Bulletin
Volume 24 Number 2
September 2001

Editor's Foreword

It is a great moment of the Society. After six years of work, we finally reached an agreement with the Royal Statistical Society of UK to have our own professional examination. The Project Coordinator, Mr. HW Fung of the Census and Statistics Department, kindly writes a report on the process how we achieved this objective. He also includes some relevant documents regarding to this issue.

Mr. O.M. Kwok of Arizona State University writes an article on the determination of sample size in ordinary least square regression. He reviews several methods in determination of sample size.

The Society will hold a seminar dinner on 21 November 2001 and the speaker is Professor Stephen Fienberg of Carnegie Mellon University. Members are welcome to attend the seminar and enjoy the dinner after the talk.

P S Chan

Editor	:	Chan, Ping-shing, CUHK	Tel.	2609 7920	Fax.	2603 5188
Associate Editors	:	Chan, Wai, CUHK		2609 6241		2603 5019
		Chan, Wai-sum, HKU		2857 8318		2858 9041
Secretary	:	Lam, John Hon-kwan, C&SD		2802 1267		2827 0551

CONTENTS

(Vol. 24/No. 2, September 2001)

President's Forum

Professor W.K. Li

I would like to thank you all for your enthusiastic turnout at the recent EGM. Your support for the change in the Society's MAA has allowed the taking over of the RSS examinations to go full speed ahead under the leadership of Mr H.W. Fung.

We had also a successful press conference in early October. The number of enquiries on the HKSS examination is increasing. We would need your continuous support in the final success of the professional examination programme, which hopefully will also enhance the professional status of statistics profession in Hong Kong.

On the lighter side, the Society has organised a Sports Day at CUHK recently. On the academic side, there will be an upcoming seminar by Professor Stephen Fienberg in November.

The Hang Seng Bank has also agreed to support the SPC for one more year. A new feature this year will be a Thematic Project on Hong Kong's Financial Industry. The winning team will be awarded the Hang Seng Bank Prize for the Best Thematic Project. As a recognition of the contribution of the SPC project to education, the Education Department of HKSAR has indicated interest to up load the past winning projects on their stat.net. We hope that there will be more collaboration with the Education Department in future in serving the local community.

Wishing you all the best.

Launching of the HKSS Professional Examination

H W Fung
Census and Statistics Department

(Mr. H W Fung is the Deputy Commissioner of the Census of Statistics Department. He is also the Project Coordinator of the HKSS Examination.)

Delight

It is with delight and honour that I am writing this article on behalf of the HKSS Project Team and also all those HKSS members who have been working on the establishment of the local professional examination. After six years of work, which commenced from my first discussion with Professor Vic Barnett (Hon Treasurer of the Royal Statistical Society (RSS)) in Beijing while attending the ISI Meeting in September 1995, we have finally achieved the objective of having our own professional examination as from May 2002, which will be the first of its type organized by a statistical society in the Asian region. The very positive response from the RSS has certainly been most helpful in firming up our confidence in pursuing with the project, and so have been the support and advice from all of you who have contributed to this seemingly formidable task.

You have been updated from time to time through this Bulletin regarding development of the examination project. So I would choose just to mention here in greater detail some of the more important work done on the examination project in the past few months and also some of the latest development work of which you should be informed and in which you may wish to participate.

The June Meetings

As you may recall, even with the financial support we have secured from the Government, and the fact that ex- or serving Presidents of the RSS had visited us in 1996 (Prof Peter Moore), 1998 (Prof Adrian Smith) and 1999 (Prof Robert Curnow), a business plan with financial details agreeable to both the RSS and the HKSS had taken much time to work out. As a result, the commencement date of the Project, originally scheduled in May 1999, had been deferred for more than once.

A breakthrough to the above finally came in June this year, when the Executive Secretary of the RSS, Mr. Ivor Goddard, visited us and had a whole week's meetings

(from 4 to 8 June 2001) with our Project Team and various Working Groups. Thanks must go to all those who actively participated in the meetings with Mr. Goddard, and also to our friends in HKU, PolyU, SPACE of HKU and the Census and Statistics Department who offered lunch and dinner meetings/briefings to Mr. Goddard. The meetings and various lunch/dinner activities have proved to be most useful in enabling the two sides to have a full exchange of views on various aspects of the Project, and in conveying to the RSS a strong message of the close collaboration among statistical professionals in Hong Kong and our determination of having our own professional examination. Thanks must also go to Mr. Goddard for offering constructive suggestions and advice on the possible alternative administrative arrangements between the two sides. Finally, a business plan acceptable to both the HKSS and RSS was worked out.

