The next three columns are simple computations based on the height and self esteem data. The bottom row consists of the sum of each column. This is all the information we need to compute the correlation.

Here are the values from the bottom row of the table (where N is 20 people) as they are related to the symbols in the formula:

$$
\begin{aligned}
N &= 20 \\
\Sigma xy &= 4937.6 \\
\Sigma x &= 1308 \\
\Sigma y &= 75.1 \\
\Sigma x^2 &= 85912 \\
\Sigma y^2 &= 285.45
\end{aligned}
$$

Now, when we plug these values into the formula given above, we get the following (I show it here tediously, one step at a time):

$$r = \frac{20(4937.6) - (1308)(75.1)}{\sqrt{[20(85912) - (1308*1308)][20(285.45) - (75.1*75.1)]}}$$

$$r = \frac{98752 - 98230.8}{\sqrt{[1718240 - 1710864][5709 - 5640.01]}}$$

$$r = \frac{521.2}{\sqrt{[7376][68.99]}}$$

$$r = \frac{521.2}{\sqrt{508870.2}}$$

$$r = \frac{521.2}{713.3514}$$

$$r = .73$$

So, the correlation for our twenty cases is .73, which is a fairly strong positive relationship. I guess there is a relationship between height and self esteem, at least in this made up data!

Inferential Statistics

Inferential statistics are used to draw inferences about a population from a sample. Consider an experiment in which 10 subjects who performed a task after 24 hours of sleep deprivation scored 12 points lower than 10 subjects who performed after a normal night's sleep. Is the difference real or could it be due to chance? How much larger could the real difference be than the 12 points found in the sample? These are the types of questions answered by inferential statistics.

There are two main methods used in inferential statistics: estimation and hypothesis testing. In estimation, the sample is used to estimate a parameter and a confidence interval about the estimate is constructed.

In the most common use of hypothesis testing, a null hypothesis is put forward and it is determined whether the data are strong enough to reject it. For any program impact study, the null hypothesis would be that the program has no effect on performance.

Inferential statistics are useful in experimental and quasi-experimental research design or in program outcome evaluation. The simplest inferential test is used to compare the average performance of two groups on a single measure. Whenever we wish to compare the average performance between two groups we should consider the t-test for differences between groups.

Most of the major inferential statistics come from a general family of statistical models known as the General Linear Model (**GLM**) that includes t-test, Analysis of Variance (ANOVA), Analysis of Covariance (ANCOVA), regression analysis, and many of the multivariate methods like factor analysis, multidimensional scaling, cluster analysis, discriminant function analysis, and so on.

The T-Test

The t-test assesses whether the means of two groups are *statistically* different from each other. This analysis is appropriate whenever you want to compare the means of two groups, and especially appropriate as the analysis for the posttest-only two-group randomized experimental design.

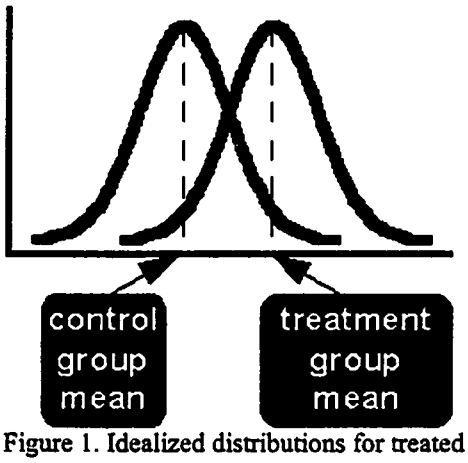


Figure 1. Idealized distributions for treated and comparison group posttest values.

Figure 1 shows the distributions for the treated (blue) and control (green) groups in a study. The figure indicates where the control and treatment group means are located. The question the t-test addresses is whether the means are statistically different.

Now, consider the three situations shown in Figure 2. Notice that three situations that the three situations don't look the same. The top example shows a case with moderate variability of scores within each group. The second situation shows the high variability and the third shows the case with low variability. Clearly, we would conclude that the two groups appear most different or distinct in the low-variability case because there is relatively little overlap between the two bell-shaped curves.

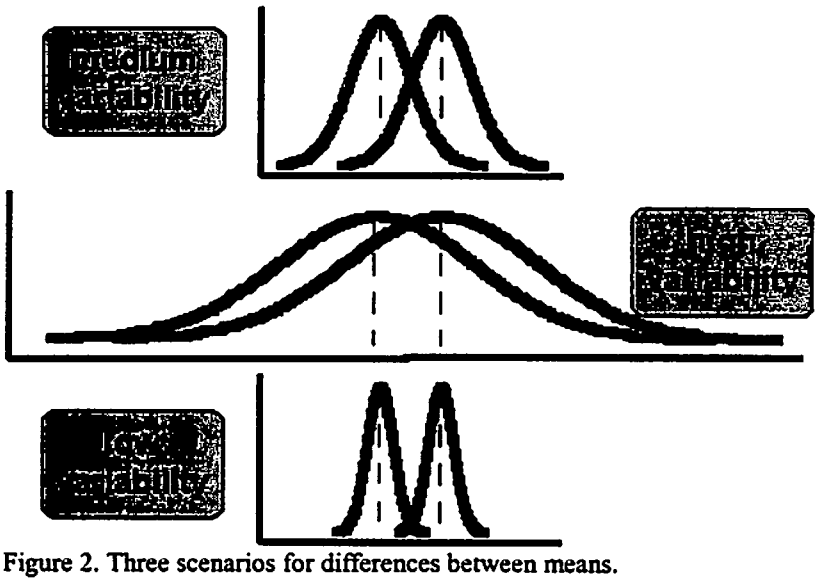


Figure 2. Three scenarios for differences between means.

Statistical Analysis of the t-test

The formula for the t-test is a ratio. The top part of the ratio is just the difference between the two means or averages. The bottom part is a measure of the variability or dispersion of the scores. Figure 3 shows the formula for the t-test and how the numerator and denominator are related to the distributions.

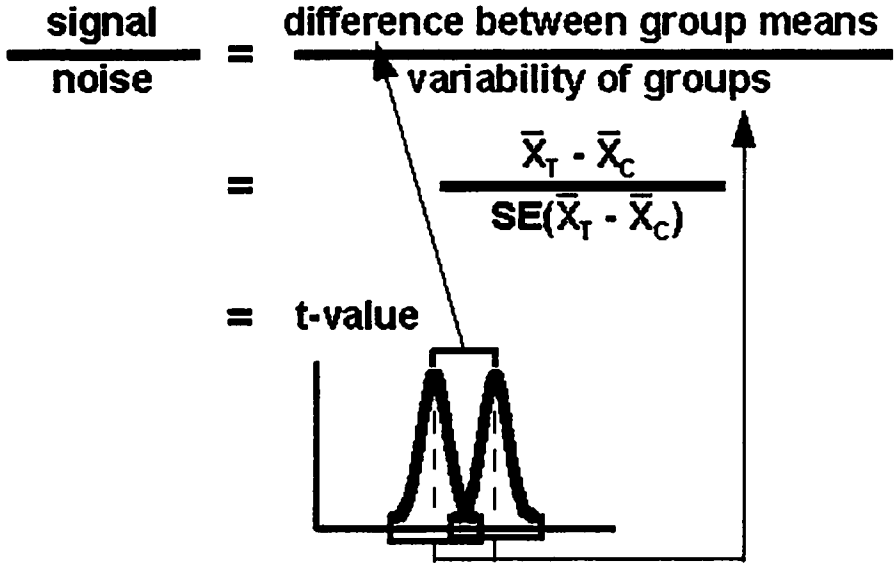


Figure 3. Formula for the t-test.

The top part of the formula is easy to compute -- just find the difference between the means. The bottom part is called the **standard error of the difference**. To compute it, we take the variance for each group and divide it by the number of people in that group. We add these two values and then take their square root. The specific formula is given in Figure 4:

$$SE(\bar{X}_T - \bar{X}_C) = \sqrt{\frac{var_T}{n_T} + \frac{var_C}{n_C}}$$

Figure 4. Formula for the Standard error of the difference between the means.

The final formula for the t-test is shown in Figure 5:

$$t = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{\frac{var_T}{n_T} + \frac{var_C}{n_C}}}$$

Figure 5. Formula for the t-test.

The t-value will be positive if the first mean is larger than the second and negative if it is smaller. Once we compute the t-value we have to look at the table of significance to test whether the ratio is large enough to say that the difference between the groups is not likely to have been a chance finding. To test the significance, you need to set a risk level. In most social research, the "rule of thumb" is to set the alpha level at .05. This means that five times out of a hundred you would find a statistically significant difference between the means.

Interval Estimation

If we want to be reasonably confident that our inference is correct, we cannot claim that μ is precisely equal to the observed $\overline{X}$. Instead, we must construct an interval estimate of confidence interval of the form:

$$\mu = \overline{X} \pm \text{sampling error}$$

The crucial question is how much will be the sampling error. Of course, the sampling error will depend on how much $\overline{X}$ fluctuates. To determine the sampling error, we must determine how confident we wish to be that our interval estimate is right. It is common to choose 95% confidence. At 95% level of confidence Z value is 1.96. Therefore, the confidence interval is:

$$\Pr(\mu - 1.96\frac{\sigma}{\sqrt{n}} < \overline{X} < \mu + 1.96\frac{\sigma}{\sqrt{n}}) = 95\%$$

$$\Pr(\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}) = 95\%$$

For small sample the 95% confidence interval for the population mean is

$$\mu = \overline{X} \pm t_{.025}\frac{s}{\sqrt{n}}$$

Interval estimates of two means from two independent samples

Two population means are commonly compared by forming their difference:

$$(\mu_1 - \mu_2)$$

A reasonable estimate of this is the difference in sample means:

$$(\overline{X}_1 - \overline{X}_2)$$

95% confidence interval for independent samples, when two populations have different and known variances:

$$(\mu_1 - \mu_2) = (\overline{X}_1 - \overline{X}_2) \pm Z_{.025}\sqrt{\frac{\sigma_1}{n_1}} + \sqrt{\frac{\sigma_2}{n_2}}$$

For small samples:

$$(\mu_1-\mu_2)=(\overline{X}_1-\overline{X}_2)\pm t_{.025}\sqrt{\frac{\sigma_1}{n_1}}+\sqrt{\frac{\sigma_2}{n_2}}$$

When population variances are unknown

$$(\mu_1-\mu_2)=(\overline{X}_1-\overline{X}_2)\pm t_{.025}s_p\sqrt{\frac{1}{n_1}}+\sqrt{\frac{1}{n_2}}$$

S_p is the pooled standard deviation. We add up all the squared deviations from both samples, and then divide by the total degrees of freedom in both samples. That is

$$s_p^2=\frac{\sum(X_1-\overline{X}_2)^2+\sum(X_1-\overline{X}_2)^2}{(n_1-1)+(n_2+1)}$$

Regression Techniques

Assumptions

- The regression model is linear in parameters
- X values are fixed in repeated sampling
- Zero mean value of the disturbance term u
- Homoscedasticity or equal variance
- No correlation between the disturbance term
- Zero covariance between u and X's

Let a Population Regression Function (PRF) is given as

$$Y_i = \beta_1 + \beta_2 + u_i$$

However, the PRF is not directly observable. We estimate the PRF from Sample Regression Function (SRF):

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned}$$

where $\hat{Y}_i$ is the estimated values of Y_i. We can express the above expressions as below:

$$\begin{aligned} \hat{u}_i &= Y_i - \hat{Y} \\ &= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \end{aligned}$$

Now given n pairs of observations on Y and X, we would like to determine the SRF in such a manner that it is as close as the PRF. To achieve this, we chose the SRF in such a way that the sum of the squared residuals

$$\sum u_i^2 = \sum \left(Y_i - \hat{Y}_i \right)^2$$

is as small as possible. Minimizing the above expression we get

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}$$

$$\text{var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{\sum x_i^2}$$

$$\text{var}\left(\hat{\beta}_1\right) = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2$$

σ^2 is population variance. However, we can estimate this from the sample data using the following formula:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2}$$

Table 1. Estimation of regression equation, numerical example

X	Y	x	y	xy	x^2
100	40	-550	-29.75	16362.5	302500
200	55	-450	-14.75	6637.5	202500
300	50	-350	-19.75	6912.5	122500
400	70	-250	0.25	-62.5	62500
500	70	-150	0.25	-37.5	22500
600	65	-50	-4.75	237.5	2500
700	72	50	2.25	112.5	2500
800	74	150	4.25	637.5	22500
900	70	250	0.25	62.5	62500
1000	88	350	18.25	6387.5	122500
1100	88	450	18.25	8212.5	202500
650	69.75		$\sum xy = 59350$		$\sum x^2 = 1430000$

$\hat{\beta}_2 = 59350/143000 = 0.415$

$\hat{\beta}_1 = 60.75 - 0.415(650) = 40$

$\hat{Y} = 40 - 0.415X$

Properties of Least Squares Estimators:

1. It is linear function of a random variable, such as the dependent variable Y in the regression model.
2. It is unbiased, that is the average or expected value, $E(\hat{\beta}_2)$, is equal to the true value of β.
3. It has minimum variance in the class of all such linear unbiased estimators, which is efficient estimator.

The Coefficient of Determination (R^2).

We now consider the goodness of fit of the fitted regression line to a set of data. We expect that the residuals around the regression line are as small as possible. The R^2 is a summary measure that tells how well the sample regression line fits the data. We can derive R^2 as follows:

$\sum y_i^2 = \sum (Y_i - \overline{Y})$ = Total variation of the actual Y values about their sample mean, which is called total sum of squires (TSS).

$\sum \hat{y}_i^2 = \left(\sum \hat{Y}_i - \overline{Y}\right)^2$ = Variation of the estimated Y values about their mean, which is explained sum of squires (ESS).

$\sum u_i^2$ = residual or unexplained variation of the Y values about the regression line, or residual sum of squires. Thus

TSS = ESS + RSS, dividing by TSS

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS} = R^2 + \frac{RSS}{TSS}$$

$$1 - R^2 = \frac{RSS}{TSS}$$

$$R^2 = 1 - \frac{RSS}{TSS}$$

Properties of R^2

1. It is a positive quantity
2. Its limits are $0 \leq R^2 \leq 1$. An R^2 of 1 means a perfect fit, that is $\hat{Y}_i = Y_i$ for each *i*.

7A.6 SAS OUTPUT OF THE EXPECTATIONS-AUGMENTED PHILLIPS CURVE (7.6.2)

DEP VARIABLE: Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB > F
MODEL	2	97.334119	48.667060	35.515	0.0001
ERROR	10	13.703158	1.370816		
C TOTAL	12	111.037			

ROOT MSE	1.170605	R-SQUARE	0.8766
DEP MEAN	7.756923	ADJ R-SQ	0.8519
C.V.	15.0911		

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HO: PARAMETER = 0	PROB > |T|
INTERCEP	1	7.193357	1.594789	4.511	0.0011
X2	1	−1.392472	0.305018	−4.565	0.0010
X3	1	1.470032	0.175736	8.363	0.0001

COVARIANCE OF ESTIMATES

COVB	INTERCEP	X2	X3
INTERCEP	2.543353	−0.388917	0.02241163
X2	−0.388917	0.09303593	−0.0344189
X3	0.02241163	−0.0344189	0.03090064

CBS	Y	X2	X3	YHAT	YRESID
1	5.92	4.9	4.78	7.3970	−1.4770
2	4.30	5.9	3.84	4.6227	−0.3227
3	3.30	5.6	3.13	3.9967	−0.6967
4	6.23	4.9	3.44	5.4272	0.8028
5	10.97	5.6	6.84	9.4505	1.5195
6	9.14	8.5	9.47	9.2785	−0.1385
7	5.77	7.7	6.51	6.0412	−0.2712
8	6.45	7.1	5.92	6.0094	0.4406
9	7.60	6.1	6.08	7.6371	−0.0371
10	11.47	5.8	8.09	11.0096	0.4604
11	13.46	7.1	10.01	12.0218	1.4382
12	10.24	7.6	10.81	12.5016	−2.2616
13	5.99	9.7	8.00	5.4466	0.5434

DURBIN-WATSON *d*	2.225
1ST ORDER AUTOCORRELATION	−0.203

Notes: The numbers under the heading PROB> |T| represent *p* values.
See Chapter 12 for a discussion of the Durbin-Watson *d* statistic and of first order autocorrelation.

7A.7 SAS OUTPUT OF THE COBB-DOUGLAS PRODUCTION FUNCTION (7.10.4)

DEP VARIABLE: Y1

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB > F
MODEL	2	0.538038	0.269019	48.069	0.0001
ERROR	12	0.067153	0.005596531		
C TOTAL	14	0.605196			

ROOT MSE	0.074810	R-SQUARE	0.8890
DEP MEAN	10.096535	ADJ R-SQ	0.8705
C.V.	0.7409469		

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HO: PARAMETER = 0	PROB > |T|
INTERCEP	1	−3.338455	2.449508	−1.363	0.1979
Y2	1	1.498767	0.539803	2.777	0.0168
Y3	1	0.489858	0.102043	4.800	0.0004

COVARIANCE OF ESTIMATES

COVB	INTERCEP	Y2	Y3
INTERCEP	6.000091	−1.26056	0.1121951
Y2	−1.26056	0.2913868	−0.0384272
Y3	0.01121951	−0.0384272	0.01041288

Y	X2	X3	Y1	Y2	Y3	Y1HAT	Y1RESID
16607.7	275.5	17803.7	9.7176	5.61859	9.7872	9.8768	−0.15920
17511.3	274.4	18096.8	9.7706	5.61459	9.8035	9.8788	−0.10822
20171.2	269.7	18271.8	9.9120	5.59731	9.8131	9.8576	0.05437
20932.9	267.0	19167.3	9.9491	5.58725	9.8610	9.8660	0.08307
20406.0	267.8	19647.6	9.9236	5.59024	9.8857	9.8826	0.04097
20831.6	275.0	20803.5	9.9442	5.61677	9.9429	9.9504	−0.00615
24806.3	283.0	22076.6	10.1189	5.64545	10.0023	10.0225	0.09640
26465.8	300.7	23445.2	10.1836	5.70611	10.0624	10.1428	0.04077
27403.0	307.5	24939.0	10.2184	5.72848	10.1242	10.2066	0.01180
28628.7	303.7	26713.7	10.2622	5.71604	10.1929	10.2217	0.04051
29904.5	304.7	29957.8	10.3058	5.71933	10.3075	10.2827	0.02304
27508.2	298.6	31585.9	10.2222	5.69910	10.3605	10.2783	−0.05610
29035.5	295.5	33474.5	10.2763	5.68867	10.4185	10.2911	−0.01487
29281.5	299.0	34821.8	10.2847	5.70044	10.4580	10.3281	−0.04341
31535.8	288.1	41794.3	10.3589	5.66331	10.6405	10.3619	−0.00299

	COLLINEARITY DIAGNOSTICS		VARIANCE PROPORTIONS		
NUMBER	CONDITION EIGENVALUE	PORTION INDEX	PORTION INTERCEP	PORTION Y2	Y3
1	3.000	1.000	0.0000	0.0000	0.0000
2	.000375451	89.383	0.0491	0.0069	0.5959
3	.000024219	351.925	0.9509	0.0031	0.4040

DURBIN-WATSON *d*	0.891
1ST ORDER AUTOCORRELATION	0.366

Management and Analysis of Household Survey Data

Prepared for the:

Training Workshop on Resource Economics and Research Methods for the Officers of IFReDI, DOF, Ministry of Agriculture, Fisheries and Forestry, Royal Government of Cambodia

23 February – 03 March 2004

Prepared by:

Ferdinand Javien Paraguas

f.paraguas@cgiar.org

http://www.ferdinandjparaguas.tripod.com/

WorldFish Center

Penang, Malaysia

Management and Analysis of Household Survey Data

Table of Contents

Management and Analysis of Household Survey Data

Introduction

This document is prepared as an introductory material for the topic on **database management and data analysis** of the training **"Training Workshop on Resource Economics and Research Methods for the Officers of IFReDI, DOF, Ministry of Agriculture, Fisheries and Forestry, Royal Government of Cambodia"** (23 February – 03 March 2004, Penang, Malaysia). As an introductory material, it provides a general overview of the topic.

It begins with the discussion of the different components of data management. The importance of data management is also emphasized. Along with this material are three separate documents illustrating the efficient use of software in database development and data management and analysis, namely:

1. The Tonle Sap Economic Valuation Database: An introduction to MS Access.
2. Using Spreadsheet (MS Excel) as a Tool for Database Management and Initial Data Processing; and
3. Analysis of Household Survey Data with SPSS

Each of the above mentioned material contained detailed illustrations supported with screen shots for easy understanding. All illustrations made use of the Tonle Sap economic valuation database - the database of the ongoing project on "Economic Valuation of Aquatic Resources of Tonle Sap Basin" which is being implemented by the Institute in collaboration of the World Fish Center.

The last section of this material discusses the structure of the said database. The list of Field and Tables names is provided as an Appendix along with the field data collection sheet (questionnaire) use by the project in collecting the data.

In general, the topic aims to give participants a general understanding of data management and its importance and to equip them with the necessary skills to efficient database development, data management and analysis.

Data Management

Most research projects involve the collection of a large volume of data through survey. For the data to be processed and analyzed efficiently, they have to be entered into the computer, checked, and organized into an appropriate form for easy retrieval and analysis. These whole process requires a well-defined data entry system or in general, system of database management.

Data management refers to any activity concerned with looking after and processing this voluminous data to generate summary information. It involves some of the following steps:

- Planning data management for a project, taking into account the objectives and planned outputs, resource and skills available
- Designing field data recording sheets (questionnaire)
- Collection of data
- Checking of raw data
- **Data entry and organization of computer files**
- Backup of data files
- **Processing of data for analysis**
- Checking of processed data
- Maintenance of a data processing log
- Archiving data for future use.

The Importance of Data Management

Data management is a key area in any research and if not done well can limit the usefulness of the data. Below are list of some importance of data management.

- A well-managed datasets ensures validity of the information that can be extracted from it resulting in incorrect conclusion- garbage in garbage out. If proper attention is not paid to data entry system, it is too easy to make a mistake in data entry.

- A well-defined database structure facilitates efficient processing. Data processing and analysis proceeds most efficiently if the data have been managed well.

- Well-documented datasets can be accessible in the future, increasing their half-life and adding value. Field data is expensive to collect so must considered as valuable, but its value is maintained only if it can be used in the future. A computer file full of numbers, the exact meaning or origin of which is not known, is worthless.

The Scope of the Topic

In this topic, we will limit our discussion on two specific components of data management: (1) **Data Entry and Organization of Computer Files** and (2) **Processing of Data for Analysis**. To facilitate our discussion, we will be using the Database of the project: Tonle Sap Economic Valuation Database, which was developed using Microsoft

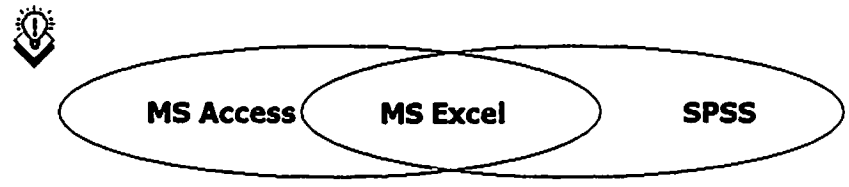
Access. Here, we discussed the different components of the database followed by detailed illustration on how each components of the database was developed. A training material was prepared for this topic: **"The Tonle Sap Economic Valuation Database: An Introduction to MS Access".**

In view of the fact that people are more familiar with Spreadsheet than Database Software, we developed a proto-type MS EXCEL version of the database and illustrate how a spreadsheet software like MS EXCEL can be used efficiently as data entry system and to do initial tabulation of data. Another training material was prepared for this topic: **Using MS Excel as a Tool for Database Management and Initial Data Processing.** We first emphasized that an erroneous data sets as a result of a poor data entry is likely to happen if MS Excel is not used efficiently.

Both versions (Access and the proto-type in Excel) enforced validation checks and provide validation messages to ensure that the sort of correct information is being entered. While we illustrate how to create query using the **Query Expression Builder (QEB)** to extract a subset of data, aggregate values of variables, merge variables; and to generate summary information in MS Access, we illustrate using MS Excel on how **Filter** works to generate subsets of data; how to merge variables using the **Vlookup** function; how to automate aggregating variables and creating cross tabulations using the **Pivot Table.**

In the document, we also mentioned the **Data Analysis** Add-ins of Excel and pointed out its limitations. Thus we recommend against the use of MS Excel when it comes to advance data analysis, which involves statistics and that proper statistical packages should be used.

This leads to another material: **"Analysis of Household Survey Data with SPSS"** which illustrates several data management and statistical techniques using the Tonle Sap Economic Valuation data. This includes: Merging two datasets/tables, Create new variables using compute and recode, Split file Filtering subsets of data, Aggregate variables, Two the data from the Tonle Sap Economic Valuation Database, Independent-Samples T Test, Paired-Samples T Test, One-Way Anova, Multiple Regression Analysis and Logistic Regression.



Relational Database management system	Most familiar, widespread and flexible **Jack of all trade-master of none**	One of the most user friendly statistical package

Database software such as MS Access and spreadsheet packages such as MS Excel are the two most common software used for data management of project surveys. MS Access is a relational database management system, which allows to link together data stored in more than one table, but lacks data analysis/statistical capability.

MS Excel to some extent can do something - though with limitations. Spreadsheet packages in general are the simplest type of package to use. They are often automatically chosen for data entry because they are familiar, widespread and flexible - but their very flexibility means that they can result in poor data entry and management. Excel offers limited form for data entry, non- relational database and does not enforce entity (primary key) or referential integrity (all of which will be discussed in Access).

Strategy

The topic is consists of 11 sessions where each session will last for 1 hour and 45 minutes. A short presentation will come first followed by illustration on "how to do it". Each of the participants will be provided a computer for them to follow what is being illustrated. This will give them real hands-on training on each of the sub topics. For this purpose, we provided the electronic copy of the necessary files to be used. These files are stored in each of the participant's computer under the folder **C:\TonleSap.** In some cases, participants are required to create files. In addition, unguided exercises will be provided at the end of each sub topic if time permits.

The Structure of Tonle Sap Economic Valuation Database

Prepared primarily for IFReDI officers, the topic made use of the data they collected in collaboration with the WorldFish center: Economic Valuation of Aquatic Resources of Tonle Sap Basin. The field collection sheet of the said study is presented Appendix A. The collected data were stored in a MS Access database: **Tonle Sap Economic Valuation.MDB**. Note that Appendix A also contains the assigned field/names (in blue) and the names of the Tables (in red) that are used in the Database (Fig 1). The database will be discussed in more detail under "**The development of Tonle Sap Economic Valuation Database: An introduction to MS Access**". A short description of the organization and structure of the data however, is deemed important at this stage.

The database comprised of 22 tables. All items in the questionnaire bearing only one response are stored in **Tbl 1 Master Table**. All items with multiple answers are stored in separate Tables. These Tables are related to **Tbl 1 Master Table.** Examples of such Tables are **Tbl 2 item8 Income, Tbl8 item40 Fish Catch, Tbl 15 item 65 Fingerlings**. Note that **Tbl** pertains to "**Table**", *item number* pertains to the corresponding item number in the questionnaire and the words *Income* and *Fish Catch* are key words of the corresponding item.

The 22 Tables are of different **Levels**. **Tbl 1 Master Table** contains household level information thus it is *Household Level*. **Tbl2 item8 Income** contains income of household by **source** thus it is *Income Level*. **Tbl8 item40 Fish Catch and Tbl15 item65 Fingerlings** contains catch and stocking information respectively of household by species thus they are *Species Level*. Hence, Tables in the database formed a hierarchical structure where **Tbl 1 Master Table** resides at the top of this structure and all other 21 Tables resides at the next level. Figures 2 and 3 below illustrates this hierarchical structure.

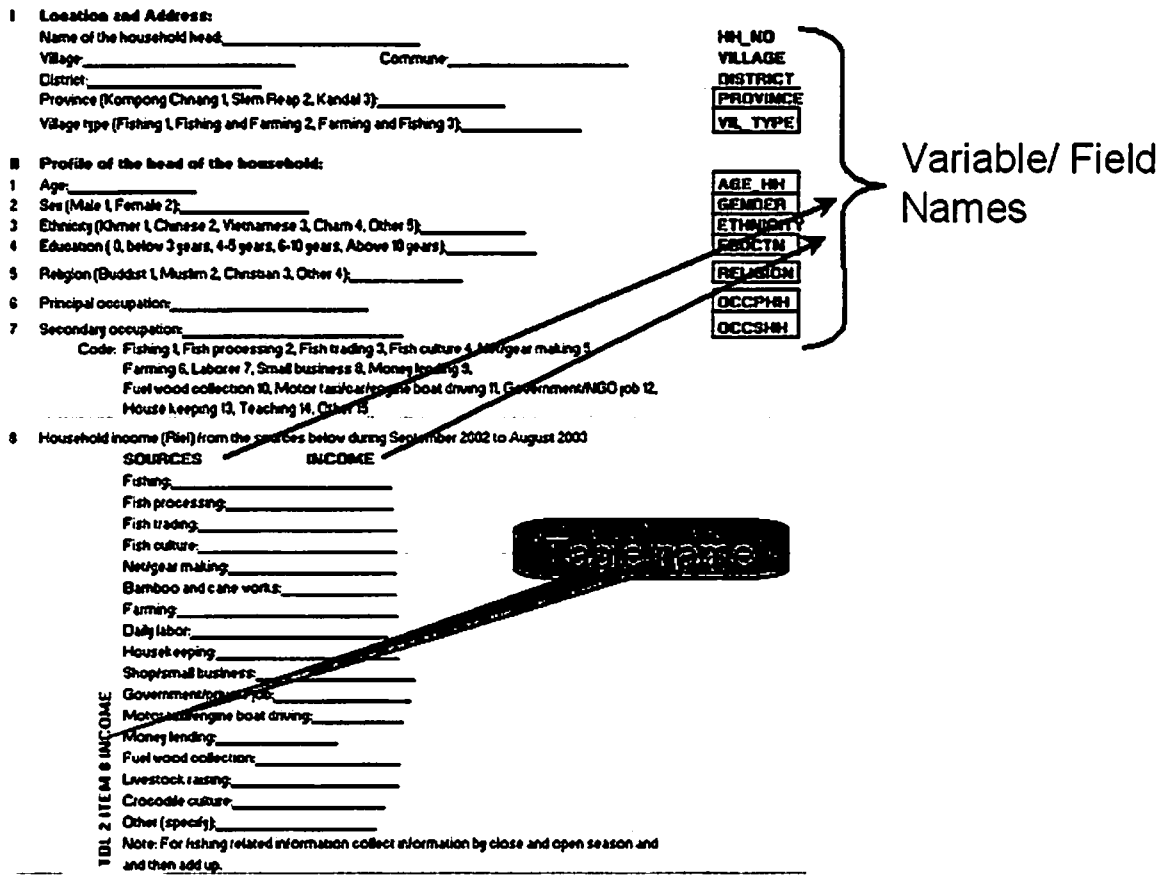
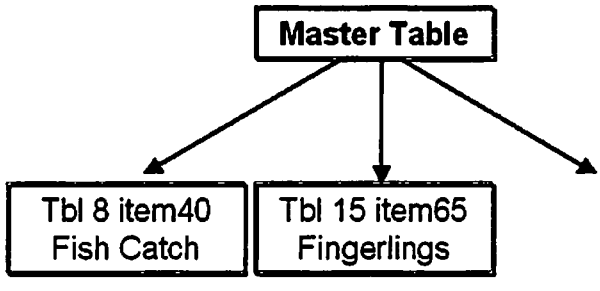


Fig.1. Screen shot of the first page of the data collection sheet bearing the assigned variable/field names and name of the Tables used in the Database.



One record exist for each household in a particular village type and particular province.

HOUSEHOLD + PROVINCE + VILLAGE TYPE is the unique identifier.

One record exist for each species within household in a particular village type and particular province and include a grid reference to a Species.

HOUSEHOLD + PROVINCE + VILLAGE TYPE + SPECIES is the unique identifier.

Fig.2. Hierarchical structure of the Tonle Sap Economic Database.

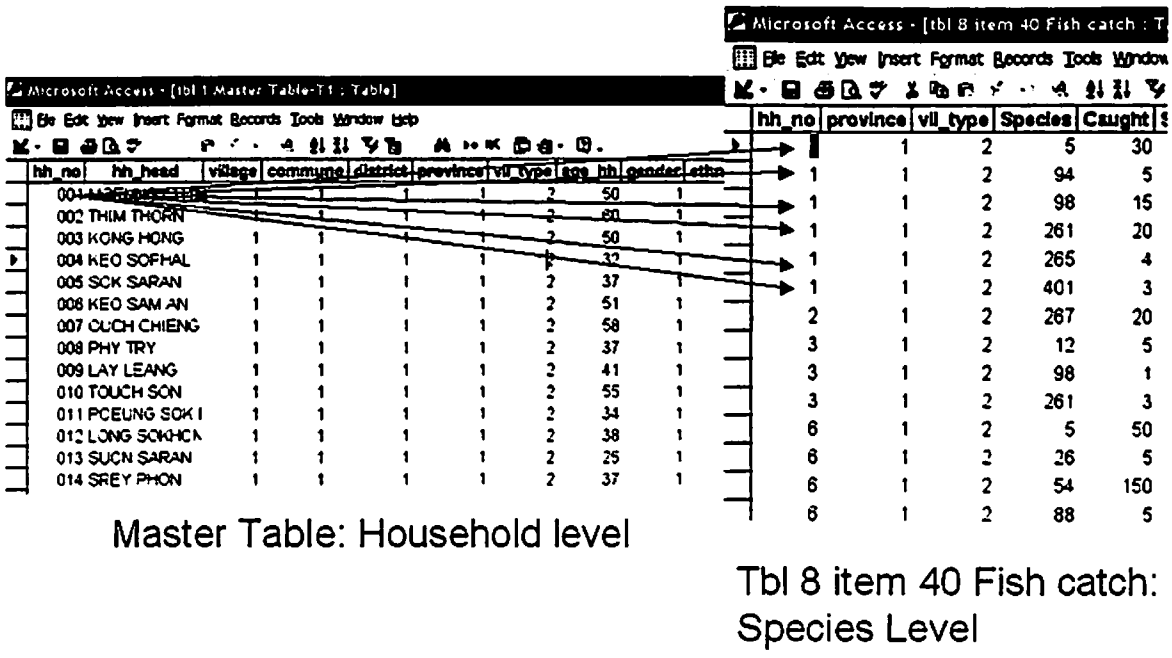


Fig.3. Relating Master Table (Household Level) with Tbl8 item40 Fish Catch (Species level)

The database contained a total of 410 households.

The Tonle Sap Economic Valuation Database: An Introduction to MS Access

Prepared for the:

Training Workshop on Resource Economics and Research Methods for the Officers of IFReDI, DOF, Ministry of Agriculture, Fisheries and Forestry, Royal Government of Cambodia

23 February – 03 March 2004

Prepared by:

Ferdinand Javien Paraguas

f.paraguas@cgiar.org

http://www.ferdinandjparaguas.tripod.com/

WorldFish Center

Penang, Malaysia

The Tonle Sap Economic Valuation Database: An Introduction to MS Access

TABLE OF CONTENTS

The Tonle Sap Economic Valuation Database: An Introduction to MS Access

Introduction

In the previous documents: *Management and Analysis of Household Survey Data*, we discussed the importance of data management and present some steps in data management. We mentioned that our focus is on two steps: (1) Data Entry and Organization of Electronic Data files and (2) Processing of Data Analysis. This document is prepared to discuss these steps.

Prepared as a training material for the ADB-IFReDI project, we will make use of the project database Tonle Sap Economic Valuation- data entry system which was developed using MS Access.

We will discuss each of the components (Tables, Queries, Forms, Reports) of the said database followed by an illustration of how to build/design such components. Specifically, we will illustrate how the IFReDI data entry system was developed. Specifically, we will illustrate how to:

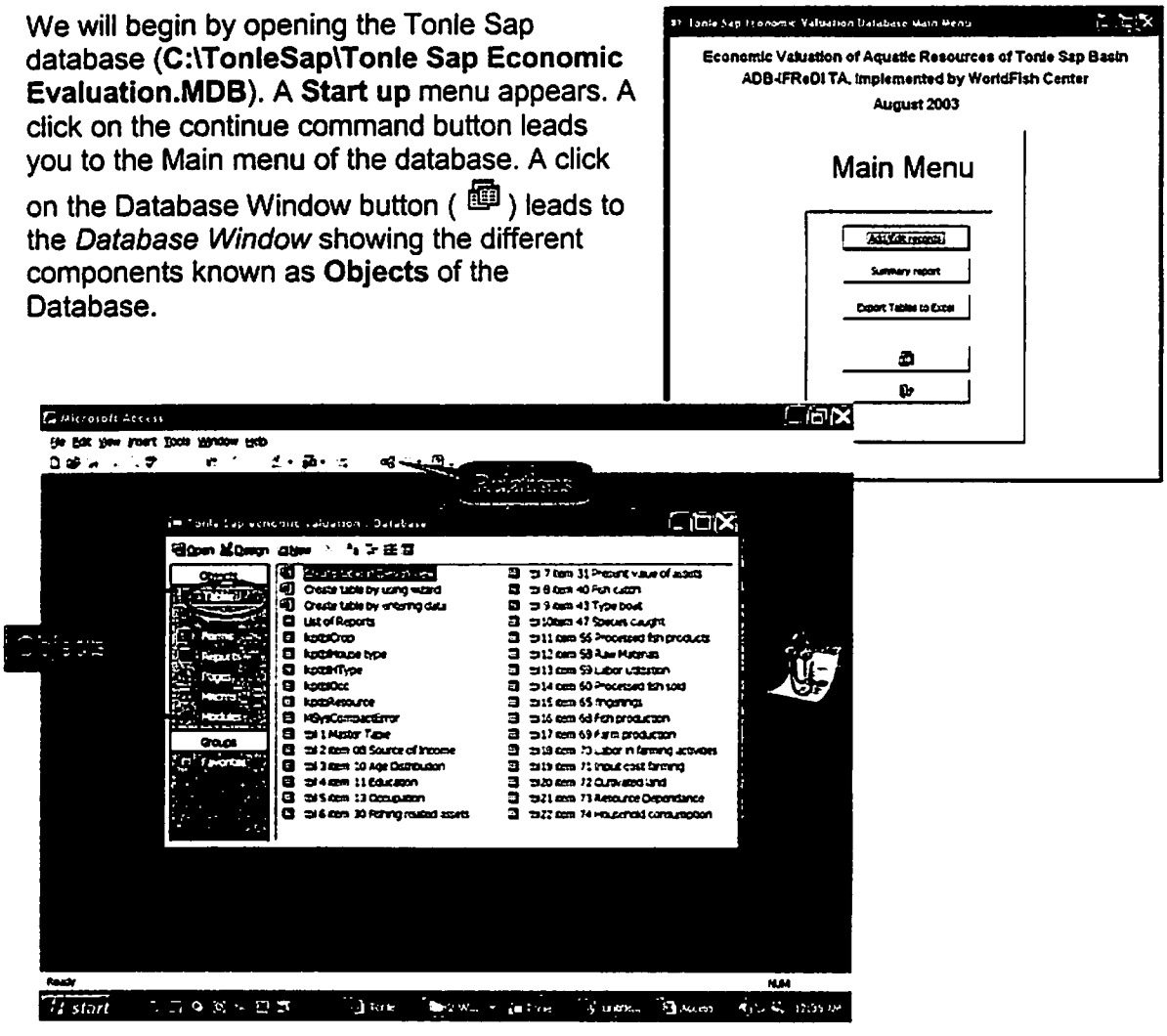
1. Relate Tables and set Primary Keys;
2. Set Validation Rules and Validation Text to ensure that the correct sort of information is being entered by accepting only entries within a specified range;
3. Extract subsets of data, aggregate data (into household level) and generate summaries using Query Expression Builder (QEB);
4. Design Data Entry Forms from scratch which will include creating pull down choice list and command buttons; and
5. Present summary information (initial analysis) using Reports.

