Living with Statistics

2012 Edition





香港特別行政區 政府統計處
Census and Statistics Department
Hong Kong Special Administrative Region

5 years+ 同心展關懷
caringorganisation

Living with Statistics

2012 Edition

Enquiries about this teaching kit can be directed to :
General Statistics Section 2(A)
Census and Statistics Department
Address : 21/F Wanchai Tower, 12 Harbour Road, Wan Chai, Hong Kong.
Tel. : (852) 2582 5054 Fax : (852) 2119 0161
E-mail : train@censtatd.gov.hk

Website of the Census and Statistics Department
www.censtatd.gov.hk

Published in June 2012

This teaching kit is available in download version only.
.

Contents

		Page
Introduction		
Chapter 1	**Population size and growth**	1:1 – 1:24
Chapter 2	**Population structure**	2:1 – 2:27
Chapter 3	**Impact of population growth on society**	3:1 – 3:16
Chapter 4	**Survey methods**	4:1 – 4:15
Chapter 5	**Uses and misuses of statistics**	5:1 – 5:22
Chapter 6	**Rate, ratio, proportion and percentage**	6:1 – 6:8
Chapter 7	**Measures of central tendency**	7:1 – 7:12
Chapter 8	**Measures of dispersion**	8:1 – 8:9
Chapter 9	**Official statistics**	9:1 – 9:27
Solutions to exercises		S:1 – S:15

Introduction

As part of the programme for promoting statistical literacy, the Census and Statistics Department (C&SD) has released five editions of the teaching kit entitled “Living with Statistics” in hardcopy format since 1990. The teaching kit mainly aims at providing secondary school teachers and students with a convenient means of accessing reference materials on commonly used statistical methods and official statistics in Hong Kong. It also touches upon possible pitfalls that should be avoided in applying statistics. This online version of the teaching kit is launched to replace the previous hardcopy version.

With a better understanding of official statistics and proper statistical methods, students will be able to enhance their awareness and appreciation of statistical information in interpreting the social and economic situations of our society in an objective and effective manner.

The following symbols are used throughout the teaching kit :

#	Provisional figures
@	Figures are subject to revision later on

1 Population size and growth

Introduction

Population estimates are widely used by the government for planning and policy formulation. The private sector and the academia also use population estimates for business or research purposes.

This chapter describes how the size of a population and its change can be measured. Births, deaths and net movement(1) are identified as the three factors that affect population growth. Various statistical indicators of their trends are introduced. Finally, a few general characteristics of population growth are outlined in relation to these determining factors.

Approaches of compiling population estimates

Basically, there are two enumeration approaches to count the population of a country/territory. They are the "de jure" enumeration approach and the "de facto" enumeration approach.

Under a "de jure" concept, all persons who usually live in a country/territory at a particular reference time-point (usually taken as the middle of a year) will be counted as the population of the country/territory.

Under a "de facto" concept, the population includes all persons who are in the country/territory at the reference time-point. That is, this method is equivalent to taking a "snapshot" of the population at a reference time-point.

In practice, mixtures of the two approaches can be used.

Previous method of compiling population estimates in Hong Kong

The "extended de facto" method was used for compiling the series of population estimates for Hong Kong up to 1995. Under the "extended de facto" approach, the Hong Kong population covers all persons who are physically in Hong Kong at the reference time-point including Hong Kong Permanent Residents and Hong Kong Non-permanent Residents as well as visitors. "Extended" relates to the fact that for a Hong Kong Permanent Resident, he/she will still be counted as part of the

Note :

(1) In the case of Hong Kong, movements from Hong Kong to overseas countries, the mainland of China or Macao for living, studying or working; and vice versa are all regarded as movements.

Hong Kong population if, at the reference time-point, he/she is not in Hong Kong but temporarily in the mainland of China or Macao.

The application of the "extended" compilation method in the past was intended to avoid fluctuations in the population estimates around major public holidays when there were enormous temporary movements of people between Hong Kong and the Mainland/Macao.

