香港統計學會

Hong Kong Statistical Society
c/o Department of Statistics & Actuarial Science,
The University of Hong Kong, Pokfulam Road, Hong Kong
http://www.hkss.org.hk

Bulletin
Volume 27 No. 1
June – September 2004

Editor's Foreword

First of all, I would like to thank all members of the Editorial Board. This is the first electronic distributed version of the Bulletin and I hope you like it. As usual, I would urge you members to contribute some articles for this Bulletin, or, you may inform us some interesting news in statistics.

In this issue, we have our President's Forum. Billy Li will show us the official statistics on forecasting the coming 30 years population of Hong Kong. Also, Frank Fong writes up the report on Statistical Project Competition for Secondary School Students. Lastly, we have an article that discusses the head counting method and survey for June forth and first of July. We expect that we shall receive a lot of comments and discussions on head counting.

L.K. Li

CONTENTS

(Vol. 27/No.1, June-September 2004)

President’s Forum

Professor Tony W.K. FUNG

Some of you may wonder why you have received this email. Yes, it is the first time that the Hong Kong Statistical Society is sending out her Bulletin to members electronically. There are two reasons in doing so.

First, the printing cost and postage of the Bulletin has become higher and higher (according to statistics, inflation has returned, and deflation has gone, hopefully forever!). It costed a few thousand dollars each time. The financial situation of the Society will be healthier if we can distribute the Bulletin electronically.

Second, the Society can communicate with her members more efficiently through email. I hope email will be used as a major way of communication in the future.

I would like to take the opportunity to introduce you the Council members of 2004-05, together with their email addresses:

Vice-President:
Mr. LEUNG Kwan-chi, Stephen (C&SD)
Email: skcleung@censtatd.gov.hk

General Secretary:
Mr. WONG Ho-fai, Howard (C&SD)
Email: hhfwong@censtatd.gov.hk

Treasurer:
Mr. TAM Chi-ho, Raymond (IVE)
Email: chtam@vtc.edu.hk

Membership Secretary:
Dr. YANG Hailiang (HKU)
Email: hlyang@hkusua.hku.hk

Publications Secretary:
Dr. LI Leong-kwan (Poly U)
Email: malblkli@polyu.edu.hk

Consultation Services Secretary:
Ms. LAW Ka-yee, Agnes (City U)
Email: msalaw@cityu.edu.hk

Programme Secretary:
Dr. CHAN Ping-shing, Ben (CUHK)
Email: benchan@cuhk.edu.hk

and I am FUNG Wing-kam, Tony (Email: wingfung@hku.hk) from HKU. You are welcome to contact us through email to express your views on Society’s matters.

Hong Kong Population Projections 2004–2033

Dr Billy LI
Census and Statistics Department

Introduction

Population projections of Hong Kong are compiled by the Census and Statistics Department (C&SD) at intervals of two to three years. Population projections are compiled after each population census or by-census; and an updating will be performed 2 or 3 years later as deemed necessary after analysis of the latest available data. The main use of population projections is to provide a common basis for Government programme planning and to facilitate business application in the private sector.

A new set of population projections, covering the period 2004–2033, with the mid-2003 population estimate as the base was released in June 2004. In compiling the projections, opportunity is taken to make use of the up-to-date information on fertility, mortality and movement patterns of the population which have emerged since the last set of population projections was produced. This paper provides a brief account of this new set of population projections, including the projection methodology, the assumptions used and the results.

Population Coverage

The Hong Kong Population is measured by the "Hong Kong Resident Population", which comprises "Usual Residents" and "Mobile Residents". In simple terms, "Usual Residents" are Residents who stay all the time or the majority of their time in Hong Kong and "Mobile Residents" are Residents who only spend the minority of their time in Hong Kong.

In more strict statistical definitions, "Usual Residents" refer to two categories of people: (1) Hong Kong Permanent Residents who have stayed in Hong Kong for at least three months during the six months before or for at least three months during the six months after the reference time-point, regardless of whether they are in Hong Kong or not at the reference time-point; and (2) Hong Kong Non-permanent Residents who are in Hong Kong at the reference time-point.

As for "Mobile Residents", they are Hong Kong Permanent Residents who have stayed in Hong Kong for at least one month but less than three months during the six months before or for at least one month but less than three months during the six months after the reference time-point, regardless of whether they are in Hong Kong or not at the reference time-point.

Projection Methodology

The standard method of compiling population projections, i.e. the ‘component method’, is adopted. Under this method, a population at a certain base year is brought forward by age and sex under separate projections of fertility, mortality and movement, year after year until the end of the projection period.

Based on the assumptions made on fertility, mortality and movement for each projection year (i.e. from mid-year of a calendar year to mid-year of the following calendar year), the size and age-sex structure of the population at the end of that projection year are worked out by applying the following algorithm:

(1) The projected forward survival ratios by age and sex are applied to the population at the beginning of a projection year to derive the surviving population at the end of that projection year.

(2) The projected age specific fertility rates are applied to the respective average numbers of women in individual childbearing ages 15–49 to obtain the total number of births. These births, after being divided into males and females by an average sex ratio at birth, are subjected to their respective projected forward survival ratios. The surviving population aged 0 by sex at the end of that projection year is then derived.

(3) The assumed net movement is added to the surviving population at the end of that projection year.