Components of MS Access Database- the Case of the Tonle Sap Data Entry System

In this section, we will review the components of an MS Access database. For illustration, to facilitate our discussion we will be using the **Tonle Sap Economic Evaluation.MDB** (hereafter referred to Tonle Sap Database).

A documentation of the Tonle Sap Economic Valuation (questionnaire) is given in **Appendix A** of the document *Management and Analysis of Household Survey Data*. Attached to each item of the questionnaire is the variable/field name used in the database. All items that take only one response are held in a table name **Tbl1 Master Table**. For items that resemble a table format are assigned to a separate table. For example item 8 of the questionnaire in page 1 (Income sources of household respondents) is assigned to a separate table name **Tbl 2 item8 Income**. **Tbl** stands for table, the 2 implies this is the 2nd table in the Database, item number pertains to the corresponding item number in the questionnaire and Income is the corresponding keyword.

We will begin by opening the Tonle Sap database (**C:\TonleSap\Tonle Sap Economic Evaluation.MDB**). A **Start up** menu appears. A click on the continue command button leads you to the Main menu of the database. A click on the Database Window button ( ) leads to the *Database Window* showing the different components known as **Objects** of the Database.



Database Window

The *Database Window* controls navigation within a particular database. A database is made up of several *objects*, grouped into a single file:

- **Tables** - hold the raw data
- **Queries** - extract part of the raw data to produce *dynasets* - dynamic sets of data which can change each time the query is run. Dynaset could be subsets of summary information
- **Forms** - determine how the data is displayed on the screen
- **Reports** - output files, ready for printing
- **Pages** - for creating/editing WWW pages
- **Macros** -lists of commands to perform particular functions

- **Modules** - programs which expert users write in a programming language called Access Basic to perform tailor-made functions not generally available

Macros and **modules** can be used to automate tasks and pull the other objects together into a user-friendly database application. We will limit our discussion in first four objects: Tables, Queries, Forms and Reports. The objects are accessed using the *Tab buttons* down the left of the *Database Window.*

Tables

Begin by investigating the **Tables**. A click on the **Tables** tab will list all the Tables in the right hand side of the Window. In this database, there are 21 **relational tables**. **Tbl 1 Master Table** is a flat table containing all items in the questionnaire with single responses. Thus this is **household level table**. Tables with "**item**" are structured tables, which are linked to **Tbl 1 Master Table** via a primary key, which will be discussed in more detail later. **Tbl 8 item40 Fish Catch** pertains to table number 8 containing information on item number 40 of the questionnaire which are fish catch related information. This is a **species level** table.

Tables with **lkptbl** are tables whose entries are used as inputs in other tables. For example, **lkptblCrop** contains the list of all crops name that are could be entries in tables **Tbl17 item69 Farm Production** and other tables related to farming.

The table **List of Reports** is special table containing all the name of reports in data database.

Lets investigate the table named **Tbl 1 Master Table** further. To examine the content of the table, check the name is highlighted, then press **<Return>** or click on **[Open]**. You can also open a table by *double clicking* on its name.

A new screen, the *Table Window*, appears revealing the data set out in a table. This method of display (known as *Datasheet View*) shows the data in columns and rows, similar to a spreadsheet.

Immediately below the data is the *status bar*, which tells you that you are positioned at Record 1 (of 410). The *current record* is indicated by an arrow in the column to the left of the data. You can move the indicator down to the next record by clicking on the button immediately to the right of the number 1 on the status bar. The next button to the right takes you to the end of the table - click on this and you should be at Record 410. Matching buttons on the left take you back a single record and back to Record 1 - try out these too. You can also move up and down using the arrow keys on the keyboard. The scroll bar down the right edge of the table window moves the display up and down.

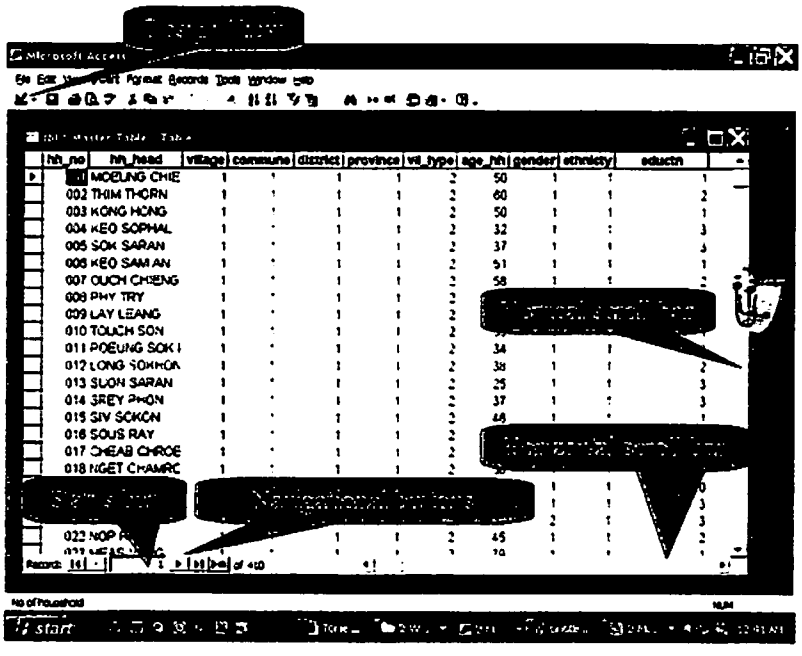
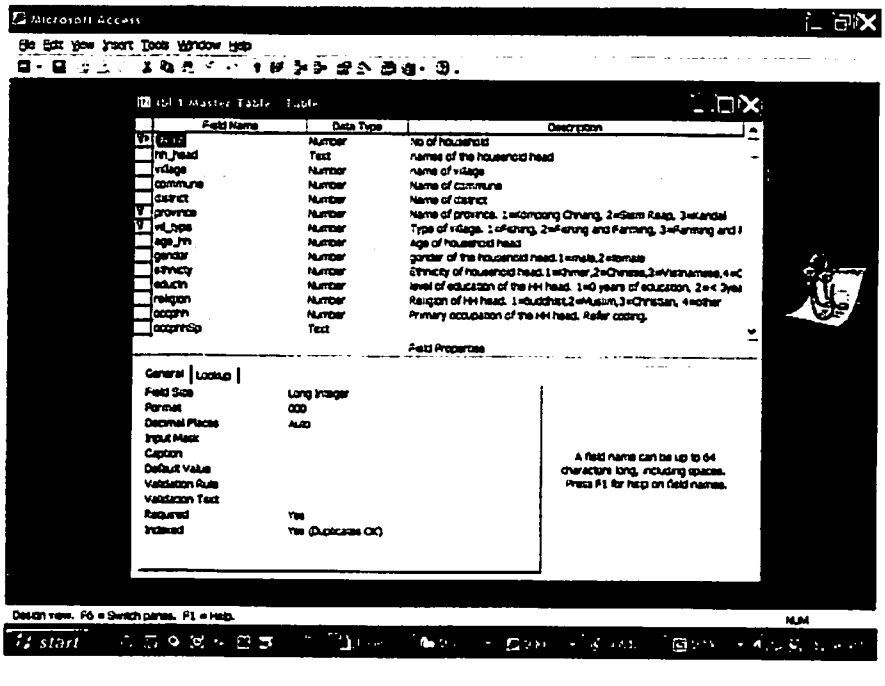


Table Design View

To see exactly what each record contains and how it has been set up:

Click on the **[View]** button ( ) on the toolbar, or use **Design View** in the **View** menu.

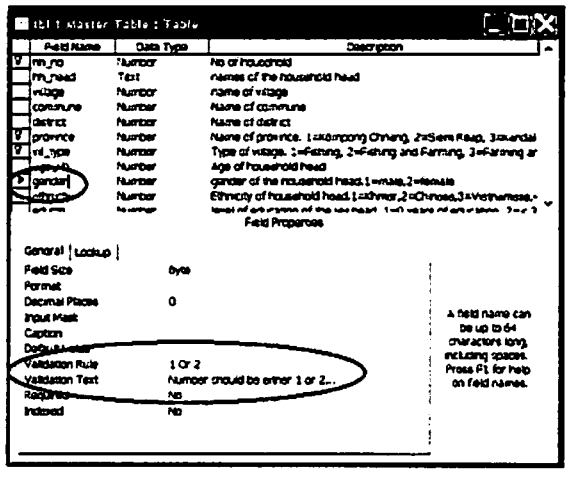
The *Table Design Window* lists the field names, indicates their data types and also shows the *field properties*. In the next section, we will illustrate in detail the setting-up of this table.



Primary key

Note the key sign (⚷) at the left of each of the fields **hh_no**, **Province** and **Vil_type**. The presence of this sign in these fields indicates that these fields are the primary key. Primary key is either a single field or a combination of fields, which acts as a unique identifier. The primary key is always a unique- MS Access does not allow for duplicates of entry in the primary key. In many cases, the choice of primary key is obvious.

Field Properties



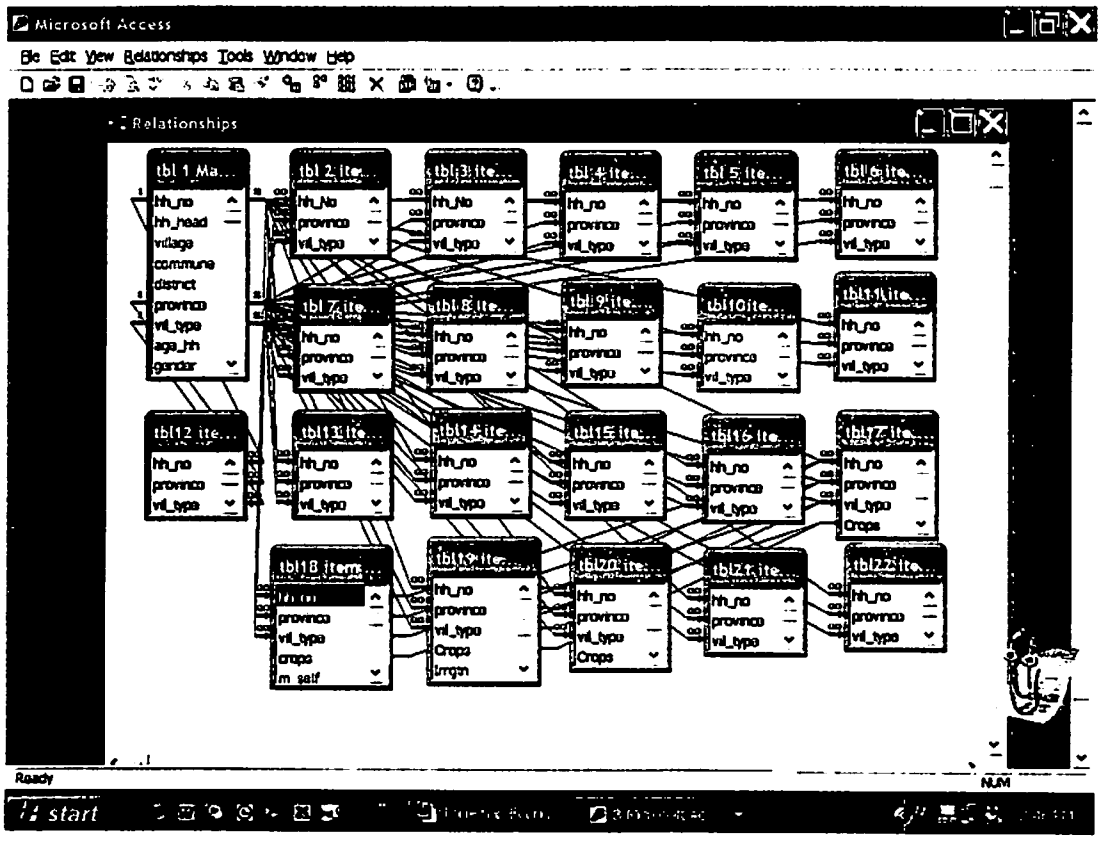
At the bottom of Table Design Window is the **Field Property Box.** Field Property Dialog Box allows you to set properties of a particular field. Properties include the **Field Size** (for a numeric Field Size includes integer, byte, decimal, etc) Format, Caption, etc. Most Properties have **list arrow** button (▾) attached on them for you to select an entry. A description of the Filed Property is also provided in the right hand side of the Field Property box (written in blue color). Take particular attention to **Validation Rule** and **Validation Text.** With **Validation Rule**, you can impose validation checks to accept only entries within a specified range, so that typing data outside the minimum and maximum values is prohibited. This is to ensure that the correct sort of information is being entered. **Validation Text** is the corresponding message that will pup-up once an invalid entry is key-in. This is equivalent to Error Alert in MS Excel.

One can also impose **Validation Rule** and **Validation Text** on the same fields during **Form Design**. However, it is a good practice to set them during **Table Design.**

Relationships.

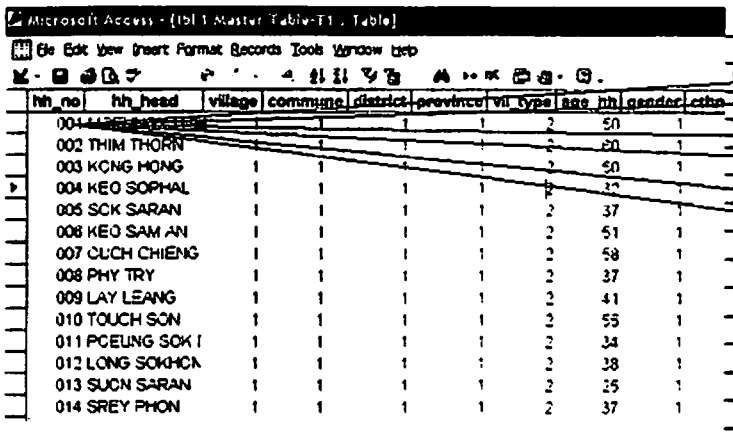
As mentioned earlier, the tables in the database are linked together. These links can be built into the design in Access, through primary key. For tables 1 and 2 to be linked, the primary key of table 1 should be present in table 2. In other words the fields should be present in both tables. This linkage is called *relationship*. To view on how the tables in the database are linked together, close the table to go back to the database window, click on relationship command button (⧉) below the menu bar, or click **Tools** menu then choose **Relationships**.

The *Relationship window* appears.



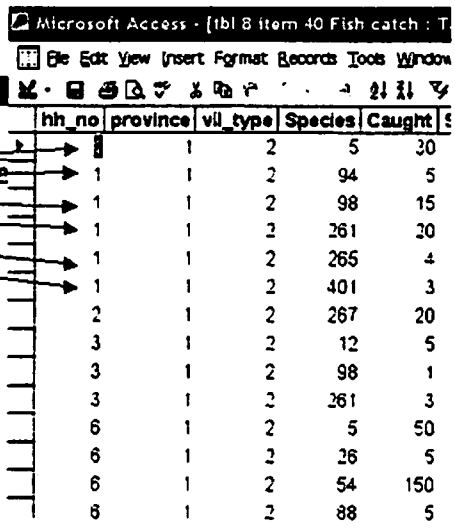
Each mini box represents a data table in the database. Inside of each mini box is the list of the corresponding fields. The tables are connected/linked to the **Tbl 1 Master Table**. This indicates relationship. This type of relationship is "*one to many*" in that a single record in Tbl 1 Master Table is related to potentially many records in other tables. Normally, when the structure is hierarchical, then the relationship is "one to many". **Access** automatically determines the relationship once you assign which fields define the relationship. Sometimes **Access** will automatically define the relationship.

For illustration purposes, we superimposed tables **Tbl 1 Master Table** and **Tbl8 item 40 Fish Catch.** In **Tbl 1 Master Table** there is only one record with **hh_no** (household) equal to 1 in **Province** 1 and 2 and while there are 6 records in **Tbl8 item 40 Fish Catch.** Needless to say these 6 records pertains to one household which is the same household in **Tbl 1 Master table.**



hh_no	hh_head	village	commune	district	province	vil_type	age_hh	gender	ethn
004 [illegible]		1	1	1	1	2	50	1	
002 THIM THORN		1	1	1	1	2	60	1	
003 KONG HONG		1	1	1	1	2	60	1	
004 KEO SOPHAL		1	1	1	1	2	37	1	
005 SOK SARAN		1	1	1	1	2	37	1	
006 KEO SAM AN		1	1	1	1	2	51	1	
007 CUCH CHIENG		1	1	1	1	2	58	1	
008 PHY TRY		1	1	1	1	2	37	1	
009 LAY LEANG		1	1	1	1	2	41	1	
010 TOUCH SON		1	1	1	1	2	55	1	
011 PCEUNG SOK I		1	1	1	1	2	34	1	
012 LONG SOKHON		1	1	1	1	2	38	1	
013 SUON SARAN		1	1	1	1	2	25	1	
014 SREY PHON		1	1	1	1	2	37	1	

Master Table: Household level

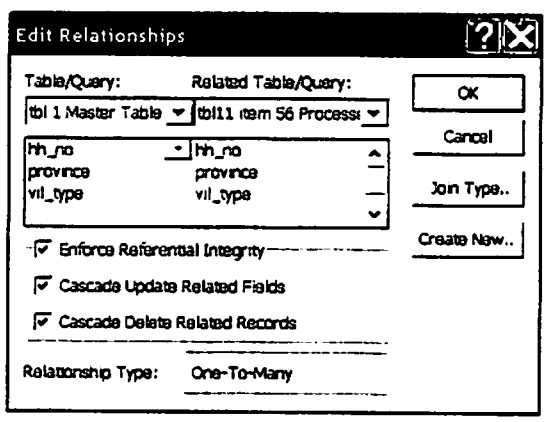


hh_no	province	vil_type	Species	Caught
1	1	2	5	30
1	1	2	94	5
1	1	2	98	15
1	1	2	261	20
1	1	2	265	4
1	1	2	401	3
2	1	2	267	20
3	1	2	12	5
3	1	2	98	1
3	1	2	261	3
6	1	2	5	50
6	1	2	26	5
6	1	2	54	150
6	1	2	88	5

Tbl 8 item 40 Fish catch: Species Level

Note that there also exit another type of relationship between **TBL18 item70 Labor in Farming, TBL19 Item12 Input Cost Farming** and **TBL20 Item72 Cultivated Land**, are all connected to **TBL17 Item69 Farm Production**, via the fields, **hh_no, province, vil_type and Crops**. This type of relationship is one to one.

You can view the relationship of two tables by double clicking in one of the lines that connect them. This will open the *Edit Relationship dialog box*, showing the Tables involved and the fields that defined the relationship. Note that **referential integrity** was also enforced in these tables. **Referential integrity** helps validate relationships by not allowing encoder to enter a record in a table on the "many" side of a relationship where there is no corresponding record in the **Tbl 1 Master Table** ("one" side). For example, with referential integrity, you would not be able to enter details of stocking (Tbl8 item40 Fish Catch) without going through page 1 first (until there was a household for that stocking). Also, with referential integrity, deleting records in Tbl 1 Master Table will delete all related records in all tables.



Unlike a spreadsheet, where 7 tables with data would be confusing, this is quite a simple structure for a database. A database would typically have between 5 and 50 tables.

Learning by doing – Table construction

Tables can be constructed in many ways: (a) *Datasheet View* creates a dummy datasheet; (b) *Design View* lets you do everything yourself, from scratch; (c) *Table Wizard* has predefined *Sample Tables* (e.g. Employees, Addresses, Inventories) for both Business and Personal application and (d) *Import Table* and *Link Table* allow you to import data from or link the table to another file, respectively. This file could be an Excel spreadsheet.

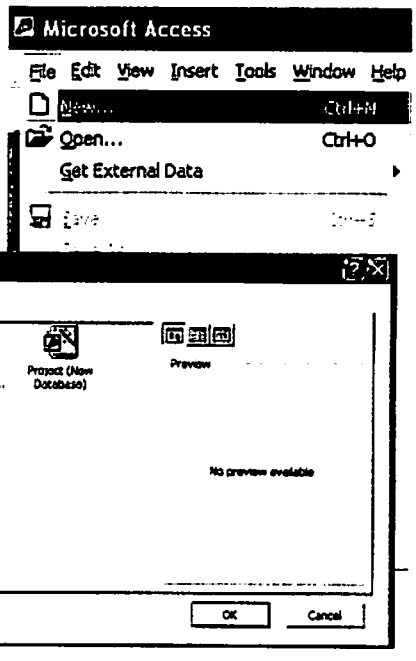
In this next section, we will illustrate how to create Tables. To understand the fundamentals of table design, we will design Tables from scratch using Design View. Specifically, we will create a new database. Then create a Table with the following fields.

Item number	Data type	Remarks
1. HH_No	Number (Long Integer)	Part of Primary key
2. PROVINCE	Number (Long Integer)	Part of Primary key
3. VIL_TYPE	Number (Long Integer)	Part of Primary key Coded: 1 if Fishing 2 if Fishing and Farming 3 if Farming and Fishing Set Validation Rule and Validation Text
4. AGE	Number (Integer)	
4. GENDER	Number (Binary)	Coded: 1 if Male 2 if Female Set Validation Rule and Validation Text

In order for us to illustrate making relationships among Tables, an exercise to create another Table will be given. This will also allow us to illustrate how to create subform later.

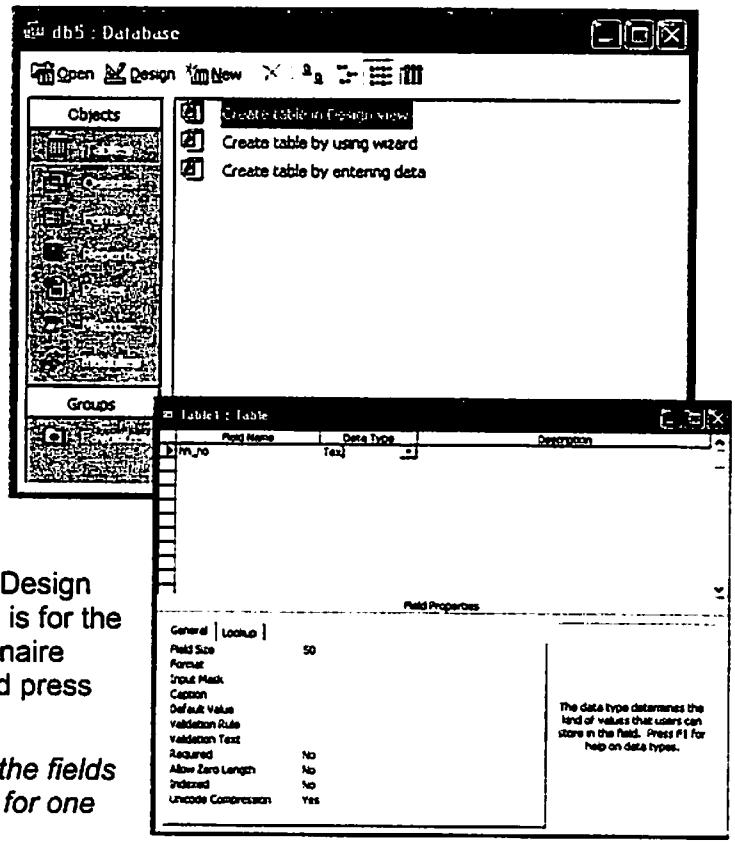
Create a New Database

1. First, create a new table by clicking on **New** command button () or use the **File** Menu to select **New...**
2. **New Database Window** appears. Click on  Database icon  and click on **Ok** command button.



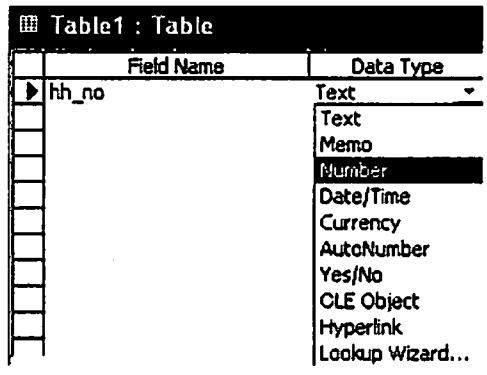
Create a New Table

3. A new **Database** window will appear. Under **Tables**, click on **Create table in Design View.**



4. A blank Table (Table 1) in Design View appears.The first field is for the household hold or questionnaire number - type in **hh_no** and press **<Return>**

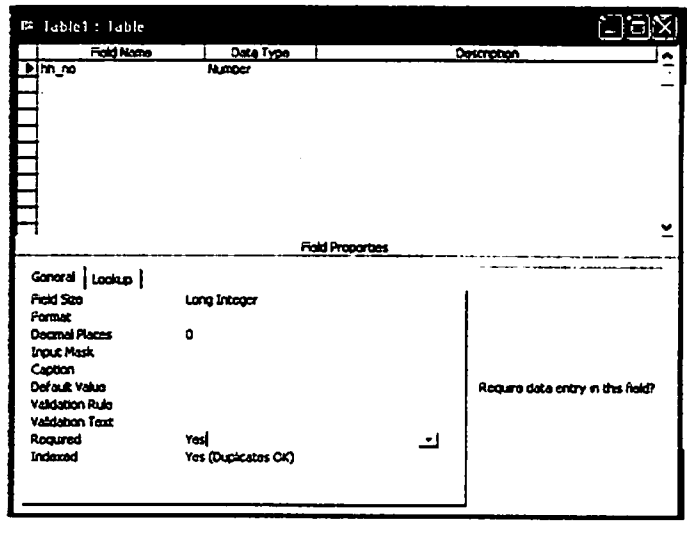
Note: It's important that you name the fields exactly as specified in these notes for one of the later exercises to work.



5. The corresponding *Data Type* is **Text** by default – use the list arrow button (▼) attached to *Data Type* to change it to **Number**.
6. The *Description* is optional – but it is a good practice to write something describing the field for future use. Type Household size.
7. Move down to the *Field Properties*. Set:
   - **Decimal places** to **0**
   - **Required** to **Yes,** since it is part of the Primary Keys.
   - **Indexed** to **Yes, (Duplicate OK)**, since it is part of the Primary Keys.

> **Tip:** The key **<F6>** can be used to *Switch* panes - or you can use the mouse. The default Set *Field Size* is Long Integer. Appendix A lists all the field sizes of a Number data type with their corresponding decimal precision and storage size.

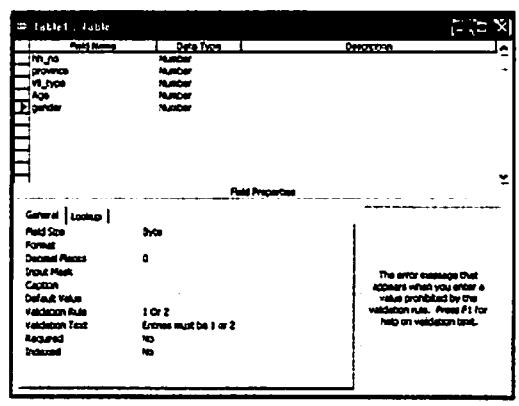
8. Move to the second field. Set the *Field Name* to **Province** and the *Data Type* to **Number.**



You could now fill in the *Description* and set some *Properties* but, to speed things up, just leave the settings for this (and subsequent fields) as they are.

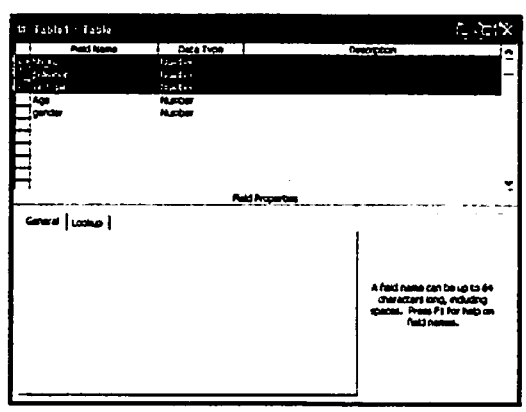
9. Move down to the *Field Properties*. Set:
   - **Decimal places** to **0**
   - **Required** to **Yes,** since it is part of the Primary Keys.
   - **Indexed** to **Yes, (Duplicate OK)**, since it is part of the Primary Keys.

10. Move to the third field. Set the *Field Name* to **Vil_type** and the *Data Type* to **Number**
11. Move down to the *Field Properties*. Set:
    - **Decimal places** to **0**
    - **Validation Rule: 1 or 2 or 3**
    - **Validation Text: Entries must be 1 or 2 or 3**
    - **Required** to **Yes,** since it is part of the Primary Keys.
    - **Indexed** to **Yes, (Duplicate OK)**, since it is part of the Primary Keys.
12. Move to the third field. Set the *Field Name* to **Age** and the *Data Type* to **Number.**
13. Move to the fourth field. Set the *Field Name* to **Gender** and the *Data Type* to **Number**
14. Move down to the *Field Properties*. Set:
    - **Decimal places** to **0**
    - **Validation Rule: 1 or 2**
    - **Validation Text: Entries must be either 1 or 2**



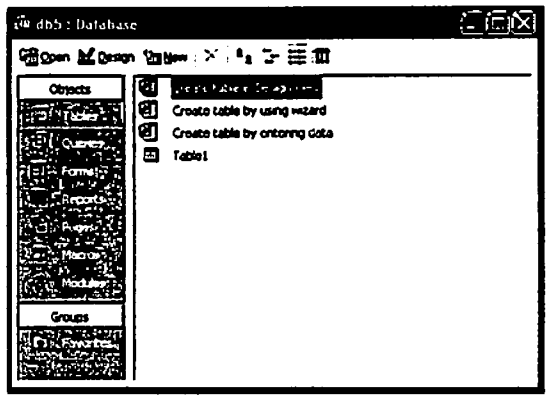
The table must look like the one on the right:

Setting up a Primary Key



15. Highlight Rows **hh_no, province and Vil_type.**
16. Click on the **Primary Key** button - a key symbol appears in the **field indicator** column.
17. Now close the table by clicking on close button. You will be asked to save the table. Click on **Yes** and retain the table name **Table1.**

Table1 must appear in the Database Window.



Exercise:

1. Create a table similar to Tbl2 item 8 income which will contain the following fields:

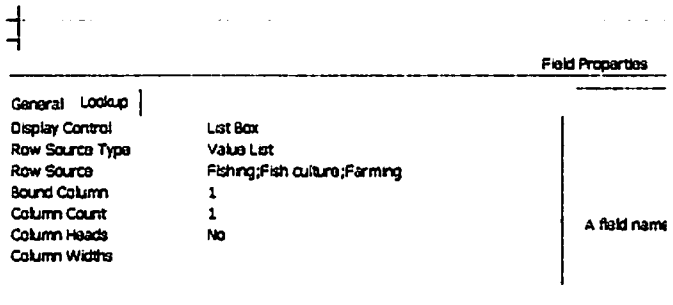
Item number	Data type	Remarks
1. HH_No	Number (Long Integer)	Part of Primary key
2. PROVINCE	Number (Long Integer)	Part of Primary key
3. VIL_TYPE	Number (Long Integer)	Part of Primary key
4. SOURCE	TEXT	Part of Primary key
5. INCOME	Number	

2. Under the Field properties of **Source**, click on **Lookup** tab set the following properties:

Display control: **List Box**

Row Source Type: **Value List**

Row Source: Fishing; Fish culture; Farming

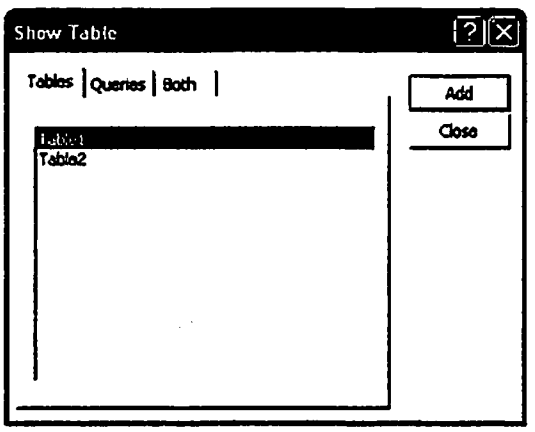


3. Save the table as **TABLE2.**

Relationships

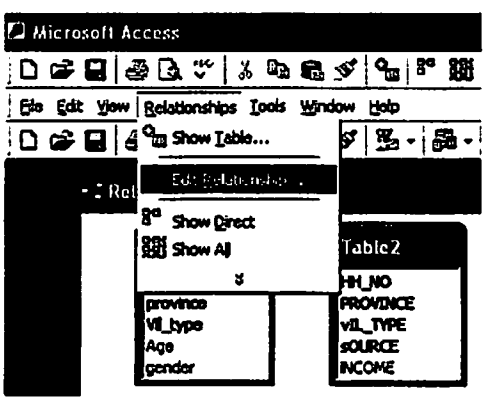
In this section, we will illustrate how to create relationship between **Table1** and **Table2**. Note that **Table2** include **SOURCE** as primary key. Hence, it is lower level than Table1. Thus, expect that the relationship is one to many.

1. From the *Database Window toolbar*, click on *Relationship* button ( ).
2. You are now going to select (add) the Tables you want create relationships. Highlight **Table1** and click on **Add** command button and do the same for **Table2**.

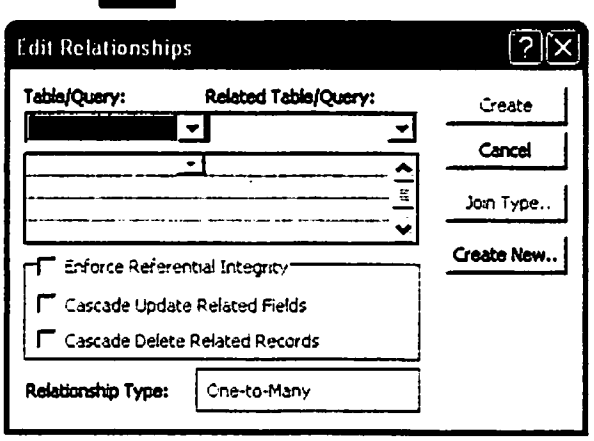


The two tables will be represented by mini-boxes (with the corresponding list of fields) in the Relationship Window.

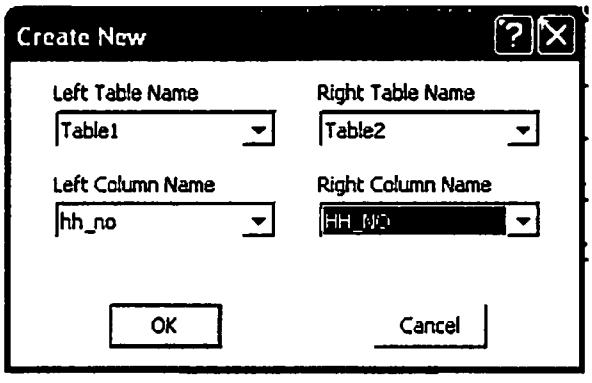
3. You are now going to establish/edit relationship. Click on **Relationship** menu then select **Edit Relationship**.



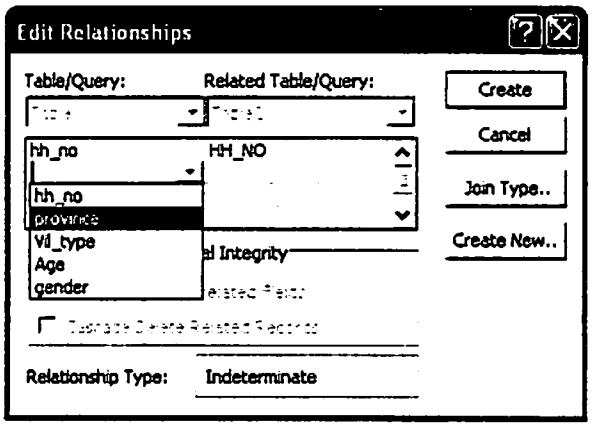
4. *Edit Relationship dialog box* appears. Click on **Create New..** command button.