Revision to the method of compiling population estimates in Hong Kong

As the social and economic ties between Hong Kong and the Mainland have become closer and closer in the past few years, the residency and mobility patterns of Hong Kong people are undergoing changes. The living and working boundaries of many people cover Hong Kong, the Mainland/Macao or even overseas. The trend for Hong Kong Permanent Residents to stay in the Mainland/Macao on a long-term basis for reasons of retirement, work or marriage is on the increase. Also, a considerable number of Hong Kong residents are staying overseas for work, tour or study at the reference time-point. As a result, using the "extended de facto" method for compiling population estimates of Hong Kong has gradually become less appropriate. Besides, according to the definition of "de facto" population, visitors to Hong Kong are included but, strictly speaking, they are not part of the Hong Kong population.

After carefully considering factors including the residency and mobility patterns of the Hong Kong people, international trends, statistical conceptualisation of statistical indicators and development of information technology, the Census and Statistics Department (C&SD) has, since August 2000, replaced the "extended de facto" method used previously by the "resident population" method for compiling population estimates(2). Since the required data are not available for earlier periods, revised population figures compiled based on the new approach are backdated to 1996 only.

Replacing the "extended de facto" method by the "resident population" method is a technical revision. The resulting change in the population estimates is not large. As the revised series will better reflect the actual situation of the Hong Kong population and provide richer information on the residency and mobility patterns, they will contribute better to government's forward planning.

Note :

(2) In the early days, many countries and territories applied the "de facto" method for compiling population estimates. The main reason is that the compilation procedure is simple and the required data are easier to obtain. With statistical developments and advancements, advanced countries and territories have gradually changed over to using the "resident population" method.

The "resident population" method – the current method of compiling population estimates in Hong Kong

The "resident population" method for compiling population estimates was introduced and announced in August 2000 to replace the previously used "extended de facto" method.

"Resident population" is a clear-cut concept in international statistical standard but the practical definitions adopted vary from place to place, as the residency and mobility patterns unique to each place need to be given adequate consideration. International statistical organisations have pointed out in particular that owing to business and social development, the mobility of residents of certain countries/territories is rather high. In handling the population statistics of these countries/territories, the authorities concerned should consider the situation in depth. In the case of Hong Kong, studies have shown that the "resident population" of Hong Kong (which is referred to as the "Hong Kong Resident Population") should be defined to include "Usual Residents" and "Mobile Residents".

"Usual Residents" include 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.

Hong Kong Non-permanent Residents are grouped under "Usual Residents". This is because for the duration that they hold the status of "Non-permanent Resident", they can be expected to be usually staying in Hong Kong.

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.

The amount of time of stay in Hong Kong of "Mobile Residents" is less than that of the "Usual Residents". Nevertheless, the "Mobile Residents" have a close link with Hong Kong and most probably they have a regular residence in Hong Kong and utilise much of Hong Kong's facilities and services. In this regard, they should be considered as part of the Hong Kong population.

Under the "resident population" approach, visitors to Hong Kong are not included in the Hong Kong population.

For details of the compilation methodology of population estimates, please see the feature article entitled "Compiling Population Estimates of Hong Kong", which was published in the February 2002 issue of *Hong Kong Monthly Digest of Statistics* compiled by C&SD.

Population data system

To furnish users with the latest information on the position of the Hong Kong population, it is the standing practice of C&SD to update and release the population estimates every half-year. The updated estimates relate to the mid-year and year-end positions.

In Hong Kong, the compilation of population estimates is supported by a comprehensive population data system. The main component of the system is population censuses and by-censuses which provide benchmarking population data, while also serving as the prime sources of data for small areas and population sub-groups. Apart from population censuses and by-censuses, the population data system also covers sample surveys of smaller scale and statistical data compiled based on information from administrative systems such as birth, death and passenger movement records. Taking all information together, they provide a population statistical database for compiling various types of population figures.

Population censuses and by-censuses taken in Hong Kong

It is established practice in Hong Kong to conduct a population census every ten years and a population by-census in the middle of the intercensal period. Population censuses were conducted in 1961, 1971, 1981, 1991, 2001 and 2011 and population by-censuses in 1966, 1976, 1986, 1996 and 2006. Following this practice, the next population by-census will be conducted in 2016.

