Chapter 1
Introduction

1.1 Background

Before learning statistics, one should know the scientific method. The ultimate goal of science is to understand and explain the natural and social phenomenon based on the conclusion of valid experiments and comprehensive observations. Observation and experimentation are the main two ways of generating "knowledge" about the natural world. In addition to observation and experimentation, scientific method also includes identification, description, and the theoretical explanation. In contrast, traditional knowledge is teachings and experiences passed on from generation to generation that is deeply rooted and developed as culture, customs, mythology, and language of the people as a way of living. Most traditional knowledge passes verbally from person to person across generations in the forms of stories, legends, folklore, rituals, songs, even regulation and laws. It refers sometimes to the matured traditions and practices in certain local communities that may differ from one community to another and may serve as a unique identity of particular communities. Most traditional knowledge is very valuable, but some needs to be tested in new contexts.

In many circumstances, observations and measurements are not possible; therefore, people have to imagine or hypothesize based on the limited available knowledge, which might not be true. Standard methods or procedures have been developed and are in use to carry out scientific inquiry or research. Depending on the nature of the research, outcomes can be broadly categorized into two groups: discovery and invention. Finding out things that already exist in the universe is called *discovery*, whereas creating or designing something new that never existed before is *invention*. Many scientific discoveries and inventions have played a significant role in changing the world and making human life a lot easier; for example, the invention of the bicycle, steam and jet engines, telephone, television, modern information technology, and so on. In agriculture, invention of new high-yielding varieties has brought about the green revolution, which is helping in feeding the ever-increasing population. Similarly, the development of high-milk-yielding dairy cattle has resulted in the white revolution. In aquaculture, we often talk about bringing about a blue revolution, but it has never taken place. People in this field are working hard in various ways to make it a reality.

A farmer grew five fish of a new strain in a single tank with excessive feeding. His fish grew 600 g in 3 months, and he compared the growth rate of his fish against the rates published in literature. He quickly noticed that the growth rate of his fish was almost double. He started advertising about his strain of fish around the world via e-mail, claiming that he had developed a new strain that grows two times faster than any other strains. Should we believe this claim? There is a similar story of a journalist who tested a new variety of rice, sowing a single seed in a well-managed plot to find a solution to the chronic shortage of rice in the Philippines in the 1950s. He harvested 1,000 grains from a single rice plant and, after extrapolating the yield, found 50 $t{\cdot}ha^{-1}$. Compared with 1·2 $t{\cdot}ha^{-1}$, the national average, he thought that, if that new variety was distributed to all the farmers, his country would not have any problems with rice shortages and could export and earn millions (Gomez and Gomez 1984). Is this a scientifically valid comparison and conclusion? The answer is "No." The scientific method involves a long process (Figure 1.1) that starts with observation then passes through all the way from imagination or hypothesis, designing and conducting experiment, data collection, analysis, and interpretation or reasoning. If the original imagination or hypothesis is proved, then it becomes a theory. A theory is not only a set of findings but also a set of well-developed themes and concepts that logically explain particular phenomena. Once a theory is widely accepted and applied, it becomes a universal law. Basically, a theory should be based on supporting data. This is often called a grounded or substantive theory, which is based on reality. It is derived from data gathered or generated systematically and analyzed through a research process. The process includes data collection from reliable sources or well-designed experiments, compilation of data, analysis and interpretation, presentation of findings, and theorizing or building a theory. Theory derived simply from phenomena is called formal theory, and if a theory is not empirically grounded in research, it is called speculative.

Scientific research is a long process and hard work. It can be boring but can also be exciting with the full joy of discovery or invention. Developing a career in research is climbing a steep ladder. Most researchers are either about to enter it or on the way to becoming researchers or great scientists, which means no one is 100% perfect. The knowledge and skills in doing research are reflected in their publications and presentations of research findings. The author has experienced several examples of lack of basic knowledge in statistics. For example, while attending a number of conferences and seminars, several researchers present their results, which say that fish fed with supplementary vitamin C or other nutrients had higher growth rate, survival, feed conversion ratio, meat yield, etc. than the fish fed from control treatment. However, these values were not statistically significant ($P > 0.05$). They even conclude and recommend that vitamin C or the other nutrients should be supplemented in the diet to increase the yield, which is wrong and misleading. Similarly, in survey type of research, researchers would