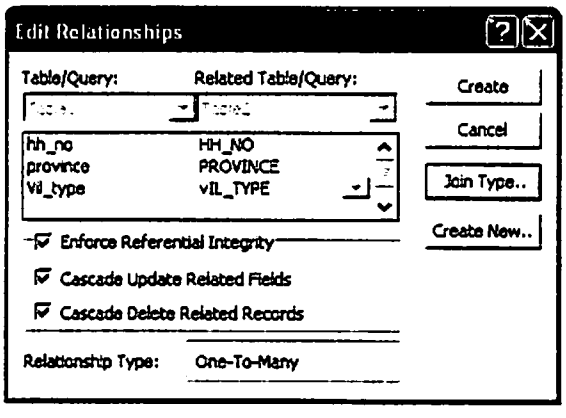
5. You are going to specify which table is "Left" (one side) and which Table is "Right" (many side) in the relationship. Then specify a field common to the Tables that define the relationship (**hh_no**). Click **Continue** to return *to Edit relationship dialog box.*



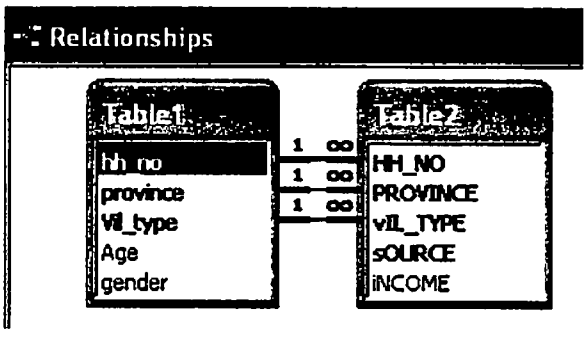
6. Edit the relationship by adding **Province** and **vil_type.** Use the list arrow button (▾) attached to each cell to select these fields.



7. Click on **Enforce referential integrity** option so with the **Cascade Updates and Delete Related Records options.**
8. Click on **Create** command button.



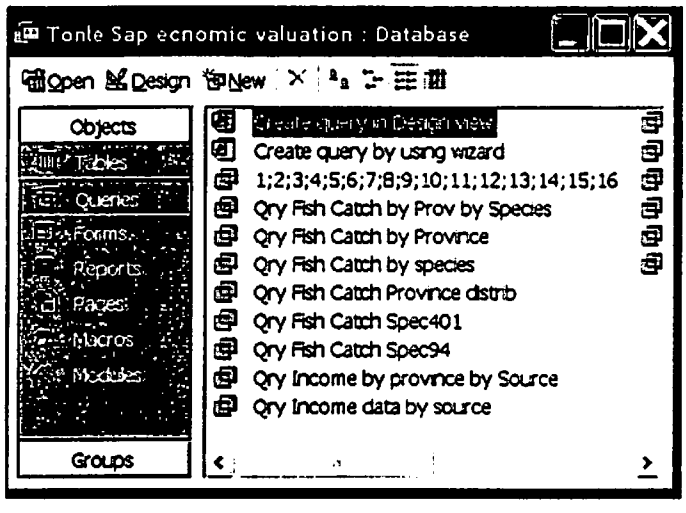
9. The relationship is created. Click close button ☒ to close the *Relationship Window.*



Query

Lets turn our attention to another object in our Database- **the Query**. A query is a question that we ask the datadase in a language called SQL (Sequential Query Language). The **recordset/dynaset** is the answer that we get back to the database. A recordset could be a subset or part of the raw data or it could be a summary statistics of the fields in the raw data in a form of corsstabulation table. Often, the recordset/dynaset are referred to as the query. Hence also apply in this document.

Let us now view some of the queries in our database.



The query **Qry Fish Catch401** subset of **Tbl8 item40 Fish Catch** table that contained catch and disposal information of households that catch species coded 401 (household level).

The query **Qry Fish Catch data by species** is a household level data of fish catch information by species.

Qry Fish Catch by Province contains average fish catch by province.

Qry Provincial Distribution is another query containing the distribution of sample by province.

To view the query, double click on their names in the *Database window*. Let us view some of the mentioned queries.

Microsoft Access

File Edit View Insert Format Records Tools Window Help

Qry Fish Catch Spec401 : Select Query

hh_no	province	vil_type	Caught	Sold	Consu
1	1	2	3	0	
11	1	2	2	2	
23	1	2	10	9	
30	1	1	50	43	
31	1	2	3	0	
48	1	1	40	0	
52	1	1	35	35	
58	1	1	1	0	
61	1	1	5	3	
62	1	1	2	0	
65	1	1	7	5	

Record: 4 of 94

Microsoft Access

File Edit View Insert Format Records Tools Window Help

Qry Fish Catch by Province : Select Query

province	AvgOfSumOfCaught	AvgOfSumOfSold	A
1	48.10	32.98	
2	115.12	81.49	
3	53.96	39.92	

Record: 1 of 3

Microsoft Access

File Edit View Insert Format Records Tools Window

Qry Province distribution : Select Query

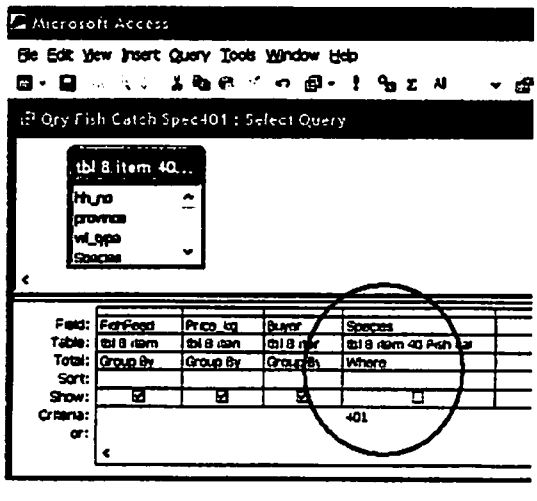
province	numprov	Percent
1	135	32.93
2	140	34.15
3	135	32.93

Record: 1 of 3

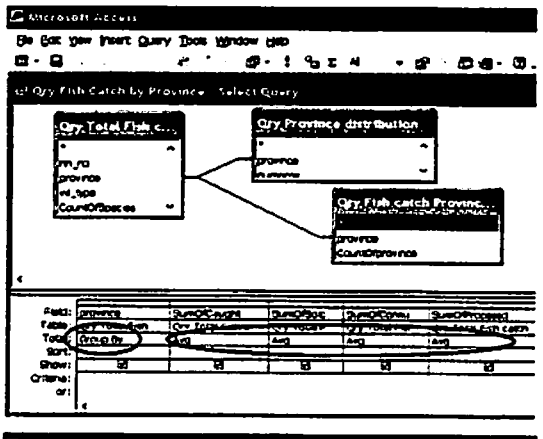
Query capability is the combined capabilities of MS Excels' **Pivot table, Vlookup and Filter.** Unlike Excel, when you save queries and reports you do not generally save the results. Instead you save the instructions that produce the results. Whenever a query or report is run the data are taken from the underlying table(s). Thus the results always reflect recent changes in the data. This is a little like "refreshing" a Pivot table in Excel so it reflects any changes in the data. Because Access stores the instructions to run the queries and reports it is possible to do a pilot survey, or just collect a few records initially and develop all the queries and reports you want, based on just these few records. The data are just used to check that you are producing the right looking table or

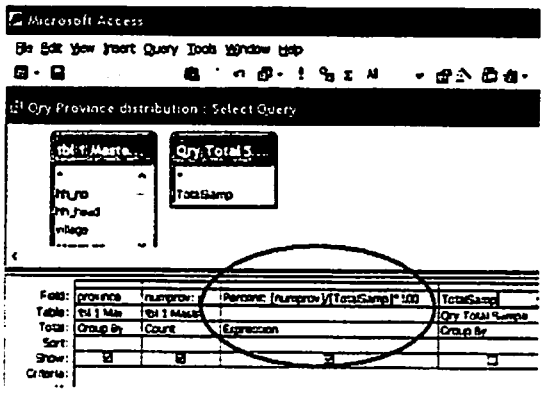
There are three viewing options available to us in the **View Menu** while working with queries. The view above is the **Datasheet view,** which is very similar to the datasheet view of Tables. The other views are the **Design view** and **SQL** (Sequential Query Language) view. The **Design view** uses the **Query Expression Builder** (QEB) utility to create the build. SQL option allows us to view the equivalent SQL that the QEB grid produces in a query window. The datasheet View allows us to view the recordset that our query produces.

The next screenshots, are the equivalent Design view of the queries above.

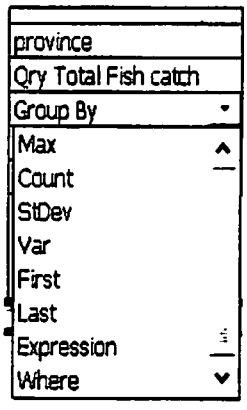


The mini boxes represent the data TABLE and pre defined QUERIES in the database. Inside of each box is the list of corresponding fields. Below the mini boxes are slots defining the query. The **FIELD** holds the all the fields you want to extract/use/see (answer the question what information). The **TABLE** is indicative of to which **Table** the field belongs to (answers the question where to get the information).





The **TOTAL** indicates the **FORMAT** (List of record or summary e.g. Average, Count, Variance, etc) of the particular field (answers the question what format).



The **"Group by"** option will list the values of the field in the Query.

The **"Where"** option (**Qry Fish Catch Spec401**) limits the result to only certain types of records in the database. In this case you need to add a **"Criteria"**. This can be done on any query...whether you have a **group by** or not. And you can limit your results based on criteria in more than one field.

The **"Expression"** option is usually used when you create a new field base on the predefined fields. The format (**TOTAL**) of the field **PERCENT** in **Qry Provincial Distribution** is **"Expression"** since it is derived by dividing the number of household in a particular province (**numprov**) by the total number of Sample (**TotSamp**) multiplied by 100. Note that the corresponding **TABLE** is blank. Note also that the **SHOW** option box of **TotSAmp** field is not ticked so with **SHOW** option box of **Species** field in **Qry Fish Catch Spec401**. This tells the query not to show these fields in the result.

Tools in Building a Query:

Sigma Button – Σ – located along the top toolbar. It will only be present when you have a query open and in the design view. This is used to do a "summary" or trend query. This is where you want to "group" records together to find out "how many" or "how much."

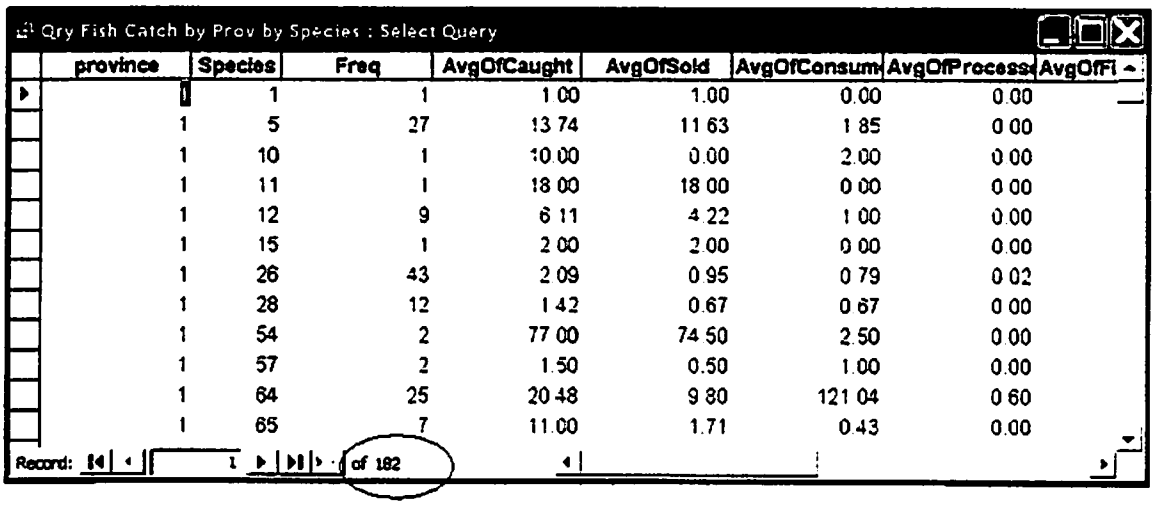
Run Button - ! – Colored red. Located along the top toolbar. This is used to "run" the query and get your results.

Add Table – — this is a button, located next to the Run button (on the toolbar). It has a yellow plus sign and a table symbol. Use this to add a table to your query...in cases where you need to pull information from more than one table. Remember to join the tables by drawing a line between the "matching" fields (if the computer doesn't automatically join them)

Datasheet Button and **Design Button** located at the rightmost corner of the toolbar. Used this view looking at the results of your query and to go back to the grid to change the parameters. An alternative is to use the **View** menu.

Creating a Query in Design View

For illustration, we will redo the query on **Qry Fish catch by Province by Species**, which by its name is the province and species-wise average quantity of fish catch and disposal (sold, consumed, processed, etc) including average unit price (Riel/kg) and the number of household.


Qry Fish Catch by Prov by Species : Select Query

province	Species	Freq	AvgOfCaught	AvgOfSold	AvgOfConsum	AvgOfProcess	AvgOfFi
1	1	1	1 00	1.00	0.00	0.00	
1	5	27	13 74	11 63	1 85	0 00	
1	10	1	10 00	0.00	2.00	0 00	
1	11	1	18 00	18 00	0 00	0 00	
1	12	9	6 11	4 22	1 00	0 00	
1	15	1	2 00	2.00	0 00	0 00	
1	26	43	2 09	0 95	0 79	0 02	
1	28	12	1 42	0.67	0 67	0 00	
1	54	2	77 00	74 50	2.50	0 00	
1	57	2	1.50	0.50	1.00	0.00	
1	64	25	20 48	9 80	121 04	0 60	
1	65	7	11.00	1.71	0.43	0.00	

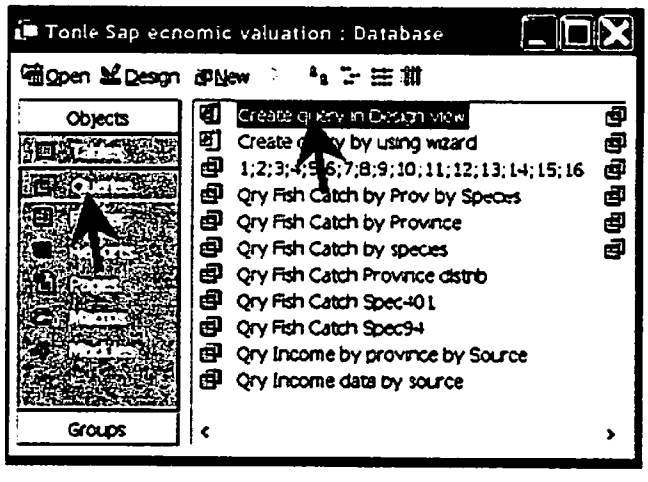
Record: 1 of 182

The table source is **Table 8 item 40 Fish Catch** and a total of 182 records were extracted.

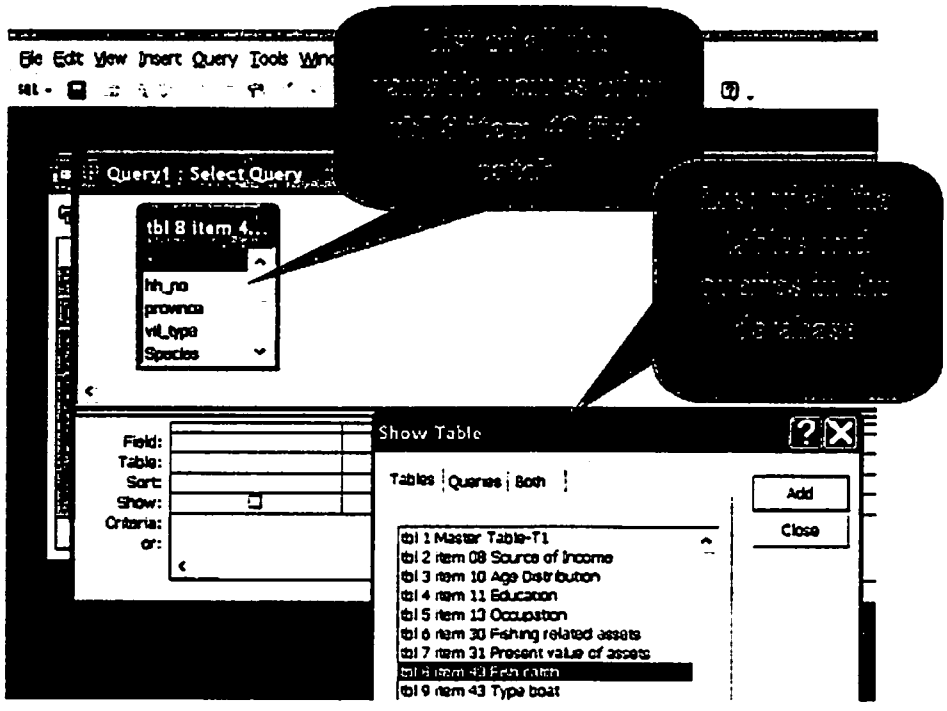
Queries can be done in several methods. For our purpose we will be using the query builder in **Design View**.

From the *Database Window*, choose the **Queries** tab (on the left of the navigation box).

1. On the right hand side of the navigation box, double click on **Create query in design view**.

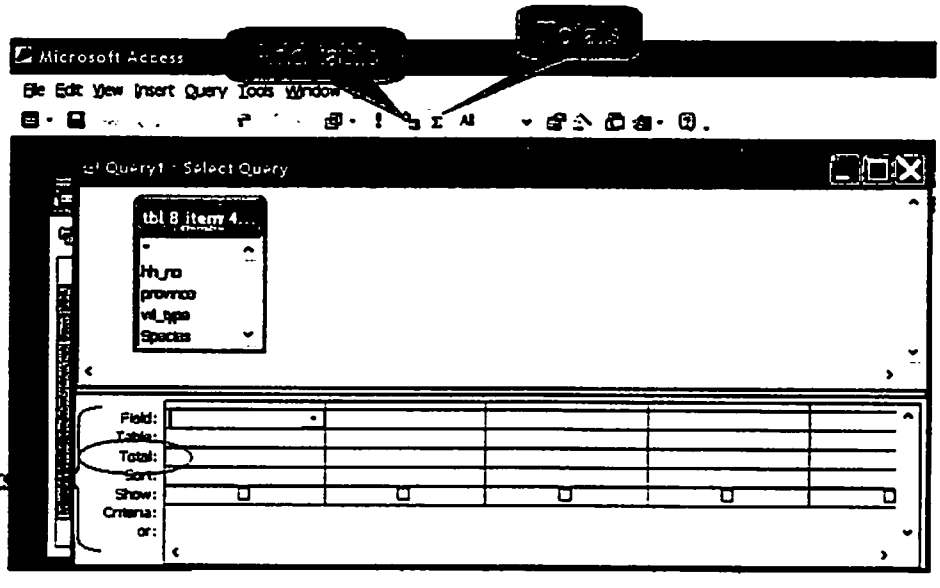


2. That will lead you to the Query Expression Builder. A **Show Table** dialog box showing all the **Tables and Queries** in the database appears in front for you to specify the table(s) you want to query. Highlight in blue **Tbl 8 item 40 Fish Catch** then click the **Add** command button.



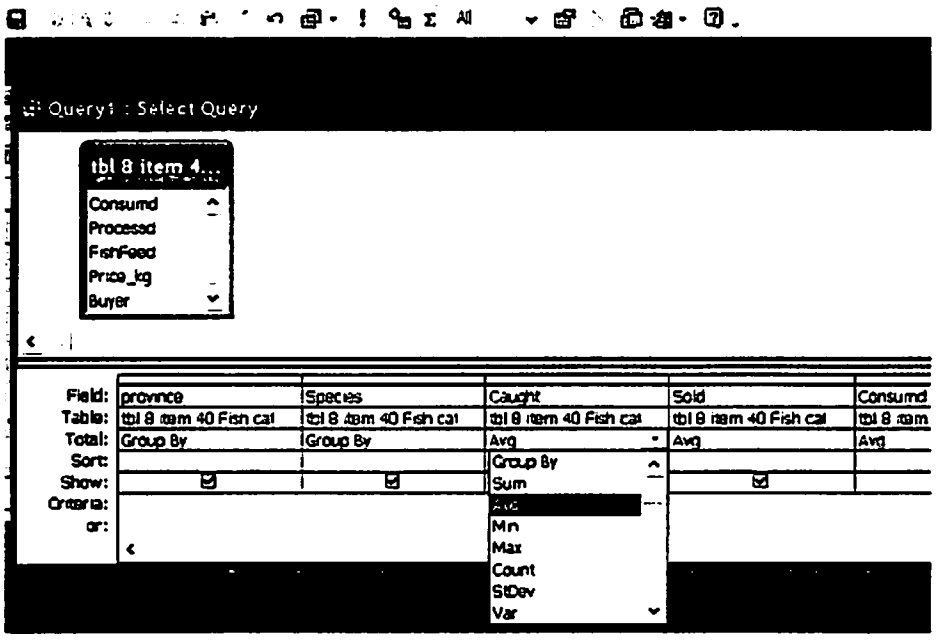
3. A mini box containing the list of all the fields of **Tbl 8 Item 40 Fish Catch** will appear at the top. Click on **Close** command button.

4. Notice the options on the left side of the grid. These are the "slots" in the grid that you need to fill in (some are optional). The three that are required for every query are **Field, Table** and **Total**. They have pull-down menus where you can choose your entries. If **Total** is not shown, click the **Totals** (Σ) command button at the top of the screen.



Slots

5. (What to see?) One by one choose the fields that you want to see by double-clicking on the field name that appears in the mini box (list of variables). The field names will then appear in the grid below on the **"field"** line.

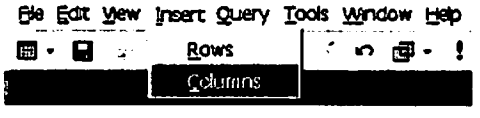


6. (What format?): Under **Total**, specify the format of the fields. Use the arrow list button (▾) in each field (column) to assign the **Format**. For **Province** and **Species**, select **"Group by"** (list province and species). For other fields select **"Avg"** (average).

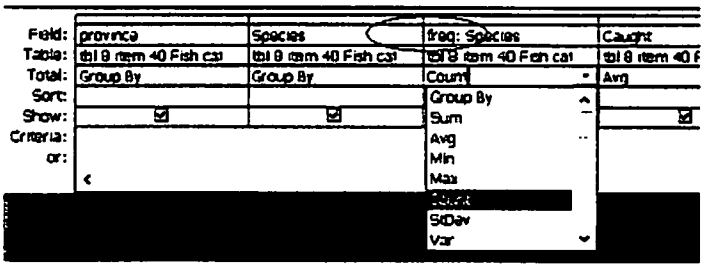
Creating a computed field

Insert/Create a **frequency** field which indicates the number of household in a particular province who reported to have caught a particular species. To do this:

7. Place the cursor anywhere under the **Caught** field.
8. Click on the **Insert** menu and select on **Columns**.

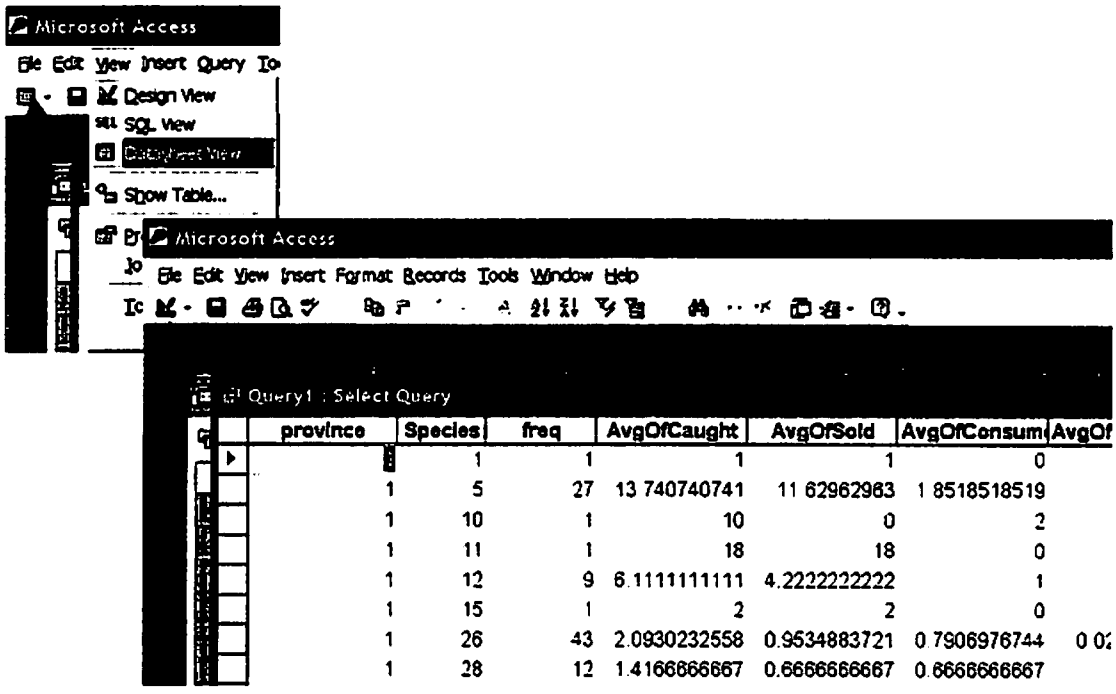


9. A blank column will be inserted between species and caught. Under the **Field**, supply any of the fields in the table (say **Species**). Also supply the name of the frequency field in the query (say **Freq**) which should be written before the table source field (**Species**) followed by a colon (:).
10. Under **Total**, use the list arrow button (▾) to select **Count**.



Viewing the Result of a Query

11. To view the result of the query click on the **Datasheet** button ( ) which can be found at the top of the screen, or use the **View** menu to select on **Datasheet View.**

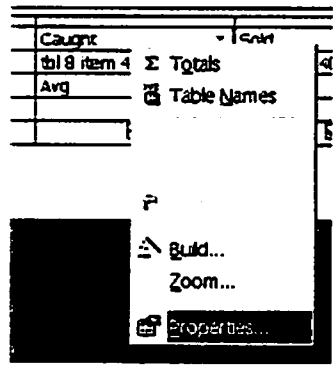


province	Species	freq	AvgOfCaught	AvgOfSold	AvgOfConsum	AvgOf
1	1	1	1	1	0	
1	5	27	13.740740741	11.62962963	1.8518518519	
1	10	1	10	0	2	
1	11	1	18	18	0	
1	12	9	6.1111111111	4.2222222222	1	
1	15	1	2	2	0	
1	26	43	2.0930232558	0.9534883721	0.7906976744	0.0
1	28	12	1.4166666667	0.6666666667	0.6666666667	

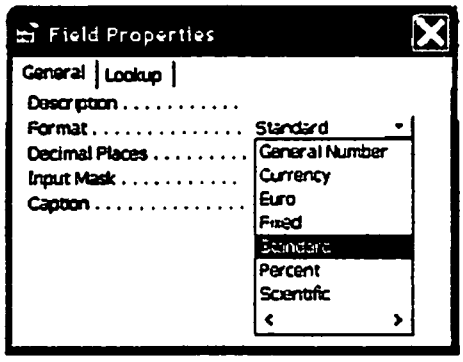
Formatting columns

The result of the query in spreadsheet format is displayed. You may want to format some of the columns e.g. say limit the number of decimal into 2 and apply comma (0,000). To do this:

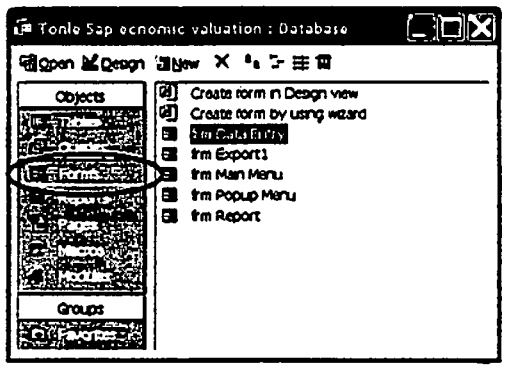
12. Switch to Design view mode by clicking on the design view button ( ) or use the **View** menu to select **Design View.**
13. Right click on Caught. A drop-down menu will appear.
14. Click on Properties.



15. A Field properties dialog box will appear. Use the list arrow ( ) attached to the Format to select Standard.
16. Close the Field Properties dialog box by clicking on close button ( ).

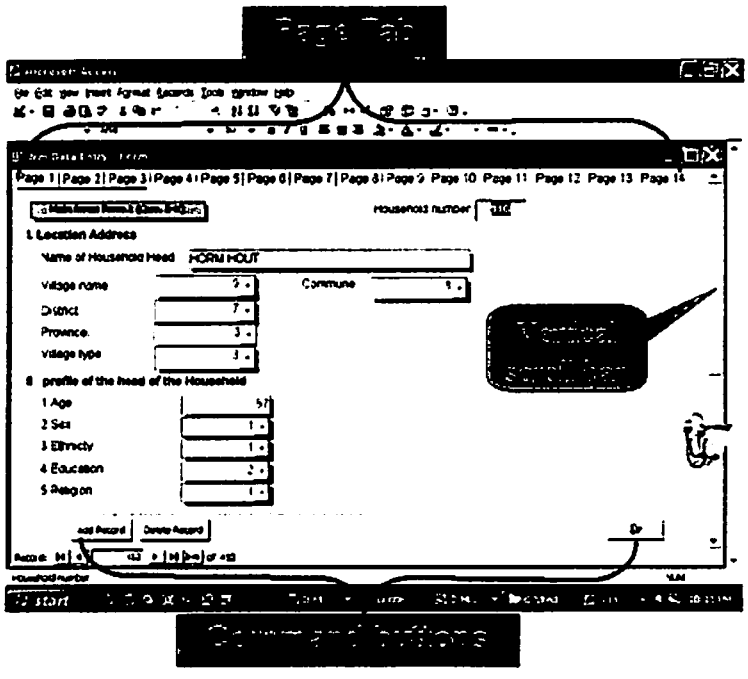


Forms

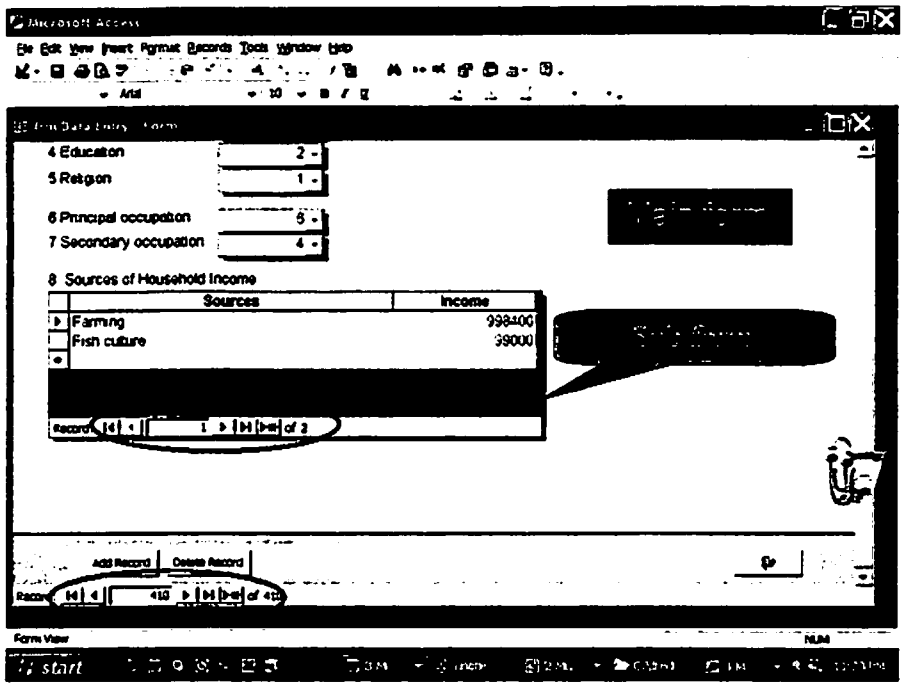


Let us now investigate the forms in our database. A click on **Forms** tab from the *Database window* lists all the forms in the database. As we can see there are 5 forms in the database. The form **"frm Data Entry"** is by its name the Data Entry form of the Database, which is a mirror image of the questionnaire. All the other forms are Menu forms, which provide general navigational functionality of the system. Earlier, we saw the **start up** and **main menu forms** when we open the database. Let us first have a closer look at the **frm Data Entry**.

The Form View



Upper window of page 1



Lower window of page 1

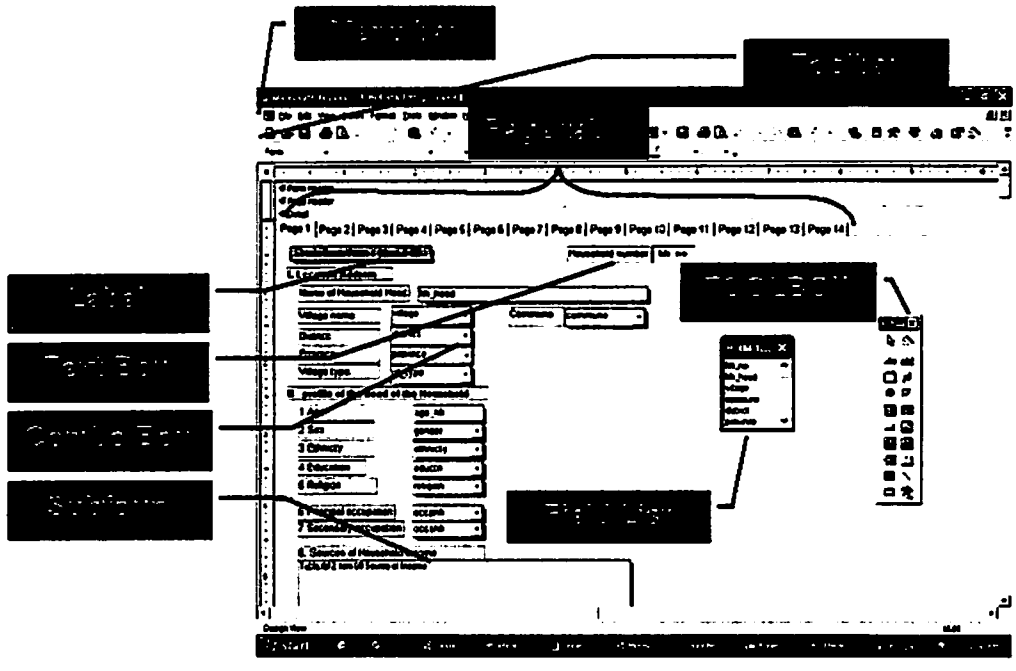
The screen shots above resemble the first page of the questionnaire. To switch from page to page, click on the **Page Tab** button, which can be found at the top of the form. Note the Sources of income (item 8) here is a sub form - a form with in a form (main form). The information entered in the main form will be stored at **Tbl 1 Master Table**, the information entered in sources of income sub from will be stored in **TBL2 item8 Income** table. Sub forms have their own navigational buttons and record selector.

At the bottom portion of the form are command buttons, **Add Record**, **Delete Record** and **Close Form** (). Note that in deleting a record in the main form, all related records in different subform will be deleted as well.

Just like a word processor environment, the form design has several menus (File, Edit View, Insert, Window and Help). The view menu allows the designer to switch from Form view to Design View or Datasheet View and to add Page and Form Header/Footer. You build the form using the Design view. You enter the data using the Form view. The Datasheet view used a spreadsheet. Page 1 cannot be viewed using Datasheet since it has a subform **Income**. But the subform **Income** is actually in Datasheet View mode. Now lets, view our Form using the Design View. You may either click View→Design View or click on the Design view button .

Design View

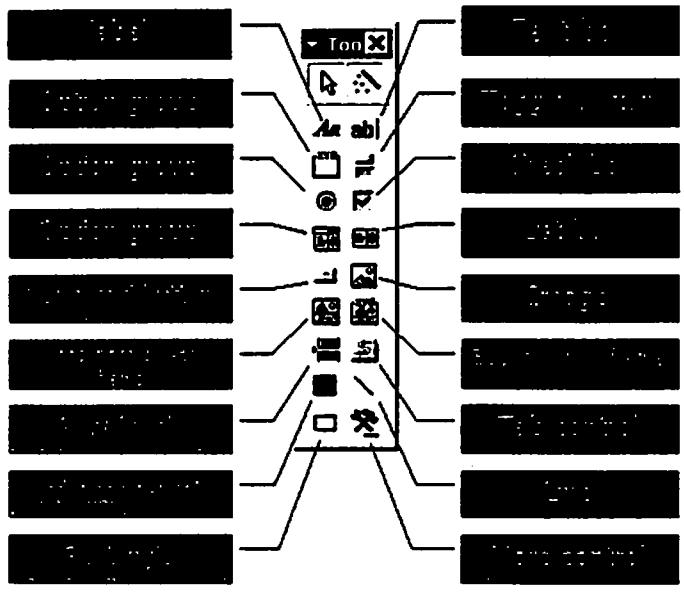
In Forms (and Reports) what you see in Design view mode is what you get in Form View model unlike Tables and Queries.



Toolbar

Note that MS Access provides you with a ***Toolbox***. The ***Toolbox*** provides many standardized elements, known as **controls** (label, text box, combo box, ext), which were used to design the form. To create a particular control on the form, click that particular control in the toolbox and point the cursor arrow to the area in the form where you want to place the control. Alternatively, you can drag and drop the field from the field list mini box to an area in the form. You can move or rearrange the controls by dragging them.

The stock controls available in the toolbox are illustrated and explained below, relating them in form for better understanding. Note that not all controls were used in the database.



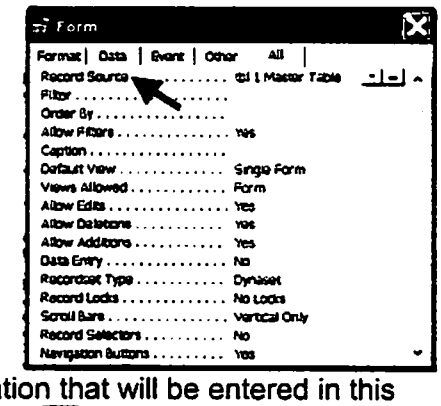
- **Select objects** – when clicked, provide the ability to select controls in the form design.
- **Control wizards** – When pressed, a wizard will be shown when adding another control to the form.
- **Label** – a control for displaying, non-editable text to the user (e.g. "I. Location Address")
- **Text box** – a control used to display data and capture textual input from the user (Name of Household, Village, Commune, etc).
- **Option group** – a control for displaying group of related controls for a data element with a limited number of options (that is Option 1 = yes, Option 2 = No Option 3 = Not applicaple).
- **Toggle Button** – A control, capable of being included in an option group, to indicate two states (depressed and undepressed) (support third state).
- **Option button** - A control, capable of being included in an option group, to indicate two states of data elements (that is, On or Off) (support third state).
- **Check box**- A control, capable pf being included in an option group, to indicate two states of data elements (i.e. Yes or No) (support third state).
- **Combo box**- A control to provide a list of values driven from a table, query or hard-coded list which drops down for the user to make a selection (District, Province, Village type).
- **List box**- A control to provide a list similar to a combo box but displays more then one element at a time and supports multiple selections.
- **Command button** – a control to provide a clickable interface in the forms' environment (Add Record, Delete Record, Close).
- **Image** – a control to display images within the form environment.