What is most worth noting under the new business plan is that the HKSS, as the controller of the examination in Hong Kong, will outsource the more routine activities of the examination to the RSS. This will enable the HKSS to focus its resources on the redevelopment of the examination syllabi and papers and the promotion of the examination. Enhancements to the syllabi will aim at making them more geared to the needs of employers and employees in the market of Hong Kong and the region. Promotion of the examination will not be confined to candidates in Hong Kong but can be extended to candidates of other places in the region such as Macao and cities in the Mainland. Such arrangement would enable both sides to make the best use of their resources. As a result, the HKSS is able to draw up a more financially viable budget for the Project.

The Agreement

Subsequent to the June meetings, the HKSS and RSS have worked out an Agreement regarding the details of the arrangement proposed above. The Agreement has been signed by the two sides in August this year. Also a Joint Announcement has been worked out by the HKSS and RSS which is to appear in the Newsletter or Bulletin of the two Societies to inform members and the public of the agreed arrangement for the examination in Hong Kong. The Joint Announcement is reproduced at Appendix 1 of this article.

The launching of the Examination

Following the above, the HKSS will formally launch its professional examination as from May 2002 which will replace the examination held by the RSS in Hong Kong. To enable this to be carried out, an EGM was held on 13 September this year to pass the necessary resolutions to enable the HKSS to admit examination associates and to agree on the subscription fees required for professional membership pursued by candidates who will

get through the HKSS examination. All the necessary amendments of the Memorandum and Articles of the Association for the HKSS were approved in the EGM.

The Press Conference

To promote the examination, which will start with the registration of examination associates, a press conference was held on 3 October. A press release was issued to the media on the same day (a copy of this is reproduced at Appendix 2). There was a rather good coverage of the press conference by the media, with six newspapers reporting it quite prominently. One newspaper also conducted an interview with me several days after the press conference and reported the subject again later in the month. In addition, a promotional poster and leaflet are being widely distributed to relevant academic institutions and organizations through programme leaders of statistics courses or human resource managers. Further promotion activities are being arranged.

The publicity efforts are taking effect and a number of enquiries have already been received from candidates and some of them have expressed keen interest in sitting for the HKSS examination. I was informed by the RSS that they had also received quite a number of enquiries about the examination in Hong Kong.

The HKSS Website

Another major development is the revamping of the HKSS Website, which is necessary in order that the public can access the relevant information about the examination and various application forms conveniently. The opportunity was also taken to update various parts of the Website. Thanks must go to our Webmaster and also various members who had kindly assisted in preparing the relevant materials and uploading them onto the Website. I would like to take this opportunity to invite all members to visit the HKSS Website and give their suggestions on what aspects further expansion or enhancement would be necessary. It is natural that with the launching of the examination and the publicity activities that have been or will be organized in recent months, more people will be browsing our Website. So it is most important that our Website is well presented and thus useful in helping to promote the image of the Society.

What will follow

The above of course just marks the beginning of the work required to be done in respect of the Examination Project. To enable various aspects of work on the project to be carried out, a number of Working Groups and Committees have been or will be formed. These include:

(a) The Working Group on Publicity and Promotion;
(b) The Professional Affairs Committee;
(c) The Examination Board; and
(d) The Syllabus Development Committee.

As you are aware, our Examination Project consists of three phases. Phase I of the project, which has just begun, mainly involves replacing the RSS examination in Hong Kong with the HKSS examination in 2002. Phase II of the project, which is to be completed in the year 2003, will aim at the introduction of new syllabi and papers to the examination. Phase III of the project will see the introduction of bilingual (English and Chinese) papers in the year 2004. At the end of Phase III, a complete accreditation and examination system will be set up which will enable the syllabi of statistics courses offered by various academic institutions in Hong Kong and the standard achieved in their course examinations to be matched with those of the HKSS examination. These will be the main tasks faced by the aforesaid Working Group/Board/Committees.

I am glad to mention here that we have got representatives from all relevant tertiary and vocational institutions and organizations participating in our Project Team and the above Working Group/Board/Committees. These include the HKU, CUHK, CityU of HK, HKUST, HK Baptist U, HK PolyU, The Open U of HK, the HK Institute of Vocational Education, the HK College of Technology and the Census and Statistics Department. No doubt you will agree with me that with the joint efforts of members of our statistical community, we will be able to make our examination project a success.

I will report to you progress of the examination project in the future issues of this Bulletin. I am sure I can continue to count on your support when you are approached by members of the above Working Group/Board/Committees for assistance and advice.

Appendix 1

Joint Announcement by
The Royal Statistical Society (RSS) and The Hong Kong Statistical Society (HKSS) on the Examination in Hong Kong for the year 2002 and onwards

The RSS and HKSS signed an Agreement in August 2001 under which the examination which has for many years been held by the RSS in Hong Kong will be replaced by the HKSS examination as from the May 2002 round of the examination. Candidates who wish to sit the May 2002 examination in Hong Kong leading to various levels of qualification in the statistical profession should join the HKSS as an Examination Associate on or before 31 January 2002, and then apply to the HKSS to sit specific examination papers within the month of March 2002. The time schedule for the HKSS examination will be the same as that announced for the RSS examination, i.e. the HKSS examination will also take place on 14-16 May 2002.

