MISTAKES AND DIFFICULTIES ENCOUNTERED BY STUDENTS PARTICIPATING IN A STATISTICAL PROJECT COMPETITION: THE TREND OF CHANGE

Shen Shir Ming, Department of Statistics, The University of Hong Kong, Hong Kong, PRC

The Hong Kong Statistical Society has 12 years of experience in running an annual statistical project competition for secondary school students. Each year the competition receives about 70 to 90 entries. Throughout the period, changes in school curriculum, technology, experience gained by teachers and exposure gained by students have had impacts on the standard of the projects submitted for the competition. The mistakes revealed in the projects and the difficulties encountered by the students are very different from those found in the earlier years. This paper aims at summarizing the changes and investigates how the students can be helped to better master the skill of application of statistics.

BACKGROUND

The statistical project competition for secondary school students has 12 years of history in Hong Kong. When this article was written, the twelfth round of competition was still underway.

When the competition was first introduced in 1986, school students had little idea about statistical projects (Shen, Li and Lam, 1991). They learned graphical presentation techniques, calculation of mean and standard deviation, and rules of probability as part of their mathematics curriculum. The element of application was either ignored or made very artificial with small and unrealistic data sets made up by authors of text books. Teaching and learning were examination oriented. Project work was rare although educational values of doing statistical projects had already been well recognized by statistics teachers; see for example Dolan (1979), Schoeman and Steyn (1979).

The Competition was considered to be a remedy to the weaknesses of the local statistical education (Cheung, Lam, Siu and Wong, 1987) and was well received. In the first round, 40 entries were received and in later years built up to receiving between 70 to 90 entries each year.

TECHNICAL ASPECTS

Technical mistakes are easier to detect and to remedy. In the first few rounds, ample technical mistakes were found in the projects submitted. Li and Shen (1992) study

the weaknesses of the participants' performance in those statistical projects. Many examples have been quoted in their study. Situations have improved in recent years; however, problems still exist and the following are some examples.

Graphical presentation

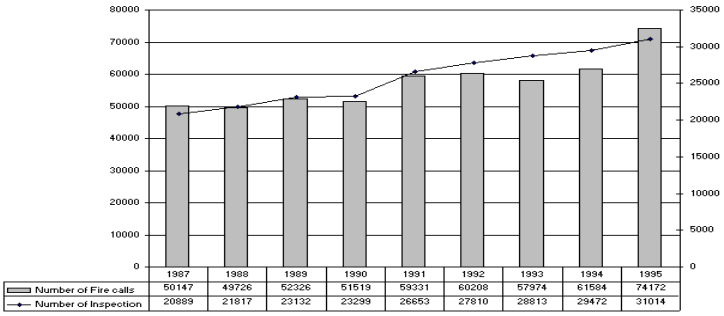
An example is quoted from a project entitled *The Deterioration of Manufacturing Industry in Hong Kong*. The participants used different graphical presentations to show the change in percentages of the workforce employed in different industries.

One graph used a bar diagram to show the workforce engaged at the secondary sector over the period 1966 - 1993, while the next two graphs used a line graph and a smoothed curve, respectively, to show the workforce engaged in food manufacturing and construction industries for the same time period. There is no justification given in the project for the use of different kinds of graphs. There are at least two possibilities.

Firstly, the participants are themselves not sure which graph is the most appropriate one for their purpose. Because all three seemed to be applicable in this situation, they made full use of the options provided by the computer software. Alternatively, participants have the perceptions that a statistical project should be made more exciting by using different kinds of graphical techniques. In fact, graphical presentation should serve the purpose of disseminating quantitative information in a quick, lasting and accurate manner. If all three graphs were of the same type, readers could easily make preliminary comparisons simply by inspecting the graphs. More information could be grasped in a shorter time. Apparently, the objectives of using statistical graphs as effective tools for giving quick information have not been made known to the students.

