BEN CHANG SHIA

HOW TO THINK ABOUT STATISTICAL CONSULTATION? LEARNING FROM DATA

The purpose of this paper is to reflect on statistics teaching and practice. To begin with, I will reflect on the present status of statistics education and its importance. Then I will discuss about seven main statistical topics: 'Descriptive statistics', 'Basic concepts', 'Statistical process control', 'Random walks', 'Introduction to statistical tools', ' How to do data analysis' and ' How to consult'. With the expansion and development of IT and the Internet, it is the e-century now, and therefore statistics education should change. Finally I suggest the idea of "statistical electronic school (statistical e-school)".

1. INTRODUCTION

Nowadays, many kinds of scientific fields need statistical analysis, and statistics plays an important role in many fields of research. Many researchers and other people use statistical methods; however, they do not know what they are using. Here is a simple and essential example about how statistics can be taught.

If you want to inform someone else about the height of your classmates, what should you tell him? I believe you would not give him the height of everyone in your class, because it is worthless and useless to tell others about each piece of data. You only have to provide him with the average height of your class. The average height is a statistic.

Furthermore, you also can present him with the variance or standard deviation of the height of your class to describe the average deviation from the centre (mean) of the data. They are also statistics. From mean and variance (standard deviation), we can roughly understand the average of the height and how it is distributed. After we have found the average and distribution, we know the structure of our data well enough to make any decision. No matter whether we want to decide a marketing policy or to forecast a future trend, we will need the underlying structure in a data set.

Maybe your students will doubt that statistics is so powerful and they will be right. It is not exactly that you cannot make decisions or carry out a prediction without statistics, however statistics will help you to make more accurate and better decisions.

Today we can gather the information and data that we want easily. Before facing up to lots of data, it is better to think about what you want, what you should do and what you are doing. I believe you must realise what you can do with the data. You try to analyse the data and want the data to help you make decisions in management. Unfortunately, you are involved in statistics now.

No matter what kind of analysis and what kind of research topic you are interested in, you will always be involved in statistics. Hence when trying to analyse the data, you are using statistical thinking. And statistics is a very useful tool for data analysis and to help you make decisions.

2. TOPICS ABOUT STATISTICS

Briefly speaking, there are seven main statistical topics: 'Descriptive statistics', 'Basic concepts', 'Statistical process control', 'Random walks', 'Introduction to Statistical tools', ' How to do data analysis' and ' How to consult'.

We know that statistics is an important tool. It can figure out what the problem is, where the problem is, and after repeated experiments, we can see the probable trend. In fact, the point that statistics emphasises is not a completely correct answer but a reasonable field of answers. Before consulting, what we have to do is integrate the other professional fields. Hence, to make a good statistical inference, we should develop a second professional ability.

Therefore, we have to understand how the data are distributed and what structure it has. We can use the elementary statistical techniques such as mean, variance and statistical charts etc to get some ideas about the distribution. From 'Descriptive statistics' we will get the key trend and structure the data. We also need to understand some 'Basic concepts' about data.

And then reasonable, appropriate and further statistical analysis tools should be provided, such as making inference and decisions. 'Statistical process control' and 'Random walks' will help us to figure out where the problem is, what trend we cannot ignore and how to control the uncertainty.

What I have aforementioned are only concepts that we should make operative, so some powerful tools for the calculations should be available. They are statistical software, such as SAS. At the end of completing the consulting process, the results should be presented in many ways, such as graphs, tables, reports and files.

3. RELATIONSHIP BETWEEN DATA ANALYSIS AND MANAGEMENT

To follow, we talk about 'data analysis' and 'management'. First, we must take some definition of 'data' and 'analysis'. 'Data' is any collection of numerical values. It must have some structure in which we are interested and which we do not know yet. 'Analysis' is a process to extract the useful information from data. The typical process of data analysis is 'summary and display of data', 'formulation of tentative models (explanations)', 'fitting models' and 'diagnostic checking of model adequacy'.

Before analysing data, the most important task of all is collecting data. The source of data is 'process'. Now, we must give a definition of 'process'. 'Process' is a number of steps or operations aimed at a single goal. Data is a sample from the process; such as data from a time series.

Then we can make the summary and display the raw data. By means of the summary and display of data, we can have the essential understanding of our information. Then we can establish some models and find useful ones among them for the managers and, therefore, we can achieve the goals of our statistical study. For time series data, we can use time series models for improvement and forecasting of processes.