Projection Assumptions

Statistical studies using modelling methods as far as possible are made on the past trends and recent developments pertaining to the socio-economic conditions in Hong Kong to generate the fertility, mortality and movement assumptions. Where government policies are involved, it is taken that existing policies will continue to apply. For example, the existing policy of a daily quota of 150 One-way Permit Holders entering Hong Kong from the mainland of China is taken to apply throughout the projection period.

Fertility Assumptions

Hong Kong’s fertility has experienced a marked and continuous decline in the past two decades. Fertility is measured by the total fertility rate, which is the number of children born to 1 000 women **during their lifetime** if they were to pass through their childbearing ages 15–49 experiencing the age specific fertility rates prevailing in a given year. The total fertility rate decreased significantly over the last 20 years from 1 722 births per 1 000 women in 1983 to 941 in 2003.

The past trends of the age specific fertility rates (AFRs) provide the basis for formulating the fertility assumptions. Yet its process is not strictly a mechanical one that follows the extrapolated trends. Particular reference is made to the following two considerations:

(1) The fertility level in Hong Kong is currently very low and further significant decline is unlikely to occur.

(2) The experience of many low fertility economies (including Denmark, Sweden and Singapore) indicates that fertility could revert to a slightly higher level after a continuous decline.

The implications of the projected AFRs for some birth cohorts are checked to ensure that they are reasonable from the cohort perspective. In this connection, the average number of children ever born and the percentage of women having at least one child as implied by the fertility assumptions are assessed in the light of past experience in Hong Kong and the experience of other economies.

The total fertility rate is projected at 925 births per 1 000 women from 2004 to 2006, to increase gradually to 993 by 2011 and then to remain at that level for the rest of the projection period.

Mortality Assumptions

Hong Kong experienced a continuous decline in mortality during the past decades, with a corresponding increase in life expectancy. In 2003, the expectation of life at birth was 78.6 years (provisional) for males and 84.3 years (provisional) for females. Compared with other economies, Hong Kong enjoys a very low mortality.

The future mortality level of Hong Kong is projected using the Lee-Carter method. The method, describing the central death rate in the equation below, is fitted to the age-sex specific mortality rates of Hong Kong from 1976 to 2002.

$$\ln(m(x,t)) = a(x) + b(x)k(t) + e(x,t)$$

where the coefficient $a(x)$ describes the general shape of the age-sex profile, the coefficient $b(x)$ describes the deviation on the rate of increase or decrease from the general age-sex profile when $k(t)$ changes. The parameter $k(t)$ is the index on the mortality level at time t. The error term, $e(x,t)$, has zero mean and constant variance.

The Lee-Carter method is a parsimonious demographic model combined with statistical time series methods. It involves no subjective judgements and projections are based on persistent long-term trends, with probabilistic confidence intervals provided for the projected figures. This method usually outperforms other methods in long range projections when all methods based on heavy subjective judgements project very cautious improvements in mortality.

The following checks are made to ensure the appropriateness and consistency of the projected mortality rates:

(1) The excess of male mortality over female mortality of the projected mortality rates is checked against the past trend, in particular its most recent pattern; and

(2) The expectation of life at birth implied by the projected mortality rates is assessed with reference to the experience and the projections of other economies.

The projected age-sex specific mortality rates will decline, leading to a further rise in life expectancy from 2004 to 2033. Over the next 30 years, the expectation of life at birth is expected to increase from 78.6 years in 2003 to 82.5 years in 2033 for males and from 84.3 years in 2003 to 88 years in 2033 for females.

Movement Assumptions

The recent trends of the residency and mobility patterns of the Hong Kong population provide the basis for formulating the assumptions on the movement.

Diagram 1 shows the flows of Hong Kong residents. Separate assumptions are made in respect of the four components of net movement:

(1) Net flow of Hong Kong Permanent Residents into the Usual Residents category;

(2) Inflow of One-way Permit Holders;

(3) Net flow of Hong Kong Non-permanent Residents other than One-way Permit Holders (including net change in number of persons from visitors status to resident status); and

(4) Net flow of Mobile Residents.

Assumptions on fertility, mortality and movement entail a certain degree of uncertainty. These assumptions will hence be revised in a roll-forward manner in each round of population projections conducted at intervals of two to three years.

Projection Results

The Hong Kong Resident Population is projected to increase at an average annual rate of 0.7%, from 6.80 million in mid-2003 to 8.38 million in mid-2033. The number of Usual Residents is projected to increase from 6.62 million in mid-2003 to 8.02 million in mid-2033. In addition, the number of Mobile Residents is projected to increase from 185 300 in mid-2003 to 365 300 in mid-2033.

The population is expected to remain on an aging trend. The proportion of the population aged 65 and over is projected to rise markedly, from 11.7% in 2003 to 27% in 2033, though the rise would be gradual up to around 2015 (when the proportion would reach 14.3%) and would be at a much faster pace thereafter. Meanwhile, the proportion of the population aged under 15 would decrease gradually from 16% in 2003 to 11% by the end of the projection period.

The aging population trend will also be revealed by the increasing median age of the population, from 38 in 2003 to 49 in 2033.

The changes in the age structure of the projected population can also be seen from the overall dependency ratio. This is defined as the number of persons aged under 15 and those aged 65 and over per 1 000 persons aged 15–64. During the projection period, the ratio would drop from 378 in 2003 to 334 in 2011 and rise to 598 in 2033.