- **Unbound Object Frame** – a control for showing OLE objects (that is, Office Documents, Charts, etc...) that are not stored in a database table, thus considered unbound
- **Bound Object Frame** – a control for showing OLE objects (that is, Office Documents, Charts, etc...) that are stored in a database table, thus considered unbound
- **Page break** – a control to introduce a physical break of the page within a form or report, navigable by the Page Up/Page Down keyboard buttons.
- **Tab Control** – a control for introducing tabbed elements into your interfacesimilar to the tabbed dialogs and forms found within the operating system (The Form itself is composed of 24 Tabs).
- **Subform/Subreport Control** – a control for introducing a secondary subordinate form or report into a parent form or report (Source of Income).
- **Line Control** – A control for creating a line in the interface of a form or report.
- **Rectangle Control** – a control for creating a box in the interface of a form or report.
- **More controls** – a toolbox option for exploring all registered Active X controls on the system.

Properties

Each and every object/control in the form has a set of **Properties** attached. The form itself is an object and has its own **Properties**. At first sight seem very complicated but which in fact aren't that bad.

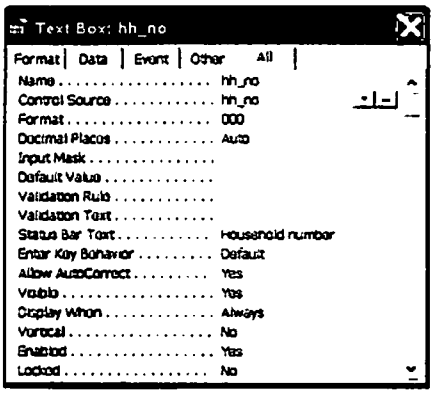
To view the **Properties** double click in the box (▪), which can be found on the upper left hand corner of the form.



While there are quite a few tabs in the **Properties Dialog Box**, they should not intimidate you. The tabs within the properties are means for logically ordering the vast array of properties by the **Access Form Object**. Access offers a large number of properties to handle the design complexities of a database, but you are not required to change each and every property. In the next section we will discuss some of the common properties, which we need to change or fill in. At this point, we check on the **Data** tab and specifically, we highlight that the *Record Source* property of the form which is **Master Table 1**. This indicates that the information that will be entered in this form will be stored in **Master Table 1** table. The list arrow (▾) attached to Record Source allows you to select the **Table** from the list of Tables in the Database.

Not only is the Form has properties but the controls as well. Lets view the properties dialog boxes of some of the controls in our database. We will not however pay in-depth review of each of the properties but only on some properties, specifically the *Control Source*.

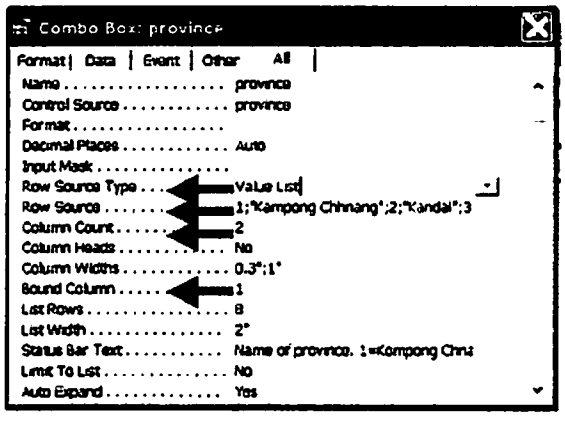
Text Box Property Dialog Box (HH_NO)

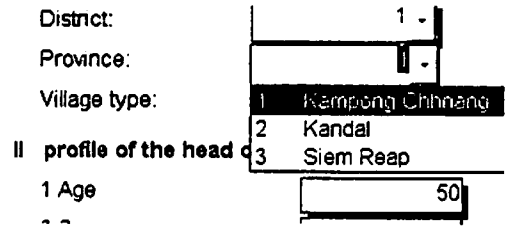


As mentioned earlier, the text box provides capability of displaying data to and accepting data input from the user. The *Record Source* property of the text box assigns the field from the table (indicated by the control source of the form in this case Master Table 1) to where the values of that field will be stored (**HH_NO**).

Note that the *Text Box Property Dialog Box* allows you also to set Validation Rule and Validation Text. But if you have already set *Validation Rule and Validation Text* for a particular field during Table Design, there is no need to set another one for the same field here. In our discussion on Table design, we suggested that you set all Validation Rule and Validation Text during Table Design not during Form Design.

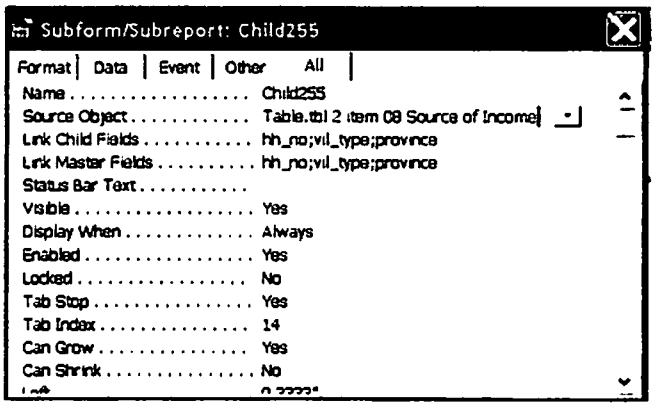
Combo Box Property Dialog Box (Province)





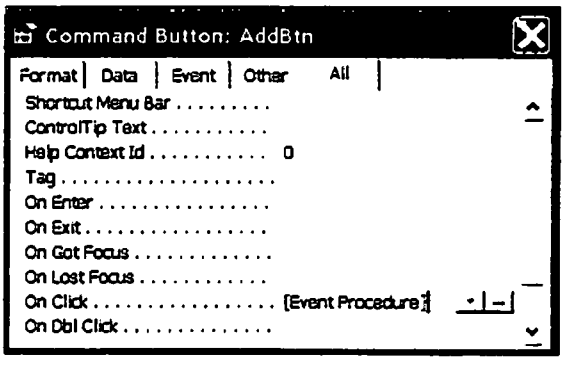
The *Combo Box Property Dialog box* is almost similar to that of a *Text Box* except for some additional Properties: *Row Source Type and Row Source*. As mentioned earlier, *Combo Box* provide a list of values driven from a table, query or hard-coded list which drops down for the user to make a selection. There are three choices for *Row Source Type*: **Table/Query, Value List and Field List**. If you choose **"Value List"**, then you have to list down the possible choices under *Row Source*:. If you choose **"Table/Query,"** then you have to specify the name of the *Table or Query* in the *Row Source* property. In the screen shot above, we use **Value List** for **Province**. Note the sequence of the List: number followed by province name. Beside the Property dialog box is the screen shot of Form View on Province. Note that there are 2 columns in the list of choices with the first column being the **"Code"** and the second column being the **"name of the province"**. This is because **"2"** is specified under *Column Count* Property and **"1"** is (first column) under *Bound Column* Property.

Dialog box of the Sources of income (subform)

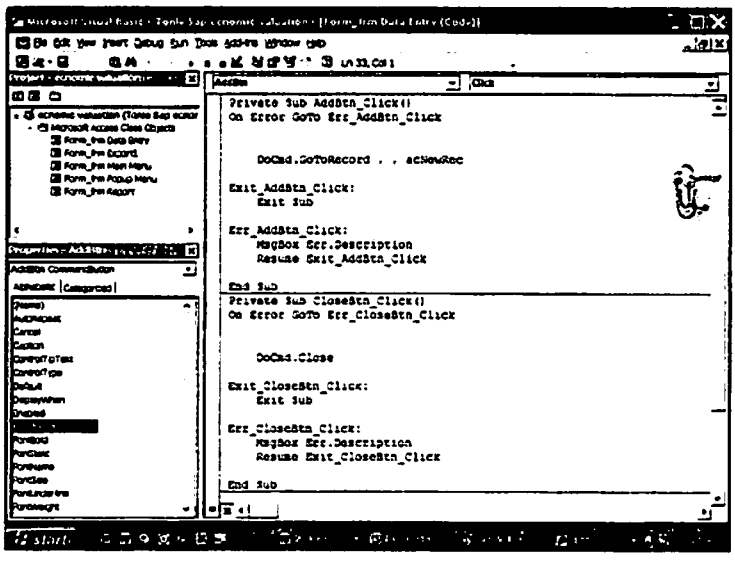


The subform (child form) control is a form with in another form (Parent form). Like the Parent form, it has a source object as shown in the property dialog box, which is a table (**tbl 2 item8 income**). The link Child Fields and Link Master Fields are the fields that links the **Source Object** (tables) of the Parent form (**tbl 1 Master Table**) and Child Form (**tbl 2 item8 income**). Normally, when you create a subform, Access will automatically identify the Link child and Link Parent fields.

Dialog box of a command button (Add Record)

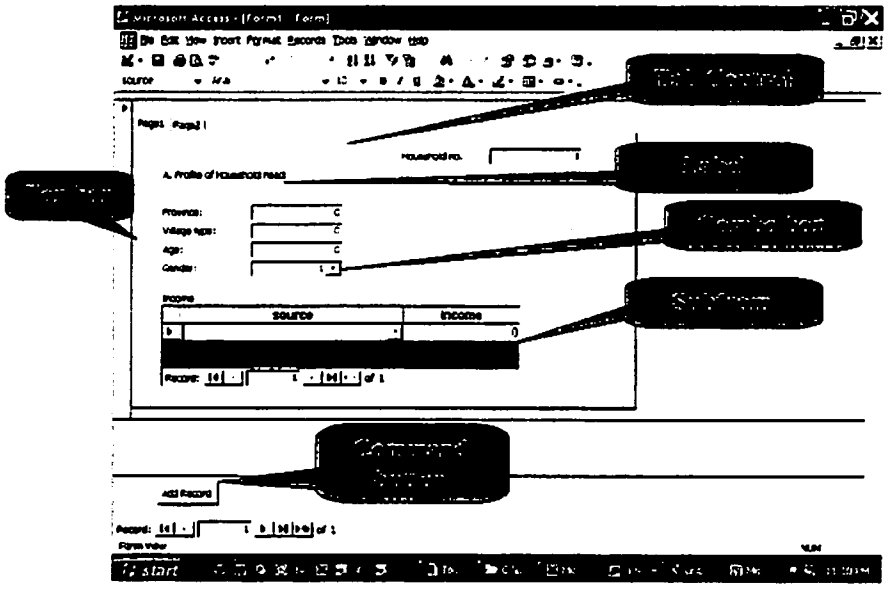


Most of the command buttons in this database were created using the wizard without altering properties dialog box and the code at all. In the next section we will illustrate how to create a command button. At this point we will show the *On Click* property of the **Add Record** Command button and the *Event Procedure Code* that was generated automatically by the wizard. This *Event Procedure Code* is a set of commands that will be run when you click the button in the *Form View*. To view the code, click on the three-doted button (...) attached to *On Click Property*.

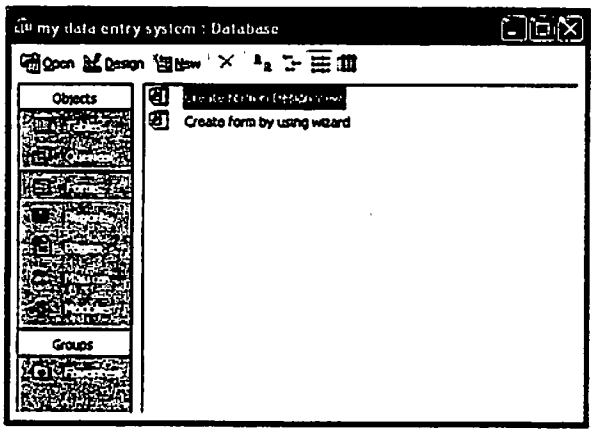


Creating a Data Entry Form in Design View

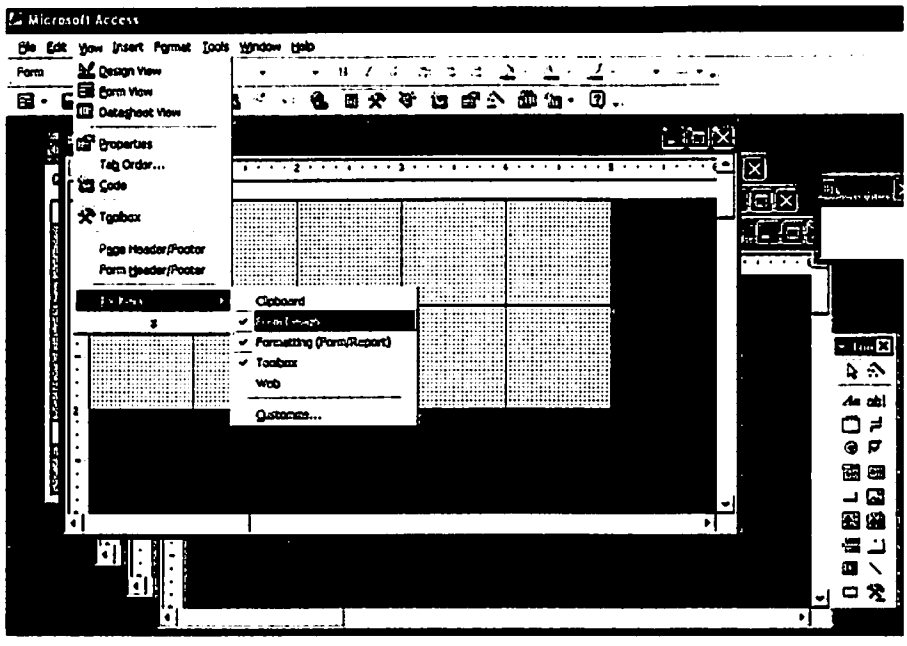
In this section we will illustrate how to design Data Entry Form from scratch. Specifically, we will illustrate how to create a Form with a **Tab control, Label, Text, Combo Box, Sub form,** and a **Command button**. We will use the Tables we created in the previous section. Specifically, we will create a form which looks like the one below:



1. Open the database **Exer1.MDB**. This database contains Tables **Table1** and **Table2** we created earlier.
2. On the right panel of the *Database Window*, choose the **Forms** tab.
3. On the left panel of the *Database Window*, double click on **Create form in design view.**



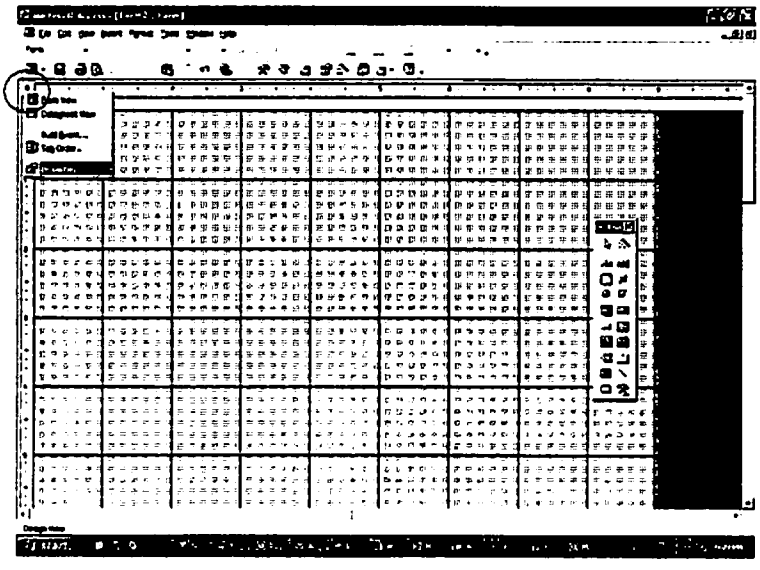
4. A blank form will appear with the **Toolbar**. If the **Toolbar** is hidden, open the **View** menu, click on **Toolbox** or simple click on the *ToolBox button* ( ) in the Toolbar. To automatically view the toolbox every time you create a form, click on **Toolbars** from the **View** menu... and then **ToolBox**.



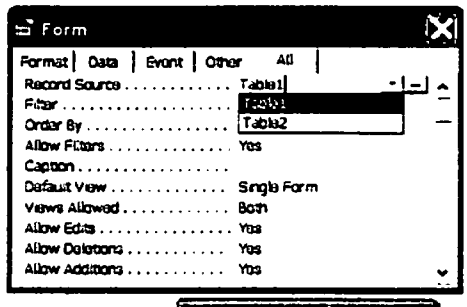
5. Click on the **Maximize Window** button (▣) to enlarge the form design window or double the title bar
6. Drag the *Detail* borders out to fill the whole screen.

Specifying the Table to where the information will be stored

7. One of the first things you do is to specify the **Table** to where the information to be entered to this form will be stored. Double click on the form button (▪), which can be found at the upper left most corner of the **Form**. Or alternatively, click on Property Button ( ) in the toolbar. Another alternative is use the **View** menu then **Property**.

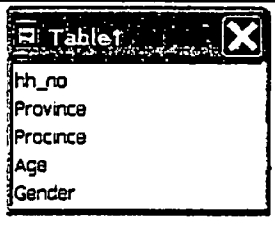


A **Form Property Dialog Box** appears.



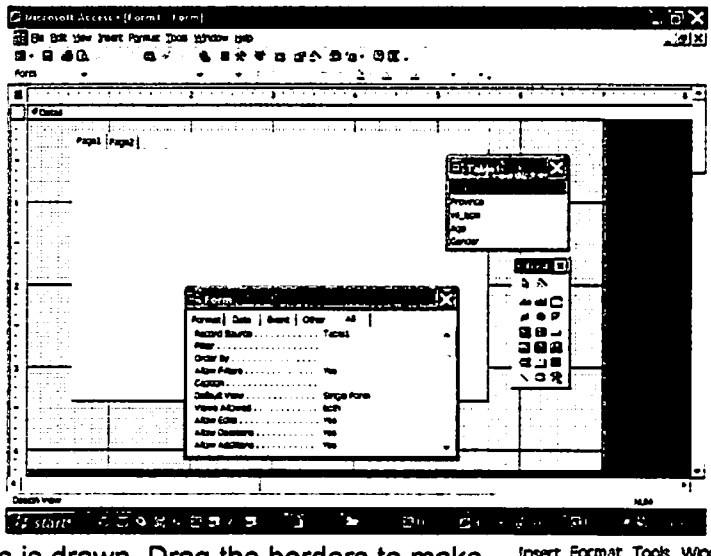
8. Click on the *List arrow* ( ) attached to *Record Source* and select **Table1**.

A mini box containing the list of fields of Table should appear. If it does not appear, click on the **List button** ( ) in the Toolbar. You are now ready to create controls with the fields. But first you need to create a **Tab control**.



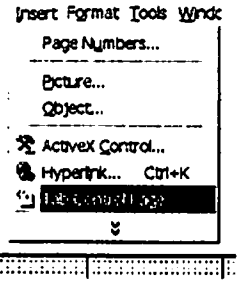
Creating a Tab Control

9. Click on the **Tab control** button ( ) in the Toolbox.



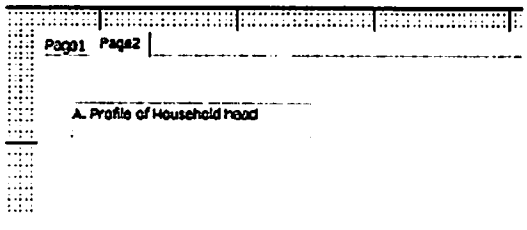
The Pointer will change to Tab control button signifying that you select an area to where you put the tab control.

10. Hold down the mouse button and select an area.
11. A Tab control having 2 Tabs is drawn. Drag the borders to make it bigger.



You can add more Tabs through **Insert** menu➔ **Tab Control Page**.

Creating a Text Control



Now First, let us create a Text Control. Say "II. Profile of Household Head".

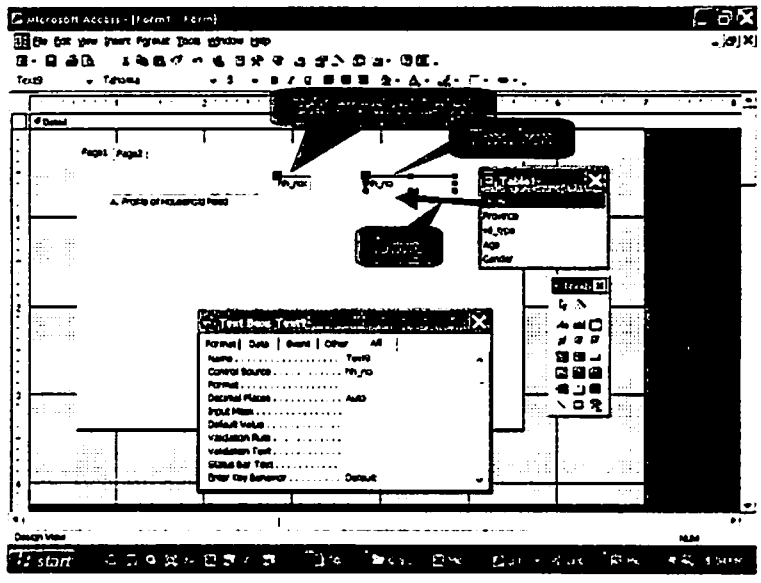
12. Click on the Label button ( )in the Tool box and select an area in the form to where you want to place the Label. A blank label will appear for you to write the **Caption**.

You may change the font, font size, font color, back ground, apply special effect on the control

Creating Text Box

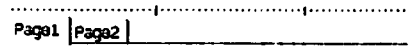
Let us create a text box for **Household no:**.

1. You may use the Toolbox or simply drag "**hh_no**" from the List Window to an area you want to put the control.



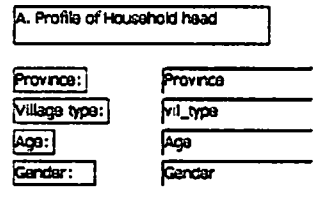
Note that the **Text Box** comes with a label with caption similar to the field name.

2. Edit the Caption of the attached Label to "**Household no:**"



Note you can however always delete the attached label.

3. Follow the steps number 13 and 14 for **Province, vil_type, Age and Gender**.

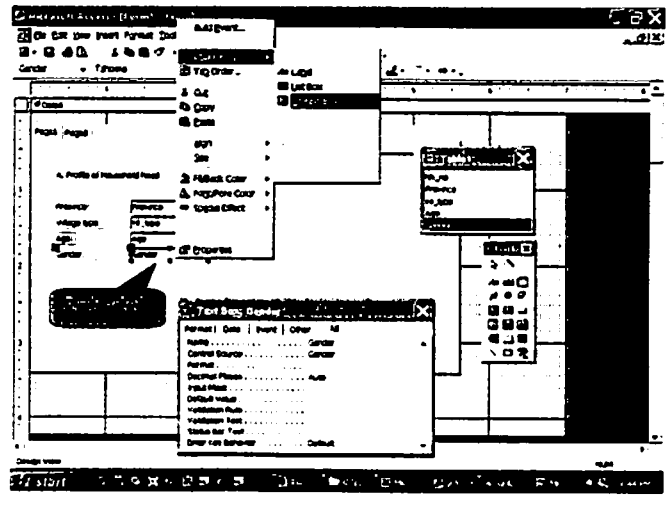


Creating a Combo Box Control by Changing a Text Box Control to a Combo Box.

Lets create a combo box for **Gender**.

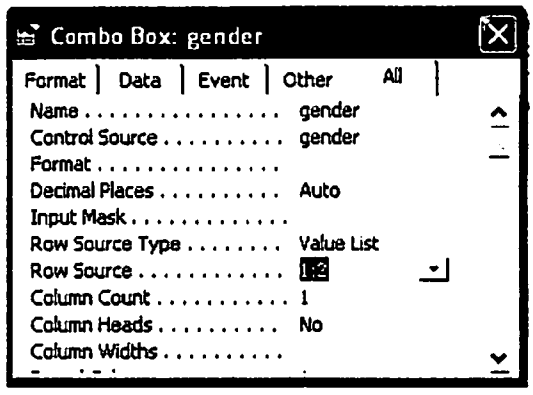
4. Right click on the text box **Gender**. Then select **Change To** followed by **Combo Box** on the drop down menu.

Now it changed to a Combo Box with the attached List Arrow.

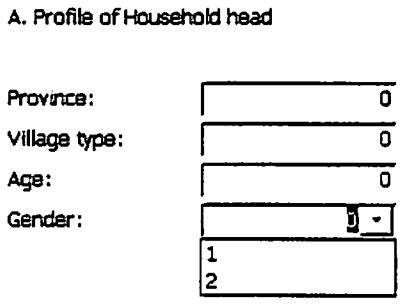


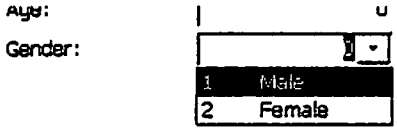
You are now going to fill-up the **Row Source Type** and **Row Sources properties** using in the property Dialog box. Make sure that the **Gender** combo box is clicked.

5. Use the List Arrow (▾) attached to **Row Source Type property** to select **Value list.**
6. For Row Source, type "1;2".

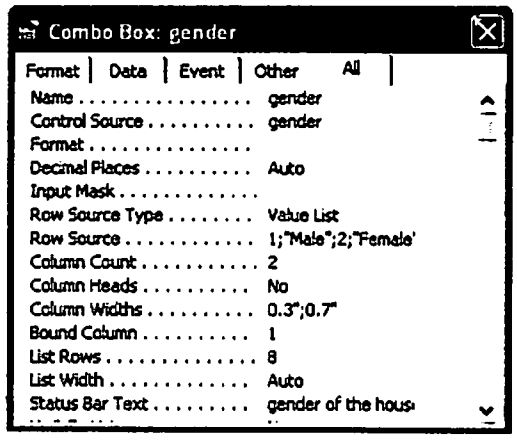


View the Form by clicking the **Form View** button ( ) found at the left most corner of the Toolbar. Click on the List arrow attached to "**Gender**". It is not clear what sex pertains to 1 and 2. Add the Value Label on the codes to make the Form look like the following screen shot:

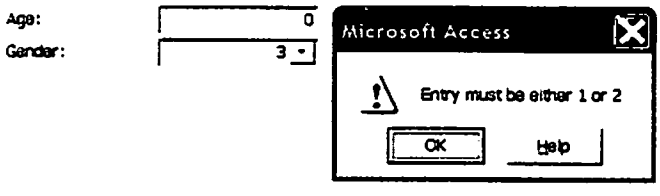




7. Click on **Design View** button ( ) to go back to Design View mode.
8. Click on gender combo box.
9. Edit **the Row Source** entry to: **1; "Male"; 2; Female.**
10. Set **Column Widths property** to **0.3;0.7**



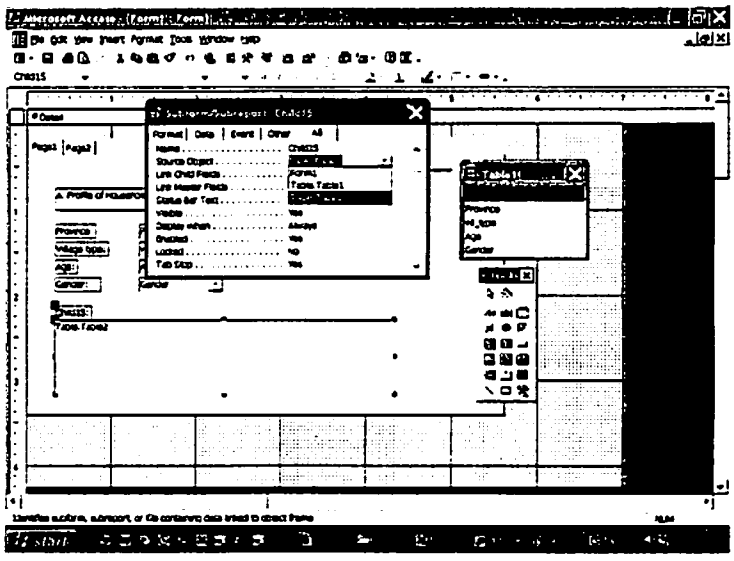
Now View you Form again and try to enter "3" under **Gender**. You will be prompted with the following message. The message is the one you wrote under **Validation Text** property during **Table Design.**



Creating a Subform

Let us create a subform whose *Source Object* is **Table 2** and *link fields* **hh_no, province and vil_type.**

1. Click on the *subform* button ( ) in the Toolbox.



The Pointer will change to subform control button signifying that you select an area to where you put the tab control.

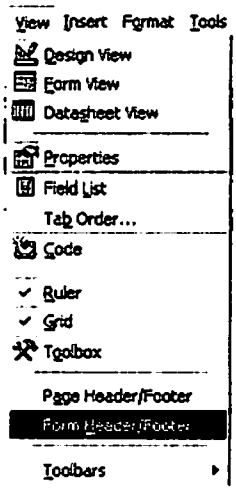
2. Hold down the mouse button and select an area (just below Gender).
3. The Subform wizard will start to launch. Click on **Cancel** button and let us use the Property box to set the properties.
4. You are going to specify the *Source Object* of the subform. In the *Property box*, click on the arrow list ( ) attached to *Source Object* and use **Table 2.**

Note that Access automatically sets the *Link Child fields and link Master field.*

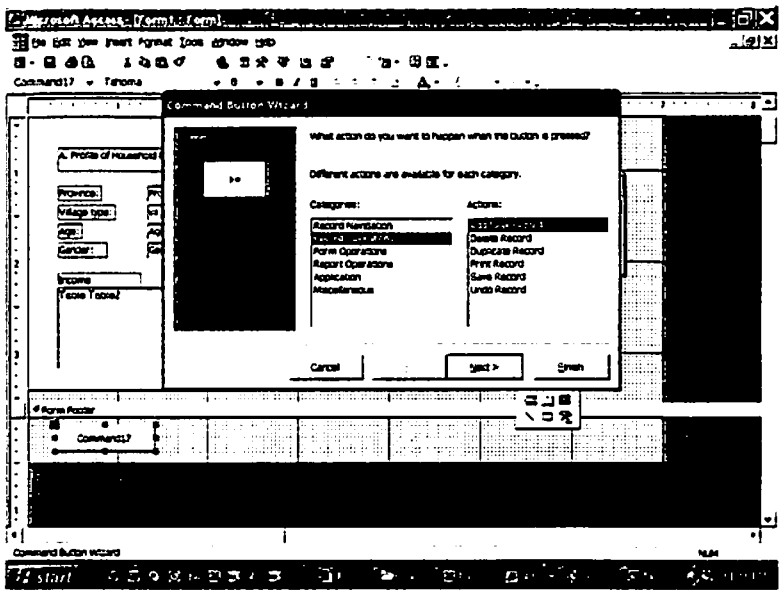
Creating a Command Button: Add record

In this section we will illustrate how to create command button using the Command Button Wizard. Specially, we will illustrate how to create a command button that Adds a Record and another Command button that Closes a Form. We will put these command buttons below on the Footer Section of the Form.

1. First check if the form has *Form Footer* under *Design View*. If no, use click **View menu then Form Header/Footer** to view form header and footer.



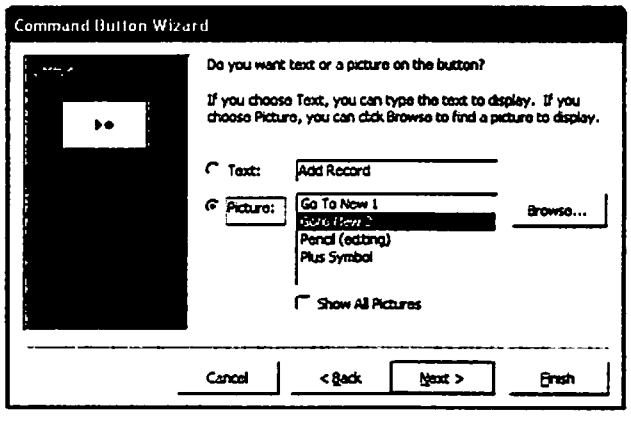
2. Click on the **Command Button Control** ( ) on the toolbox and select and area in the Footer to where you want to put the command button.



3. A **command button** will be displayed. Also, a *command button Wizard dialog box* will pop-up prompting you what action that the command button will do when it is clicked. A list of different **Categories of Action** is listed in a window on the left panel for you to select and the corresponding **Actions** on the right panel.

4. Since the command button adds a record when clicked, it belongs to **Record Operations category**. Highlight *Record Operations* in blue and *Add New Record on the corresponding Action panel.*

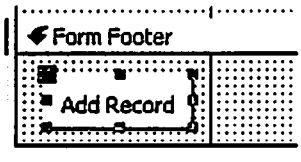
5. Click on **Next** command button.

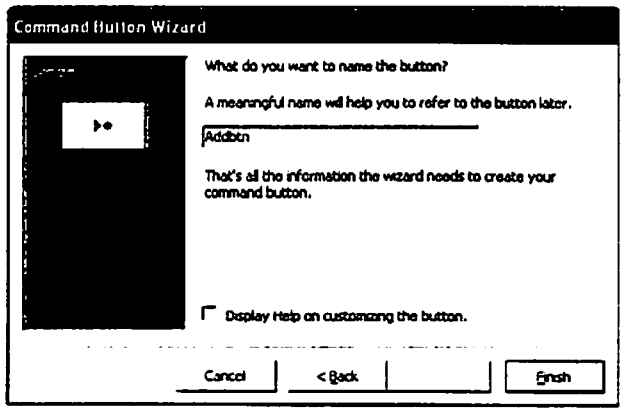


6. The Wizard will now ask you if you want a text/caption or a picture on the button. Select **Text** and click on **Next** command button.

7. Lastly, you have to give a **name** of the command button. Type **Addbtn.**

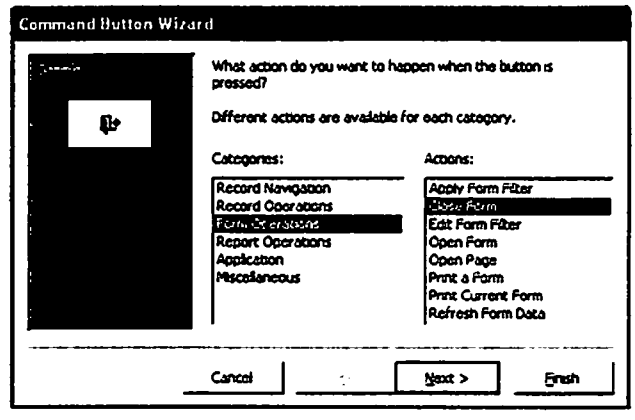
8. Click on **Finish** command button. Your command button will should look like this:





Creating a Command Button: Form Operation

Most of the **"Actions"** of the command buttons in Tonle Sap Database are to open and close a form. These **Actions** can be categorized **under Form Operations**. To create such command buttons, follow the step as detailed above and select **Form Operations** as **Categories** when the Wizard ask you for action to happen when the button is clicked.



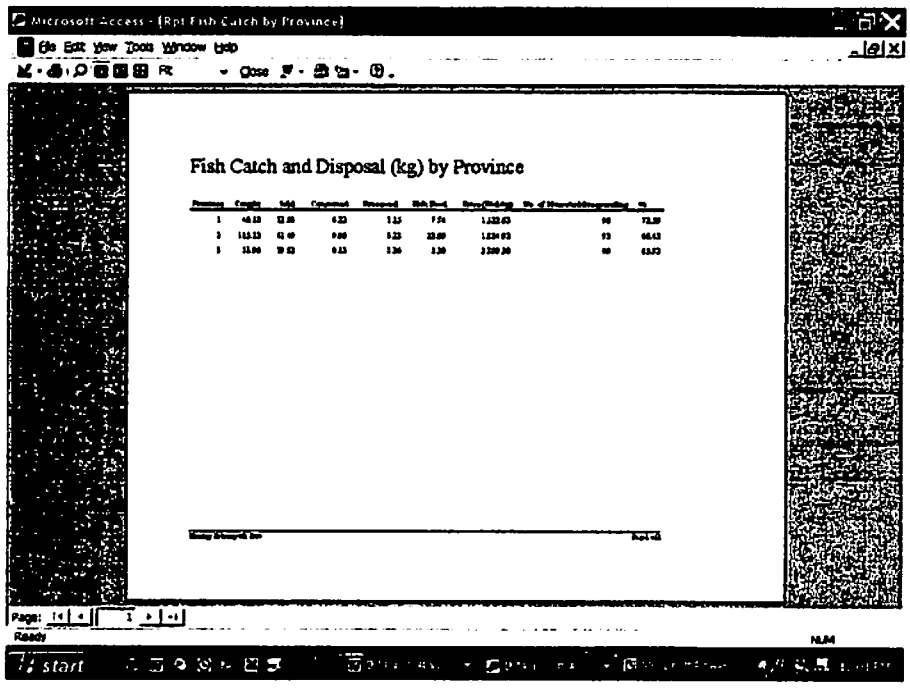
Report

You can view and print a simple list of records in your table or query to present information. But it may not be the best way of presenting information to others. Using **Reports** is an alternative to produce detailed and summary information that can be easier to read and understand; have attractive formatting (with clipart or scanned images, can be illustrated with charts and graphs and include calculations.

In this section we will show some of the simple reports in the database. These reports were all designed using **Report Wizard**. In the next section, we will illustrate how to design a report using the Wizard.