The RSS and HKSS would like to assure all candidates who are intending to sit the HKSS examination that the syllabuses and examination papers and the standard for marking examination scripts for the May 2002 round will be the same as those of the RSS examination. Candidates who have been taking the RSS examination in Hong Kong are advised to continue to pursue their professional qualification in statistics by taking the HKSS examination. The certificates issued by the HKSS to successful candidates for all levels of the professional examination (i.e. Ordinary Certificate, Higher Certificate and Graduate Diploma) will be endorsed by the RSS and fully recognized by the RSS as of a standard equivalent to those issued by the RSS.

Under the Agreement between the RSS and HKSS, the HKSS will undertake a review of the examination syllabuses with a view to proposing enhancements which will enable the syllabuses to be more geared to the needs of employers and employees in the market of Hong Kong and possibly also other places in the Asian region. New syllabuses and examination papers introduced to the HKSS examination will have the endorsement of the RSS that they are of a standard intellectually equivalent to that of the existing RSS examination syllabuses and papers. It is expected that changes to the syllabuses will be introduced to the HKSS examination as from the May 2003 round of the examination.

It is also a target of the HKSS to introduce bilingual papers (English and Chinese) as from the May 2004 round of the examination to cater for candidates who may wish to sit for the

Chinese language version of the papers. Arrangements will be made between the HKSS and RSS to ensure that the marking of the papers in Chinese will be of a standard equivalent to that in English. Again, the qualifications granted by the HKSS to candidates taking the papers in Chinese will be endorsed and fully recognized by the RSS to be of a standard equivalent to the qualifications of the same level granted by the RSS.

Both the RSS and HKSS believe that the above arrangement, particularly that for the HKSS to develop the examination syllabuses in aspects catering for the specific needs of the local Hong Kong economy, should be of benefit to the Hong Kong statistical community.

As with the RSS examination, candidates taking the HKSS examination may be eligible for exemption from taking the examination of the Ordinary Certificate, Higher Certificate or some papers of the Graduate Diploma because of the courses in statistics they have taken and passed in certain institutions. They should submit their applications for exemption early for processing by the HKSS and not later than the time they submit their applications in March for sitting specific examination papers. Candidates will be informed in April whether their applications for exemption and examination are accepted, and also other details of the May examination.

Information about the HKSS examination, including the fees to be charged for individual papers, the application forms for registration as examination associates and as candidates for specific examination papers, and past papers and solutions for reference will be available on the HKSS website at www.hkss.org.hk as from 3 October 2001. Registration as Examination Associates of the HKSS for the May 2002 examination will start in October 2001 and end on 31 January 2002, while application for sitting individual examination papers will commence on 1 March 2002 and end on 31 March 2002.

Meanwhile, anybody wishing to make enquiries about the HKSS examination or to obtain application forms from the Examination Office of the HKSS, are welcome to contact the HKSS through the following channels:

The Hong Kong Statistical Society Examination Office
c/o Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam Road, Hong Kong.
Tel: (852) 2859 2467
Fax: (852) 2858 9041
E-Mail: exam@hkss.org.hk

Appendix 2

Press Release issued by HKSS on 3 October 2001

Hong Kong will launch
its first statistical professional examination for the Asian region

Hong Kong will have its own examination for statistical professionals as from the year 2002, the Hong Kong Statistical Society (HKSS) announced in a press conference today. This will also be the first statistical professional examination offered by a statistical society in the Asian region.

"It is an important development for the statistical community in Hong Kong and the Asian region. So far examinations for statistical professionals have been provided only by the Royal Statistical Society (RSS) of the United Kingdom, which has been holding the open examination in various places around the world, including Hong Kong. After several years' discussion between the HKSS and the RSS, agreement has been reached in August this year that the HKSS examination will replace the RSS examination in Hong Kong as from the May 2002 round of the examination," said Professor W K Li, President of the HKSS.

The project of establishing a local accreditation and examination system for statistical professionals was proposed by the HKSS in 1998 and sponsored by the Services Support Fund (which has been replaced by the Innovation and Technology Fund since 2000). The coordinator of the project is Mr. H W Fung, who is a senior member of the HKSS and a senior staff in the Government Statistical Service.

"I sat for the examination in 1974, one year after I graduated from the university and just before I joined the Census and Statistics Department. I found it a challenging and rewarding task to sit for the examination and get through it, as it demanded knowledge of not only statistical theories taught in the university but also practical techniques and methods which have to be applied in real life situations. I strongly recommend students of statistics to sit for the examination," said Mr. H W Fung.