The sex ratio (i.e. the number of males per 1 000 females) of the population is projected to fall noticeably, from 939 in 2003 to 698 in 2033. There will be variations in the sex ratio by age group. In particular, the sex ratio for the age group 25–44 is expected to be much affected by the presence of foreign domestic helpers comprising mostly younger females. Also relevant is the continued entry of One-way Permit Holders in the coming years, many being Hong Kong men's wives living in the Mainland. Making reference to data which exclude foreign domestic helpers, the sex ratio of the population is higher, but still will come down from 997 in 2003 to 749 in 2033. The movement of mainland wives into Hong Kong and the fact that females live longer than males are the main reasons.

The change in population size involves different factors. Increase arises from birth and in-movement whereas decrease arises from death and out-movement.

The number of births is projected to remain at about 46 000 per annum. The number of deaths is projected to increase from about 37 000 per year currently to about 71 000 at the end of the projection period. The increase in the number of deaths is mainly attributable to the growing proportion of older persons in the population despite a longer life expectancy (which implies a lower mortality rate). It is projected that, by around 2016, the annual number of deaths will exceed the annual number of births.

Over the entire period from mid-2003 to mid-2033, the overall population is projected to increase by 1.58 million. There is a negative natural increase of 0.16 million (indicating 0.16 million more deaths than births) and a net in-movement (i.e. inflow less outflow) of 1.74 million.

Both One-way Permit Holders (OWPHs)

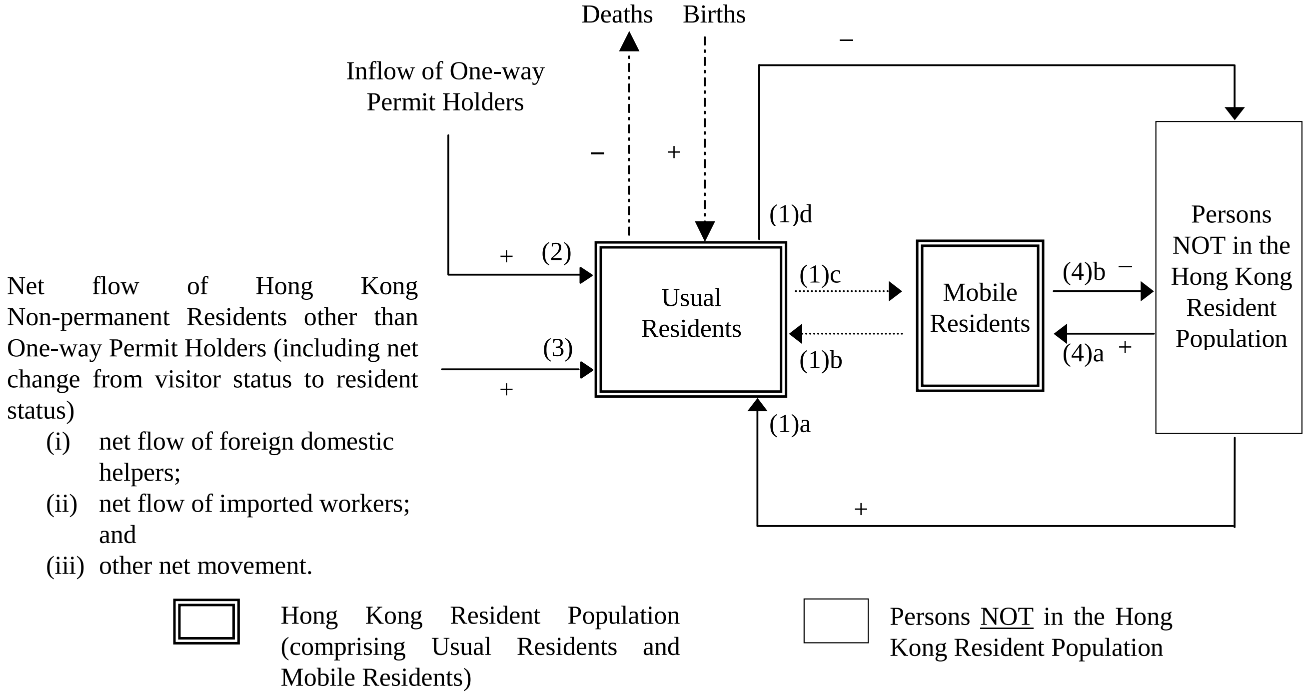
and births are important constituents of population growth. The ratios of the inflow of OWPHs and number of births to the overall population increase are 104% and 88% respectively. Besides, not counting the arrival of OWPHs mentioned above, there will be a net inflow of 100 000 persons, bearing a ratio of 6% to the overall population increase.

Deaths offset part of the population increase. The ratio of the number of deaths to the overall population increase is 98%.

Table 1 provides some summary statistics of the projected population.

Different data users make different uses of the population projections. They may require analysis and study to be conducted on different sectors of the population. For example, some studies concern only Usual Residents. Some would require analysis with foreign domestic helpers excluded. Different data users, especially different government departments, may need more detailed computations done on the basis of the basic projections, and the C&SD will render the required service.

Diagram 1: Flows of Hong Kong Residents



Notes: The numbers in brackets are in correspondence with the movement components described in the text.