The aim of conducting population censuses/by-censuses is to obtain up-to-date benchmark information on the socio-economic characteristics of the population and on its geographical distribution. They provide benchmark data for studying the direction and trend of population changes. The data are key inputs for making projections concerning population, household, labour force and employment. Population censuses/by-censuses differ from other general household sample surveys in their sizable scale which enable them to provide statistics of high precision, even for population sub-groups and small geographical areas. Such information is vital to the Government for planning and policy formulation and important to the private sector for business and research purposes.

For both the 1961 and 1971 Population Censuses, all persons in the population were counted and enquired of their socio-economic characteristics. In the 1981, 1991, 2001 and 2011 Population Censuses, there was a complete headcount of all persons by age and sex whilst the detailed characteristics of households and persons were collected from a large sample. Through the use of appropriate computation methods, statistics on the size and characteristics of the population can be compiled by combining the data from the simple enumeration and the detailed enquiry.

A by-census differs from a census in not having a complete headcount of the population but simply focusing enquiry on the detailed characteristics of a large sample of the population. The size and characteristics of the entire population are inferred from the sample results in accordance with appropriate statistical theory.

The 2011 Population Census

Similar to the past censuses, the 2011 Population Census (11C) comprises a complete but simple enumeration of the whole population and a sample enquiry on the detailed characteristics of the population. Yet the sampling fraction of the detailed enquiry was reduced from 1/7 in the 2001 Population Census (01C) to 1/10 in the 11C. This follows the adoption of a 1/10 sampling fraction for the detailed enquiry in the 2006 Population By-census (06BC).

Like the 01C, the 11C covers the Hong Kong Resident Population under the "resident population" approach. The "resident population" approach has been adopted to compile the population estimates of Hong Kong since August 2000, so as to take into account the changing residency and mobility patterns of the Hong Kong population in recent years (see also the previous section on the "resident population method").

A multi-modal data collection approach was adopted for the first time in the 11C. In addition to the traditional "face-to-face interview" method, data were also collected from households through postal return and electronic questionnaire on the Internet in the 11C.

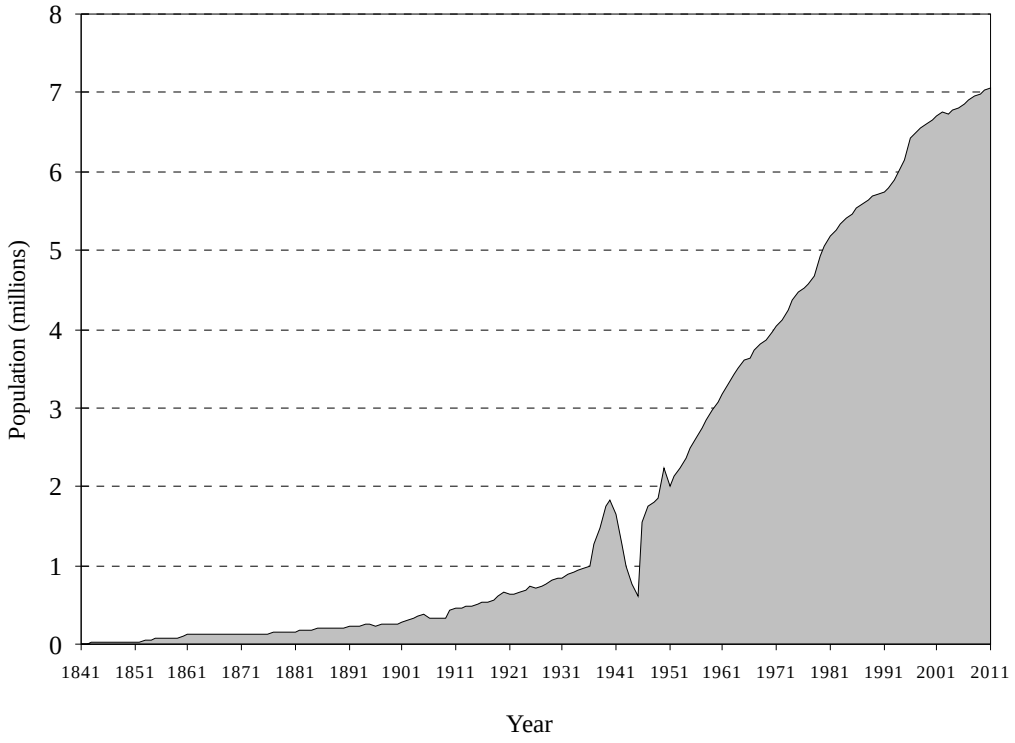
Population size

In mid-2011, the population of Hong Kong was 7.07 million. In 1841, the population was only 7 500. During the past century or so, the population experienced a substantial increase.