Report View

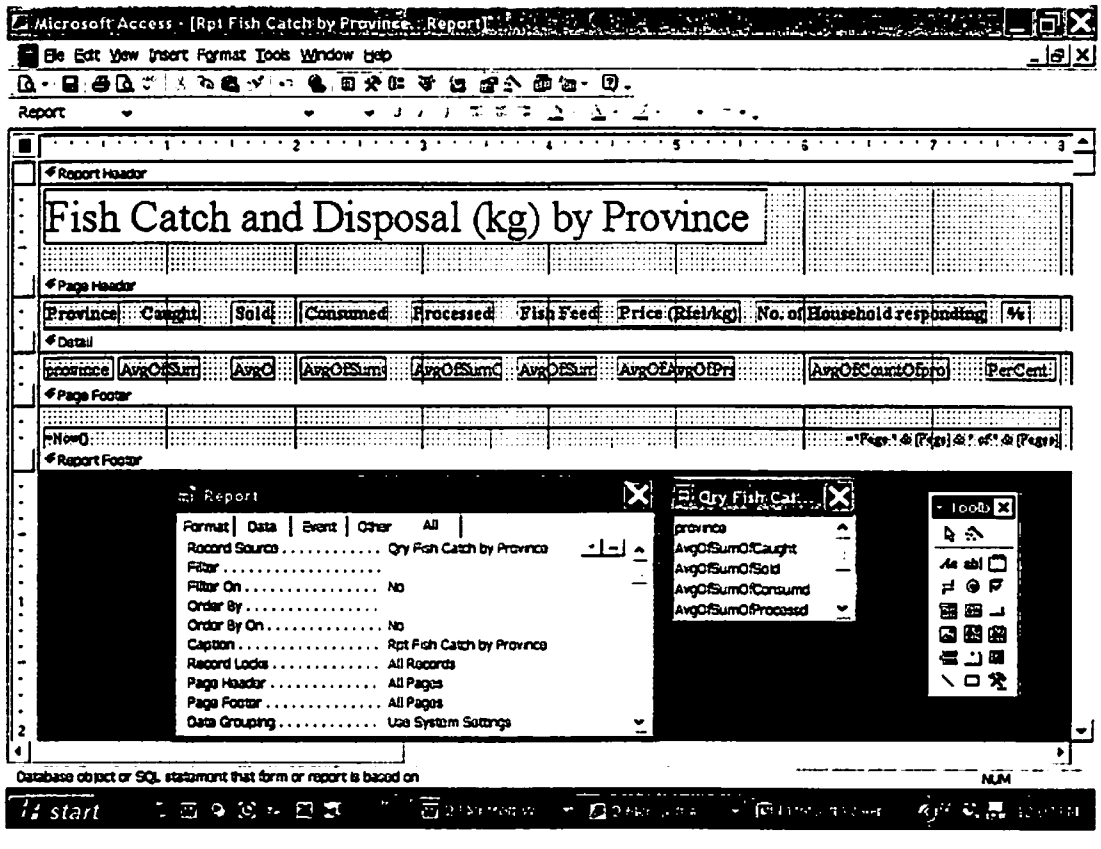
In the *Database Window*, click on the **Report** tab. The list of all Reports is shown in the right panel of the Database window. Let's view the **RPT Fish Catch by Province** report. This is a simple report showing the average fish catch and disposal of fishing household by province. Actually, this is the result of the Query **Qry Fish Catch by Province**. To print the report, click on the print button ( )in the toolbar or alternatively, click the **File** menu and Click on **Print.**



Report Design View

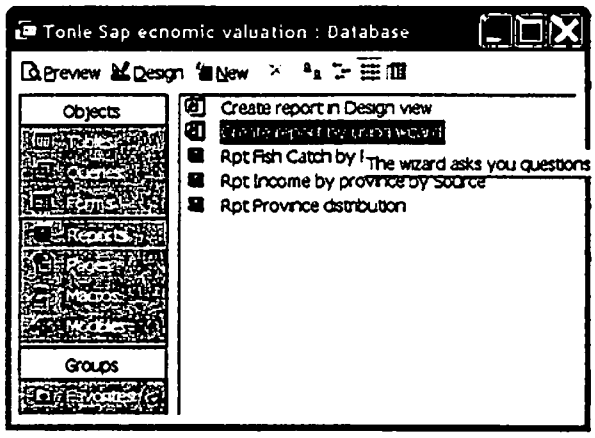
Now lets have a look on the **Design View** of the Report. The Design View Environment of Report is very similar to Form Design View. There is nothing special here. As mentioned in the Form Section, we made use of the Header and Footer in Report. In the Report Header is the title of the report, which is a **Label control**. In the Page header are the captions of the fields, which are also a **Label control** and in the Detail Section are the names of the fields in the Query which are **Text Box control** without the Label attached. The **Record Source** of this Report is the Query **Qry Fish Catch by Province** as shown in the **Report Property** dialog box.

The functions "**=Now**" in the left hand side corner of the Page Header section returns the current date under **Report View** and ""Page " & [Page] & " of " & [Pages]" at the right hand side returns the page numbers. Again, these functions are automatically generated by the wizard. If you don't like to show them, you can simply delete them.

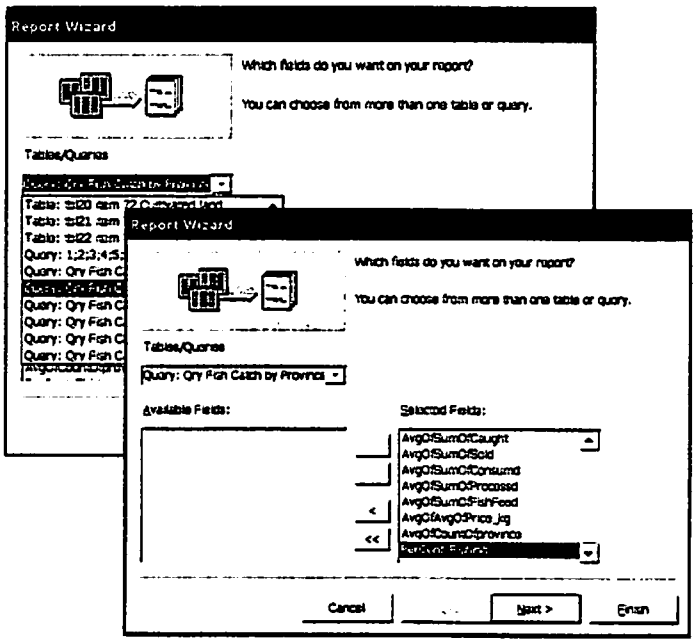


Creating Report by Using Wizard

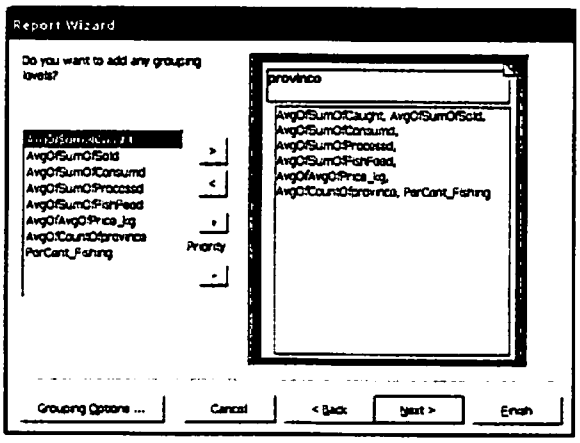
In this section we will illustrate how the **Qry Fish Catch by Province** Query was created using the Wizard.



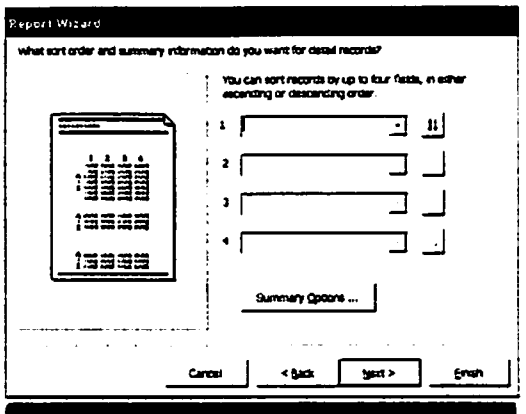
1. In the *Database Window* click on **Report.**
2. On the Right panel of the Window, click on *Create report by using Wizard.* A Report Wizard dialog box will appear.
3. Use the List Arrow button ( ) attached to *Table/Queries* to select Query: **Qry Fish Catch by Province.** The list of all fields of the Query will be shown.
4. Next you are going to select the fields that you want to include in the report. Move **All** fields into second column (Selected Fields) by clicking on the >> button.
5. Click **Next** to continue.



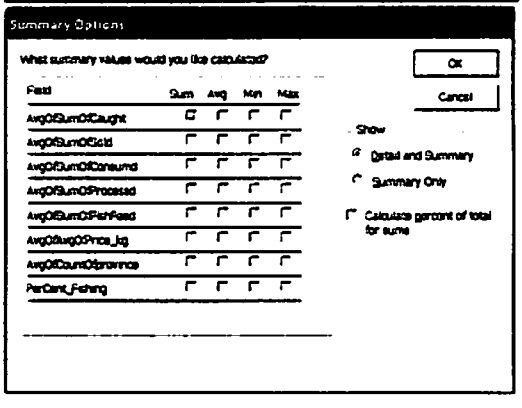
6. You will be asked if you want to add another Grouping level. In this **Province** is already set as Grouping-level field you don't need another field so just click on Next to continue.



7. Next you are going to specify any fields that you want to order in alphabetical order. Unless you want, you can select the fields from the drop down list.

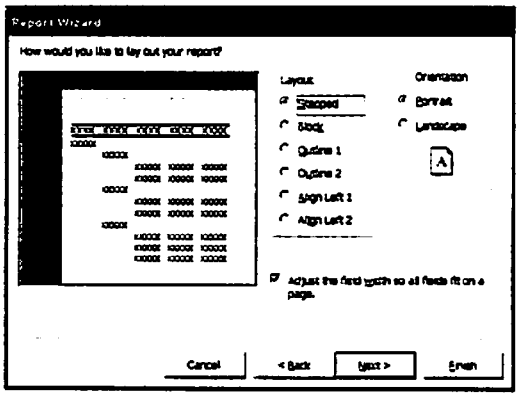


Note the **Summary Options** command button. If you want summaries on the fields other than Averages click on this button. This will lead you *to Summary options dialog Box*. If you don't tick on any of these boxes. The Averages will be displayed. Click on **Ok** to go back *to Report Wizard dialog box.*



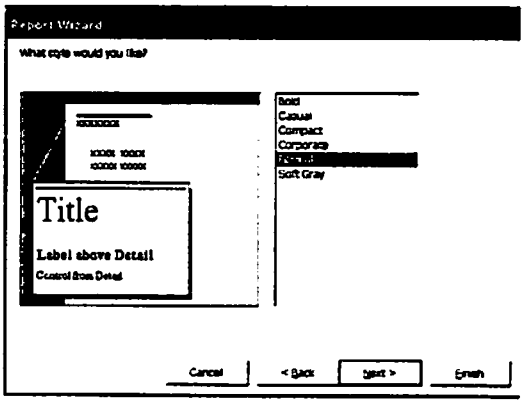
8. Click on **Next** to Continue.

9. Next step in the wizard allows you to choose page orientation and layout of report. Accept setting and click **Next** to continue.

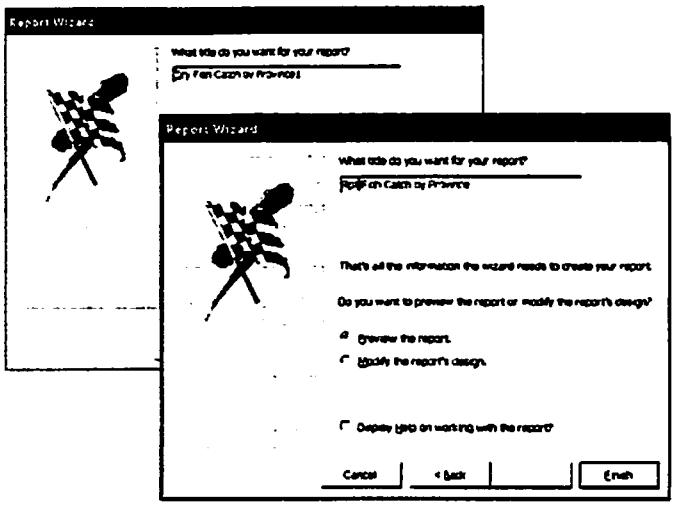


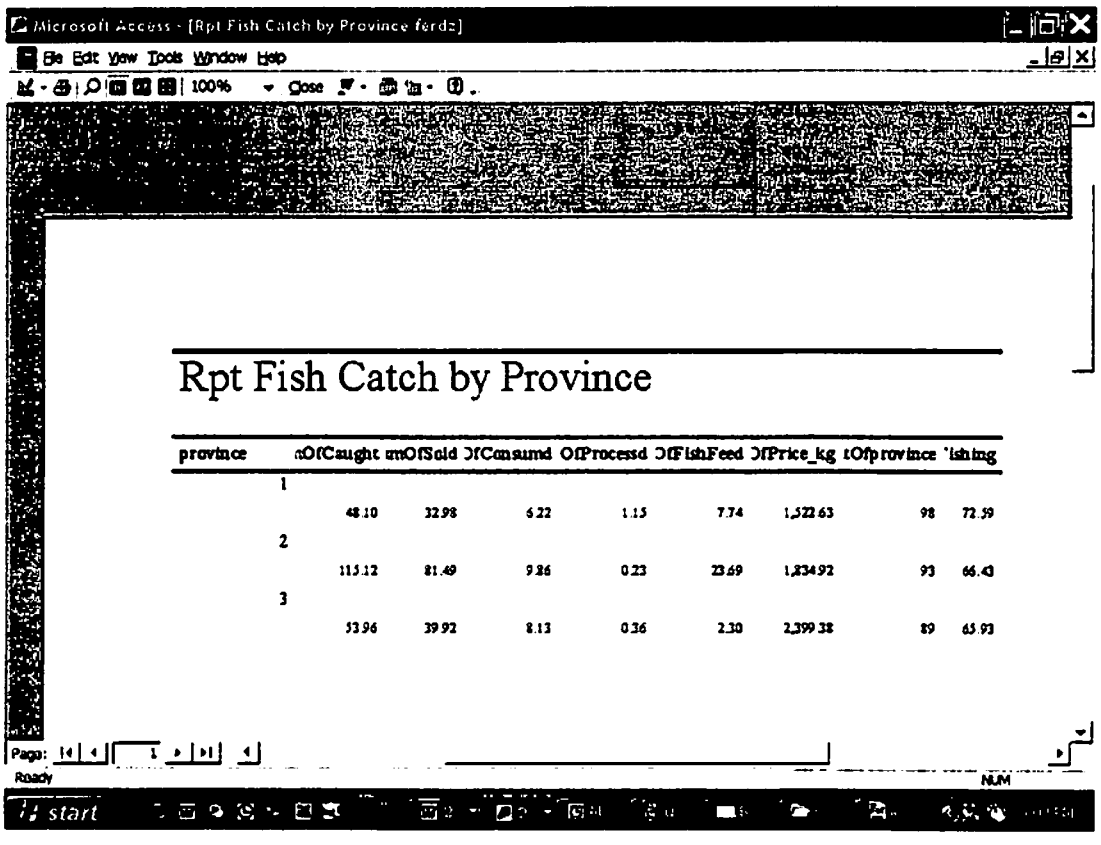
10. Next step in the wizard allows you to choose Style of report. Click on **Casual** for style of report.

11. Click **Next** to continue.



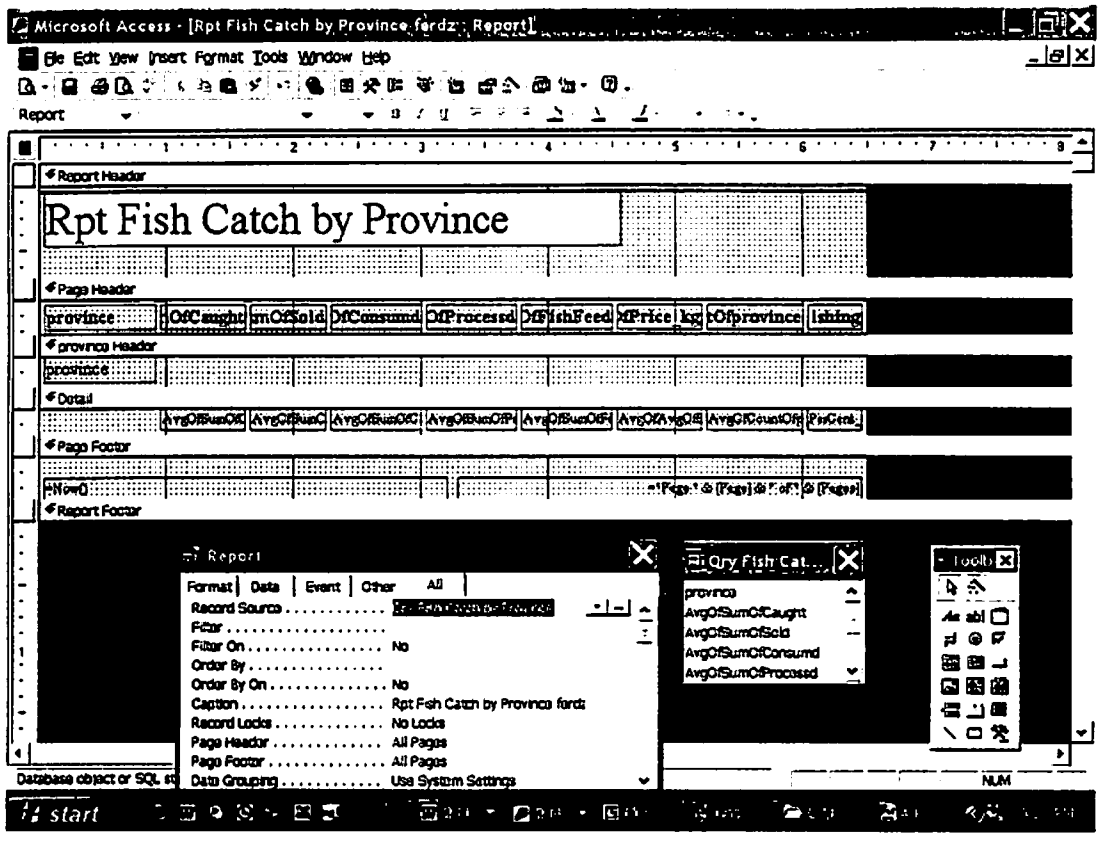
12. Finally, provide a **name** of the Report. Replace **Qry** with **Rpt** at the start of the filename. You want to preview the report so leave this setting as it is. Click **Finish.**



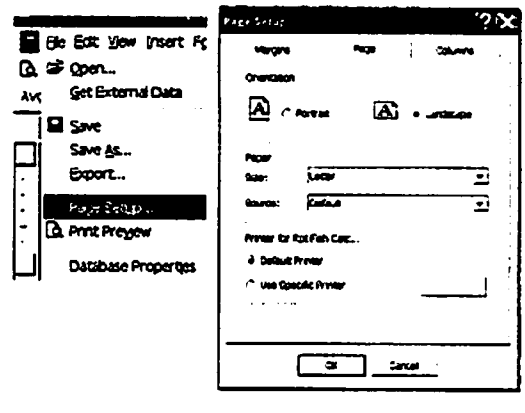


You need to make some modifications.

13. Switch to **Design View** ( ).



14. Now you can edit the ***Title of the Report***. There are two ways of doing this. First, click on the *Title control* (**Rpt Fish Catch by Province**). The Title should appear on the **Caption** Property in the Property dialog box. You can Edit the **Caption** like any other Label control in the Property Dialog Box. Once you are finish, you can click anywhere outside the property dialog box and the Title will change accordingly. Secondly, click on the Title three times and you can edit the on the Title directly. You may also change the Font size.



15. Edit the Caption/Label of the fields (e.g **AvgOf SumOf Caught** to just **Catch**) and adjust the spacing.
16. Now change the Orientation of the Report (Portrait to Landscape).
17. Click **OK** command button to finish the job.

Appendix A: FieldSize Property (copied from MS Access electronic Help)

You can use the **FieldSize** property to set the maximum size for data stored in a field set to the Text, Number, or AutoNumber data type.

Setting

If the **DataType** property is set to Text, enter a number from 0 to 255. The default setting is 50.

If the **DataType** property is set to AutoNumber, the **FieldSize** property can be set to Long Integer or Replication ID.

If the **DataType** property is set to Number, the **FieldSize** property settings and their values are related in the following way.

Setting	Description	Decimal precision	Storage size
Byte	Stores numbers from 0 to 255 (no fractions).	None	1 byte
Decimal	Stores numbers from -10^38 -1 through 10^38 -1 (.adp) Stores numbers from-10^28 - 1 through 10^28 -1 (.mdb)	28	12bytes
Integer	Stores numbers from – 32,768 to 32,767 (no fractions).	None	2 bytes
Long Integer	(Default) Stores numbers from –2,147,483,648 to 2,147,483,647 (no fractions).	None	4 bytes
Single	Stores numbers from –3.402823E38 to –1.401298E–45 for negative values and from 1.401298E–45 to 3.402823E38 for positive values.	7	4 bytes
Double	Stores numbers from –1.79769313486231E308 to –4.94065645841247E–324 for negative values and from 1.79769313486231E308 to	15	8 bytes

	4.94065645841247E–324 for positive values.		
Replication ID	Globally unique identifier (GUID)	N/A	16 bytes

You can set this property only from the table's property sheet.

To set the size of a field from Visual Basic, use the DAO **Size** property to read and set the maximum size of Text fields (for data types other than Text, the ADO **Type** property setting automatically determines the **Size** property setting).

Note You can specify the default field sizes for Text and Number fields by changing the values under **Default Field Sizes** on the **Tables/Queries** tab, available by clicking **Options** on the **Tools** menu.

Remarks

You should use the smallest possible **FieldSize** property setting because smaller data sizes can be processed faster and require less memory.

Caution If you convert a large **FieldSize** setting to a smaller one in a field that already contains data, you might lose data. For example, if you change the **FieldSize** setting for a Text data type field from 255 to 50, data beyond the new 50-character setting will be discarded.

If the data in a Number data type field doesn't fit in a new **FieldSize** setting, fractional numbers may be rounded or you might get a **Null** value. For example, if you change from a Single to an Integer field size, fractional values will be rounded to the nearest whole number and values greater than 32,767 or less than –32,768 will result in null fields.

You can't undo changes to data that result from a change to the **FieldSize** property after saving those changes in table Design view.

Tip You can use the Currency data type if you plan to perform many calculations on a field that contains data with one to four decimal places. **Single** and **Double** data type fields require floating-point calculation. Currency data type fields use a faster fixed-point calculation.

Using Spreadsheet (MS Excel) as Tool for Database Management and Initial Data Processing

Prepared for the:

Training Workshop on Resource Economics and Research Methods for the Officers of IFReDI, DOF, Ministry of Agriculture, Fisheries and Forestry, Royal Government of Cambodia

23 February – 03 March 2004

Prepared by:

Ferdinand Javien Paraguas

f.paraguas@cgiar.org

http://www.ferdinandjparaguas.tripod.com/

WorldFish Center

Penang, Malaysia

Using Spreadsheet (MS Excel) as Tool for Database Management and Initial Data Processing

TABLE OF CONTENTS

Using Spreadsheet (MS Excel) as Tool for Database Management and Initial Data Processing

Introduction

Spreadsheets are commonly used for data entry because they are familiar, in widespread use and very flexible. MS Excel is probably the simplest and most familiar spreadsheet package to use for storage of data. It is a flexible package, which allows us to manipulate our data easily. However, its very flexibility means that, if we do not use it effectively, the end result is poor data entry and management. We need to apply a great deal of rigor and discipline to ensure that this does not happen.

This document is to illustrate how MS Excel can be used effectively for data entry and initial data processing.

We begin the next section with a simple example that shows some data that have been poorly entered. This is followed by a general guideline for developing databases in MS Excel. In section 3, we will present and discuss a Proto-type version of the Tonle Sap Economic Valuation Database in MS Excel. In Section 4 we, we will illustrate how the MS Excel version of the database was developed highlighting some features in Excel that facilitate simple and reliable data entry. Specifically, using the Tonle Sap Economic valuation questionnaire, we illustrate how to:

1. impose validation checks to accept only entries within a specified range, so that typing data outside the minimum and maximum values is prohibited. This is to ensure that the correct sort of information is being entered;
2. create input message and error alert dialog box. Input message is a message that will pop-up if a cell is selected. It is very useful in giving details about a variable and guide encoders to what entries should be entered in that particular cell;
3. customize Error Alert messages to inform encoder that a wrong entry has been entered.
4. create drop-down choice list which allows the encoder to select an entry from an list of choices provided.
5. format columns and many more..

In section 5, we will illustrate how to do basic data operation using MS Excel. Specifically, we will illustrate how to extract a subset of a data using the **Filter**; how to merge data files using **Vlookup** function; how to categorize records data points using the **IF** functions and how to aggregate data using **Pivot Table**.

In Section 6, we will illustrate how to generate cross tabulations using the pivot table tool of MS Excel. Finally, we give our conclusions about the use of Excel for statistical work.

An Example

Below is a screen shot of a spreadsheet that resembles a poorly entered data set which was done in Excel. This is a simple set of data that can be entered very effectively in a spreadsheet.

	A	B	C	D	E	F
1	Fingerling Stocking Information					
2		Spcies	Quantity	Average	Source of	Total
3		code	stocked	size	Fingerling	Cost
4			(number)	(cm)	code	(Riel)
5	HH_NO	SPECIES	QTY	SIZE	SOURCE	COST
6	1	1	2000	25	Hatchery	1000
7	1	Giant snake head	1000	27	Hatchery	1300
8	1	3	10000	20	hatchery	1500
9	1	4	1000	5 pieces/kg	Hatchery	100
10	2	2	900	26	hatchery	1200
11	2	5	3000	8	own produce	400 Riel
12	3	1	5000	35	2	1400
13	3	6	5800	28	2	870
14						
15						

The data were entered by a data encoder, who - as instructed - typed what was written on the recording sheet in the field. However, this has led to errors (ringed cells). For example, **"Giant snake head"** should be replaced by its numeric code like the rest of the entries under **SPECIES**. In the first place, this will not happen if data validation had been set. The entry 5 pieces/kg under **SIZE** should be converted to centimeter without writing the unit. The Entry **"400 Riel"** under **COST** is another. The mixing lower and upper case for non-numeric entries is also an error. Such entries will cause problems during processing and when the data are transferred to a statistical package for further analysis.

Most of these errors can be avoided if some thought is given to the layout of the data in the spreadsheet. This guide attempts to show how to avoid these problems.

Guidelines for data entry in Excel

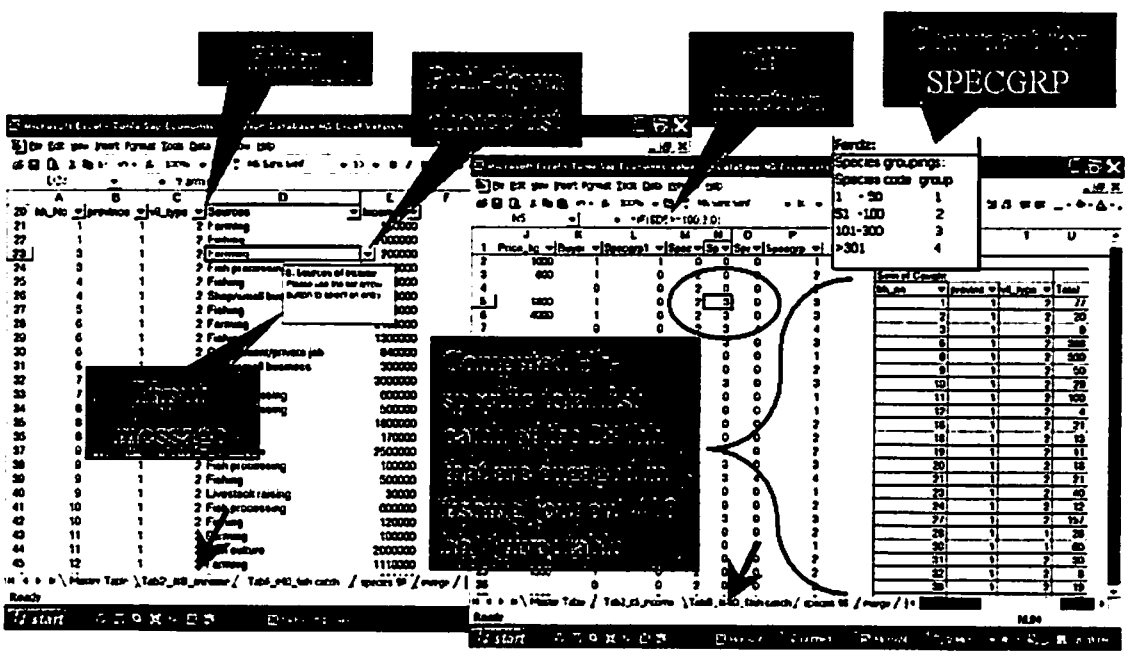
Below are some important points to remember when designing a data entry system in Excel.

- Don't mix numeric and alpha numeric entries for a particular variable. Avoid using alpha numeric/text values/or non-numerical values as variable type. If the use of alpha numeric is unavoidable, use drop down list to avoid wrong spellings. Note that MS Excel and most statistical software are sensitive to lower and upper cases of data values. Thus, the data values **"Hatchery"** and **"hatchery"** are considered different values.

- Measurement of numeric values must be uniform in all samples. If respondents gave different units, you may have entries with different unit provided that the you create another variable for the unit or the conversion rate.
- Do not enter the measurement and unit in one variable. If units are uniform, then the name of the unit as an entry is negligible.
- No calculation or conversions by hand. The raw data should be entered directly into the spreadsheets.

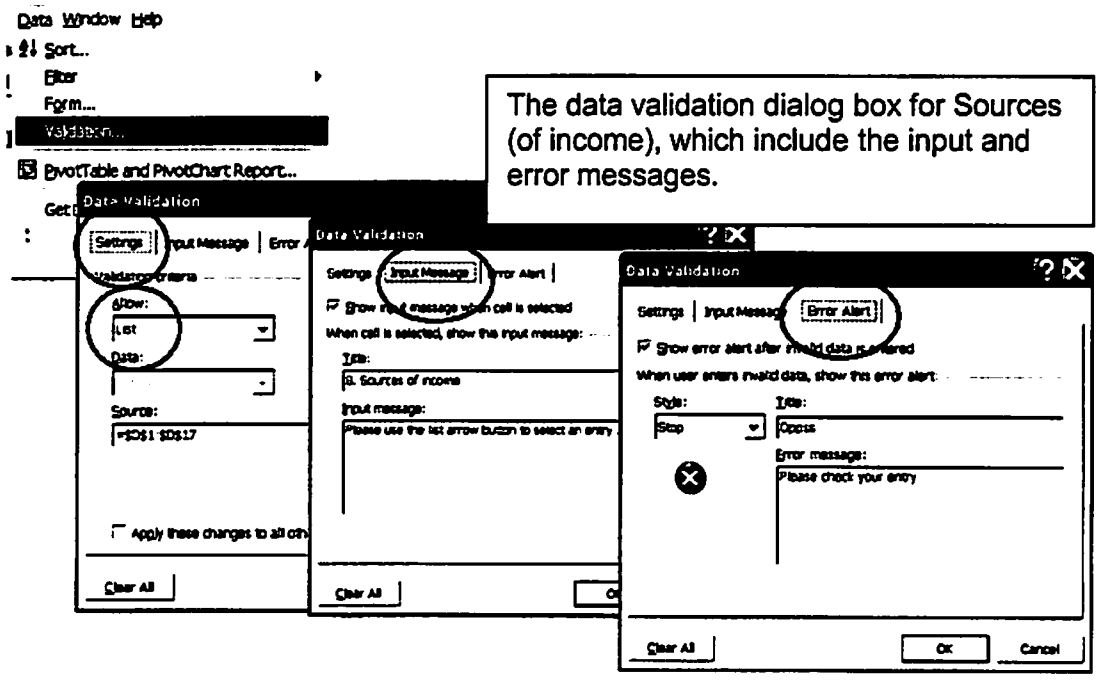
Tonle Sap Economic Evaluation Database- A Proto-type in MS Excel

For demonstration purposes, we built a proto-type workbook database version of Tonle Sap Economic Valuation database using MS Excel (Proto Type Tonle Sap Database in Excel.XLS). Our purpose is to emphasize the effective use of the software, thus we only included three tables (**Master Table 1, Tab2 Item 8 Income** and **Tab8 item 40 fish catch**). Each table is stored in a separate worksheet.



This is actually a **Workbook** where a *Table* is placed in one **Worksheet**. The workbook becomes a database. Though it only provides minimal data entry forms, the validation checks, input and error messages proved to be very effective and ensure that the correct information is being entered. Its drop-down list feature allows the user to select entries from a specified choice list. This is superimposed by the variable **Sources** (of income) in the figure above under worksheet **Tab2 it8 Income**. As will be discussed in more detail data validation, supplying input and error messages can be done using the **Validation** option of the **Data** menu.

However, It has no capability for disallowing double entry. An encoder might mistakenly enter one record twice. Another limitation of this database or (or any database developed using MS Excel) is that it does not enforce entity or referential integrity.



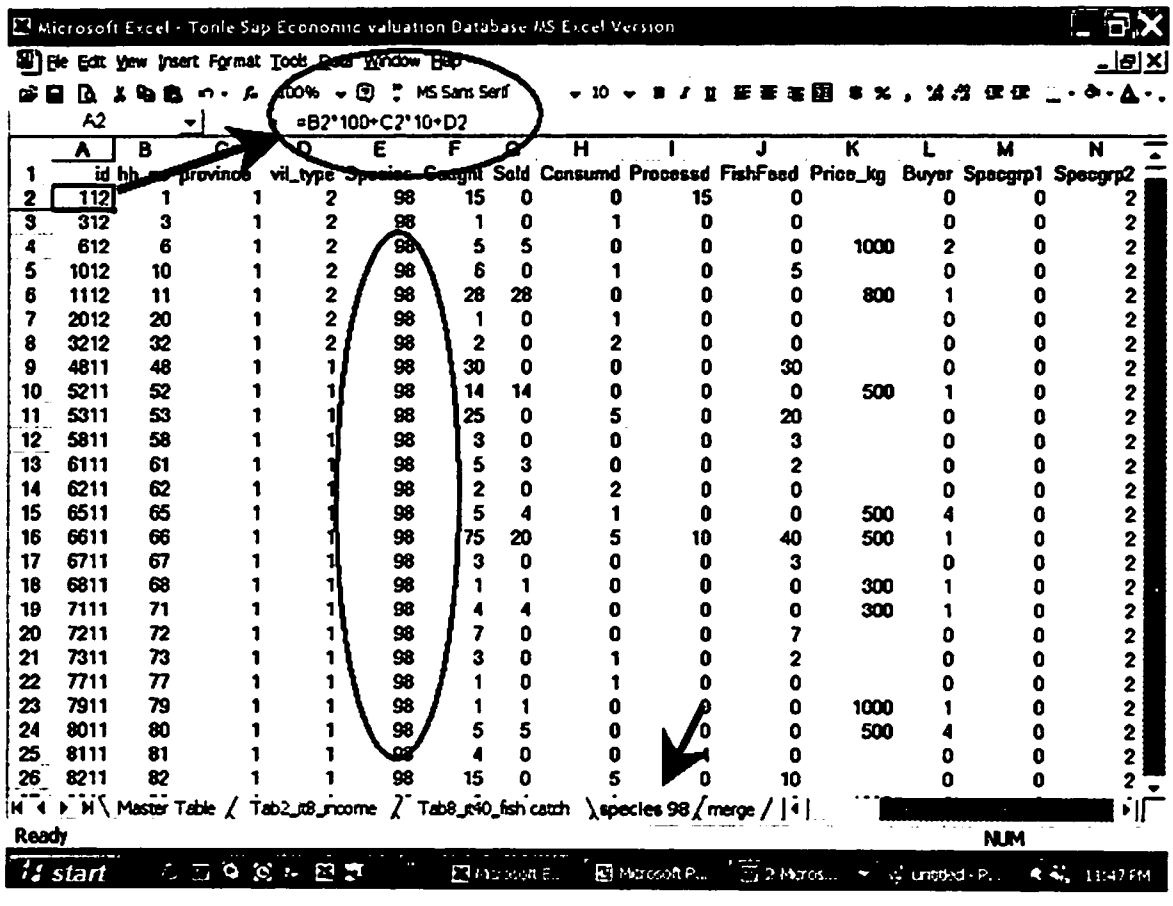
The data validation dialog box for Sources (of income), which include the input and error messages.

To facilitate easy data processing we utilized some built in function of MS Access. In the original MS Access version of the database which was discussed in a separate document, we found out that there were around 200 fish species reported being caught.

The **Specgrp** field in the **Tab8 it8 Fish catch** group these species into four (4) categories. **A cell comment** is placed to the field that show which species belong to which group. The whole process involves using the **IF** function to first generate the fields **Specgrp1** to **Specgrp4.** The field **Specgrp** is generated using the Max (maximum) function and take the maximum value of **Specgrp1** to **Specgrp4** – the true species category.

Also in **Tab8 it8 Fish catch** are the computed household-specific total fish catch for 281 households (out of 410 total households) who are engaged in fishing. This subset of data is generated using **Pivot table**, which is accessible using the **Data** menu. Pivot table is a powerful tool of MS access for generating cross tabulation. Pivot table is discussed in more detail in the next section.