(1) = Net flow of Hong Kong Permanent Residents into the Usual Residents category
= (1)a + (1)b – (1)c – (1)d

(4) = Net flow of Mobile Residents
= (1)c – (1)b + (4)a – (4)b

Table 1: Key Summary Statistics

A. Significant Characteristics of the Population for Selected Years

	Mid-2003 (Base)	Mid-2008	Mid-2013	Mid-2018	Mid-2023	Mid-2028	Mid-2033
Population	6 803 100	7 058 900	7 386 900	7 691 800	7 970 200	8 202 200	8 384 100
Usual Residents	6 617 800	6 843 600	7 141 600	7 416 500	7 664 900	7 866 900	8 018 800
Mobile Residents	185 300	215 300	245 300	275 300	305 300	335 300	365 300
Average annual growth rate over a 5-year period	0.8%	0.7%	0.9%	0.8%	0.7%	0.6%	0.4%
Sex ratio (males per 1 000 females)	939	890	842	800	763	730	698
Percentage of population							
Aged 0–14	16%	14%	12%	12%	12%	11%	11%
Aged 15–64	73%	74%	75%	72%	69%	65%	63%
Aged 65 and over	12%	12%	13%	16%	19%	24%	27%
Dependency ratio							
Child dependency ratio	216	182	163	163	168	171	171
Elderly dependency ratio	161	163	178	219	282	362	428
Overall dependency ratio	378	346	341	383	449	533	598
Median age	38	40	42	44	46	47	49

B. Components of Population Growth

Reference time-point (Mid-year)	Hong Kong Resident Population	Usual Residents	Mobile Residents	Compared with the previous reference time-point	From the previous reference time-point to this reference time-point		
				Population growth	Births	Deaths	Net movement (Inflow *less* outflow)
2003 [1]	6 803 100	6 617 800	185 300	16 100	48 700	35 200	2 600
2004	6 842 500	6 651 200	191 300	39 500	45 100	37 400	31 800
2005	6 888 800	6 691 500	197 300	46 300	45 900	37 900	38 300
2006	6 939 700	6 736 400	203 300	50 800	44 700	38 600	44 700
2007	6 996 200	6 786 900	209 300	56 500	44 700	39 400	51 200
2008	7 058 900	6 843 600	215 300	62 700	45 200	40 200	57 600
2009	7 123 000	6 901 700	221 300	64 100	45 800	41 000	59 300
2010	7 188 500	6 961 200	227 300	65 500	46 400	41 900	61 000
2011	7 255 400	7 022 100	233 300	66 900	47 000	42 900	62 800
2012	7 321 700	7 082 400	239 300	66 200	47 400	43 900	62 800
2013	7 386 900	7 141 600	245 300	65 200	47 500	45 000	62 800
2014	7 449 700	7 198 400	251 300	62 800	47 500	46 000	61 300
2015	7 511 600	7 254 300	257 300	61 900	47 700	47 000	61 300
2016	7 572 600	7 309 300	263 300	61 000	47 800	48 100	61 300
2017	7 632 700	7 363 300	269 300	60 100	47 900	49 100	61 300
2018	7 691 800	7 416 500	275 300	59 200	48 000	50 100	61 300
2019	7 750 000	7 468 600	281 300	58 100	48 000	51 100	61 300
2020	7 807 000	7 519 700	287 300	57 100	47 900	52 100	61 300
2021	7 862 800	7 569 500	293 300	55 800	47 800	53 200	61 300
2022	7 917 300	7 617 900	299 300	54 400	47 600	54 400	61 300
2023	7 970 200	7 664 900	305 300	52 900	47 300	55 600	61 300
2024	8 020 100	7 708 800	311 300	49 900	46 900	56 800	59 800
2025	8 068 300	7 751 000	317 300	48 200	46 500	58 000	59 800
2026	8 114 800	7 791 500	323 300	46 500	46 100	59 300	59 800
2027	8 159 500	7 830 200	329 300	44 700	45 600	60 700	59 800
2028	8 202 200	7 866 900	335 300	42 800	45 100	62 100	59 800
2029	8 243 000	7 901 700	341 300	40 700	44 600	63 600	59 800
2030	8 281 700	7 934 400	347 300	38 700	44 200	65 200	59 800
2031	8 318 200	7 964 900	353 300	36 600	43 700	66 900	59 800
2032	8 352 600	7 993 300	359 300	34 400	43 400	68 800	59 800
2033	8 384 100	8 018 800	365 300	31 500	43 100	71 300	59 800

Note: (1) Base year population estimates.

2003/04 Statistical Project Competition for Secondary School Students

Frank Fong
Organizing Committee of
2003/04 Statistical Project Competition

With the joint effort of wholehearted members of the Society and contributions from patrons and sponsor, the 2003/04 round of the Statistical Project Competition has successfully been concluded.

In this 18th round of the Competition, a total of 158 entry projects were received, with 97 for senior section and 61 for junior section. These projects were compiled by 717 students from 57 secondary schools. The spectrum of study was very wide, with the themes of the projects closely related to the livelihood and economy of Hong Kong. Topics of study covered some hot topics of current interest such as ageing of the population, effects of the atypical pneumonia outbreak, rise of the unemployment rate and development of tourism.