Today we all know that information is very important. It can help us make the correct decisions and do the right things. Where does the information come from and who generates it? You can get the information from the news or research papers written by someone else. But is it suitable for you? I do not think so. We can use the information collected by others as reference; however we should also create new information by ourselves. Therefore, we must use the process of data analysis to create the information that is more suitable for us and for our decisions.

There is now a new information technology (IT), namely "data mining", which means a series of actions to select, discover and model from the mass of data, and find out the unknown patterns to transfer to the people responsible for decision making in the company and improve the profits.

The series of actions involved in data mining includes five major steps: sample, explore, modify, model and assess. Maybe you will think that data mining is just information technology. However, we can find above the five major steps that are almost always involved in statistics. Data mining can also include regression analysis, decision trees, neural networks and other statistical analyses. The process of data analysis is just the extension of the process I described previously and the kernel of data mining is data analysis. The only difference between them is the target, since the target of data mining is the mass of data from a lot of databases.

On the other hand, what role should a researcher or a business manager play in the statistical analysis? He cannot understand the results of data analysis if he has not been involved in it. A researcher or business manager must think about 'what kind of data to collect and analyse', 'how the data should be analysed' and 'what special inquiries and actions are needed'. In other words, he must provide his professional knowledge for the statisticians. We can make more useful and correct decisions by combining professional knowledge and the statistical analysis process.

To sum up, what is statistical control of the trend? There are different ways of thinking with different variables through time. No matter what way we follow, the most important of all is that we have to understand the trend and control it, to make correct conclusions.

4. FUTURE OF STATISTICS EDUCATION

With the expansion and development of IT and the Internet, it is the e-century now. There are over 250 million people using the Internet. Everyone can use the Internet from the school, office and home easily. We will not be able to live without it in the future. It probably means everyone may become a distance learner. So statistics education should be different.

According to the University Law, education has three tasks, which are teaching, research and service. The roles of Universities are creating knowledge, preserving knowledge, integrating knowledge, transmitting knowledge and applying knowledge. Of course these are also the roles of education. Especially in the e-century, education is no longer restricted to the traditional in-campus type. A new kind of channel should be developed to allow everyone to get knowledge from any corner of the world.

We should also expand the range of education in statistics, that should be combined with the Internet, and diversify the ways and channels of education. In doing this, we will let students from anywhere and from any level understand the importance of statistics and its application. We can then help them acquire the correct understanding of statistics science and so we can achieve the goal of expanding statistics science.

A virtual University will include distance teaching, distance education, online course, asynchronous learning network (ALN), virtual campus and classroom. This new virtual University has many competitive advantages such as responsiveness, accessibility, convenience and quality at a reduced cost (Chen, 1998). Aoki and Pogroszewski (1998) describe the infrastructure needed for providing students with a learning experience and related support services to complete a degree program partially or totally online and for providing faculty members with resources for teaching and

doing research effectively online. They present the Virtual University Reference Model (VURM), which is reproduced in Figure 1. This model is intended to be a guideline or a framework for colleges and universities that plan to deliver instruction and support services to distance learners, and to be a checklist to evaluate existing distance learning programs.

In the model, a virtual university is broken down to four major components: administrative services, student services, resource services, and faculty services. Each component has a different purpose and provides students with different facilities.

Figure 1. Virtual University Reference Model
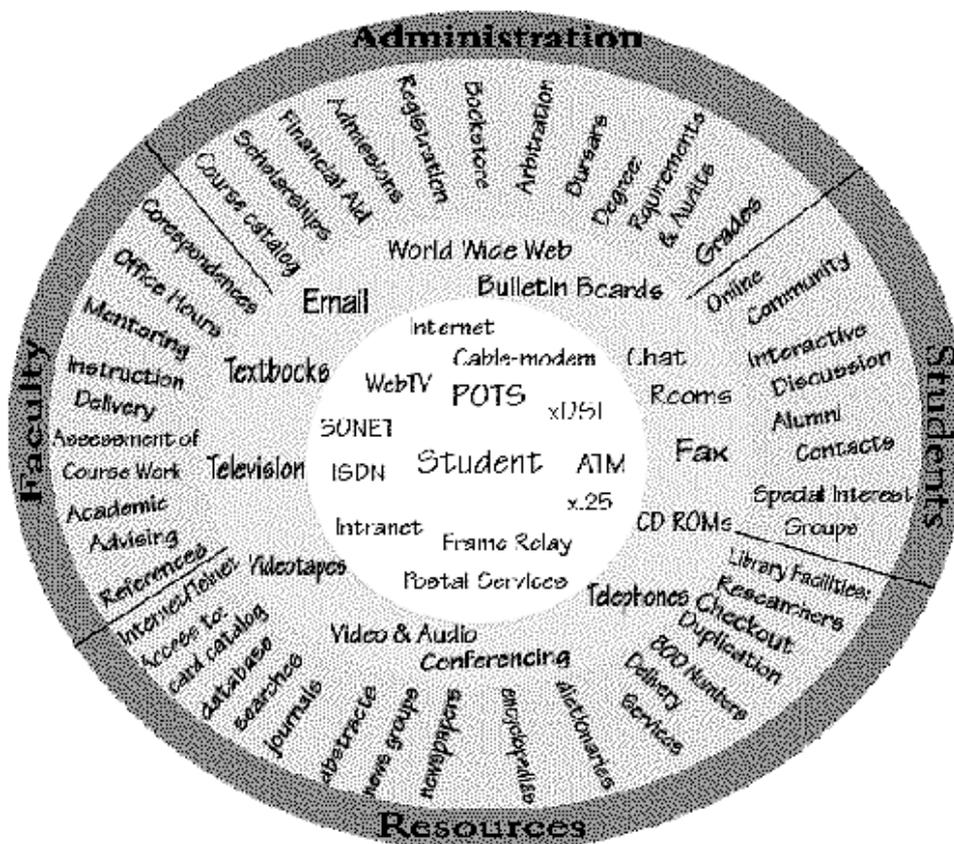
The second outer ring in the model shows the types of services a student receives from each of the four noted component areas. The inner three rings represent (from the innermost):