The structure of the HKSS examination, similar to the RSS examination, consists of three levels: the Ordinary Certificate level, the Higher Certificate level and the Graduate Diploma level. The Graduate Diploma qualification granted by the RSS is equivalent to an Honours Degree in British universities and is accepted by universities in Hong Kong as

meeting the entry requirements to Masters degree programmes in statistics.

"The growing importance of data gathering and data analysis in various fields of business points to a strong demand for persons with sound knowledge in both statistical theories and their practical applications. The examination system serves to provide a convenient yardstick for employers to ascertain that their employees are professionally qualified to carry out the data gathering and analysis work. We in Hong Kong are most glad to be entrusted by the RSS, a highly prestigious professional statistical society, to take over the examination in Hong Kong. All the certificates issued by the HKSS will be endorsed by the RSS and fully recognized by the RSS as of a standard equivalent to the certificates issued by the RSS for the same levels of qualifications," said Mr. H W Fung.

The HKSS examination, which has mutual recognition with the RSS examination, is a recommendable avenue for pursuing qualifications at various levels of the statistical profession. Structured in 3 stages, the examination allows flexibility in the pursuit of studies and examination-taking so that people can proceed in accordance with their individual circumstances in regard to study, work and family situations.

Commenting on the new arrangements, Professor Peter Green, President of the RSS, stated: "I am delighted with the agreement concluded between the RSS and HKSS. The RSS has been proud to be associated with the statistical community in Hong Kong over the past decades. However, it is surely right that statisticians in Hong Kong, through the HKSS, should now take responsibility for their own examination system. This will enable them to refine the syllabuses and examinations to meet the particular needs of the Hong Kong economy and the business and academic communities there. The RSS will of course be happy to continue to work in partnership with HKSS to ensure that the standards of the examination remain equivalent to the international standards of the RSS."

Candidates in Hong Kong would benefit financially from the taking over of the examination by the HKSS as they will have to pay less for each examination paper they wish to sit, given that the examination is organised locally in Hong Kong.

For candidates in Hong Kong who have been sitting the RSS examination and preparing to take the RSS examination in May 2002, they are assured by the HKSS that the syllabus and examination papers and the standard for marking examination scripts for the May 2002 HKSS examination will be the same as those for the RSS examination. The certificates issued by the HKSS to successful candidates for all levels of the professional examination will be endorsed by the RSS and fully recognized by the RSS as of a standard

equivalent to those issued by the RSS.

"Under the Agreement between the HKSS and RSS, the HKSS would in the coming years undertake a review of the examination syllabus with a view to bringing about enhancements which would enable the syllabus to be more geared to the needs of employers and employees in the market of Hong Kong. It is also the plan of the HKSS to introduce in due course bilingual papers (English and Chinese) to cater for candidates who may wish to take the Chinese version of the papers," said Dr. W K Chiu, Executive Secretary of the HKSS Examination Office.

"Candidates wishing to sit for the HKSS examination will have to register as Examination Associates of the HKSS first. The annual subscription for an Examination Associate is HK$250. The deadline for registration as Examination Associate for taking the May 2002 examination is end January 2002. Acceptance for registration will commence in October 2001", added Dr. W K Chiu.

After registering as Examination Associates, candidates wishing to take the HKSS examination in May 2002 should submit applications in the month of March 2002 with regard to the papers they wish to sit. Some candidates may be eligible for exemption from taking the examination of the Ordinary Certificate, Higher Certificate or some papers of the Graduate Diploma because of the courses in statistics they have taken and passed in certain institutions. They should submit their applications for exemption early for processing by the HKSS and not later than the time they submit their applications in March for sitting specific examination papers. Candidates will be informed in April whether their applications for exemption and examination are accepted and also other details of the examination. The 2002 HKSS examination will take place on 14-16 of May 2002.

"Details of the fees required, application forms for registration as Examination Associates and for sitting specific examination papers, and past examination papers and solutions for reference will all be available on the HKSS website at www.hkss.org.hk as from 3 October 2002," said Dr. W K Chiu.

If candidates have any enquiries concerning the HKSS examination or wish to obtain application forms from the HKSS Examination Office, they are welcome to contact the HKSS through the following channels:

The Hong Kong Statistical Society Examination Office
c/o Department of Statistics and Actuarial Science

The University of Hong Kong
Pokfulam Road, Hong Kong.
Tel.: 2859 2467 Fax .: 2858 9041
E-mail: exam@hkss.org.hk

(Annex : Introduction to the
Hong Kong Statistical Society Examination)

Introduction to the Hong Kong Statistical Society Examination

The structure of the HKSS examination consists of three levels viz. the Ordinary Certificate level, the Higher Certificate level and the Graduate Diploma level.