The adjudication panel, formed under the leadership of Ms Teresa NG of the City University of Hong Kong, comprised 28 academics from local tertiary institutions and statisticians working in the government. The panel scrutinized the project reports according to a set of criteria through various meetings and discussions. The more distinguished teams were then invited for an interview with the panel before concluding the results. Comments from the panel were recorded and subsequently provided to individual groups of participants after the competition.

The prize presentation ceremony of the Competition was held on 24th April 2004 at the City University of Hong Kong. Honourable patrons and other officiating guests hosted the ceremony and delivered enlightening speeches to the audience. At the ceremony, the first, second, third and three distinguished prizes for each of the junior and senior sections, as well as a special prize on the best thematic project among all, were awarded to the winning teams. Winners of the first prize for both the junior and senior sections were invited to present their winning projects and shared with the audience their experience in preparing their projects. Apart from cash prizes, trophies and statistical publications to the winning teams, all participants of the Competition were also awarded a certificate of appreciation as an appreciation of their participation.

Opportunity is taken to express, on behalf of the Organizing Committee, our sincere gratitude to Mr Eric K. C. LI, former Member of the Legislative Council; Mr. Chris WARDLAW, Deputy Secretary for Education and Manpower; Mr. Frederick W. H. HO, Commissioner for Census and Statistics for being the patrons of the Competition; to Mr. H. W. FUNG, Deputy Commissioner for Census and Statistics for being one of the officiating guests of the prize presentation ceremony; to the Hang Seng Bank for being our sole sponsor; and to the enthusiastic support from interested members, academics and colleagues in the government. The Competition would not have been a success without their joint effort and support.

The Statistical Project Competition aims at promoting a sense of civic awareness of secondary students and the proper use of statistics. The Competition provides a good opportunity for youngsters to learn to better use figures in understanding our community and to present them in a group work project. I have no doubt in believing that the coming rounds of the project will continuously be well-supported by secondary schools.

Counting at Rally and March

Dr. Jennifer CHAN, The University of Hong Kong
Dr. Leong-kwan LI, The Hong Kong Polytechnic University

Note: This article represents the view of the authors and not their respective universities

Part one: Static Mass Counting at Rally and March

Following the demonstration on 11th April, the turnover of the July 1st rally has raised a debate again. As statistics scholars, we would like to comment on this issue. We will first talk about methods of counting a static mass in part one of the articles. Then in part two, we will focus on counting a moving mass as well as reporting our estimation of the turnover of the July 1st rally this year.

We have to declare in the first place that we have no intention to get involved in any political dispute. As a matter of fact, how many people joined the rally is not necessarily related to the strength and quality of citizens' aspiration. In a democratic and open society, competition in figures on a political level usually appears in elections or other related system of decision making by the public, or at best extended to the result of public opinion surveys. Taking the turnover figure as the benchmark of the strength of opinion apparently will just appear in a relatively closed society. It is not our wish that Hong Kong is heading this way.

From another point of view, when society begins to base the decision making on figures from the public, different people, according to their own rationales, inevitably may interpret, or even distort, in different ways the figures and the social phenomenon behind.

For many years, after a large-scale rally has been held, the organizer would announce a figure that is considerably different from the statistics by the government. For scientific spirit to be rooted in our society, we believe, we need the courage to inquire into the truth and explode this myth. We take this our duty as professional statisticians.

After preliminary studies, we have every reason to believe that the difference in figures by the organizer and the government comes from different quantitative definitions and statistical methods. For example, does the number of marchers include people who just stayed for a short while? Does it include those who assembled at the starting point but did not set out with the crowd? If the organizer estimates the number of marchers based on the number of pamphlets distributed, while the police just estimates the number of people according to the area covered by the crowd at a certain period of time, there will certainly be a great difference in the figures obtained by them. Both parties estimate the number based on their actual needs and operation but may differ from the truth.

In order to inquire into this issue, a group of friends coming from different institutions, including Clement York-kee So and Winnie Yuen-fung Kwok from the Chinese University of Hong Kong, Robert Ting-yiu Chung from the University of Hong Kong as well as the two authors, prepared in May to count the turnover in the June 4 Vigil this year. After summarizing that experience, in late June we formed a team with volunteer students to count the number of marchers in this year's July 1st rally. We just have one objective – to seek the truth.

Review of various methods

First of all, concerning the statistical methods of counting the number of marchers, we classify them into two big categories according to their nature. The first category consists of taking bird's-eye view photos to estimate the density of the marchers occupying the roads, which is to be multiplied by the area of the site or the roads to estimate the total number of people occupying the site at a certain moment. This method is often used for a static mass. When applied on a moving mass such as a march, we can estimate the total number of marchers by having the number of people occupying the routes multiplied by the number of trips that can be made during the rally period.

Another method is counting at a fixed reference point, i.e. directly count the number of marchers passing through a certain point. Because of the limitation of personnel, usually this method will just be conducted by systematic sampling - counting the number of marchers during sampled intervals. Because this method cannot count the marchers who join or drop out midway, some studies will conduct another sample survey, so as to assess the percentage of the number of marchers passing through the point over the total number of marchers. For example, if a study shows that that point can only count 80% of the marchers, in order to estimate the total number of marchers, we have to multiply the number of people passing through that point by the adjusted basis of 1.25 (the quotient of 100 divided by 80).