P1: SFK/UKS P2: SFK/UKS QC: SFK/UKS T1: SFK
BLBS023-Bhujel 97814051XXXXX October 26, 2008 15:39

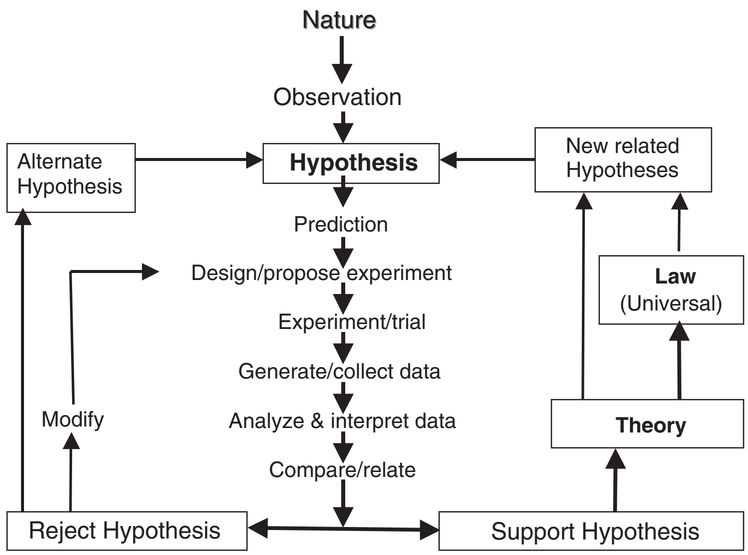

Figure 1.1 Scientific method.

say that farmers in district A had relatively bigger farms and higher fish production; however, these figures were not significantly different ($P > 0.05$) when compared using statistical tools. To claim this is wrong if the statistical analysis does not show any difference. Plenty of examples of this kind are found even in scientific literature, especially in aquaculture, which shows that there is a need for enhanced understanding of statistics among aquaculturists. This handbook has been written to help those researchers who are encountering problems, considering the fact that statistics is a must for researchers. However, commenting on the present status, Galton rightly said, "Some people hate the very name statistics, but I find them full of beauty and interest whenever they are not brutalized, but delicately handled by the higher methods, and are warily interpreted, their power of dealing with complicated phenomena is extraordinary."

1.2 History and definition of statistics

The word "statistics" originated from the Latin word "state," which means government. The states or the militaries were the first users of statistics and other advanced equipment or technologies, e.g. computers, remote sensing and geographical information systems, for the purpose of keeping records on the number of soldiers who died and returned alive during or after wars, the population of a city or state, and so on. Now, statistics is widely used by common people, e.g. football statistics, labor statistics, student enrollment statistics, and so on. As a plural noun, statistics means computed or estimated quantities, e.g. FAO statistics on production of rice (mt), aquaculture production (mt) of carps, catfish, tilapias, and so on. Statistic as a singular noun means a datum or numerical fact.

As statistics basically deals with numerical facts, it is considered a branch of applied mathematics. In fact, it is not only the mathematics; it is more about critical thinking and reasoning. Various scholars have tried to define statistics differently to reflect its processes and increasing roles. The simplest definition considers statistics a branch of mathematics that deals with the collection, analysis, and interpretation of data. The collection of data includes a good plan or design for a trial or questionnaire or survey, a clear procedure or method, materials or equipments to be used, and data compilation and storage. Data analysis means locating the central tendency, analyzing variability, exploring relationships or trends, and so on. The final part is the interpretation of the results and then making conclusions and recommendations. Therefore, the above definition was thought to be incomplete. Other definitions have been proposed. For example, statistics is the scientific study of numerical data based on variation in nature or the science of analyzing data and drawing conclusions, taking variation into account. This definition grasps the variation in data as the main characteristic. Variation or diversity is universal; For example, weights of fish vary, even if they are from the same age group, raised in the same cage, tank, or pond, and fed the same amount and type of feed. Even identical twins can differ in many attributes. If there is no variation in data, statistical analysis is not necessary. For example, Table 1.1 shows that variations (standard deviations) are zero in both sets of data in Trial 1. It can be seen that Treatment B resulted in higher survival of fish compared with Treatment A. But in Trial 2, replication 3 of Treatment A had higher survival than replication 4 of Treatment B. Due to this overlap, it is difficult to determine whether Treatment B results in higher survival of fish. Use of a statistical tool is not necessary in the case of Trial 1, whereas for Trial 2, a statistical tool is very important to make the right and confirmed decision.