The database also contained two other Worksheets: The **species 98** and **merge.**

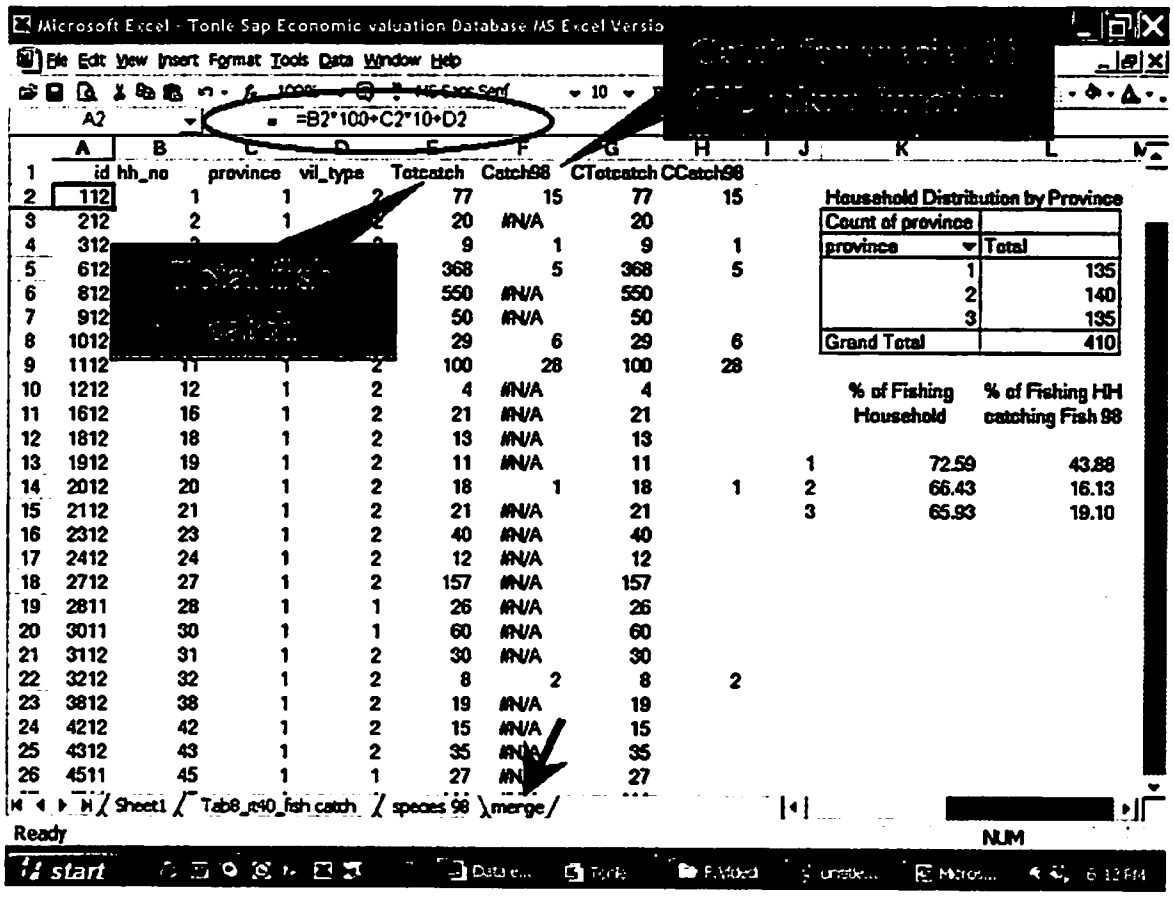


	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	id	hh_no	province	vil_type	Species	Caught	Sold	Consumd	Processd	FishFeed	Price_kg	Buyer	Specgrp1	Specgrp2
2	112	1	1	2	98	15	0	0	15	0		0	0	2
3	312	3	1	2	98	1	0	1	0	0		0	0	2
4	612	6	1	2	98	5	5	0	0	0	1000	2	0	2
5	1012	10	1	2	98	6	0	1	0	5		0	0	2
6	1112	11	1	2	98	28	28	0	0	0	800	1	0	2
7	2012	20	1	2	98	1	0	1	0	0		0	0	2
8	3212	32	1	2	98	2	0	2	0	0		0	0	2
9	4811	48	1	1	98	30	0	0	0	30		0	0	2
10	5211	52	1	1	98	14	14	0	0	0	500	1	0	2
11	5311	53	1	1	98	25	0	5	0	20		0	0	2
12	5811	58	1	1	98	3	0	0	0	3		0	0	2
13	6111	61	1	1	98	5	3	0	0	2		0	0	2
14	6211	62	1	1	98	2	0	2	0	0		0	0	2
15	6511	65	1	1	98	5	4	1	0	0	500	4	0	2
16	6611	66	1	1	98	75	20	5	10	40	500	1	0	2
17	6711	67	1	1	98	3	0	0	0	3		0	0	2
18	6811	68	1	1	98	1	1	0	0	0	300	1	0	2
19	7111	71	1	1	98	4	4	0	0	0	300	1	0	2
20	7211	72	1	1	98	7	0	0	0	7		0	0	2
21	7311	73	1	1	98	3	0	1	0	2		0	0	2
22	7711	77	1	1	98	1	0	1	0	0		0	0	2
23	7911	79	1	1	98	1	1	0	0	0	1000	1	0	2
24	8011	80	1	1	98	5	5	0	0	0	500	4	0	2
25	8111	81	1	1	98	4	0	0	4	0		0	0	2
26	8211	82	1	1	98	15	0	5	0	10		0	0	2

Species 98 worksheet is a subset of **Tab8 it40 Fish catch** and which as its name contained catch and disposal information of only those 78 households catching species 98 (out of 281 fishing household; out of 410 total households). This is generated using the Filters – a very simple mechanism (autofilter) for selecting data subsets. Autofilters can be set up to choose specific values or a range of values. Several filters can be used, each acting further on the current data subset. An advanced filter is provided for more complicated selections.

Note the field "**id**" at column "**A**" of the worksheet. This serves as a single variable unique identifier. Though the "**hh_no**" in this case can be used since each of its values are unique. The **id** field is an example of creating a single variable primary key whose values are the concatenated values of the original sets of primary keys (hh_no, province and village type). This is done by using the formula **hh_no*100 + province*10 + vil_type**. Such single variable primary key is needed in merging variables using the **Vlookup** function, which will be discussed next.

The values of the field "**id**" can also be created using the **Concatenate** function. However, the result of a **concatenate** function is text string, not numeric.

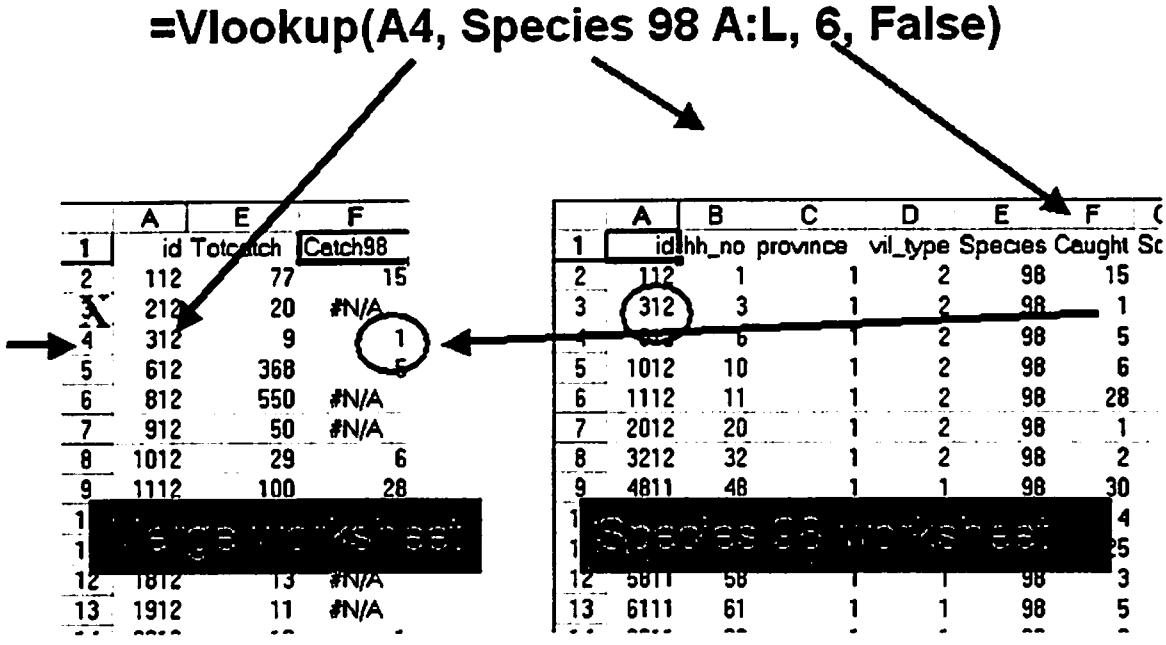


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	id	hh_no	province	vil_type	Totcatch	Catch98	CTotcatch	CCatch98					
2	112	1	1	2	77	15	77	15					
3	212	2	1	2	20	#N/A	20						
4	312	3	1	2	9	1	9	1					
5	612	6	1	2	368	5	368	5					
6	812	8	1	2	550	#N/A	550						
7	912	9	1	2	50	#N/A	50						
8	1012	10	1	2	29	6	29	6					
9	1112	11	1	2	100	28	100	28					
10	1212	12	1	2	4	#N/A	4						
11	1612	16	1	2	21	#N/A	21						
12	1812	18	1	2	13	#N/A	13						
13	1912	19	1	2	11	#N/A	11						
14	2012	20	1	2	18	1	18	1					
15	2112	21	1	2	21	#N/A	21						
16	2312	23	1	2	40	#N/A	40						
17	2412	24	1	2	12	#N/A	12						
18	2712	27	1	2	157	#N/A	157						
19	2811	28	1	1	26	#N/A	26						
20	3011	30	1	1	60	#N/A	60						
21	3112	31	1	2	30	#N/A	30						
22	3212	32	1	2	8	2	8	2					
23	3812	38	1	2	19	#N/A	19						
24	4212	42	1	2	15	#N/A	15						
25	4312	43	1	2	35	#N/A	35						
26	4511	45	1	1	27	#N/A	27						

Finally, the "**merge**" worksheet as it name imply, contained merged data. Specifically, it is a merge data on total fish catch (**Totcatch**) of the 281 fishing households and Catch quantity of specie 98 (Catch98) of the 76 households (of the 281 fishing households) who claimed of having caught the fish. Another way of looking at it is to match to fishing household their catch for species 98.

The process involved using the **Vlookup** function. The **Totcatch** information, which was generated using pivot table in **Tab8 it8 Fish Catch** worksheet, was copied in this worksheet and was used as the base file. The values of **Catch98** field is lookup from **Species 98** worksheet using **id** field (**A2** in formula bar) as the matching field (common to both data. The figure below illustrates how the **Vlookup** function works. At Cell F4 of the merge worksheet is, we write the formula: **=Vlookup(A4, Species 98 A:L, 6, False)**.

This means put in this cell (**F4**, the value which can be found in the 6th column (column **F**) in the range **A:L** of **Species 98** worksheet with corresponding First column value equal to **312** (A4 in merge worksheet). The value "#N/A" was found in cells F3, F6, F7 with corresponding id 212, 812 and 912, respectively, due to these **ids** are not found in the **Species 98** worksheet.



Vlookup from the Help menu of Excel

VLOOKUP

Searches for a value in the leftmost column of a table, and then returns a value in the same row from a column you specify in the table. Use VLOOKUP instead of HLOOKUP when your comparison values are located in a column to the left of the data you want to find.

Syntax

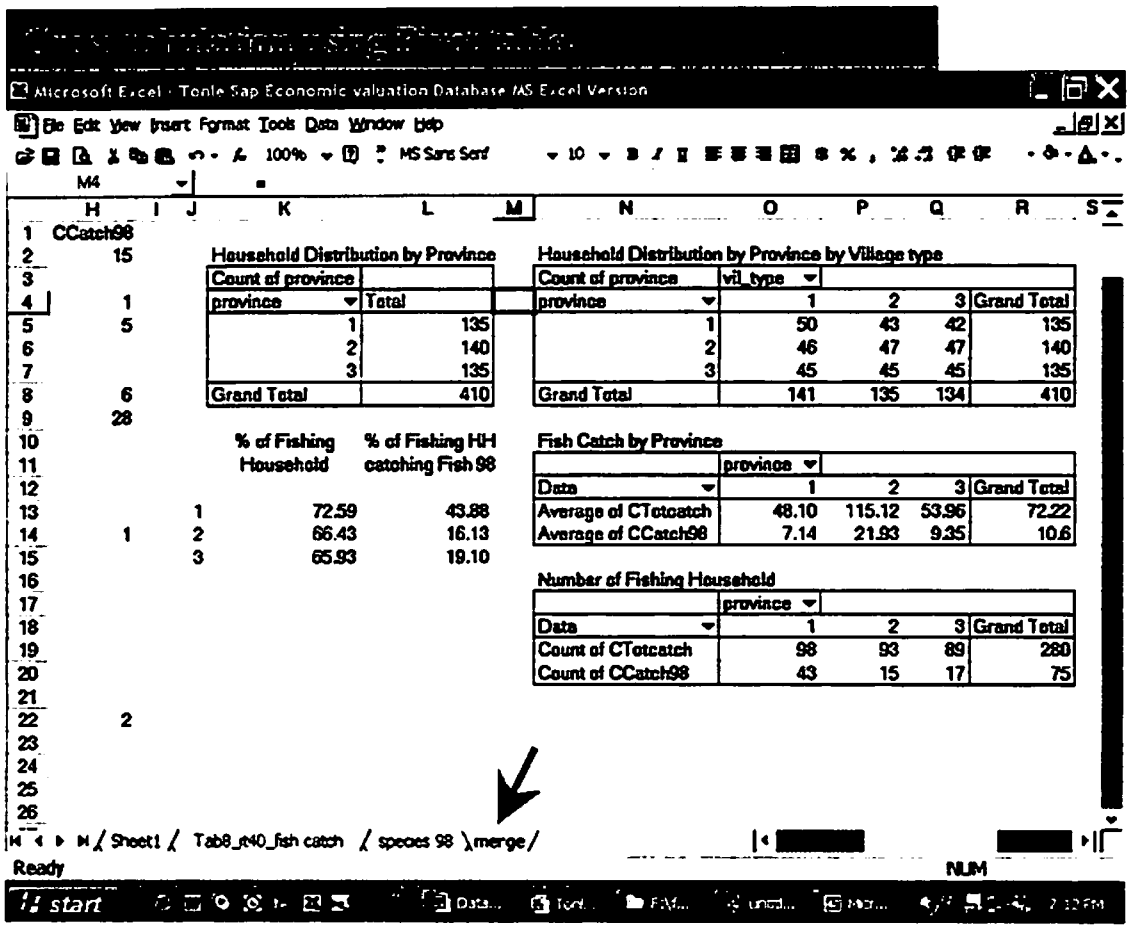
VLOOKUP(lookup_value,table_array,col_index_num,range_lookup)

Lookup_value is the value to be found in the first column of the array.
Table_array is the table of information in which data is looked up. Use a reference to a range or a range name, such as Database or List.

Col_index_num is the column number in **table_array** from which the matching value must be returned.

Range_lookup is a logical value that specifies whether you want VLOOKUP to find an exact match or an approximate match. If TRUE or omitted, an approximate match is returned. In other words, if an exact match is not found, the next largest value that is less than lookup_value is returned. If FALSE, VLOOKUP will find an exact match. If one is not found, the error value #N/A is returned.

At far right columns of the merge worksheet are four summary tables.



Microsoft Excel - Tonle Sap Economic valuation Database MS Excel Version

File Edit View Insert Format Tools Data Window Help

M4

CCatch98 = 15

Household Distribution by Province

Count of province	
province	Total
1	135
2	140
3	135
Grand Total	410

Household Distribution by Province by Village type

Count of province	vil_type			
province	1	2	3	Grand Total
1	50	43	42	135
2	46	47	47	140
3	45	45	45	135
Grand Total	141	135	134	410

		% of Fishing Household	% of Fishing HH catching Fish 98
1	1	72.59	43.88
	2	66.43	16.13
	3	65.93	19.10

Fish Catch by Province

	province			
Data	1	2	3	Grand Total
Average of CTotcatch	48.10	115.12	53.96	72.22
Average of CCatch98	7.14	21.93	9.35	10.6

Number of Fishing Household

	province			
Data	1	2	3	Grand Total
Count of CTotcatch	98	93	89	280
Count of CCatch98	43	15	17	75

2 = 22

Sheet1 / Tab8_it40_fish catch / species 98 \merge/

Ready NUM

start

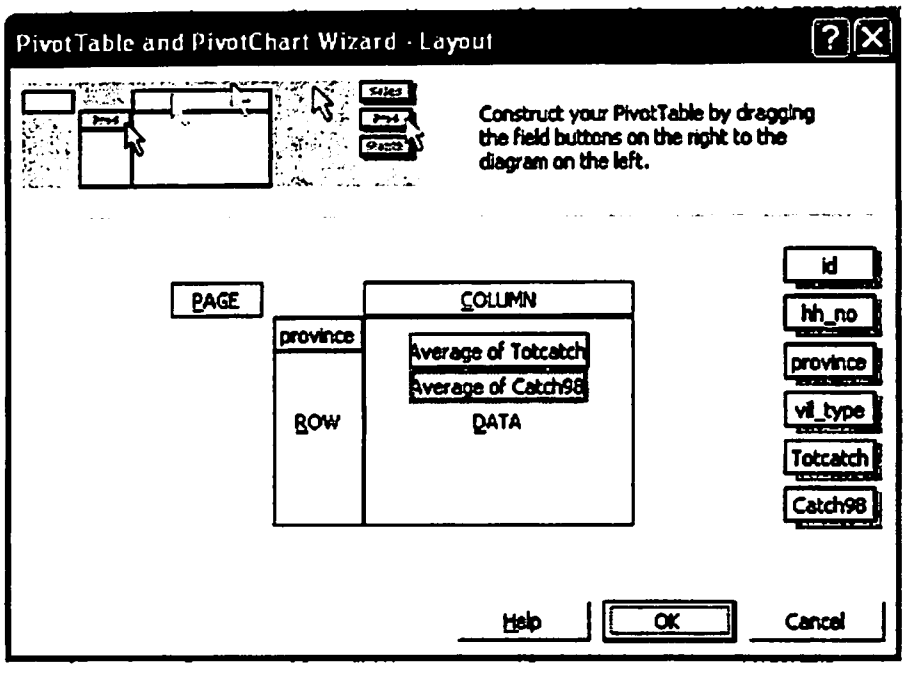
The tables **Household distribution by province, Fish catch by province, Number of Fishing Households** are all *one-way tables* while the **Household Distribution by Province by Village type** is *a two-way table.* All these tables were generated using the **Pivot Table,** which is accessible under the **Data** menu. **Pivot table** is one of the most powerful tools of MS Excel, even more powerful than the cross tabulation features of the major statistical software. The tables can be modified interactively to do any of the following:

(1) add or delete other variables

(2) add/change summary statistics (minimum, Maximum, Standard Deviation, Variance) and percentages

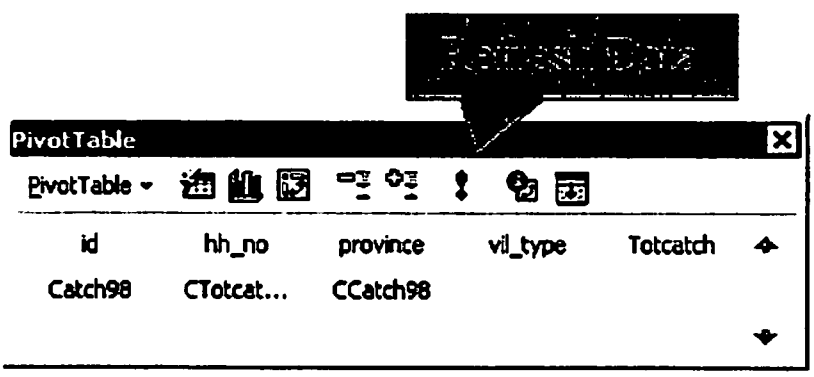
(3) change the level of the cross tabulation table (two-dimension, three-dimension) and orientation

All these can be done under the **Wizard-Layout** dialog box of the Pivot Table:



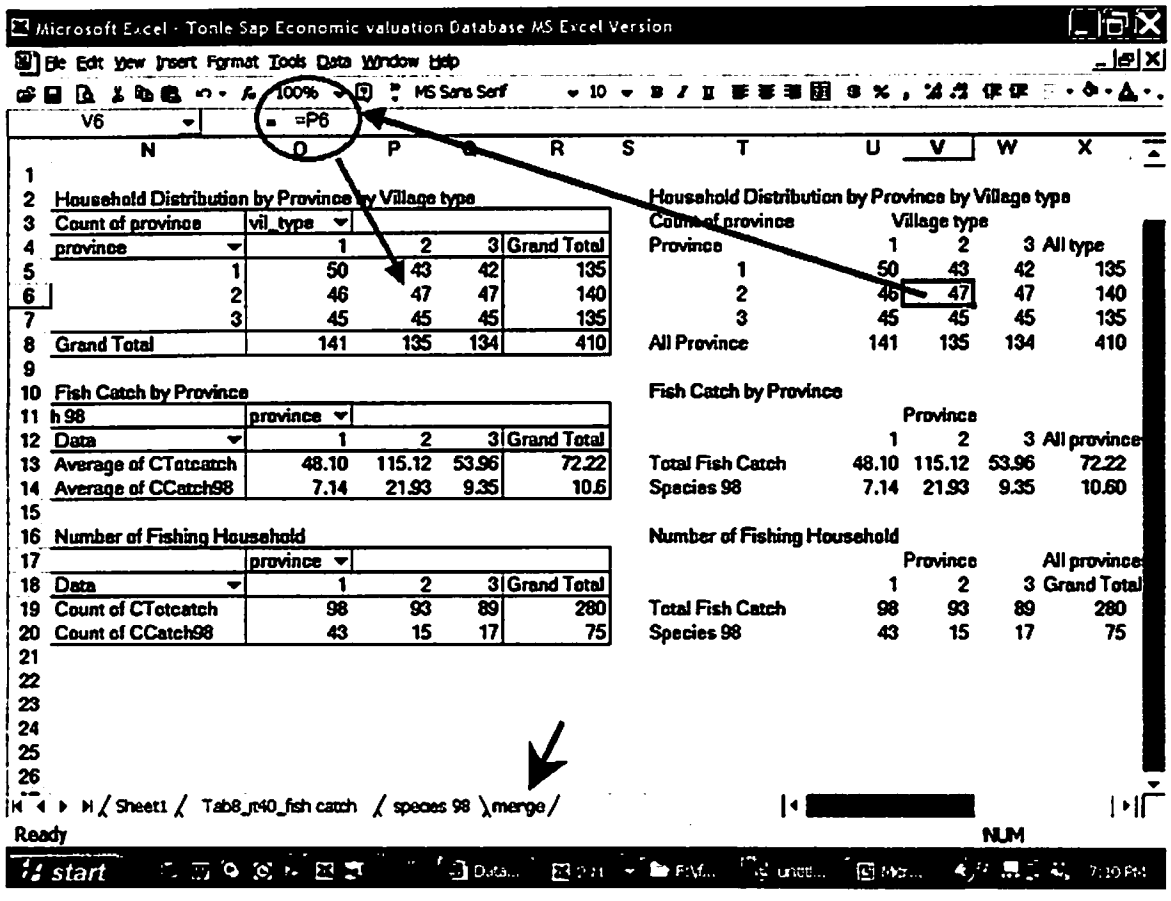
In this layout, the field names appear as a set of buttons on the right. Pay particular attention to the **Row, Column and Data** Sections of the layout. Under row, field province is used. Under **Data, Totcatch** and **Catch98** fields are used. To specify what fields will be used to which section, simply drag and drop the fields one after another to the section where you want them to be. Note that the fields used under **Row** and **Column** are normally categorical variables. Using measurement variables in these sections will produce a large table of nonsense. The fields under the **Data** section are the variables you want to summarize. These fields are normally numeric variables.

The layout above will produce the **Fish Catch by Province** table above. It is worth mentioning in this stage that if you updated data after you have set up your pivot table, just click on the Refresh data button to update your table.



In the next section, we will discuss in detail how to construct pivot tables including changing layout to produce two and three dimension tables and change the summary statistics.

Adjacent to the Pivot Tables are **formatted** Tables.



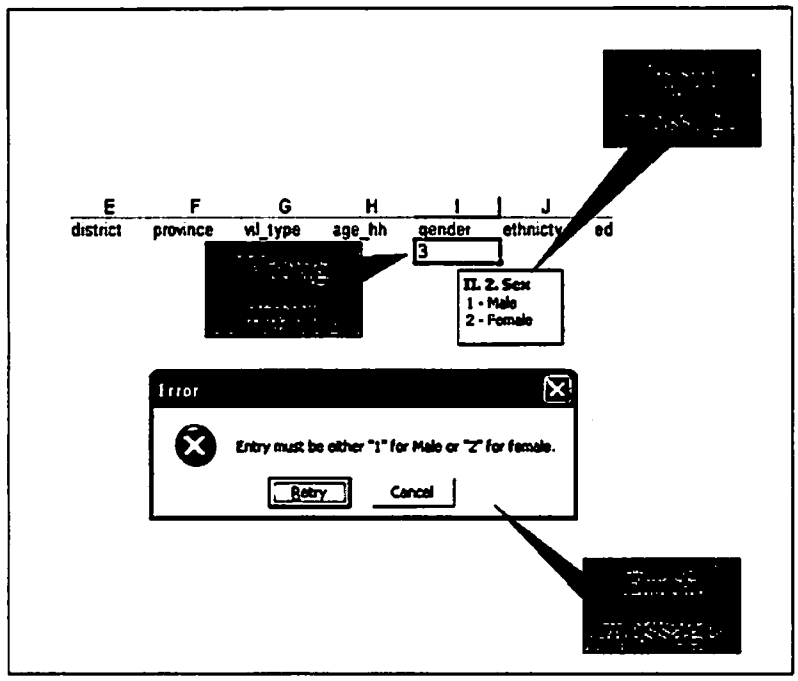
Formatted Tables are copies of the Pivot Tables. They are created since labels of Pivot Tables cannot be formatted. For example, you may want to change the word "**Average of Ccatch98**" to the actual name of the fish (say Giant catfish). Note that the values under the formatted (copy) table are equate/linked to the corresponding cell of the original pivot table. Linking will automatically update the figures in the formatted table every time you update your Pivot table using the **Refresh** command button.

Building theTonle Sap Economic Database using Excel

In this section we will reconstruct the database for illustration and better understanding.

Setting a Validation Rule, Input and Error Alert messages on a Variable

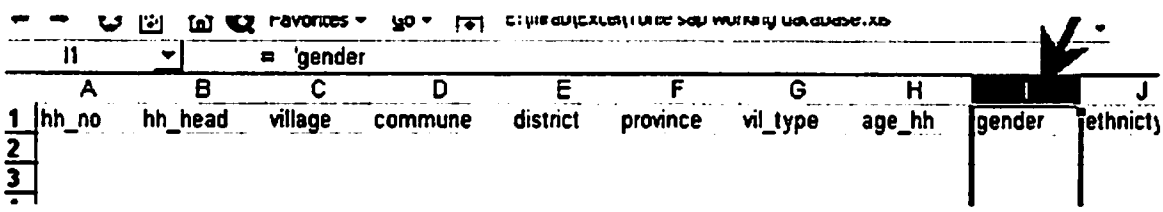
We first illustrate how to enforce validation checks, create input and error messages in MS Excel. Data validation, input and error messages can be accessed and done using the **Data** Menu of Excel. To save time, we will assume that the variables were already created.



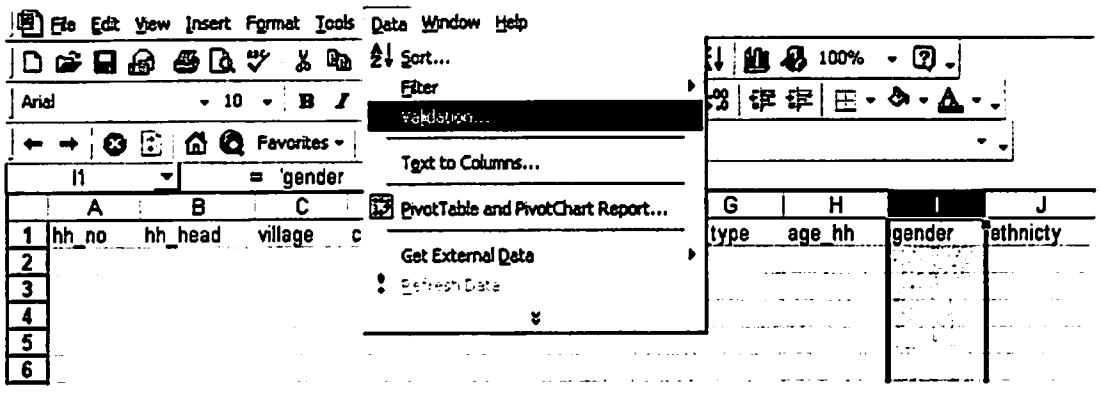
As an illustration, we will impose validation check on variable **Gender** (item 8) which will take values "1" if Male and "2" if Female. If encoder will mistakenly enter values other than these values (say "3"), an **Error alert** box will pop-up. We will also create **Input message** showing the details of the variable (from what item number in questionnaire and the possible values it take).

Setting a Validation Rule

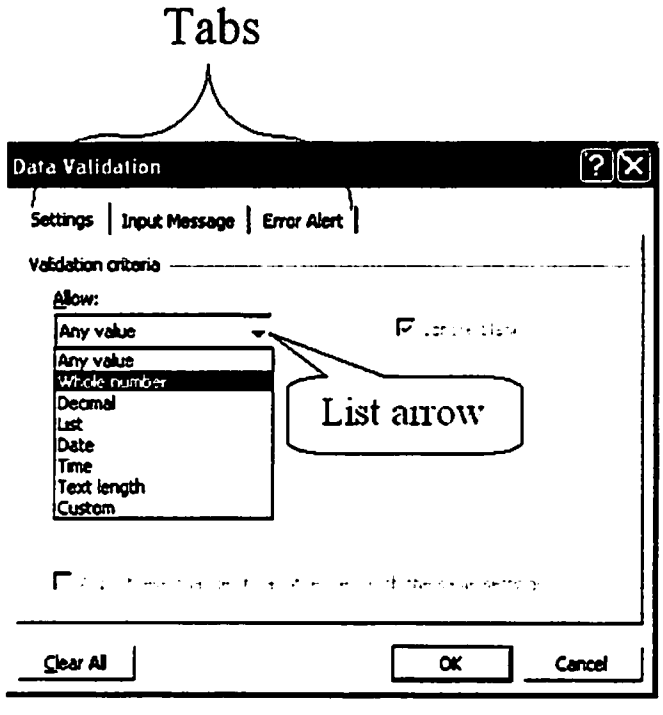
1. Highlight the column of the Variable (GENDER) by clicking on its column letter (I).



2. Click on **Data** menu, choose **Validation...** - a *Data Validation* dialog box will appear:

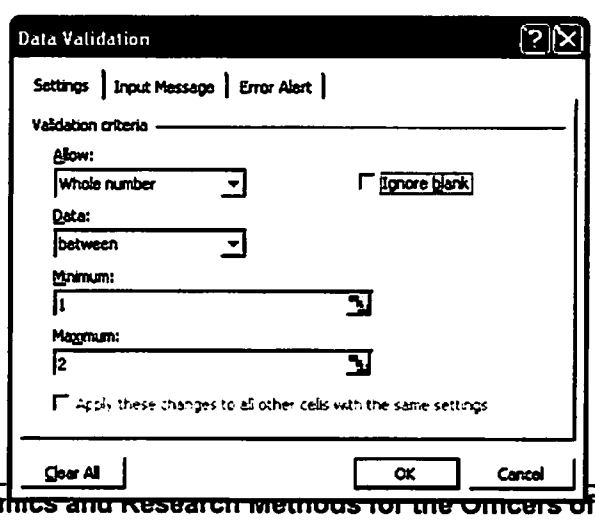


Note: There are three tabs in a **Data Validation** dialog box. **Settings** allows you to set validation rule/criteria. **Input message** allows you to write the message about the cell or in general the variable and **Error Alert** allows you to the write message once validation rule is not met (wrong entry).



3. Under **Validation Criteria** on the **Settings** tab, click on the *list arrow* attached to **Allow:** and Click on the option **Whole number** from the list provided.

4. Further settings appear: type in a *Minimum:* of **"1" (if male)** and a *Maximum:* of **"2"** if Female



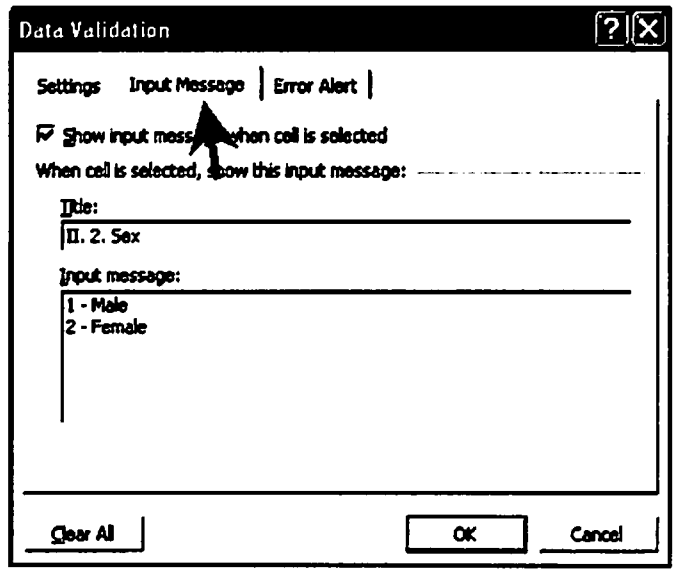
Note: You have to set up maximum/minimum values - Excel doesn't allow you to leave these blank. These need not be fixed values, as here, but could be references to other cells.

Setting the Input message

5. Click on the **Input Message** tab of the *Data validation* dialog box. Type "II. 2. Sex" under *Title*.
6. Under *Input message*:, type:

   "1 - Male

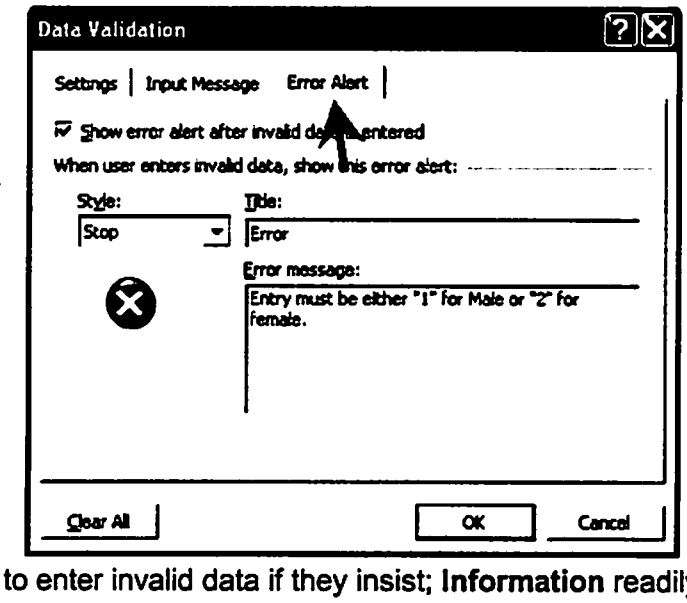
   2 - Female"



Customizing the Error Alert message

At this point, you can actually close the **Data Validation** dialog box without completing the **Error Alert** tab. Once the **Validation Criteria** is set, the **Error alert** message will be automatically activated. However, the warning message will not very helpful as it stands. It tells you there is an error but doesn't tell you what you need to type. You can customize the message as follows:

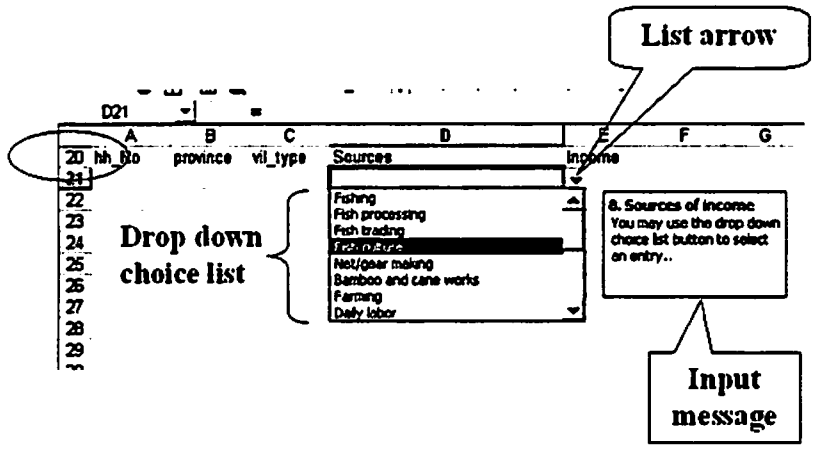
7. Click on the **Error Alert** tab of the Data validation dialog box.
8. Type "Error" under *Title*.
9. Under *Error message*:, type: "Entry must be either "1" if male or "2" if female.
10. Close the dialog box by clicking on **Ok** command button.



Note that Excel provides three levels of warning: **Stop** forces the user to retry until valid data is entered; **Warning** allows the user to enter invalid data if they insist; **Information** readily accepts invalid data.

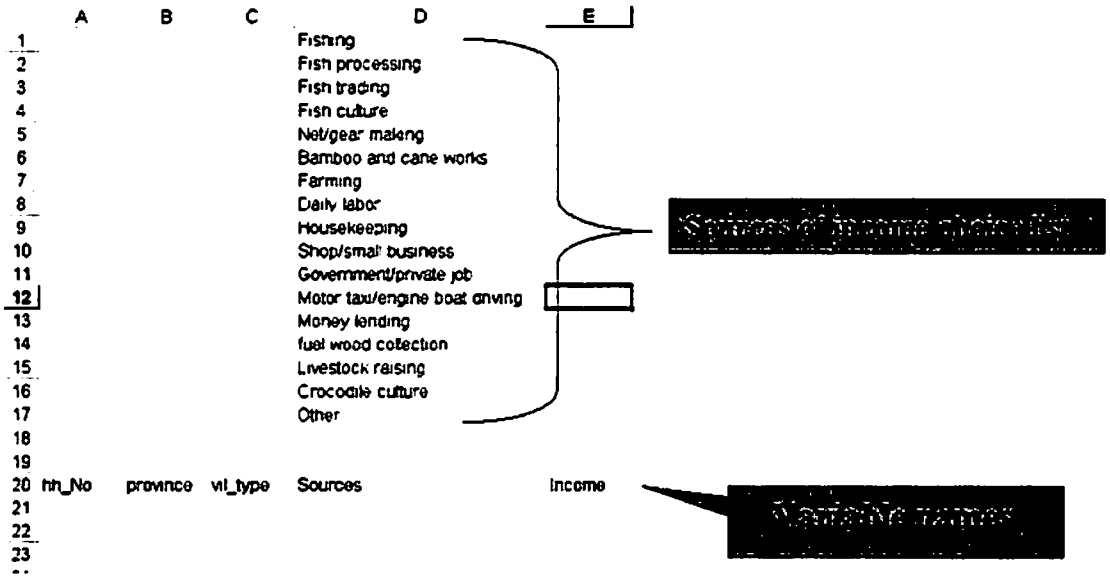
Setting Drop-down Choice List

So far you have only looked at categorical, indeed whole number, validation. You can similarly check for interval and numbers with decimal points. Other possibilities are dates/times and text up to a certain number of characters. Another option allows data entry from a fixed list of values (numeric or non-numeric). We will illustrate next how to create drop down choice list. We will use sources of income as our example. The figure below will set as our target output.