Headcount at June 4th Candlelight Vigil

We tried to use different methods to calculate the number of people taking part in this year's June 4th Candlelight Vigil, including "first videotape and then count", "count all" and "sample headcount". All these methods belong to the first category of static mass counting. However, because the quality of the videotape image was not satisfactory, it was difficult to count with the tape. And the count-all method was not reliable either.

Figures from different groups of students differed a lot as they commented that it is difficult to count a large area of people with some people blocking the view of others. Thus this method was abandoned as well. Therefore, sample counting was the main method, calculation of which was based on density multiplied by area.

The standard of calculating density is an area of 3m X 3m. In the six football pitches in Victoria Park, we systematically sampled and counted the number of people contained in this standardized area at peak hours, obtaining an average of 25.3 people, i.e. the crowd density per square metre is 2.81 persons. On site we did an experiment as well, finding that in an area of 2m X 2m, 10 persons could be closely packed together, i.e. 2.5 persons per square metre, which is close to the figure above.

We also measured the surface area covered by the crowd in the six pitches at the June 4th Candlelight Vigil, the figure obtained being 14,917m^2. Thus, having the density multiplied by the area, the number of people obtained is 41,900 persons. But this is just the number of people inside the pitches, not including those outside. We also sent three groups of students to count the number of people outside, including those sitting at the stands and on the grass at the back, but excluding the police and other helpers. The figure obtained was about 9,400 persons. Thus, our estimation of the total number of participants by density and area, people joined or dropped out midway not included, was 51,000.

Static Mass Headcount at July 1st Rally

Concerning headcounts at the July 1 rally, the police, the Civil Human Rights Front and Mingpao Daily all based their calculation on the site area, multiplied by the density of the moving crowd at a certain moment and the number of trips that can be made during the rally period, to estimate the total number of marchers.

The police based on the density of the moving crowd photographed, which was multiplied by the time and speed required for the march, obtained a total number of marchers of approximately 200,000. The police have told the Civil Human Rights Front that if the routes are full of people, there should be 170,000 people. The Civil Human Right Front multiplied this density figure provided by the police by the number of trips calculated by them, plus some adjustment to reveal the number of people who left midway, and found a total number of 530,000.

Mingpao Daily used bird's-eye view photos to calculate the number of people with a computer, finding out that on average an area of 100m^2 can contain 100 persons. The paper then estimated that the area of the rally routes was 70,000 m^2 and thus put the number of people of the whole trip at 72,000. The rally period was set at 5.5 hours and each trip took 1.5 hours, i.e. 3.67 trips altogether. Finally it estimated the number of marchers at 264,000.

The accuracy of these calculation methods depends on three factors - area, density and time frame. There is not much controversy over area, but density is a big problem. Based on the figure by the police - 170,000, the crowd density per square metre is 2.36 persons. But if it is estimated by Mingpao's figure, the crowd density is just 1 person. Our research team, on the day of the July 1st rally, found that the flow rate of the march was approximately 700 persons/minute. Assume that the movement speed of people was 50m/min and the average width of the roads was 20m, the density per square metre obtained would be 0.7 person. If the flow rate of crowd was 35m per minutes, the density per square metre obtained would be 1 person.

From the simple experiments above and the actual situation at the June 4 Candlelight Vigil, we know that a crowd density per square metre of 2.5 persons could just happen in an environment as crowded as somewhere in Causeway Bay where two marches joined. Normally the average crowd density of a moving mass should be lower than this, and 1 person/m^2 should be more reasonable which is close to Mingpao's figure.

The third factor is time frame. If the rally period is set at 5.5 hours and each journey took 1.5 hours as used by Mingpao, then there should be 2.7 trips counting only 4 hours instead of 3.7 trips because the density reached 72,000 over the whole area only after the first trip. Thus, if Mingpao's figure of 72,000 is adopted and multiplied by 2.7, the result will be 194,000 persons instead of 264,000 persons. The difference of the numbers of marchers is significant. How these figures are different from the result obtained by our fixed-point headcount will be discussed part two.

Part two: Moving Mass Headcount at July 1st Rally

We have introduced methods to headcount the static mass at the June 4th Candlelight Vigil and the July 1st rally. We mentioned the relationship between area, density and length of time of rally. Then in this part, we will focus on counting a moving mass and the turnover result obtained by this method.

Moving Mass Estimation at July 1st Rally

Due to the limitations of personnel, we adopted a fixed-point counting method to calculate the number of participants in the July 1st rally. We took the pedestrian bridge at the junction of Hennessey Road and Arsenal Street as the headcount station. Six volunteer students each counted the flow of people on one traffic lane, in the manner of systematically counting for one minute in every five minutes. The number of people counted in that minute was multiplied by five to represent the number of people marching through in that five minutes. In this way, the number of people passing through that headcount station was about 149,000 persons. To evaluate the adjusted basis, the research team, during July 2nd – 11th, randomly sampled, and successfully telephone interviewed 3, 512 citizens aged 18 or above. The mean response rate was 63.8%, in which 231 persons had joined the July 1st rally. They answered the following question: Have you passed under the pedestrian bridge at the junction of Hennessy Road and Arsenal Street, in the direction from Wanchai to Admiralty? Amongst those citizens, 77.4% said they had passed under that bridge at Hennessy Road, so the adjusted basis is 1.29 (the quotient of 100 divided by 77.4). Having the adjusted basis multiplied by the number counted - 149,000, and including a pure sampling error of 6%, the total number of marchers will be between 165,000 and 228,000.