Similarly, if all the fish spawned when a new hormone was used repeatedly, we could say that the new hormone is effective. But if only 90% of fish spawned among the fish injected with the new hormone, even in only one trial out of five, then statistics is necessary for making any decision. Variation in data means there are gray areas. In order to express the gray areas, researchers frequently use

Table 1.1 Hypothetical data showing with and without variation.

	Survival of Fish (%)			
	Trial 1		Trial 2	
Replication	Treatment A	Treatment B	Treatment A	Treatment B
1	75	100	85	100
2	75	100	86	96
3	75	100	91	92
4	75	100	75	90
Mean	75.0	100.0	84.3	94.5
Standard deviation	0.0	0.0	6.7	4.4

the terms *almost*, *higher*, *lower*, *many*, *few*, or *relatively more*, and so on. But the results of the research supposedly conveyed by these words are not clear as these are vague and general words/statements. Those who have statistical knowledge would at least use percentage or probability. Statistical skill and knowledge therefore give individuals the skill or the power of interpretation and reaching conclusions. Furthermore, it teaches the techniques of presenting research results correctly and also helps in critical evaluation of literature published or planned to be published. Considering the various uses and importance of statistics, it has been defined as the science of decision making under uncertainty, as a body of methods and theory applied to numerical evidence in making decisions in the face of uncertainty, as a toolkit for problem solving, and so on.

Statistics is categorized as descriptive, which means use of data to report or describe the present status or the situation. It can be in either tabular or graphical form for the purpose of facilitating explanation. Selection of appropriate descriptive statistics is important. Another category of statistics is inferential, which means data are used to make inference, decisions, or conclusions based on the characteristics of the samples or parts of a whole. Knud–Hansen (1997) considers statistics as an inductive process where attempts to understand the whole are based on examining representative parts (or samples) through sampling and experimentation. Therefore, statistics can also be considered an art of collecting, presenting, describing, and interpreting data to understand our world and solve the problems.

There are several benefits of researchers having statistical knowledge and skill. According to Knud–Hansen (1997), it provides:

- skills of establishing and testing (proving/disproving) hypotheses
- knowledge about what and how much data to collect and not to collect
- confidence in results and interpretations
- power to critically review literature or others' work

In conclusion, statistics should not be considered only a branch of mathematics, but also an essential background for researchers, which ultimately becomes a way of their life. More importantly, it is a logical way of thinking that is necessary for everyone; therefore, according to H.G. Wells, "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."

1.3 Scope and application

1.3.1 In general

Attempts to define statistics have also been made based on its application. Statistics for biological sciences is often defined as "biometry," derived from two words: *bio* meaning life and *metron* meaning measure. In other words, it is the measurement of living organisms. The biological phenomena are so diverse and affected by many causal or environmental factors, and the factors themselves are variable, uncontrollable, and often unidentifiable; therefore, a fish pond is considered a black box! These are probabilistic in nature or statistical thinking, which means there is nothing absolutely right and nothing absolutely wrong! Some scholars express that it should be considered as a separate discipline. It is also referred to as "Bio-statistics," which means application of statistical methods to the solution of biological problems. Francis Galton, the cousin of Charles Darwin, has been considered the father of biometry. Other contributors and great scholars of biometry include Karl Pearson (1857–1936) and Ronald Fisher (1890–1962). Now, statistics has also been incorporated in various other disciplines, e.g. sociometrics, which means statistics combined with sociology. Similarly, when combined with economics, it is called econometrics; with psychology, psychometrics; with chemistry, chemometrics; and with forestry, forest biometrics. There are many more fields that are using statistics as an essential component in their disciplines. This clearly shows that the use of statistics exists in almost every field.