(1) Ordinary Certificate in Statistics

The aim of this first level certificate is to provide a sound grounding in the principles and practice of statistics with emphasis on practical data collection, presentation and interpretation. It is intended both as a first qualification, an end in itself, and also as a basis for further work in probability and statistics, as for example in the more advanced syllabuses of the Society.

The examination consists of two 3-hour written papers. The entry level for the Ordinary Certificate is a credit (i.e. C or above) in Mathematics of the Hong Kong Certificate of Education Examination (HKCEE), or an equivalent qualification or relevant work experience.

(2) Higher Certificate in Statistics

Entry to the Higher Certificate assumes knowledge of the material in the Ordinary Certificate. Candidates must have passed or been given exemption from the Ordinary Certificate before sitting examinations for the Higher Certificate. This award is intended both as a more advanced qualification in statistics, an end in itself, and as a basis for further work in statistics up to the highest undergraduate level, as for example in the Society's Graduate Diploma.

The examination consists of three 3-hour written papers on theoretical and applied statistics.

(3) Graduate Diploma in Statistics

The Graduate Diploma is a qualification in applied statistics at a level equivalent to a good Honours Degree in British universities and is accepted by universities in Hong Kong as meeting the entry requirements to Masters degree programmes in statistics. Award of the Graduate Diploma entitles the holder to apply for the professional qualification of Graduate Statistician. Entry to the Graduate Diploma assumes knowledge of the material in the Higher Certificate. Candidates must have passed or been given exemption from the Higher Certificate before sitting examinations for the Graduate Diploma.

The Diploma is examined by five 3-hour written papers, two on Statistical Theory and Methods, two on Applied Statistics and one or an Options Paper. The subjects in the options paper take the form of half-syllabuses in several subject areas:

Statistics for Economics	Econometrics
Operational Research	Medical Statistics
Biometry	Statistics for Industry and Quality Improvement

Candidates have to answer questions from <u>two</u> of these half-syllabuses.

Candidates for the Graduate Diploma may take all five papers at one sitting. Alternatively, they may take the two Statistical Theory and Methods Papers at one sitting (with or without the Options Paper) and the two Applied Statistics Papers at a separate sitting (again with or without the Options Paper), provided that the Options Paper is taken at one of these two sittings.

Reviewing Sample Size Estimation in OLS Regression

Oi-Man Kwok
Arizona State University

The main purpose of this paper is to review research on the estimation of sample size from the OLS regression. Different rules of thumb on sample size estimation in regression analysis have been proposed. However, those rules of thumb have been shown to underestimate the necessary sample size needed to detect an effect (Green, 1991; Maxwell, 2000). Power analysis (Cohen, 1988) is now commonly used to estimate the necessary sample size. Three values are necessary to estimate sample size: alpha level, desired power, and effect size. However, choosing a realistic effect size in the population is difficult. Ideally, this parameter should be based on researchers' knowledge of that research area rather than simply assuming an arbitrary value (Jaccard, Turrisi, & Wan, 1990; Maxwell, 2000). Maxwell (2000) suggested an alternative approach to determine sample size, which is to specify the zero-order correlations among predictors and the dependent variable through an exchangeable correlation structure. This approach provides an easier way to find the necessary sample size. Additionally, a method for estimating the necessary sample size to detect interaction effects in OLS regression, and the effect of measurement error on sample size estimation are discussed.

Introduction

Estimation of Sample Size: Overview

How many subjects should be collected for a study is always an important issue. Historically, researchers have relied on rules of thumb to estimate the necessary sample sizes, such as Tabachnick and Fidell's (1989) 5 (observations) to 1 (predictor) rule, and Harris' (1975) $N > 50 + m$ rule (Green, 1991). However, as shown by Green (1991) and Maxwell (2000), these rules of thumbs always underestimate the necessary sample size for detecting the proposed effects. Instead of using rules of thumb, power analysis provides a far stronger basis for sample size estimation (Cohen, 1988; Green, 1991; Maxwell, 2000).

Power is the probability of making a correct decision of rejecting the null hypothesis when it is false (Cohen, 1988; Stevens, 1992). Basically, two types of power estimation can be identified: a priori estimation and post hoc estimation (Stevens, 1992). A priori estimation of power provides the researcher with an estimate of how many subjects will be needed in a study for adequate power. Post hoc estimation can help provide evidence about the likelihood that a non-significant finding may be due to the absence of the effect versus insufficient power to detect an effect of the specified size. Because the main concern of this paper is on sample size estimation, my focus will be primarily on a priori power estimation.