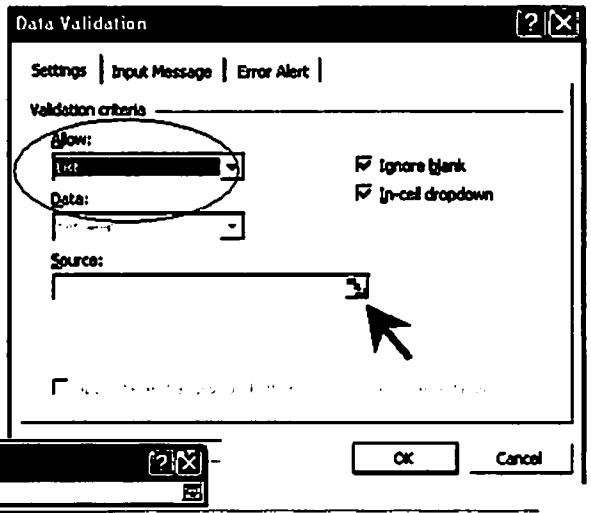
The list of possible sources of income is written in rows 1 to 20, which are hidden. The variable names are written in row number 20. The step in creating this is as follows:

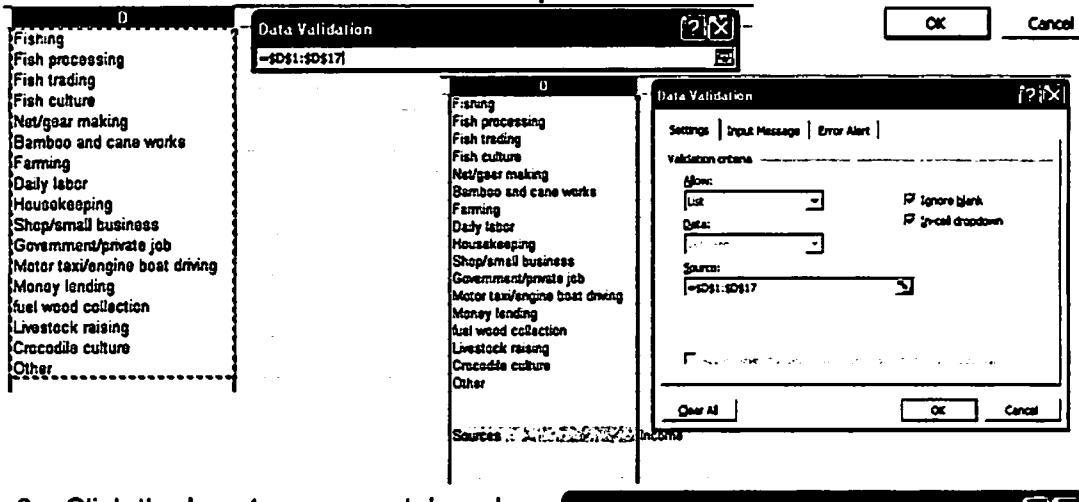
1. List down the possible sources of income above the their variable name (Sources). Each row corresponds to a source.



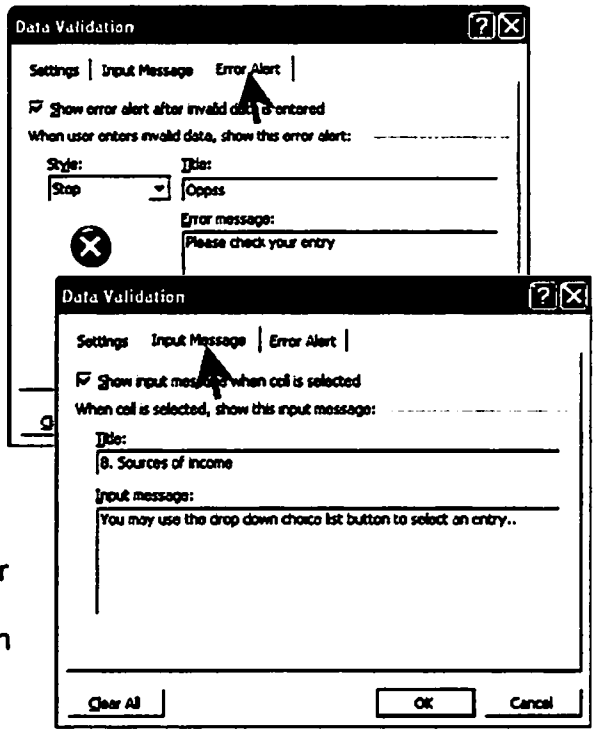
2. Highlight the whole column where the variable Sources is written by clicking on the column D then open the Data Validation dialog box clicking the validation option under Data menu (step 1 and 2 above).

3. Under **Validation Criteria** on the **Settings** tab click on the **list arrow** attached to **Allow:** and Click on the option **List** from the list provided.
4. Click on the **list arrow** attached to **Source** to specify the source of the choice list.
5. Highlight the list of the possible sources of income (D1:D17) then hit the *Enter* key.





6. Click the **Input message** tab and fill in the necessary input message and title:
7. Click the **Error Alert** tab and fill in the necessary input message and title:
8. Click on **OK** command button to close the Data Validation dialog box.



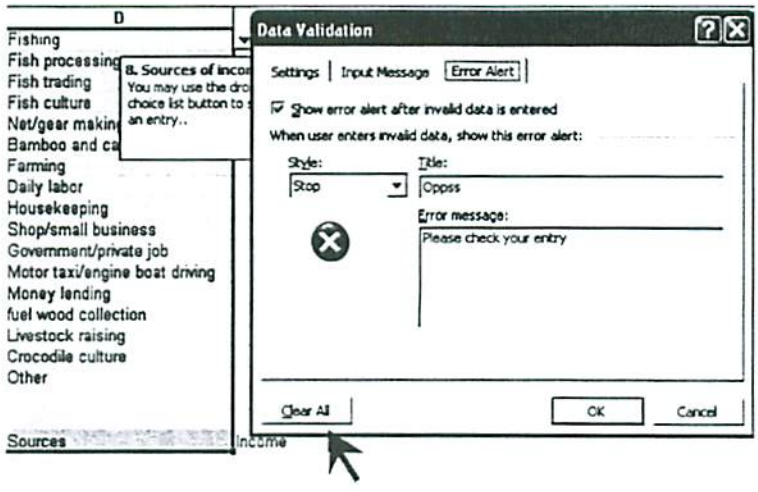
Note: Validations are made on data entry and not on calculated values. Validation can be set on the cells used in the calculation.

Sometimes, use has to be made of the **OR** or **AND** function. These offer alternative/joint tests respectively. Such type of validation can be done using the *Custom* option under Validation criteria.

Deleting validation rule and input and error alert messages

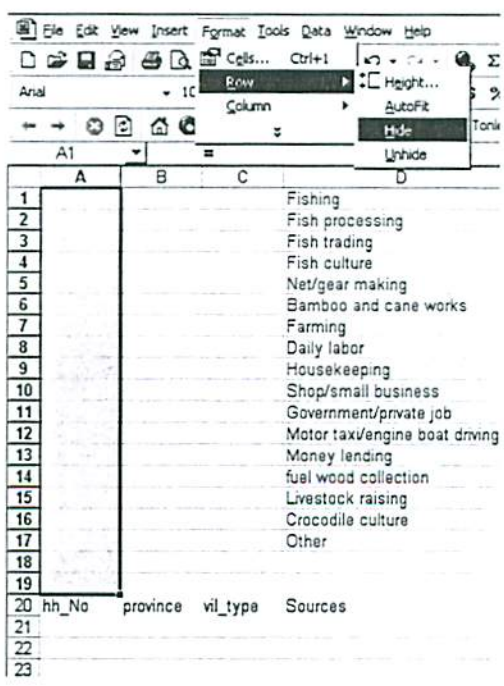
You can also delete (or copy) **validation rules** and **input and error alert messages**. For illustration, we will delete the validation rules including the input and error alert messages of cells D1:D20.

1. Highlight cells D1:D20.
2. Access the Data Validation dialog box by clicking on **Validation** under the **Data** Menu.
3. Click on the **Clear All** command button in found in the Data Validation dialog box.



Hiding rows/columns

1. To make the worksheet neat, you might opt not to display the list of possible sources of income. This can be done by hiding Rows 1 to 19. The following steps below illustrate how to do this task.
2. Highlight the row you want to hide (1 to 19).
3. Under the ***Format*** menu, select ***Row*** then ***Hide***.

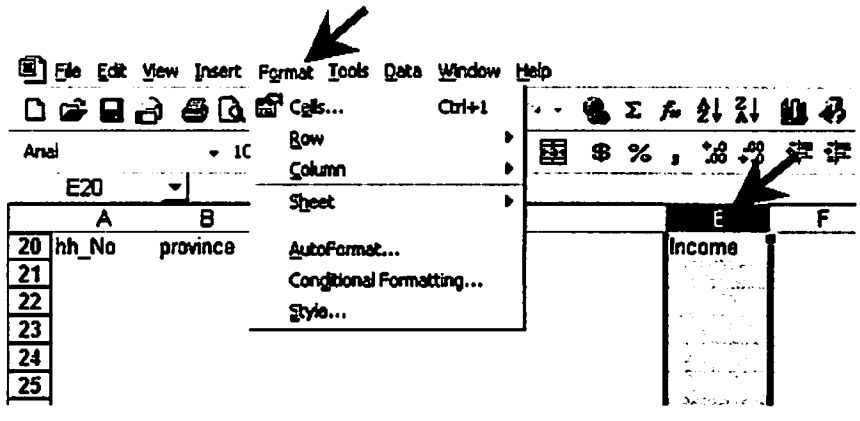


Formatting a cell/Column

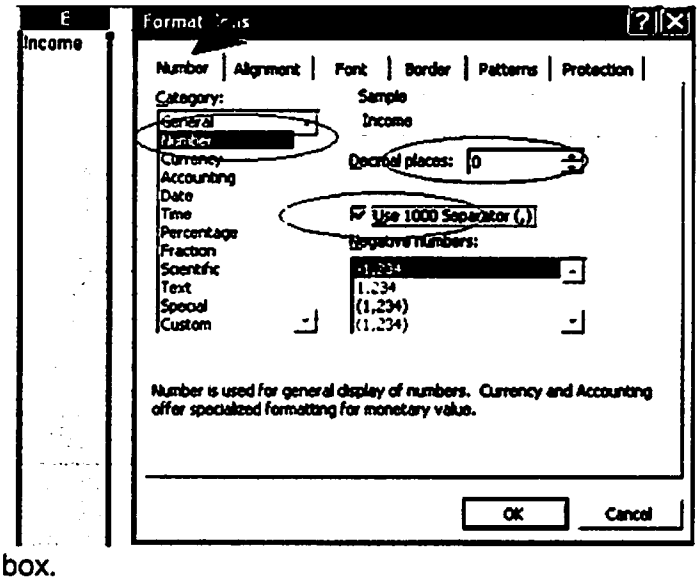
MS Excel allows you to format a number (e.g. how many decimal places) or date/time (e.g. to display the month as *August* or *Aug* or */08/*). Formatting makes a worksheet neat and more organized. With formatted values of a variable, it is easy to detect mistakes (if any). By applying a format you do not alter the contents of the cell in any way - just how it is displayed.

As an illustration, we will format the possible values of income variable in such a way that it will be an integer (no decimals) and use a comma (,) separator for values greater than or equal to 1,000. The steps below illustrate how it is done.

1. Highlight the column of the Variable (Income) by clicking on its column letter (E).
2. Under the **Format** menu, select **Cells...**



A format cells dialog box will appear.



3. Make sure that you are at *Number* tab. Click on the *list arrow* attached to *Category*. Click on the option **Number** from the list provided.
4. Set "0" for *Decimal places*.
5. Tick the Use 1000 Separator (,)
6. Click on OK command button to close the dialog box.

Selecting data subsets using Filter

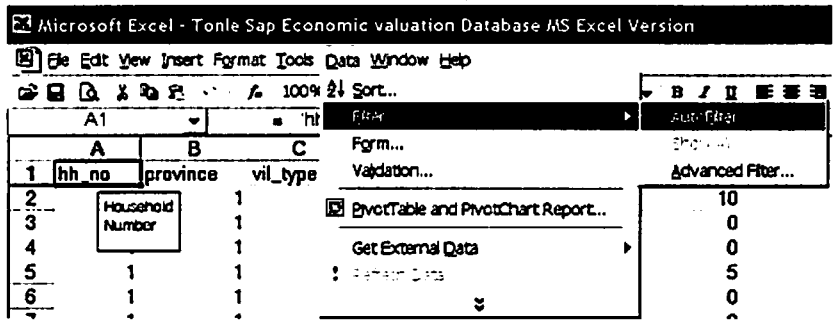
MS Excel provides a very simple mechanism (**AutoFilter**) for selecting data subsets. **AutoFilters** can be set up to choose specific values or a range of values. Several filters can be used, each acting further on the current data subset. An **Advanced Filter** is provided for more complicated selections.

In this section we will discuss in detail how to select subsets of data using **Filter**. Lets focus our attention on **Species 98** which is a result of **Filter** containing only the household who reported having caught **Species 98**. This set of data is thus a subset of Fish catch data under **Tbl8 it40 Fish Catch** worksheet.

It is important to note here that not all households catch a specific species (say fish 98) as not all households in the data are fishing households.

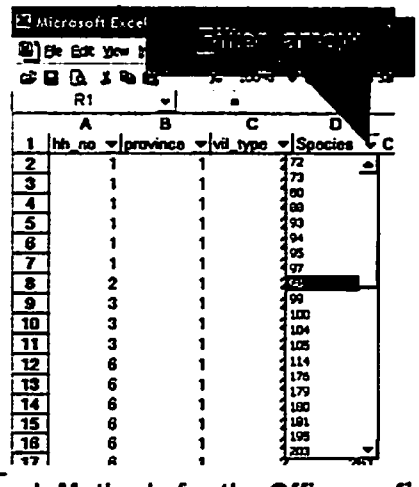
The following steps illustrate how this is done.

1. Under the **Tbl8 it40 Fish Catch** worksheet, click on **Data** menu then click on **Filter** followed by **AutoFilter**



Filter arrows are now attached to the to each filed names.

2. Click on the *filter arrow* attached to Species (D1) and choose **98.**



You now only have the data rows whose **Species** matches "**98**". Note the row numbering down the left hand side of the screen, which has turned blue, as has the filter arrow attached to **Species** (**D1**). These changes indicate that a filter is in operation. There are 76 number of filtered records.

Click on the *filter arrow* attached to **Species** field and choose (**All**). But before redisplaying all the data, copy the filtered data into a blank sheet and rename it (say **species 98**).

Microsoft Excel - Tonle Sap Economic valuation Database MS Excel Ver

File Edit View Insert Format Tools Data Window Help

100% MS Sans Serif 10

A1 'hh_no

	A	B	C	D	E	F
1	hh_no	province	vil_type	Species	Caught	Sold
4		1	2	98	15	0
10		1	2	98	1	0
16		1	2	98	5	5
26	10	1	2	98	6	0
32	11	1	2	98	28	28
59	20	1	2	98	1	0
101	32	1	2	98	2	0
123	48	1	1	98	30	0
149	52	1	1	98	14	14
158	53	1	1	98	25	0
177	58	1	1	98	3	0
201	61	1	1	98	5	3
208	62	1	1	98	2	0
238	65	1	1	98	5	4
243	66	1	1	98	75	20
251	67	1	1	98	3	0
256	68	1	1	98	1	1
268	71	1	1	98	4	4

Household Number

This is a very simple filter operation involving only one criteria. You can also filter subset of data base several criteria using the **AdvanceFilter** option.

Aggregating data using Pivot Table

The three different tables in this database are of different "levels". The Master Table is household level. The **Tbl2_it8_Income** is income "level" and the **Tbl8_it40_Fish Catch** is species level.

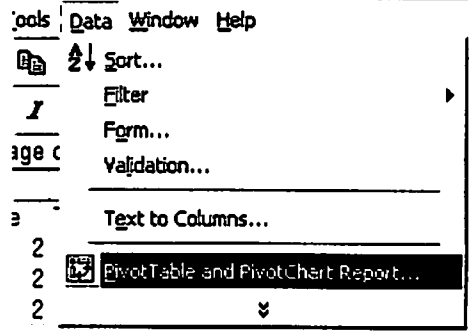
Of interest maybe is to aggregate species level data to produce a household level data containing household total catch, total fish consumed, total fish sold, etc. Then match this aggregated data to the households who catch fish coded 98. This is exactly what is contained in **merge** worksheet. In this section, we will illustrate how the aggregation was done using **Pivot table**. Here pivot Table is used to organize data. Matching the resulting aggregated data to the 76 filtered household who catch species 98 is illustrated in the next section.

Basically, we will generate a household level total catch for those fishing households, which will look like the screen shot figure below which can be found in **Tbl8_it40_Fish Catch** worksheet.

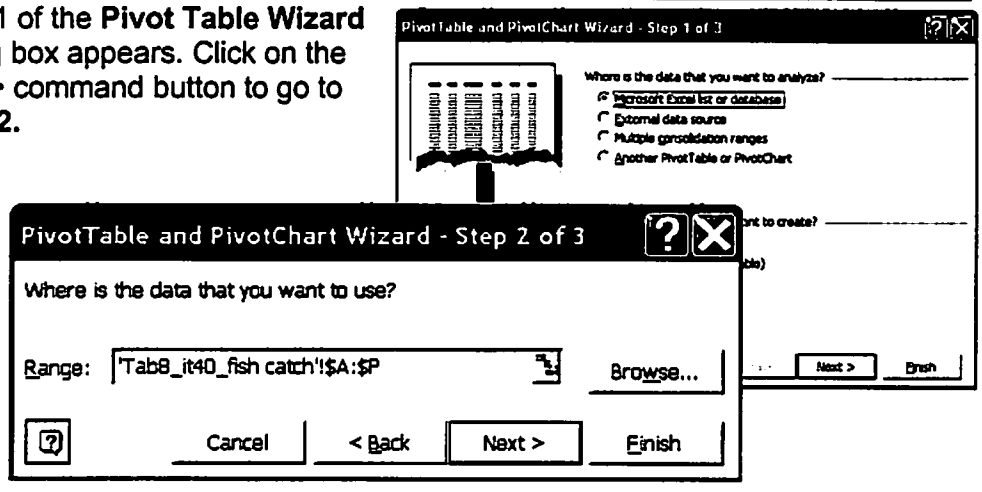
Sum of Caught			
hh_no	provinc	vil_type	Total
1	1	2	77
2	1	2	20
3	1	2	9
6	1	2	368
8	1	2	550
9	1	2	50
10	1	2	29
11	1	2	100
12	1	2	4
16	1	2	21
18	1	2	13
19	1	2	11

The following steps illustrate this job:

1. Under **Tbl8_it40_Fish Catch** worksheet, click on **Data** menu then Chose **Pivot Table...**

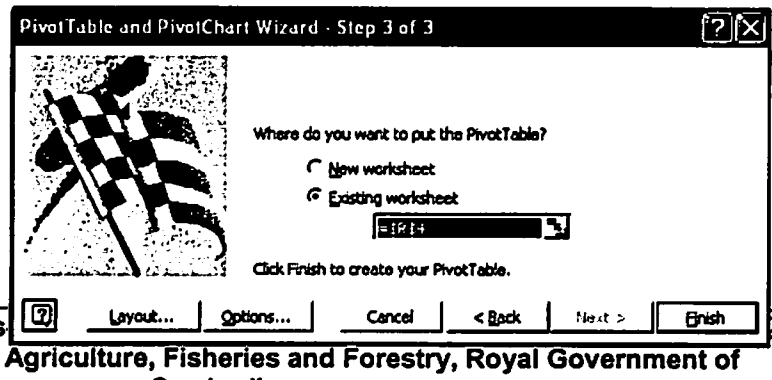


2. Step 1 of the **Pivot Table Wizard** dialog box appears. Click on the **Next>** command button to go to **Step 2.**



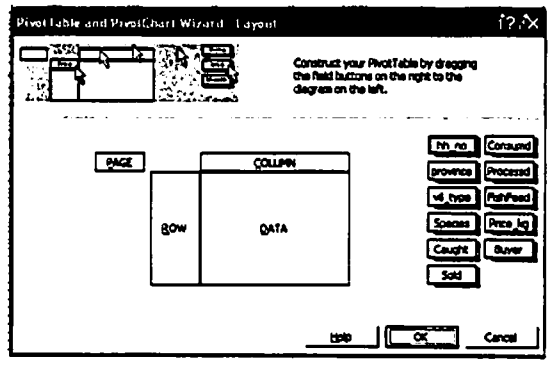
3. Step 2 of the **Pivot Table Wizard** dialog box ask you to specify the range of the data you want to use. Click the arrow attached to *Range:*, then specify the range by highlighting after which hit the **Return** (press **Enter**) key then click on **Next>** command button.

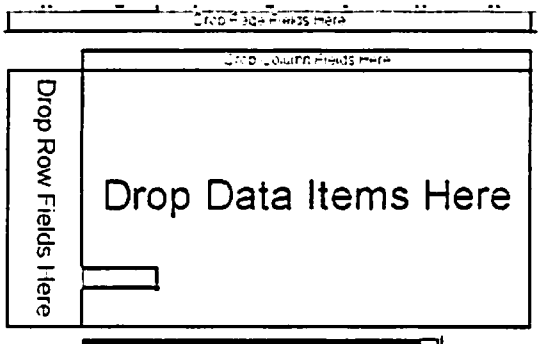
4. The 3rd and final step dialog box of Pivot Table appear asking you where



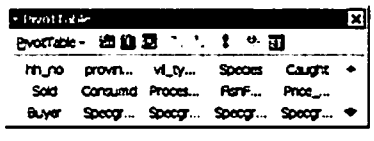
to put the Pivot Table.

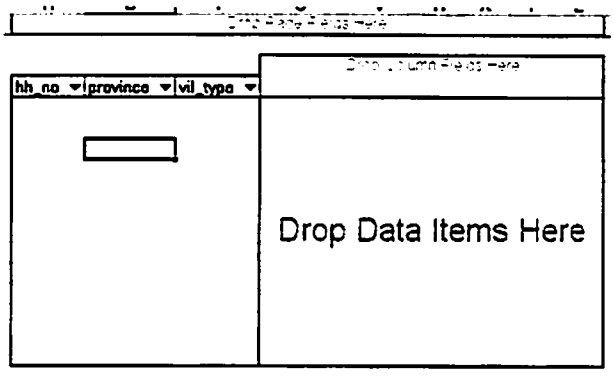
If you click on the **Layout** command, a **Wizard Layout** dialog box will appear as shown earlier.

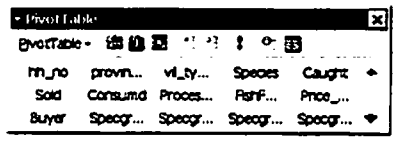




If you click on the **Finish** command, a **skeleton** Pivot table is drawn and Pivot Table toolbar and Field List appears.





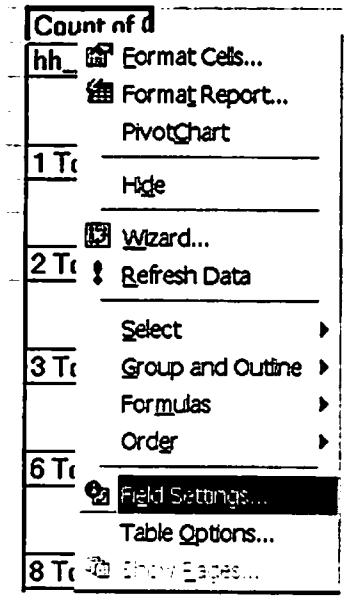


5. For the sake of illustration, click on the **Finish** command button to use the **skeleton** Pivot table.
6. Drag and drop **hh_no**, **province** and **vil_type** fields to the **Row (Drop Row Fields Here)** section of the skeleton one after another after which drag and drop **Caught** field under **Data (Drop Data Section Here)** section.

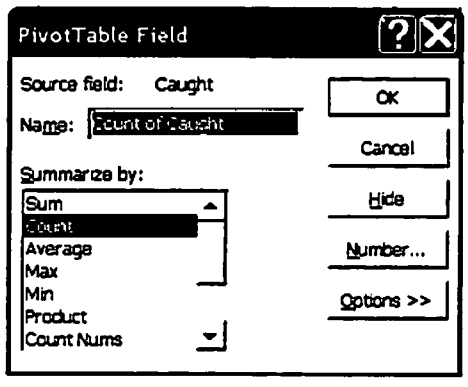
As soon as you drop the Caught field, the skeleton will transform to the Pivot Table shown below. Note that what is shown is the Count or the number of species caught by each of the fishing households. Note also that there are unnecessary rows total in the table. The next step is change the count statistics to SUM and hide the row totals.

Count of C			
hh_no	province	vil_type	Total
1	1	2	6
	1 Total		6
1 Total			6
2	1	2	1
	1 Total		1
2 Total			1
3	1	2	3
	1 Total		3
3 Total			3
6	1	2	7
	1 Total		7
6 Total			7
8	1	2	3

Changing summary stat in a Pivot table



7. Right click on the **Count of Caught** button in the top-most portion of the Table. A drop down menu will appear.
8. Select **Field Settings...**
9. A PivotTable of dialog box of the **Caught** Field appears. Select on **Sum** under *Summarize by*:



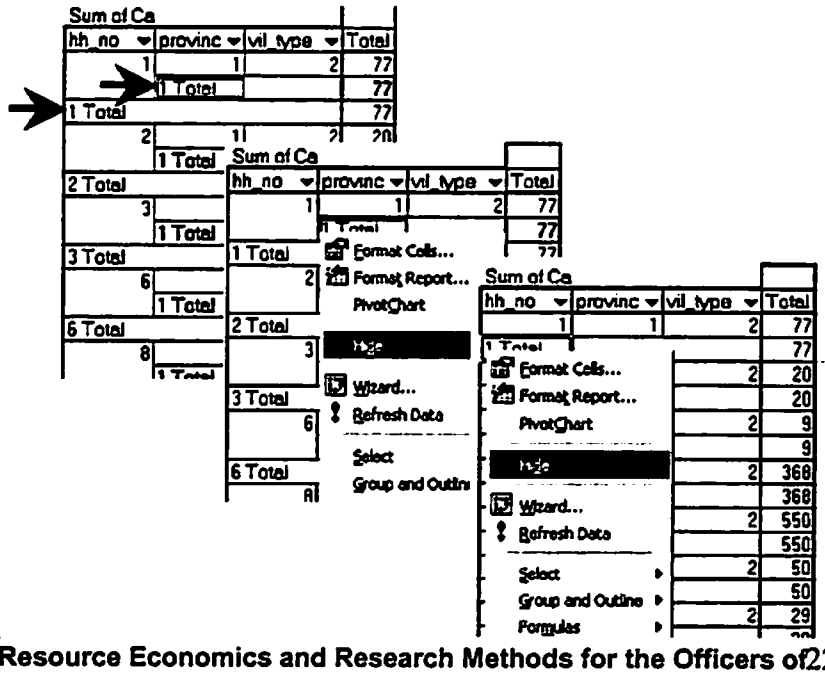
The **Count of Caught** button changes to **Sum of Caught.**

Hiding Row/Column total

The next step is hide the **Row Totals.**

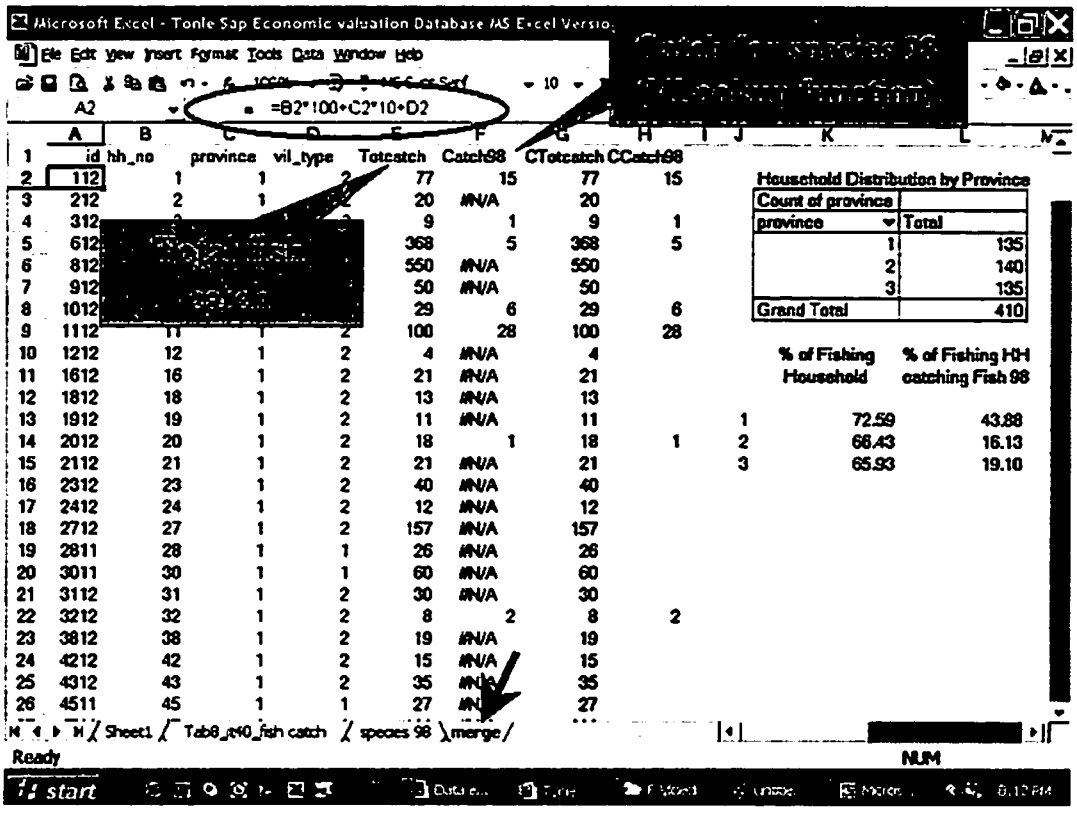
10. Right click one of the **Total** under **Province** heading.
11. A drop-down menu appears. Click on **Hide.**

All the **Row Total** under **Province** disappears. Do the same step on the **Row Total** under **hh_no.**



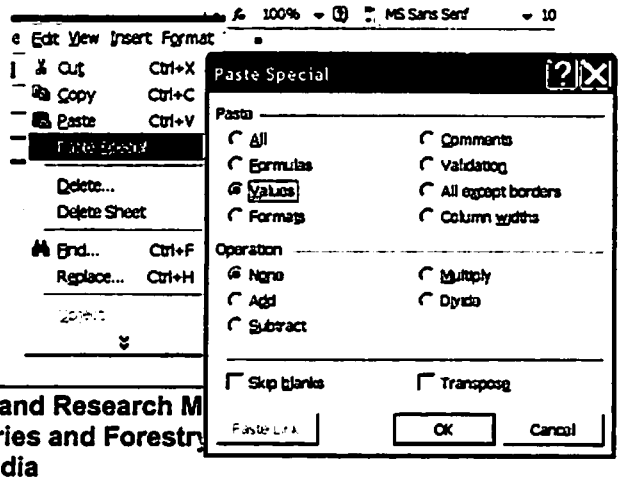
Merging variables using the Vlookup Function

In the previous sections, we illustrated how to aggregate **species level data** to come up with **household level** of the 218 fishing households using **Pivot Table**. Earlier, we also illustrate how to extract all the 76 households catching species 98 using **Filter**. This subset of data is saved in worksheet **species98**. In this section we will illustrate how to merge the 218 aggregate data of fishing households and the subset of 76 households who catch species 98, in such a way that the catch for species 98 (catch98) of a particular household will correspond to the same household in the merge file (e.g. same households will go to same row). We now again show the resulting merged data, which is actually the **merge** worksheet in our database.

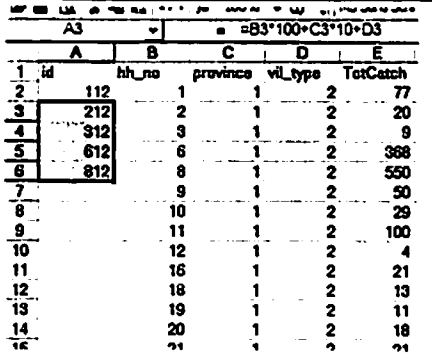


Initial step: First organize the data

1. **Copy** and **Paste special→ values** the entire table into another worksheet (say **merge**).
2. Delete the first few unnecessary rows leaving the column headings in Row1.



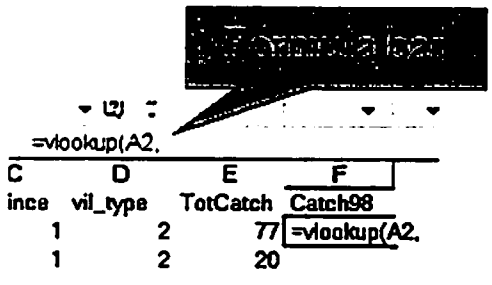
3. Supply **single field primary key (id)** in column **A** and type "=b2*100 + c2*10 + d2 in cell **A2**".
4. Copy **A2** to the rest of the cells in column **A**.
5. Rename **Total** (Column **E** heading) to **TotCatch**

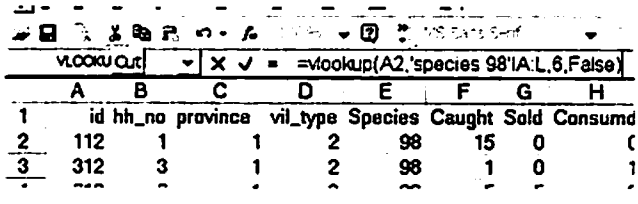


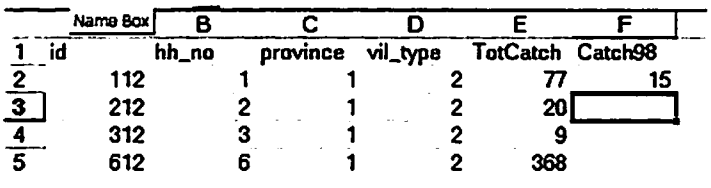
Now you are ready to do **Vlookup**:

The task remaining here is to lookup (copy) the values of catch of **species 98** in column **F**. Using **Vlookup** function, this can be accomplished by setting-up the formula **=Vlookup(A2, ,6, False)** in cell **F2** and copy **F2** in the remaining cells of column **F**. For illustration, this can be done efficiently following the steps below:

1. Type the heading **Catch98** in cell **F1**.
2. Set cell **F2** as the active cell by clicking on it.
3. In the formula bar, type **=vlookup(A2,**
4. Supply the **table array** by switching to **species98** worksheet and highlight the **columns A to L**. Note the formula in the formula bar updates correspondingly.
5. Type "," (comma) followed by "**6**" "," "False" then ")".
6. Press **Enter** key.

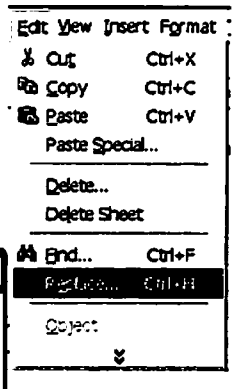


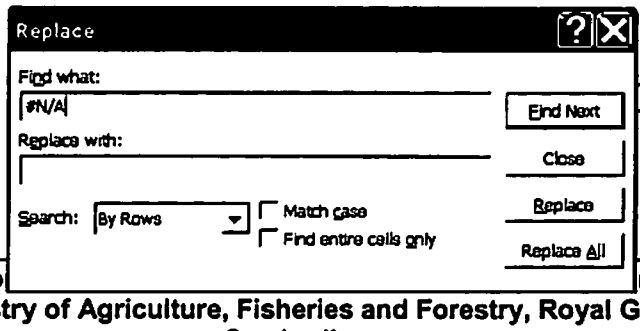




7. Copy **F2**, to the rest of the cells in column **F**.

Note that for those fishing households who do not catch **Species 98**, the values are "**#N/A**". Such values will create problem during analysis so you must get rid of them. Deleting them one by one is a solution but its not efficient. Duplicate the fields **TotCatch** and **Catch98**. Use **Copy→Paste Special→Values** to copy the columns. Then Used **Edit →Replace→Replac e All** to replace "#N/As" with missing values.





Pivot Tables to Summarize Information

In the previous section, we illustrated the use of **PivotTable** to aggregate data with the aid of skeleton pivot table. In this section we will illustrate the use of **PivotTable** to do cross tabulation/summarize information. For comparison, we will be making use of **Layout Wizard**. We will show how to produce multiple tables in the same sheet by just simply copying another table and modify to present the desired data.

One-WayTable: Household Distribution by Province

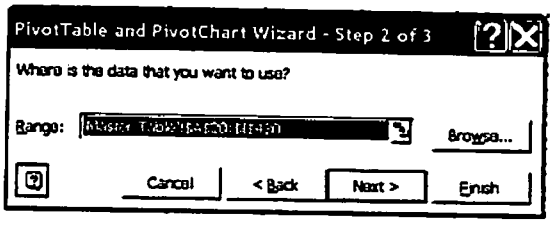
First, let's redo the first Table "**Household Distribution by Province**". Note that the table used a household level data in **Master Table** worksheet.