We conducted a post-hoc sample survey because if the survey were conducted on that day, the interviewees could not foresee from where they would leave. Even if the survey were conducted at the destination, only those who joined midway and marched to the destination could be counted, while those who left midway could not be included. Anther research team led by Dr. Paul S. F. Yip, Senior Lecturer of the University of Hong Kong department of statistics and actuarial science had headcount stations at Causeway Bay and Admiralty respectively, and also conducted a sample survey at the bridge near the Pacific Place at Admiralty so as to calculate the adjusted basis. However, their adjusted basis cannot include those who joined and left between the headcount stations at Causeway Bay and Admiralty, and those who joined after the Admiralty headcount station. They added 10% as an adjustment for the total number of participants. Acceptable though, taking 10% as the adjustment basis is considerably subjective.

Another survey into the number of marchers on July 1st led by Dr. John Bacon Shore of Social Sciences Research Centre of the Faculty of Social Sciences, The University of Hong Kong estimated that 105,000 to 120,000 people marched along Queensway on that day. They set a video camera on the pedestrian bridge over Queensway from Lippo Centre, filmed all marchers and examined 4 seconds of video from every 30 seconds of the video to obtain a sample mean of about 32 people for 4 seconds. Then multiplying the hourly rate of about 30,000 people to the march period of 4 hours, they obtained the estimate without making any allowance for marchers joining or leaving midway away from their fixed-point. Our experience of counting people in the through the "slow down" video tape in the July 1st rally last year shows that counting this

930 samples of 4 second tape altogether is difficult and lots of personal judgments have to be made on whether including marchers walking along edges of traffic lane or time frame. Systematic errors may be aroused in the process.

The limitations of different ways of estimation

Actually all the estimation methods discussed have their own limitations. The estimation of the total number of marchers by the density of people occupying roads is a bit coarse, because it is based on lots of assumptions. Thus according to different assumptions, the numbers estimated can vary a great deal. In addition to the possible inaccuracy in estimating the density of the moving march, the density can also vary a lot due to the situation of the roads, different intervals of time and locations. And the estimation by the number of trips made will also be affected by the weather on that day and the age of the marchers. The ideal method to deal with it is to collect more data samples at different time intervals and locations, so as to have a more accurate estimation. However, this will make the calculation very complicated.

Even so, this estimation method cannot calculate the participants joining or dropping out midway. On July 1st at a very high temperature of 35℃, assuming that marchers who were with elderly people or children could just finish one-third of the trip, then the number of trips of the whole rally period should be increased because to these participants, the actual trip was reduced. Indeed how many marchers were of this kind? On average how long have they walked? All these factors would affect the accuracy of the estimation which is hard to evaluate because it is not easy to contact them.

Even if a fixed-point headcount method is adopted, how to count the marchers joining or leaving midway is still a difficulty. Theoretically, a post-hoc sample survey can reach different marchers. However, because the sampled targets were all citizens aged 18 or over who have joined the march, it takes time and resources to accumulate a sample of a considerable size. Readers may also be worried that the interviewees may have misreported, such as those who did not join would say they have participated in. However, even so, such a bias apparently is not directional as to whether they have passed the fixed-point or not. Yet because the survey only included participants who were 18 or above, there is no way to reveal the situation of the marchers aged under 18 who are expected to be less likely to pass the fixed-point.

Readers may have already noticed that, not just the estimations by the density of the moving crowd differ a lot, there are also great differences amongst direct headcounts at locations close to each other. For example, the headcount station of our research team at the bridge in Arsenal Street, close to Admiralty, recorded 149,000 marchers past. Yet the other research team led by Paul Yip at the bridge near the Pacific Place in Admiralty found 110,000 persons. The two headcount stations were not far away yet there is a difference of 40 thousand people. Similarly, Yip's research team at Radio City in Causeway Bay recorded 120,000 marchers. Singtao Daily News reported that Taipo Tertiary Student Association Research Team also had their headcount in Causeway Bay. They opted for videotaping one minute of the march in every 15 minutes, and then played the video and counted the turnover. They found that the number of marchers passing through that headcount station was 90,000. The two figures also have a difference of 30,000. If it is not because of inaccurate counting, then the number of people joining or leaving midway is indeed considerable. Say the figure by Yip at the bridge near the Pacific Place in Admiralty may not include a group of people who entered Pacific Place before the bridge to seek shelter. Because of the extreme hotness, I believe such number would be substantial.

At last, the authors cannot emphasize enough that when facing numerous estimation results, we should not just focus on an estimate, but have to deeply understand the reliability and limitations of different estimation methods, such as whether the assumption is reasonable, etc. Amongst the research estimations mentioned in this article, only our research team and the teams led by Yip and Bacon Shore [????]have reported error bound, allowing the readers to evaluate the reliability of the estimation. If the error bound is large, the reliability will be low. For example, if the error bound is 100,000 persons; relative to an estimation of 200,000, the number lies between 100,000 and 300,000, the upper and lower limits of which are considerably different. Moreover, the error bound reported by our research team just reflects the error coming from the adjusted basis. Systematic sampling error, i.e. estimating the total number of marchers in five minutes by the number of marchers in one minute, as well as other non-sampling errors all cannot be reflected.