As indicated by Neter, Kutner, Nachtsheim & Wasserman (1996), researchers can obtain more precise estimation of the needed sample size to control both Type I and Type II errors through power analysis. The ability to reject null hypothesis, Ho, given that it is false is always increased by a large sample. At the same time, it is wasteful for researchers to collect more observations than are necessary to detect a large effect, especially under conditions of limited resources. Thus, power analysis is a better approach for accurate sample size estimation. In power analysis, three factors determine the necessary sample size: alpha level, power, and effect size (Cohen, 1988; 1990). Alpha is the type I error rate, which is generally set to .05 in social sciences. Power is defined as one minus type II error rate, which is the probability of rejecting the null hypothesis when it is false. Cohen (1988) suggested power = .80 represents an adequate power. Some researchers (e.g. Topol et al., 1997) have even called for higher levels of power when estimating the necessary sample size. Effect size, in general, is the magnitude of the standardized difference between the alternative hypothesis and the null hypothesis.

Sample Size Calculation under Regression Framework

Generally, there are two separate classes of parameters that are of interest in a regression model: the overall prediction (R^2) and the individual regression coefficients. In addition, the contribution of a single predictor or a set of predictors in a regression model can be evaluated through the comparison between the full model (with all the variables) and the reduced model (without the target set of predictors).

Thus, Cohen (1988) proposed the following equation for calculating the effect size[1] based on the equation for testing the gain in prediction from a set of variables (Cohen, 1988; 1990; Cohen & Cohen, 1983):

$$f^2 = (R^2 - R^2_{(-j)}) / (1-R^2) \qquad (1),$$

where f^2 is the effect size of the parameters being tested, R^2 is the multiple correlation of the full model, and $R^2_{(-j)}$ is the multiple correlation of the reduced model (with jth predictor or set of predictors excluded from the model). Thus, effect size measures "the strength of a particular effect, specifically the proportion of systematic variance accounted for by the effect relative to unexplained variance in the criterion" (Aiken & West, 1991, p.157).

The noncentrality parameter (λ), which measures the extent of difference

[1] In fact, for examining a particular regression model with a set of predictors, the calculation of effect size is equal to $f^2 = R^2 / (1-R^2)$, where R^2 is the proposed multiple correlation of the model.

between the central (null hypothesis) F distribution and noncentral (alternative hypothesis) F distribution, is a function of effect size, number of predictors, and sample size, as shown below:

$$\lambda = f^2 (u+v+1) \qquad (2)$$

where f^2 is the effect size, u is equal to the number of predictors, p, and v is equal to N-p-1 (N is the overall sample size).

Thus, the estimated sample size is equal to

$$N = \lambda / f^2 = \lambda (1-R^2)/(R^2 - R^2_{(-j)}) \qquad (3)$$

The estimated sample size can be calculated if researchers can specify the noncentrality parameter and the value of the relevant multiple correlations. Cohen (1988) provided some guidelines for the value of effect sizes in terms of small (f^2 = .02 or $R^2_{yj.-j}$=.02), medium (f^2 = .15 or $R^2_{yj.-j}$=.13), and large (f^2 = .35 or $R^2_{yj.-j}$=.26). Therefore, the specification of noncentrality parameter is the central task for sample size estimation. Green (1991) proposed an equation to approximate the value of λ. Given alpha equals to .05 and power equals to .80, λ is a simple function of number of predictors, as shown below:

$$\lambda = 6.4 + 1.65\ K - .05\ K^2 \qquad (4)$$

where K is the number of predictors (and this equation is applicable when K < 11). With a different number of predictors and different effect size (small, medium and large), the estimated sample size can be calculated through Green's and Cohen's equations.

As mentioned at the beginning, deciding how large an effect size is expected to be prior to data collection can be difficult, especially when previous research in the area is limited. Instead of using Cohen's norms as a guideline, Maxwell (2000) provided an alternative method. In this method, sample size is estimated by specifying the correlations between the predictors and between the predictors and the dependent variable though the following equations:

$$f^2 = (R^2 - R^2_{(-j)})/(1-R^2) = (\rho^2_{y(xj\,.\,x(-j)})/(1-R^2) \qquad (5)$$

where $\rho^2_{y(xj\,.\,x(-j)}$ is the squared semi-partial correlation of variable j (or the R^2 difference between model with and without variable j). Thus, by substituting the above equation back into Cohen's equation (2) we have[2]

[2] This equation is slightly different from equation 3, in which this equation includes the number of predictors (K) but not in equation 3. The difference between these two equations will be very slight unless N is very small (Maxwell, 2000).

$$N = (\lambda (1-R^2) /(\rho^2 _{y(xj \,.\, x(-j))})) + K - 1 \quad (6)$$

Applying the property of fully exchangeable correlation structure[3] (Maxwell, 2000), the sample size estimation equation reduces to[4]:

$$R^2 = (K \rho^2 xy)/[1+(K-1) \rho xx] \quad (7)$$

and

$$\rho^2 _{y(xj \,.\, x(-j))} = (\rho^2 xy (1 - \rho xx)) / [(1 + (K-1) \rho xx) (1 + (K-2) \rho xx)] \quad (8)$$

where ρxx is the average correlation among predictors; ρxy is the average correlation between predictors and between each predictor and the dependent variable. K is the number of predictors in the proposed model. By substituting equation 7 and 8 back into equation 6 with λ value[5] = 7.85, Power = .80, alpha = .05, and N >120, we have:

$$N = \{7.85[1+(K-1) \rho xx - K \rho^2 xy][1+(K-2) \rho xx]/(\rho^2 xy(1- \rho xx)\} + K - 1 \quad (9)$$

Following equation 9, sample size can be estimated by specifying the number of predictors (K), average correlation between predictors (ρxx) and average correlation between predictors and dependent variable (ρxy).