Dual axes is an option provided by many computer software packages and appear frequently in mass media. Very few intelligent uses of the technique have ever been demonstrated apart from saving space which could be beneficial to some articles appearing in the newspapers or magazines. The technique certainly provides a tool to manipulate statistical facts because very different impressions can be produced in a dual axes diagram by adopting different scaling of the two axes. Many a time, such graphs are totally unnecessary. For example, in the project, *Fire and Fire Services in Hong Kong,* the participants tried to present the number of fire calls and the number of fire inspections in the same graph. They produced a graph as shown in Figure 1. The vertical axis on the

right hand side is redundant, and the two variables should have been presented in the same graph without this second axis. The scaling of the second vertical axis such that the number of inspections appear to be close to the top of each of the bars showing the number of fire calls is entirely artificial and could be subject to query.



	1987	1988	1989	1990	1991	1992	1993	1994	1995
Number of Fire calls	50147	49726	52326	51519	59331	60208	57974	61584	74172
Number of Inspection	20889	21817	23132	23299	26653	27810	28813	29472	31014

Figure 1. Use of dual axes.

In spite of statistics being taught as part of mathematics, only a few projects had applied simple mathematical skill in the projects. The calculation of proportions and percentages which often help a lot in digging out information stored in data sets is the most commonly applied technique. In the Fire and Fire Services project, for example, a table was shown to present the number of classified fire inspections from 1987 to 1995. Of the last three rows of the table giving percentage increase in the number of inspections, the first two rows showing the percentage increase in two four-year periods were obtained from existing government publications. The students went on to produce the last row which gives the percentage increase over the entire eight year period. To compare these with the yearly number of fires and to study the development in fire services, it would have been much more informative if the students had produced the annual percentage increase.

Quantitative analysis and interpretation

The deficiency in statistical education being taught as part of mathematics has been revealed. The recognition of patterns of number sequences, for example, is one of the favorite exercises in mathematics. Those exercises may give number sequences with missing values like

(a) 2 , 4 , 6 , _ , 10

(b) 1 , 2 , 4 , 7 , __

and students are expected to give “8” as the answer to sequence (a) and “11” to sequence (b). Such training deals with deterministic mathematics but not the non-deterministic statistics, and students being drilled by those exercises could reach a quick and misleading conclusion. In a project entitled *The Promising Future - New Towns,* for example, the participants conclude, after reviewing three years of population growth in a new town Yuen Long, that the population in that town will keep on growing. A statistical trend has little chance to be identified by such few observations. The essence of statistics as a science of dealing with uncertainty and variability has not been appreciated.

NON-TECHNICAL ASPECTS

The Competition conducted in Hong Kong requires participants to choose a topic of study that can be on various aspects of the Hong Kong Society like the Hong Kong economy, population, education, housing, industry and so on. In this way, the project is combining statistics with other social areas. Therefore students have an opportunity to apply the statistical techniques they learned in the classroom to the real life data.

Selection / collection of data

In recent years, participants have become more experienced in looking for the data they need. The advancement in technology has certainly played a part because instead of relying solely on the hard copies of data they can now make use of the world wide web to search for data. The intellectual property right is also respected, and they have started to build up the tradition to quote the source of the data they used.

Inadequate information, however, still prevails. For example, in a project entitled, *Miracle of the World - Blooming Property Market in Hong Kong,* the students used different data sets to explain the demand for housing units. Figure 2 (a) and (b) reproduced here are the population growth rate and population growth due to migration presented by the participants.

A natural query arising from these graphs is why did the students choose a different time period for the two graphs. This is an important consideration because based on Figure 2(b), the students say that the population growth due to migration in year 94-95 is almost nine times that of 90-91. The project goes on to emphasize that the population has a growing trend and hence the property market will continue to bloom. Had the students been able to obtain a longer series such as the one in Figure 2(a), though still not

long enough to demonstrate a long term trend, they might be able to see that the population growth may go up or down, not necessarily going up all the time. In fact, hasty and simplistic conclusions drawn from inspecting inadequate data are common.