Household Distribution by Province

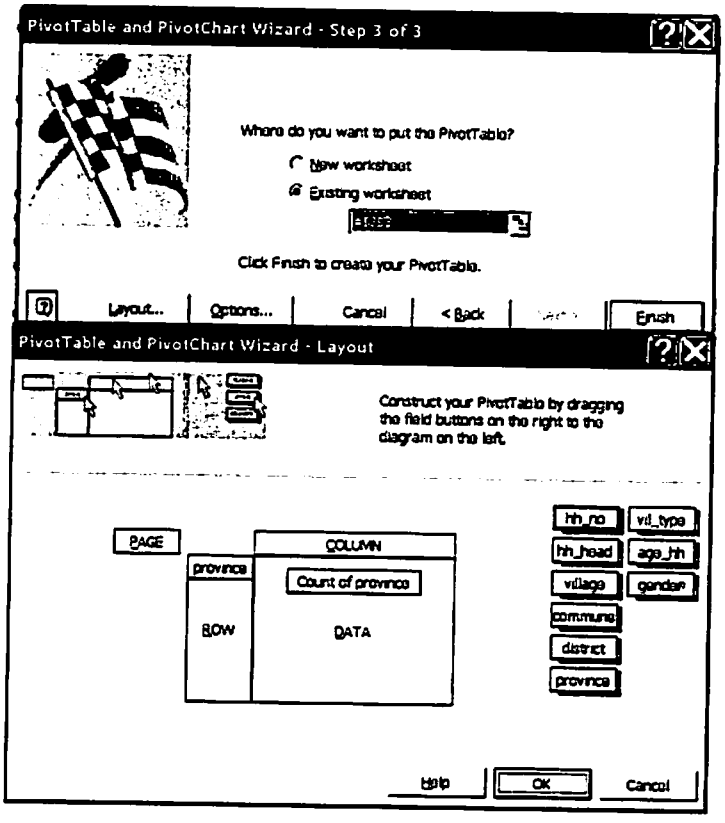
Count of prov	
province	Total
1	135
2	140
3	135
Grand Total	410

Under the **merge** worksheet, launch layout wizard by:

1. Clicking on **Data** menu followed by **PivotTable...**
2. When ask on the location of the data to use, switch to **Master Table** worksheet and highlight the range accordingly.



3. Click on **Next>** command button to go to **Step 3** of the **PivotTable Wizard.**
4. Choose ***Existing worksheet*** and specify a blank cell (**J3**) in **Merge** worksheet.
5. Click on **Layout...** command button.



The layout wizard appears.

6. Drag and drop **Province** button under **Row** section of the **Layout.**
7. Again, drag and drop **Province** button under **Data** section of the layout.

The button will change to **Count of Province** signifying frequency. In the next section we will show how to modify summary statistics under layout wizard.

8. Click on **OK** command button.

Two-Way Table: Household Distribution by Province by Village type

In this section, we will illustrate how to modify a table by adding one dimension. Specifically, we will redo the table **Household Distribution by Province by Village Type** by copying and modifying the first table- **Household Distribution by Province** table.

Household Distribution by Province by Village type

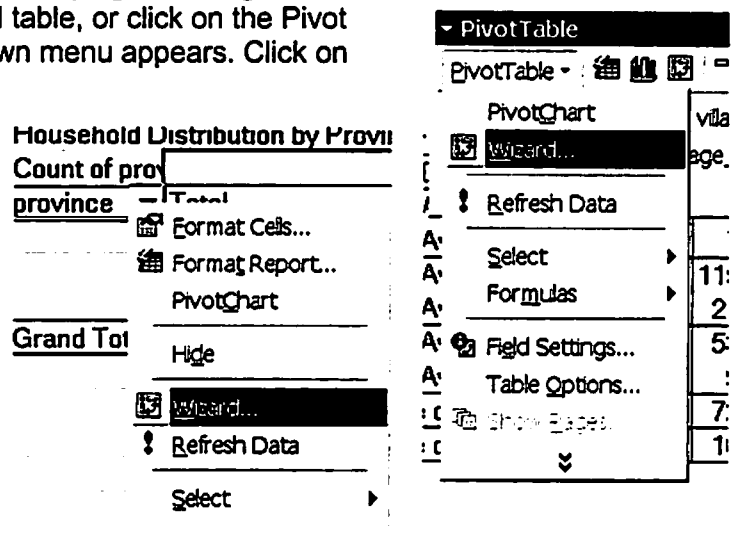
Count of province	vil_type			
province	1	2	3	Grand Total
1	50	43	42	135
2	46	47	47	140
3	45	45	45	135
Grand Total	141	135	134	410

1. Copy by highlighting and paste **Household Distribution by Province** Table to a blank area (say cell N2) in the worksheet.

Household Distribution by Province

Count of pro	
province	Total
1	135
2	140
3	135
Grand Total	410

2. Launch the **Layout Wizard** by right-clicking anywhere on the copied table, or click on the Pivot table Toolbar. A pull down menu appears. Click on **Wizard...**



Step 3 dialog box of the **Pivot Table Wizard** will popup.

3. Click on **Layout...** command button to go to **Layout wizard.**
4. Drag and drop **vil_type** to **COLUMN** section of the layout.
5. Click on **OK** command button.

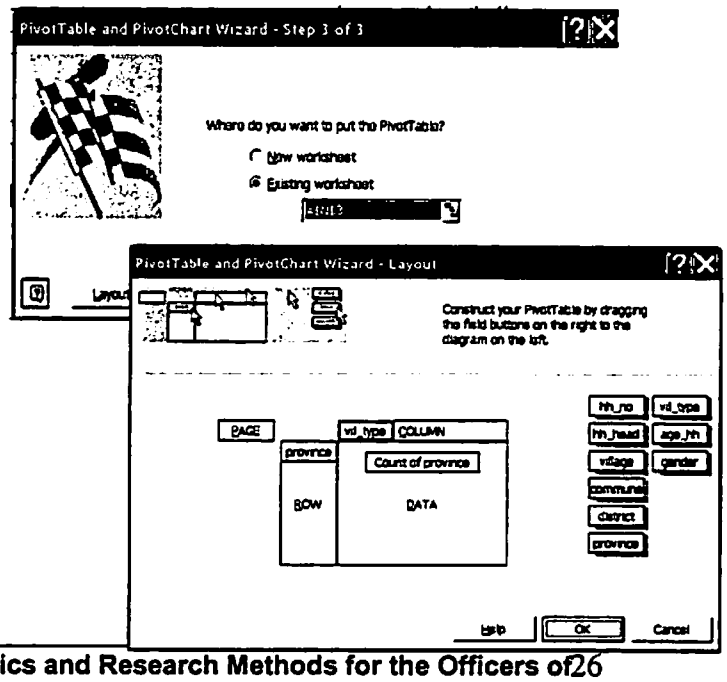


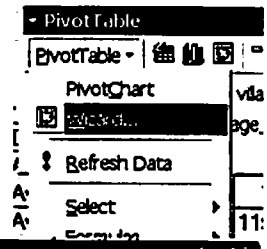
Table of averages

In this section we will illustrate how to change summary statistics of a given table using the **Layout Wizard**. Specifically, we will show how table **Fish Catch by Province** was created.

Fish Catch by Province

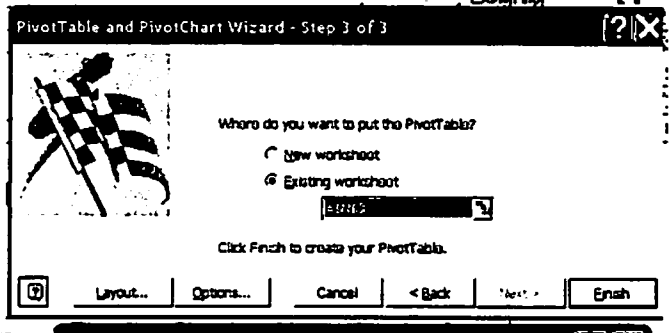
	province			
Data	1	2	3	Grand Total
Average of CTotcatch	48.10	115.12	53.96	72.22
Average of CCatch98	7.14	21.93	9.35	10.6

1. Copy by highlighting and paste **Household Distribution by Province** table to a blank area (say cell J14).
2. Launch the **Layout Wizard** by RIGHT clicking anywhere on the copied table. A pull down menu appears. Click on **Wizard**.



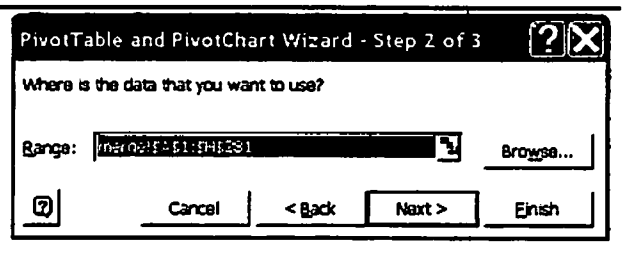
Step 3 dialog box of the **PivotTable Wizard** will popup.

3. Change the location of data to be used by clicking on **<Back** command button.



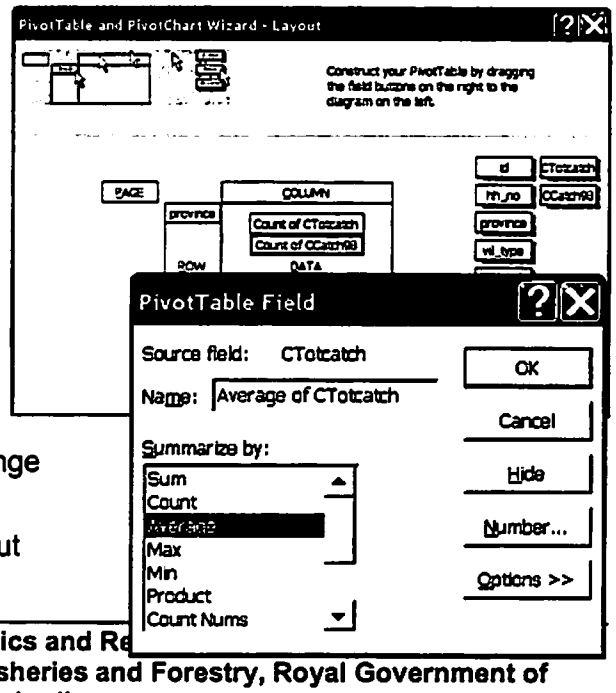
The **Step 2** dialog box of the **PivotTable Wizard** will popup.

4. Change the range accordingly after which click on **Next>** command button to go back to **Step 3** dialog box and click on **Layout** command button to launch the **Layout Wizard**.



5. Drag and drop **Province** button into the **Row** section and **CTotCatch** and **CCatch98** into the **Data** section.

Note that the **CTotCatch** and Catch98 button are changed to **Count of CTotCatch** and **CCatch98**, respectively.



6. Change the **Count of CtotCatch** and **Count of Ccatch98** buttons to **Averages** by double clicking on them one after another.
7. A **PivotTable Field** dialog box appear. Under *summarize by*: Change **Count** to **Average**.
8. Click on **OK** to go back to the Layout Wizard then and again click on OK.

Exploring the data

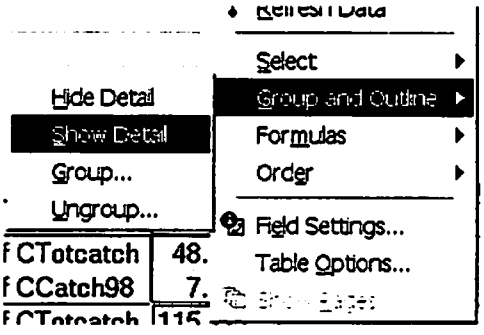
Fish Catch by Province

	province			
Data	1	2	3	Grand Total
Average of CTotcatch	48.10	115.12	53.96	72.22
Average of CCatch98	7.14	21.93	9.35	10.6

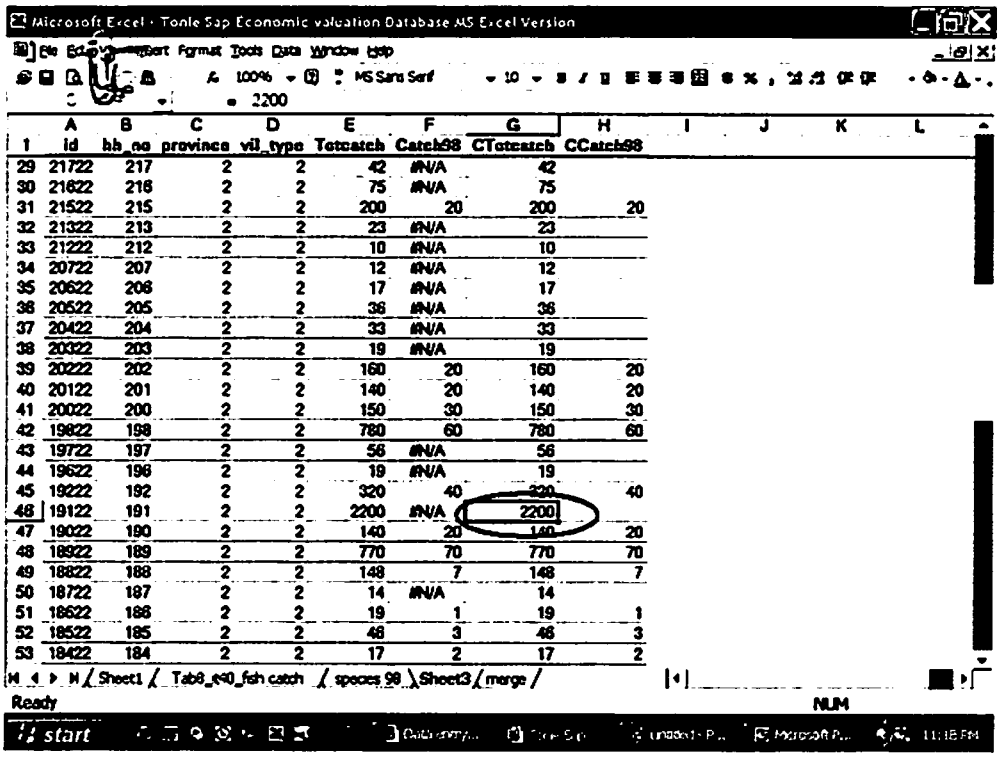
It is sometimes useful to look at individual observations in more detail. You might find some categories interesting and may indicate oddities in the data. Take for example the **Average of CTotcatch** under **province 2 (115.12)**, which is 100% higher than the averages in the other two provinces.

In this section we will illustrate how to explore and extract the household that make up a cell in the pivot table. Specifically, we will extract the households that make up **115.12.**

1. Right click on the cell containing **115.12.**
2. Select on **Group** and **Outline** then **Show Detail.**



Alternatively, double click on the cell. A new sheet is created containing the households.



	A	B	C	D	E	F	G	H
1	id	hh_no	province	vil_type	Totcatch	Catch98	CTotcatch	CCatch98
29	21722	217	2	2	42	#N/A	42	
30	21622	216	2	2	75	#N/A	75	
31	21522	215	2	2	200	20	200	20
32	21322	213	2	2	23	#N/A	23	
33	21222	212	2	2	10	#N/A	10	
34	20722	207	2	2	12	#N/A	12	
35	20622	206	2	2	17	#N/A	17	
36	20522	205	2	2	36	#N/A	36	
37	20422	204	2	2	33	#N/A	33	
38	20322	203	2	2	19	#N/A	19	
39	20222	202	2	2	160	20	160	20
40	20122	201	2	2	140	20	140	20
41	20022	200	2	2	150	30	150	30
42	19822	198	2	2	780	60	780	60
43	19722	197	2	2	56	#N/A	56	
44	19622	196	2	2	19	#N/A	19	
45	19222	192	2	2	320	40	320	40
46	19122	191	2	2	2200	#N/A	2200	
47	19022	190	2	2	140	20	140	20
48	18922	189	2	2	770	70	770	70
49	18822	188	2	2	148	7	148	7
50	18722	187	2	2	14	#N/A	14	
51	18622	186	2	2	19	1	19	1
52	18522	185	2	2	46	3	46	3
53	18422	184	2	2	17	2	17	2

Basically,

these are the households belonging in Province number 2. Note that a huge total fish catch (**2200**) is reported in **row 46** (Household number 191; province number 2; village type 2).

Automatic Updating and Linking Pivot Table to a Formatted Table

It is common that during analysis, you will find some mistakes in the data and you need to correct them. After correcting some entries, as mentioned earlier you do not need to redo the pivot table to update the figures. A click on the **Refresh** command button in the Pivot table Tool Bar will automatically update the pivot table. The **Refresh** command button will also allow you to design the pivot tables in advance before or during the data entry process and once data entry is done, instant summary tables can be generated.

The format of a Pivot table however is not really appealing. You may want to change the word "**Average of Ccatch98**" to the actual name of the fish instead (say Giant catfish). You cannot however do that in Pivot Table. To have maximum formatting flexibility, make a copy of the entire table and link the values all the figures of formatted table to the corresponding cell of the original pivot table. Linking will automatically update the figures in the formatted table every time you update your Pivot table using the **Refresh** command button.

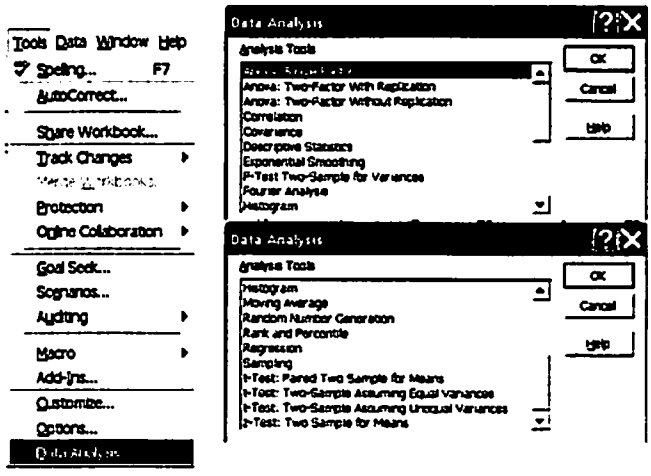
Use the **Paste Special→ Paste Values** in copying. The copy can be formatted like any other Excel range. To automatically update the figures in your formatted (copy) table, equate each of the figures in the formatted table the corresponding cells in the original pivot Table. Example: click on cell U5 then type "**= O5**". This imply that cell U5 which is the intersection between Province 1 and Village type 1 will take the value in O5 which is the intersection between Province 1 and Village type 1 in the original Pivot table. Then you can copy this cell to the rest of the cells containing figures.

U5 = =O5

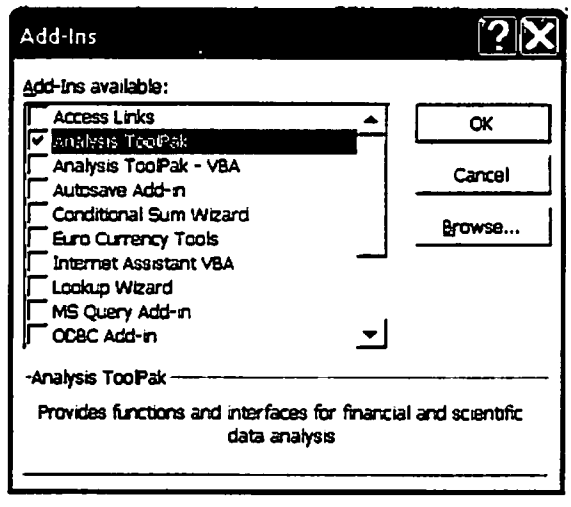
	N	O	P	Q	R	S	T	U	V	W	X
1											
2	Household Distribution by Province by Village type						Household Distribution by Province by Village type				
3	Count of province	vil_type					Count of province	Village type			
4	province	1	2	3	Grand Total		Province	1	2	3	All type
5	1	50	43	42	135		1	50	43	42	135
6	2	46	47	47	140		2	46	47	47	140
7	3	45	45	45	135		3	45	45	45	135
8	Grand Total	141	135	134	410		All Province	141	135	134	410
9											

MS Excel as a Statistical Tool???

MS Excel offers wide range of statistical tools, which are accessible through **Tools** menu➔ **Data Analysis.**



Data Analysis is an Add-in in Excel. If **Data Analysis** option does not appear in the pull down menu, click on the **Add-Ins..** option to view the list of Available Add-Ins window. Tick the box on **Analysis tool** to activate Data Analysis.



In this section we will not illustrate how these different Statistical methods can be used. Instead, we will point out the weaknesses of using Excel for statistical tasks. The Data Analysis feature of Excel gives the impression that it does a lot of statistical method when in fact its capabilities are very limited and undesirable.

We focus our attention to **Regression Analysis** and **Analysis of Variance (ANOVA)**. In regression analysis, the ANOVA (Analysis of variance) and regression statistics in a regression analysis without constant term are incorrect. Even if you include constant term, the standardized residuals seemed to have calculated using unknown formula.

For ANOVA, Excel neither supports unbalanced data (if the number of replicates are not equal for all treatment) neither allow missing values.

Thus, we recommend against the use of MS Excel when it comes to advance data analysis, which involves Statistics. We recommend that proper statistical packages (SPSS, SAS, Minitab, etc) should be used. We have another document which illustrates the used of SPSS in doing statistical analysis and data management using the Tonle Sap Economic Evaluation data.

Appendix A

Economic Valuation of Aquatic Resources of Tonle Sap Basin
ADB-IFReDI TA, Implemented by the WorldFish Center
Questionnaire
August, 2003

Location and Address:

Name of the household head:_____ HH_NO

Village:_____ Commune:_____ VILLAGE

District:_____ DISTRICT

Province (Kompong Chnang 1, Siem Reap 2, Kandal 3):_____ PROVINCE

Village type (Fishing 1, Fishing and Farming 2, Farming and Fishing 3):_____ VIL_TYPE

Profile of the head of the household:

Age:_____ AGE_HH

Sex (Male 1, Female 2):_____ GENDER

Ethnicity (Khmer 1, Chinese 2, Vietnamese 3, Cham 4, Other 5):_____ ETHNICITY

Education (0, below 3 years, 4-5 years, 6-10 years, Above 10 years):_____ EDUCTN

Religion (Buddist 1, Muslim 2, Christian 3, Other 4):_____ RELIGION

Principal occupation:_____ OCCPHH

Secondary occupation:_____ OCCSHH

Code: Fishing 1, Fish processing 2, Fish trading 3, Fish culture 4, Net/gear making 5, Farming 6, Laborer 7, Small business 8, Money lending 9, Fuel wood collection 10, Motor taxi/car/engine boat driving 11, Government/NGO job 12, House keeping 13, Teaching 14, Other 15

Household income (Riel) from the sources below during September 2002 to August 2003

TBL 2 ITEM 8 INCOME

SOURCES	INCOME
Fishing:	
Fish processing:	
Fish trading:	
Fish culture:	
Net/gear making:	
Bamboo and cane works:	
Farming:	
Daily labor:	
Housekeeping:	
Shop/small business:	
Government/private job:	
Motor taxi/engine boat driving:	
Money lending:	
Fuel wood collection:	
Livestock raising:	
Crocodile culture:	
Other (specify):	

Note: For fishing related information collect information by close and open season and and then add up.

Profile of the members of the household including household head

Total number of members in the household: Male: MALEN Female: FEMN

Age distribution of the members (years): AGEDISTN N_MALE N_FEMALE

TBL 3 ITEM 10 AGE DISTRIBUTION

Below 5	Male:	Female:
6 to 10	Male:	Female:
11 to 15	Male:	Female:
16 to 30	Male:	Female:
31 to 45	Male:	Female:
46 to 60	Male:	Female:
Above 60	Male:	Female:

Level of education of the eligible members of the households in years:

EDUC N_MALE N_FEMALE

TBL 4 ITEM 11 EDUCATION

0	Male:	Female:
1 to 3	Male:	Female:
4 to 5	Male:	Female:
6 to 10	Male:	Female:
Above 10	Male:	Female:

Number of members in the household eligible to work: Male: ELGWM Female: ELGBF

Occupation of the eligible members of the household:

TBL 5 ITEM 13 OCCUPATION

OCCP	POCCM	POCCM	SOCCM	SOCCM
Fishing	Male:		Female	
Fish processing	Male:		Female	
Fish trading	Male:		Female	
Fish culture	Male:		Female	
Net/gear making	Male:		Female	
Bamboo and cane works	Male:		Female	
Farming	Male:		Female	
Daily labor	Male:		Female	
Housekeeping	Male:		Female	
Shop/small business	Male:		Female	
Government/private job	Male:		Female	
Motor taxi/engine boat driving	Male:		Female	
Money lending	Male:		Female	
Fuel wood collection	Male:		Female	

Environment, Sanitation and Energy Sources

Type or status of latrine/toilet T_TOILET

Open latrine above water/land.....1
No latrine/toilet.....2
Sanitary latrine.....3
Other (specify).....4

Sources and nature of drinking water S_H2O S_H2O1

River/lake water.....1
Bottled water.....2
Tubewell water.....3
Pond water.....4

Other (specify).....6	
If river/lake/pond, do you purify or boil the water (yes 1, No 2)	B_H2O
Sources of cooking and washing water	S_CWH2O
River/lake water.....1	S_CWH2O1
Pond water.....2	
Tubewell water.....3	
Other (specify).....4	
Sources of bathing and cleaning water	BCH2OS
River/lake water.....1	BCH2O
Pond water.....2	
Tubewell water.....3	
Other (specify).....4	
Sources of Medical care:	MEDICS
Village quack.....1	MEDIC
Medical clinic (public).....2	
Medical clinic (private).....3	
Traditional herbs.....4	
Other.....5	
How far (km) is the nearest hospital from your house? (km)_____	HOSPD
Is your village connected with electricity grid? (Yes 1, No 2)	ELCON
Do you have electricity in your home? (Yes 1, No 2)	ELHOME
If yes, sources of electricity:_____	ELSOURCE
Own generator.....1	
Connected with electricity grid.....2	
Connected with private generator.....3	
If no, what are the sources of household energy needs?_____	
For lighting:	LIGHTEN
Kerosene Lamp.....1	LIGHTEN1
Candle.....2	
Battery.....3	
Other (specify).....4	
For cooking and fish processing:	
Forest wood.....1	COOKEN
Bio-gas burner.....2	COOKEN1
Charcoal.....4	
Other (specyfy).....5	
If firewood, from where do you collect firewood and how much (kg)?	
Flooded forest:_____	QWDFLD
Mountain forest:_____	QWDMNT
Homestead forest:_____	QWDHSTD
Other (specify):_____	QWDOTH

Household asset and land ownership

Land ownership status

Land type	Area in sq.m	
Homestead		LHSTD
Agricultural land		LAGRI
Pond land		LPOND
Orchard land		LORCH
Fellow land		LFELLOW

Housing type. H_TYPE

Floating: H_TYPE1

Small floating house made of wood and tin/tiles roof.....1
Medium floating house made of wood and tin/tiles roof.....2
Big floating house made of wood and tin/tiles roof.....3
Small floating house made of cane/bamboo and palm leaves.....4
Medium floating house made of cane/bamboo and palm leaves.....5
Big floating house made of cane/bamboo and palm leaves.....6

On land:

Small house made of wood and tin/tiles roof.....7
Medium house made of wood and tin/tiles roof.....8
Big house made of wood and tin/tiles roof.....9
Small house made of cane/bamboo and palm leaves.....10
Medium house made of cane/bamboo and palm leaves.....11
Big house made of cane/bamboo and palm leaves.....12

Present value of the house (in Riel):_____ V_VALUE

Present value of household assets:

Electronic appliances:_____ V_EAPPL
Transport equipments(boat, bicycle, motorbike, motor boat):_____ VTRANSE
Furniture/fixtures:_____ VFURN
Electricity generator:_____ V_GEN
Battery:_____ V_BATT
Tube well:_____ V_TUBE
Hand phone/Radio transmission:_____ V_HPHONE

Fisheries and fishing related assets and present value (Riel)

TBL 6 ITEM 30 FISHING RELATED ASSETS

FISHING EQUIPMENT	NUMBER	PRESENT VALUE (RIEL)
Fishing equipment	Number	Present value (Riel)
EQPMNT	**NUMBER**	**PRVALUE**
Harpoon		
Bamboo trap		
Castnet		
Liftnet		
Gillnet		
Sein net		
Hook long line		
Bamboo fence		
Bagnet (non-motorized)		
Funnel trap		
Other (specify)		

Present value of assets related to fish processing

TBL 7 ITEM 31 PRESENT VALUE OF ASSETS

Processing equipment	Number	Present value (Riel)
PEQPMNT	**NUMBER**	**PRVALUE**
Barrel		
Cube		
Smoke griller		
Jar		
Others (specify)		

Present value of farm equipment

Traditional:_____ **V_TFEQ**

Modern:_____ **V_MODEQ**

Fishing and related activities

How many people in your household do fishing? Male: **MFISH** Female: **FMFISH**

What time of the year do you fish? **F_TIME**

(Closed season 1, Open season 2, All season 3, Occassional 4)

Why do you fish? (For sale 1, consumption only 2) **F_REASON**

Did you fish last week?(Yes 1, No 2) **F_FREQ**

If yes, how many days?_____ **F_DAY**

How many hours per day?_____ **F_HOURS**

How many persons?_____ **F_PERSON**

How much fish (in kg) did you caught, consumed and sold during last week? (use species code list)

Species	Quantity (kg)					Price/kg	Buyer*
	Caught	Sold	Consumed	Processed	Fish feed**		
SPECIES	CAUGHT	SOLD	CONSUMD	PROCESSD	FISHFEED	PRICE_KG	BUYER

*Buyer: Fish collector on site 1, Middle man/trader at the landing site 2, Fish processor 3, Cage farmers 4, Animal/Crocodile farmer 5

**include animal feed also

Where did you catch fish? **FPLACE1-3**

(Rice field 1, Great Lake 2, River 3, Canal 4, Flooded forest 5, Fishing lot 6, Pond 7)

What type of fishing gear did you use for fishing?:_____ **FGEAR1-3**

Gill net 1, Cast net 2, Bamboo fenced trap 3, seine net 4, long line 5, Scoop net 6, Bagnet 7, Bamboo fenched trap 8, net 9)

What type of boat did you use for fishing at what cost?

TBL 9 ITEM 43 TYPE BOAT

Type of boat	Number of boats	Owned	Renting cost	Fuel cost (Riel)
T_BOAT	N_BOAT	OWNED	R_COST	F_COST
Motorized				
Non-motorized				
Other (specify)				

How much fish (in kg) could you catch, consume and sell during open season in 2002?

Caught (kg):_____ **QCATCHO**

Consumption (kg):_____ **QCONSO**

Sold (kg):_____ **QSOLDO**

Processed (kg):_____ **QPRCO**

Used as fish feed (kg):_____ **QFFEEDO**

How much fish (in kg) could you catch, consume and sell during close season in 2003?

Caught (kg):_____ **QCATCHC**

Consumption (kg):_____ **QCONSC**

Sold (kg):_____ **QSOLDC**

Processed (kg):_____ **QPRCC**

Used as fish feed (kg):_____ **QFEDC**

Where did you sell fish during last open and close seasons? **SOLDWHM1**

To fish collector on site.....1

To middle men/trader at the landing site.....2

To fish processor.....3

To cage farmer.....4

To animal/crocodile farmer.....5

Other (specify).....6

What species did you catch?
(use species code list)

Open season 2002		Close season 2003	
Species	Rank by quantity caught	Species	Rank by quantity caught
O_SPECIE	**O_RANK**	**C_SPECIE**	**C_RANK**

How do you market your fish? (Yes 1, No 2)

Carry to the market to sell directly to the consumers	CONSMKT
Carry to the landing site and sell to wholesaler	WSALER
Middlemen collect from fishing ground	FGROUND
Wholesale buyer from whom you borrowed money	WSLENDER
Other (specify)	OTHERS

Did you borrow money during this year? (Yes 1, No 2) B_MONEY

If the answer is "yes", from whom did you borrow money? (yes 1, No 2)

Friends and relatives.....1	S_MFRIND
Financial institutions.....2	S_MFINST
Local money lender.....3	S_MLEND
Fish trader/whole saler.....4	S_MFTRAD
Other (specify).....5	S_MOTH

Are you obliged to sell fish to your trader/money lender? (Yes 1, No 2) MLENDERS

If yes, do you get market price for your fish? (Yes 1, No 2) MPRIC

If no, how much less per kg?_____ LESSMPR

Fish Processing Activities

Do you process fish? (Yes 1, No 2) PROCFS

If yes, purpose of processing: (yes 1, No 2)

Self consumption	PROCONS
For sale	PROCSL

What fish products did you process during September 2002 - August 2003?

TBL11 ITEM 56 PROCESSED FISH PRODUCTS

Products	Quantity (kg)	Sold (kg)	Consump-tion (kg)	Price/kg
PRODUCTS	**QTY**	**SOLD**	**CONSUMP**	**PRICE**
Salted dried fish				
Dried fish				
Smoked fish				
Fermented fish				
Fish filleting				
Prahoc				
Semi-final prahoc				
Fisah sauce				
Fish ball				
Other (specify)				

Where do you process fish? PLPROC1

Within house.....1	PLPROC2
River bank.....2	
Other (specify).....3	

What kind of raw materials did you use for processing?

TBL12 ITEM 58 RAW MATERIALS

Raw materials	Quantity (kg)	Price/kg
R_MAT	**QTY**	**PRICE**
Fish		
Salt		
Sugar		
Firewood		
Charcoal		
Fuel		
Sun dried materials (sq m)		
Other (specify)		

Labor Utilization

TBL13 ITEM 59 LABOUR UTILIZATION

	Peak Season			Non-peak season		Wage
Sex	Self	Hired	Wage rate	Self	Hired	rate
SEX	**P_SELF**	**P_HIRED**	**P_WRATE**	**N_SELF**	**N_HIRE**	**N_WRATE**
Male						
Female						

How much processed fish did you sell during September 2002 - August, 2003

TBL14 ITEM 60 PROCESSED FISH SOLD

Type of processed fish	Amount sold (kg)	Price/kg	Buyer**	Amount consumed
TYPE	**SOLD**	**PRICE**	**BUYER**	**CONSUMED**
Salted dry				
Dried fish				
Smoked				
Fermented				
Pra hok				
Fish sauce				
Other				

*Buyer: Collector on site 1, Middle man/trader in the nearest city 2, Sell directly to the consumer 3 Animal/Crocodile farmer 4

Fish Cage/Pond Culture

Do you culture fish in cage or pond? (Yes 1, No 2) **CGCULT**

If yes, in how many cages/ponds:_____ **NOCAGES**

Total area of cages/ponds (sq m):_____ **AREACAG**

Present value of the cage/pond (Riel):_____ **VALCAG**

Type of species and number of fingerlings:

TBL15 ITEM 65 FINGERLINGS

Species	Number	Size (cm)	Source	Cost per fingerling (Riel)
Giant snake head				
Pangusius				
SPECIES	**NUMBER**	**SIZE**	**SOURCE**	**COST**

Operating cost per day:

Labor (hours):

-maintenance:_____ **OC_MAIN**

-fish collection:_____ **OC_FCOLL**

Fish bought (kg):_____ **OC_FB**

Fish caught (kg):_____ **OC_FC**

Other cost:_____ **OTHER_OC**

Rearing period (months):_____ **REAR_P**

Total production (kg) and selling price:

TBL16 ITEM 68 FISH PRODUCTION

Species	Consumption	Production	Sold	Price/kg	Buyer*
Giant snake head					
Pangusius					
SPECIES	**CONSUMED**	**PROD**	**SOLD**	**PRICE**	**BUYER**

*Buyer: Fish collector on site 1, Middle man/trader at the landing site 2, Fish processor 3, Money lender/collector 4, Animal/Crocodile farmer 5

Farming Activities

Farm production during September 2003 - August 2003

TBL17 ITEM 69 FARM PRODUCTION

Crops and vegetables	Production (in kg)	Total value (Riel)
CROPS	PROD	VALUE

Labor utilization in farming activities during 2002 -2003

TBL18 ITEM 70 LABOR IN FARMING

Crops /vegetables	Male			Female			Wage rate	
	Self	Exchange	Hired	Self	Exc	Hired	Male	Female
CROPS	M_SELF	M_EXCHG	M_HIRE	FM_SELF	_EXC	FM_HIR	M_WRATE	FM_WRATE

Input costs (Riel) of farming activities during 2002 - 2003.

TBL19 ITEM 71 INPUT COST FARMING

Crops/ vegetables	Inputs cost				
	Irrigation	Chemical fertilizers	Cow dung	Seeds	
CROPS	**IRRGTN**	**FERT**	**COWDUNG**	**SEEDS**	

How much land did you cultivate during 2002 - 2003? (sq m)

TBL 20 ITEM 72 CULTIVATED LAND

Crops /vegetables	Land cultivated	
	Irrigated	Non- irrigated
CROPS	**IRRGTD**	**N_IRRGTD**

Dependence of the Households on the Resources within Lake and Flooded Forest

What resources do you collect from the Lake and flooded forest and how do you value them in your daily life?

TBL21 ITEM 73 RESOURCE DEPENDENCE

Resources	Yes/No	Quantity (kg)		Value	How important*
		Sold	Consumption		
RESOURCES		**SOLD**	**CONSUMP**	**VALUE**	**IMPRTNCE**
Firewood					
Bamboo/canes					
Mat making materials					
Wild animals/birds					
Transportation					
Animal grazing					
Duck grazing					
Fruits collection					
Sundaune (fruit)					
Sundaune (leaves)					
Phkasnor					
Water Lilly					
Lotus/ lotus roots					
Troyraug					
Kanchet					
Saomaoprey					
Morning glory					
Snails/crab collection					
Mollusk					
Rat					
Toat					
Turtles					
Tortoises					
Swamp eel					
Snake					
Traditional medicine					
Recreation					

* Very important 4, Important 3, somewhat important 2, Not important 1

Household Consumption during last 7 days

How much of the following items your household consume during last week?

TBL22 ITEM 74 HOUSEHOLD CONSUMPTION

Consumable items	Quantity (kg)		Price/kg
	Self	Market	
ITEMS	**SELF**	**MARKET**	**PRICE**
Rice			
Corn			
Noodle			
Vegetables			
Fish			
Chicken			
Meat			
Bread			
Salted dry fish			
Pra hoc			
Fermented			
Smoked fish			
Fish egg			
Fish sauce			
Fish ball			

Interviewed by:_____ Varified by:_____

Date:_____ Date:_____