This approach can provide more precise estimation of sample size when the effect size is relatively small and the necessary sample size is large (N>100). However, it may also overestimate the necessary sample size when the effect size is large because of the pre-chosen value of λ (7.85), which corresponds to an overall sample size[6] of 120.

---

[3] A Full exchangeable correlation structure implies that all predictors correlate equally between themselves and with the criterion variable. See Maxwell (2000) for more information on exchangeable correlation structure.

[4] For the details of the mathematical derivation, see the appendices of Maxwell's (2000) article.

[5] The adoption this value (7.85) implies that a moderate or large sample size is estimated.

[6] As pointed out by Cohen (1988), there is a problem when using equation (2) to estimate the sample size, because both λ and v (degree of freedom of the denominator in F ratio) are functions of N. As shown in the λ tables in Cohen's book (1988, p.448-455), larger λ values correspond to smaller v values, which are directly related to the estimated sample size. Cohen suggested using an "iterative" method to evaluate the accuracy of the estimated sample size, by comparing the estimated sample size

and the chosen v value. If the discrepancy between these two values is small, the estimated sample size is acceptable. The following equation (p.445) is also used for calculating the λ value regard to an exact N:

$$\lambda = \lambda_L - [(1/v_L - 1/v)/(1/v_L - 1/v_U)](\lambda_L - \lambda_U)$$

where v is the degrees of freedom of the denominator of the F-ratio; λ_L is the lower bound λ (from Cohen's table); λ_U is the upper bound λ value; v_L is the v value according to the lower bound λ, where v_U is the v value according to the upper bound λ. For example, suppose we want to estimate the necessary sample size for detecting the effect of a set predictors (u=5) with alpha = .05, power = .80, and $R^2_{Y.B}$ = .10 (the posited R^2 for the set of predictors). If we first set v = 120 and look up the corresponding λ from Cohen's tables based on the above information, we get λ = 13.3. Substituting this value and the above information back into equation (2) we get N = [13.3(1 - .10)]/.10 = 120. Thus, the v value based on this estimated N is equal to v = N - u - 1 = 120 - 5 - 1 = 114. The difference between the chosen v (120) and the estimated v (114) is not large enough to affect the planning of the study. For further examination, we can substitute the estimated v into the above equation, with the corresponding lower bound v (v_L = 60) with λ_L = 14.0 and higher bound v (v_H = 120). The calculated λ is equal to 13.3 (which is close to the one we used at the beginning), and we can use this value to recalculate the necessary sample size and compare it to the previous estimated sample size. The chosen v is acceptable if there is no obvious discrepancy between the two lambdas (or the two estimated sample sizes).

Thus, some adjustment of λ may be necessary before applying this equation, especially when estimating the required sample size of a study with a large effect size.

The above approaches focus on the calculation of necessary sample size in a regression model without any higher order terms (e.g. interaction or quadratic terms). By applying the concept of "gain in prediction of a set of predictors" (Cohen, 1988; 1990; Cohen & Cohen, 1983), we can estimate the necessary sample size for detecting a specific interaction effect using equation 3. The reduced model (-j) is the model without the interaction term (Aiken & West, 1991; Jaccard et al., 1990). One thing should be kept in mind when estimating the sample size for detecting the interaction term: the X_1 and X_2 product term will not be normally distributed, even though the two original variables, X_1 and X_2, are normally distributed. Hence, the estimated sample

size may be slightly smaller than the necessary sample size because the procedure overestimates the effect size and power (Aiken & West, 1991; Jaccard et al., 1990).

In the previous discussion, only the interaction between continuous variables has been discussed. In fact, there is one thing that should be kept in mind when estimating the sample size of the interaction term between a categorical variable and a continuous variable. When there is a categorical variable in the model, assuming equal size for each group can simplify the sample size estimation procedure because only one sample size has to be estimated. Additionally, this approach can also maximize the precision of the comparisons (Jaccard et al., 1990).