How can the number of marchers be more accurately and effectively studied is a very interesting research topic. However, the authors think that it is much more important to understand the public's opinions rather than debate on the crowd size. Besides, citizens should fairly and objectively evaluate different statistical results. Do not just look at the figures because every statistical method has its own limitations and cannot be fully relied on.

Appendix

A. *Estimate of the total number of marchers passing the headcount station and its standard error based on systematic sample of headcounts in each lane*

Systematic samples of headcounts $x_{i,j}$ in the i-th 5-minute interval in traffic lane j are collected in each of the 6 traffic lanes. Let X_j^* denote the number of marchers passing the headcount station in traffic lane j during the July 1st rally, X^* denote the total number of marchers passing the headcount station during the July 1st rally and n_j denote the number of 1-minute intervals during which headcounts are observed or predicted. Then, neglecting the finite population correction factor, estimate of total number of marchers passing the headcount station during the July 1st rally and its standard error estimate using the *successive difference* method are

$$X^* = \sum_j X_j^* = \sum_j \sum_i 5x_{i,j} = 149{,}507$$

and

$$se(X^*) = \left(\sum_j \mathrm{var}(X_j^*) \right)^{\frac{1}{2}} = \left(\sum_j \frac{s_{d,j}^2}{2n_j} \right)^{\frac{1}{2}}$$

$$= \left[\sum_j \frac{\frac{1}{n_j - 1} \sum_i (x_{i,j} - x_{i-1,j})^2}{2n_j} \right]^{\frac{1}{2}} = 1{,}772.29 \, .$$

Hence the 95% confidence interval for the total number of marchers X^* passing the headcount station is (145,512, 152,602). Figures are summarized in the table below.

	Lane 1	Lane 2	Lane 3	Lane 4	Lane 5	Lane 6	Total
X_j^*	27,480	27,465	26,035	25,615	25,168	17,294	149,057
$\mathrm{var}(X_j^*)$	533,147	430,704	867,141	792,013	299,051	218,944	3,140,999

B. *Estimate of ratio of marchers passing the headcount station out of all marchers and its standard error estimate*

Let x_i and y_i denote the indicator variables of whether citizen i in the sample had joined the July 1st rally and whether he/she had passed the headcount station on that day respectively. It is known that the population size of citizens aged 18 or above, N=5,620,000 (population estimate in 2001), the sample size of citizens aged 18 or above, n=3512, the number of citizens aged 18 or above in the sample who had joined the July 1st rally, $\sum_i x_i = \sum_i x_i^2 = 231$ and the number of citizens aged 18 or above in the sample who had passed the headcount station during the July 1st rally, $\sum_i y_i = \sum_i y_i^2 = \sum_i x_i y_i = 179$. Then the proportion of citizens aged 18 or above who had joined the July 1st rally is

$$\overline{X} = \frac{X}{N} = \frac{192{,}358}{5{,}620{,}000} = 0.034 \quad \text{(refer to Appendix C for } X = 192{,}358\text{)}$$

and the estimate of ratio of marchers passing the headcount station out of all marchers in the July 1st rally is

$$r = \frac{\sum_i y_i}{\sum_i x_i} = \frac{179}{231} = 0.775.$$

Neglecting the finite population correction factor, the standard error of the ratio estimate r is

$$s.e. = \left(\frac{s_r^2}{n\bar{X}^2}\right)^{\frac{1}{2}} = \left(\frac{(\sum_i y_i^2 - 2r\sum_i x_i y_i + r^2 \sum_i x_i^2)}{n(n-1)\bar{X}^2}\right)^{\frac{1}{2}}$$

$$= \left(\frac{179 - 2\times 0.775\times 179 + 0.775^2 \times 231}{3512\times(3512-1)\times 0.034^2}\right)^{\frac{1}{2}} = 0.0528$$

Hence a 95% confidence interval for the ratio estimate r is (0.67, 0.88).

C. *Estimate of the total number of marcher in the July 1st rally and its standard error estimate*

Combining the two set of results, estimate of the total number of marchers X in the July 1st rally and its 95% confidence interval are

$$X = \frac{X^*}{r} = \frac{149{,}057}{0.775} = 192{,}000$$

and

$$(\frac{145{,}512}{0.88}, \frac{152{,}602}{0.67}) = (165{,}000, 228{,}000)$$

correct to nearest thousand.

News

University of Hong Kong

The Department of Statistics and Actuarial Science has played the role of host in two international conferences in the summer, one entitled "Insurance Mathematics, Ruin Theory and Monte Carlo Methods," which was co-organized with the Institute of Mathematical Research of HKU, June 28-30, and the other entitled "Threshold Models and New Developments in Time Series," which was held in July 12-14, in honour of Professor Howell Tong's 60th birthday. Details and pictures are available in the Department's website: http://www.hku.hk/statistics/.

The University has accepted Professor Howell Tong's resignation, effective September 22, 2004. The Department of Statistics and Actuarial Science thanks Prof. Tong's important contribution during his term as the structural Chair of Statistics and wishes him all the best in his career in the London School of Economics and Political Science, London, the U.K.

There are four academic posts in the Department of Statistics and Actuarial Science being advertised. Details are given in the Department's website: http://www.hku.hk/statistics/.

Professor Tony W.K. Fung was elected Council Member, International Statistical Institute (ISI), 2005-2009.