1. The student and his or her relationship to each of these four areas;
2. Transmission systems with which the services can be accessed by students;
3. Applications and tools to be used in offering the service elements in the outer ring.

It is important to note that in this model students are the centre of the model and all the service components and elements are depicted in relation to the students (Kumiko Aoki and Donna Pogroszewski, 1998).

So far the WWW is a very efficient tool for presenting statistical information in new ways that contribute to attracting the interest of ordinary people: The classic social facts

(economy, demography, etc) are often presented by exploiting the "technological appeal" as much as possible (Galmacci, 2001). From Galmacci's paper we can realise how many kinds of powerful statistical resources are spread in the WWW, at different web sites in each corner of the world. They should be integrated to a complete and efficient library of statistics education resources in the Internet.

This is the reason why I bring up the idea of "statistical electronic school (statistical e-school)". This new type of educational institution will not be based on the traditional campus and classrooms. In other words, it may not have a physical presence consisting of buildings and departments, but it might exist in the cyberspace (Aoki & Pogroszewski, 1998).

The statistical e-school is not only an e-book database, an introduction site, and the channel where Video Education is provided. The statistical e-school is an integrated site in the Internet world, where we are willing to include: statistical classroom for 3 levels, e-book database, Video Education database, survey and economics database, statistical consulting centre, internet survey and other services. These services include virtual personal assistant, planning learning programs, searching within the e-school site, inquiring about the main library, introducing new and good books, bulletin, game zone, and service interface.

4.1. STATISTICAL CLASSROOM

First level

The aim is to teach students to understand the application of elementary statistics and the meaning of descriptive statistics, such as percentile, percentage, mean, variance and other statistical charts etc. Using interesting examples in ordinary life we will let students learn how to make preliminary analyses.

Second level

The aim is to teach students and researchers to learn more advanced techniques of statistical analysis, such as regression, analysis of variance, time series analysis, sampling survey, contingency table, categorical data analysis, etc. In addition, students can try to use statistical software, such as SPSS, Statistica, SAS, S-plus, etc.

At this level, which is relevant to the professional statisticians and researchers, we will provide the Video Education and hire a teacher to answer the questions.

Third level

This is the upper level of statistical education, including mathematical statistics, Bayes inference, generalised linear model etc. The emphasis is on deriving the theory and teaching correct statistical concepts.

At this level we also let students know how other sciences use statistics, such as sociology, biology, information and industry technology, etc. Like the second level, we will provide Video Education and hire a teacher to answer the questions. And we will set up a group belonging to the consulting centre to solve problems from any other science.

4.2. E-BOOK DATABASES, AND STATISTICAL CONSULTING CENTER

Statistical e-school provides not only e-book and Video Education but also the functions of inquiring and consulting. Hence we will establish the databases for e-book and Video Education. Besides, using the techniques of data warehouse and data mining

we will establish the statistical survey databases, including government survey and economic data.

On the other hand, the e-school can also solve statistical problems through the consulting centre. The members of this centre should include some lecturers and graduate students. Of course, the leader should be a lecturer. In addition, we charge for using the databases and consulting. The fee will be the cost of maintaining this e-school.