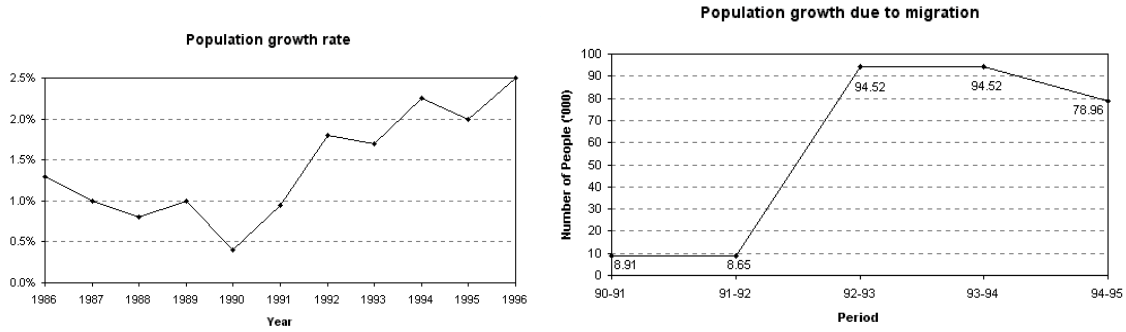


Figure 2(a) and (b). Inadequate data

Many a time more data may not imply the data used are adequate. Some students, for example, collected from the web the numbers of injuries and deaths due to household fires in different countries and made comparison. The additional data from outside Hong Kong can be a plus if used appropriately. This particular group of students, however, compared those numbers without making reference to the population density and hence arrived at unsound conclusions.

Definitions

In recent years, most of the students can put down definitions clearly in their projects. In home purchasing, for example, students were able to define what is meant by "affordability" Sometimes putting down definitions clearly does not imply fully understanding it. In a project, *The Poverty Puzzle*, poverty is defined to be people with income below the "poverty line" which is defined to be half of the median income. The project then went on to say that the government should aim at "eliminating poverty". From the definition, it is clear that as long as variation in income exists, no matter how wealthy a society gets, there will be people having income under the poverty line such defined, and hence poverty will never disappear.

DIFFICULTIES AND THE TREND

As a whole, we have witnessed remarkable improvement in the standard of the statistical projects submitted over the years. We started with receiving quite a number of

unqualified entries simply because the participants did not understand what is meant by a statistical project which rely heavily on facts and information provided by data. Some students, for example, explained government policy and produced photographs of school buildings of different times to impress the readers that the education input had increased and educational opportunities had increased. With little statistical support, such projects were all disqualified. Now, we are receiving decently written statistical reports with data tabulated, analyzed, graphically presented and results interpreted and summarized.

Interviews with participants revealed that they have benefited a lot from participating in the competition. The participants valued their experience because they found that in the process of preparation, they learned how to use statistics and present data in a systematic way, they also understood various aspects of the society better. A winner in the 87/88 round of competition said, "The competition uncovered my interest in statistics." He went on to do a degree in statistics and is now a government statistician in the Census and Statistics Department.

We expect the project competition will become even more popular. The activity has already received more and more support from school teachers since a new statistical curriculum was introduced in the high school in 1992 (Shen, 1995). Statistics began to get its identity and teachers found the project competition useful in arousing the students' interest in studying statistics.

We believe the competition is serving its educational role well. As time goes by, the message that statistics is a useful subject, applicable in real life problems, that it enhances the understanding of the society, and that it is a most useful tool in this information explosion age will be thoroughly understood.

REFERENCES

Cheung, P. K., Lam. K., Siu, M. K. and Wong, N. Y. (1987). An appraisal of the teaching of statistics in secondary schools of Hong Kong. *Proceedings of the Second International Conference on Teaching Statistics*, 241-4.

Dolan, D. (1979). Learning statistics through project work. Teaching Statistics, 1, 34-44.

Li, K. Y. and Shen, S. M. (1992). Students' weaknesses in statistical projects. *Teaching Statistics*, *14*, 2-8.

Schoemann, H. S. and Steyn, A. G. W. (1979). Statistical project work: the Pretoria experience. *The Best of Teaching Statistics* by P. Holmes (Ed.), 66-68. Teaching Statistics Trust.

Shen, S. M. (1995). New Development in School Level Statistical Education in Hong Kong. *Bulletin of the International Statistical Institute*. 1085-6.

Shen, S. M., Li, K. Y. and Lam, K. (1991). Statistical competition for secondary school students - a Hong Kong experience. *Proceedings of the Third International Conference on Teaching Statistics*, Vol. 1, 214-8.