All the predictor variables in the above discussion are assumed to be error free, which is typically unrealistic. Dunlap and Kemery (1987) pointed out that to have reasonable levels of statistical power, the reliability of both variables in the product terms have to be .80 or greater. Thus, in addition to estimate the necessary sample size, researchers should also consider the reliability of their research tools in which the necessary sample size can be underestimated if their tools have low reliabilities.

Additionally, the estimated sample size for detecting the effect of the overall model (or a set of predictors) is not the same as for detecting the individual predictor (coefficient). In psychology, researchers are typically more interested in the effects of individual coefficients rather than the effect of the overall model. In fact, under the regression framework, increase the number of predictors will generally increase the power because of the reduction of the critical F-value as the number of predictors increase, given constant sample size and proportion of explained variance. Thus, the estimation of sample size for detecting a particular effect instead of the overall effect is relatively more preferable and meaningful in psychology. When determining the necessary sample size for detecting two or more particular effects, the estimated sample size for the smallest effect is usually used as the final sample size.

Conclusion

When estimating the sample size necessary to data hypothesized effect under the regression framework, existing rules of thumb are not recommended because they underestimate the necessary sample size. Power analysis provides a better approach for estimating sample size because it can control both desired alpha and power level. The main difficulty of this approach is to decide the magnitude of the effect size. Maxwell (2000) has provided an alternative method to estimate the sample size by specifying the correlations between the predictors and the correlations between predictors and the dependent variable. However, there is still

some minor difficulty of applying Maxwell's approach because of the unclear λ (non-centrality parameter, see footnote 5). Green (1991) provided a simple formula to calculate the corresponding λ, which is useful when the number of predictors is less than 11. Thus, the more efficient way to calculate the necessary sample size is by the Maxwell approach, with Green's formula to determine the λ value.

The sample size for detecting the interaction effect can be estimated through the "gain in prediction of a set predictors" approach, which is to determine the increase of R^2 after adding the interaction term in the model. In fact, if the distribution is symmetrical, the interaction is an “isolated term” (Aiken & West, 1991). Maxwell's approach can also be used if the correlation between the interaction term and other predictors, and the correlation between the interaction and the dependent variable can be determined.

Moreover, research tools should have substantial reliability (α > .80) for more accurate estimation of the necessary sample size. Low reliability of the research tools causes the underestimation of the necessary sample size. Sample size estimation for detecting individual effects is more preferable than for detecting overall effect.

Reference

Aiken, L. S., & West, S. G. (1991). Multiple Regression: Testing And Interpreting Interactions. Newbury Park, CA: Sage.

Cohen, J. (1988). Statistical Power Analysis For The Behavioral Sciences (2nd ed.). Hillsdale, NJ: LEA.

Cohen, J. (1992). Things I have learned (so far). American Psychologist, 45(12), 1304-1312.

Cohen, J., & Cohen, P. (1983). Applied Multiple Regression/Correlation Analysis For The Behavioral Sciences (2nd ed.). Hillsdale, NJ: LEA.

Dunlap, W. P., & Kemery, E. P. (1987). Failure to detect moderating effects: Is multicollinearity the problem? Psychological Bulletin, 102(3), 418-420.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? Multivariate Behavioral Research, 26(3), 499-510.

Jaccard, J., Turrisi, R., & Wan, C. K. (1990). Interaction Effects In Multiple Regression. Newbury Park, CA: Sage.

Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. Psychological Bulletin, 96(1), 201-210.

Maxwell, S. E. (2000). Sample size and multiple regression analysis. Psychological Method, 5(4), 434-458.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied Linear Statistical Models (4th Ed.). Chicago, IL: Irwin.

Stevens, J. (1992). Applied Multivariate Statistics For The Social Sciences (2nd Ed.). Hillsdale, NJ: LEA.

Topol, E. J., Califf, R. M., Van de Werf, F., Simoons, M., Hampton, J., Lee, K. L., White, H., Simes, J., & Armstrong, P. W. (1997). Perspectives on large scale cardiovascular clinical trials for the new millennium. Circulation, 95, 1072-1082.

News

1. Statistical Seminar

Professor Stephen Fienberg, Professor of Statistics and Center for Automated Learning and Discovery, Carnegie Mellon University, has kindly agreed to give a talk on "The Interplay between Research Innovation and Federal Statistical Practice" on 21 November 2001 at 6:30 p.m. in the Hong Kong Immigration Officers Mess. A Chinese style dinner will be held in honor of the speaker after the talk.

Contact Miss Clora Chan, General Secretary of the Society (Fax. no.: 2802 1101; Tel. No.: 2582 5041) for confirmation of attendance.

2. Promotion

Department of Statistics and Actuarial Science, The University of Hong Kong

Professor Li Wai Keung has been appointed as Chair Professor of Statistics dating back to July 2000.

Department of Statistics, The Chinese University of Hong Kong

Drs. Minggao Gu and Wai Yin Poon have promoted to Reader since October 2001.