4.3. INTERNET SURVEY AND OTHER SERVICES

If the website only contains the above functions, it may not interest all the intended audience. Therefore, the e-school should provide Internet surveys to increase the interest in studying statistics. We will make the topics of these Internet surveys change and will try to find some interesting ones often.

Moreover, the on-line students can use and study in the statistical e-school more conveniently and easily. The e-school provides them with lots of kinds of friendly functions and interfaces, like intranet map, virtual personal assistant, planning learning programs, searching within the e-school site, enquiries about the main library, introducing new and good books, bulletin, game zone and service interface.

5. CONCLUSIONS

The purpose of education is to spread knowledge and to expand statistics education. With the expansion and development of IT and the Internet, we can achieve the target easily by each kind of channel. It is possible to set-up a statistical e-school to provide statistical resources and consultant services for students, teachers, statisticians, researchers, business and general people. When doing this, however, we cannot forget that the essence of statistics is analysing data. Data is coming from everywhere and has any kind of pattern. It contains a lot of information and unknown patterns, that should be discovered by statistical analysis and thinking. The statistical tools of analysis and discovery should also be improved to adapt to each kind of data, and this is why we speak about "learning from data".

APPENDIX: SOME IMPORTANT WEBSITES

Asynchronous Learning Networks, including Magazine Journal of ALN: www.alnn.org

On Line Journal of Distance Learning Administration: www.westga.edu

Web-Course tools
   I-CARE: http://Pride-sun.poly.edu/icare
   Socrates Forum: http://www.environmentslintal.com/socrates
   WebServer: http://www.madduck.com
   WebMentor: http://www.avilar.com
   Learnspace: http://www.lotus.com/learningspace
   *Top class:*http://www.wbtsystems.com
   Webct: http://www.webct.com

REFERENCES

Aoki, K., & Pogroszewski, D. (1998). *Virtual University reference model: A guide to delivering education and support services to the distance learner.* Available online at http://www.westga.edu/~distance/aoki13.html

Chen, W. S. (1998) *Establishment and management of Virtual University.* Taiwan: NKFU.

Consultant Centre of NCKU (1993). *Case reporter.* Taiwan: Department of Statistics, NCKU.

Consultant Centre (1994). *1994 Price survey.* Taiwan: Department of Statistics, FJU.

Consultant Centre (1995). *Price survey.* Taiwan: Department of Statistics, FJU.

Consultant Centre (1996), *Quality survey of Nissan new car.* Taiwan: Department of Statistics, FJU.

Galmacci, G. (2001). The impact of the internet on researchers' training. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 159-169). Granada: International Association for Statistical Education and International Statistical Institute.

Hand, D. J., & Everitt, B. S. (1987). *The statistical consultant in action.* Cambridge University Press.

Lin, H. L., & Chen, C. C. (1996). *Statistics method and application I.* Taipei: Yeh-Yeh Publishing.

Lin, H. L., & Chen, C. C. (1996). *Statistics method and application II.* Taipei: Yeh-Yeh Publishing.

Statistics Science. November, 1995. Taiwan: NSC.

Survey Research. April 1996. Taiwan: Office of Survey Research Academia SINICA.

The Statistical Consultant, (1995) Volume 12, No. 2.

The Statistical Consultant, (1996) Volume 13, No. 2.

Ben-Chang Shia
Department of Statistics, Fu Jen University,
510 Chung Chen Rd. Hsin Chuang, 24205,
Taipei Hsien, Taiwan, China.
E-mail: stat1001@mails.fju.edu.tw

You will learn how data scientists exercise statistical thinking in designing data collection, derive insights from visualizing data, obtain supporting evidence for data-based decisions and construct models for predicting future trends from data. More about this course. What you'll learn. Skip What you'll learn. Data collection, analysis and inference. Data classification to identify key traits and customers. Conditional Probability-How to judge the probability of an event, based on certain conditions. How to use Bayesian modeling and inference for forecasting and studying public opi Hosted by The Department of Statistics, LSE. Learning from data: the art of statistics. David Spiegelhalter. Chair of the Winton Centre for Risk & Evidence Communication, University of Cambridge. ex-President, Royal Statistical Society (2017-2018). LSE 2019. out March 28th! The traditional statistics course. • Describing data with summary statistics. o dull. • Probability theory for drawing random observation from a population distribution. o difficult and mathematical. • Probability theory for distributions of summary statistics. o mathematical and incomprehensible. • Formulae for statistical tests. o mathematical, unmotivated, just a bag of tools. • (If lucky) Examples of using statistical models in real life. A 'modern' statistical course.