More importantly, this age is the era of information technology. The majority of organizations understand the value of data. They have maintained databases and stored a lot of data, even though they may not have used it yet. The number of such organizations is increasing daily. One organization may have several products and activities; all of them need to be recorded or maintained. Therefore, people who have skill in using the data are always in demand. The demand lies not only in maintaining the databases but also in analyzing them. Most organizations have started doing this, and they are used for making decisions or policies and formulating strategic plans. Large corporations have large volumes of data and require very skillful people to handle, analyze, and interpret that data. *Data mining* is the handling of such large volumes of data to explore, analyze, and discover meaningful patterns or trends so that forecasting is possible. Until now, most organizations have only stored data, but the time is coming for data mining to help policy making of these organizations. This shows that there will be a huge demand for statisticians in the near future.

1.3.2 In aquaculture

Although fish farming dates back about 4,000 years, FAO data show that its actual growth started only after the 1980s. It has now become the fastest growing food production sector. As it is the only alternative to compensate for the decline of capture fishery, it is expected to grow even faster to meet the 80 million mt (almost double the current level) production demand by the year 2050. However, currently, numerous challenges in this field have created an urgent need for more research within various disciplines, for example:

- increasing environmental problems caused by aquaculture development
- introduction of new aquaculture species causing threat to indigenous ones
- developing techniques of breeding, nursing, and culture of indigenous species
- increasing disease problems as a result of transboundary movement of aquatic species and intensification of culture systems

- development of low-cost feeds from locally available ingredients
- replacement for fish meal and fish oils, which are also used for livestock feeds
- economic studies for its viability or comparative studies with other sectors
- little is known about the roles of micronutrients and their interactions, e.g. minerals, vitamins, and fatty acids
- more studies on technology transfer or adaptive research and participatory on-farm trials
- food safety and quality

There is so much research to do for the full-fledged development of the aquaculture sector; however, most of the aquaculture scientists/researchers lack statistical knowledge and skills. Statistical background is essential for researchers to be able to design proper scientific experiments, analyze and interpret data correctly, and present them appropriately. Most aquaculture researchers have to find experts in their Statistics Departments or even outside to get help with designing experiments and analyzing and interpreting the data. They also face big challenges in publishing research articles. As a result, many of them often end up not disseminating the results after carrying out research, even when they have very fruitful findings; because of this, the whole aquaculture industry is suffering.

1.4 Questions

Q1. Why do science and statistics have such a close relationship?
Q2. Why do you think statistics is so important?
Q3. How will you apply statistical knowledge and skills in the future?
Q4. Debate whether one can or can't be a researcher without statistical knowledge.
Q5. Write an essay on the applications of statistics in aquaculture.

1.5 Practical exercise

Ex. 1. During the first practical session, instructors should guide students or trainees in developing basic skills of efficient spreadsheet data handling techniques. The following exercise would be useful:

- select, insert, and delete rows/columns
- insert and rename a worksheet
- enter numbers or texts
- create series of numbers, alphabets, dates, and their combinations
- perform data sorting and use of formula, functions, e.g. sum, etc.
- make good tables and different types of graphs

Data are given in Table 1.2 for practice.

P1: SFK/UKS P2: SFK/UKS QC: SFK/UKS T1: SFK
BLBS023-Bhujel 97814051XXXXX October 26, 2008 15:39

Table 1.2 Batch weights of 15 fish from a trial at the Asian Institute of Technology, Thailand.

Treatments	Feeding Rate (%)	Replication	Batch Weight (g) Stocking	Batch Weight (g) Final Weight
A: Normal fish (3 months old)	1	1	663.3	784.2
	1	2	595.7	840.6
	1	3	581.6	814.9
	2	1	548.4	1005.1
	2	2	636.3	1304.8
	2	3	643.2	1259.1
	3	1	609.0	1513.9
	3	2	636.1	1432.4
	3	3	661.6	1291.3
B: Stunted fish (12 months old)	1	1	588.7	790.6
	1	2	493.5	734.8
	1	3	549.0	750.6
	2	1	505.5	1179.5
	2	2	517.8	1007.2
	2	3	549.3	1151.7
	3	1	560.2	1612.9
	3	2	526.6	1370.2
	3	3	572.8	1608.7

Note: Normal fish (Group A) were only 3 months of age, whereas stunted fish (Group B) were 12 months old but raised at high density under limited feeding conditions.