Amidst the continuously increasing trend (in sharp contrast to the trend occurred during the Second World War), the rate of change in the population was not always the same.

The 2011 Population Census conducted from June to August 2011 provides a benchmark for revising the population figures compiled since the 2006 Population By-census. As a result, population-related figures from 2007 to 2010 have been revised accordingly.

Chart 1.1 Mid-year population of Hong Kong, 1841–2011



Growth rate of the population

To measure the speed of population change, the growth rate can be calculated by dividing the increase in population during a period by the population size at the beginning of that period.

Thus, given that the population increased from 7 024 200 at mid-2010 to 7 071 600 at mid-2011, the population grew at a rate of

$$\frac{7\,071\,600 - 7\,024\,200}{7\,024\,200} \times 100\% = \underline{\underline{0.7\%}}$$

during the period from mid-2010 to mid-2011.

If the population decreased in the period, the growth rate would have been a negative number.

Compound average growth rate of the population

For a span of time over a number of years, it is common practice to measure the average annual growth rate on a compound basis.

The population was 7.07 million at mid-2011 and 6.86 million at mid-2006 (i.e. five years ago). If the unknown compound average growth rate is r per cent per annum during the 5-year period from mid-2006 to mid-2011, then

$$\begin{bmatrix}\text{Population at}\\ \text{mid-2011}\end{bmatrix} = \begin{bmatrix}\text{Population at}\\ \text{mid-2006}\end{bmatrix} \times \left(1 + \frac{r}{100}\right)^5$$

or $$7.07 = 6.86 \times \left(1 + \frac{r}{100}\right)^5$$

hence $$r = \left(\sqrt[5]{\frac{7.07}{6.86}} - 1\right) \times 100 = \underline{\underline{0.6}}$$

That is, the compound average growth rate of the population was 0.6% per annum during the 5-year period from mid-2006 to mid-2011.

Three factors affecting population growth

There are three factors that affect population growth, namely births, deaths and net movements.

At different times, some factors may have greater effects than others. To study their contribution to population growth, it is useful to express the change in population size in terms of its components :

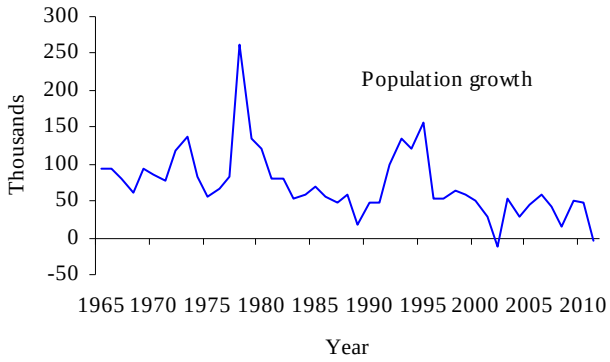
$P_1 - P_0 = (B - D) + (E - L)$

where P_1 = population at end of the period
P_0 = population at beginning of the period
B = live births during the period
D = deaths during the period
(B – D) = natural increase during the period
E = number of entrants during the period
L = number of leavers during the period
(E – L) = net movements during the period

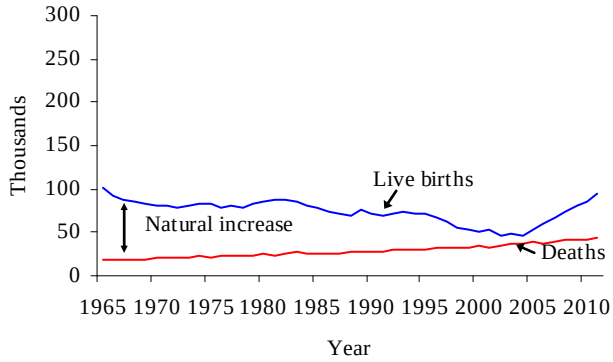
Chart 1.2 Annual change in population of Hong Kong, mid-1965 to mid-2011[3]

(a) Annual change in population size

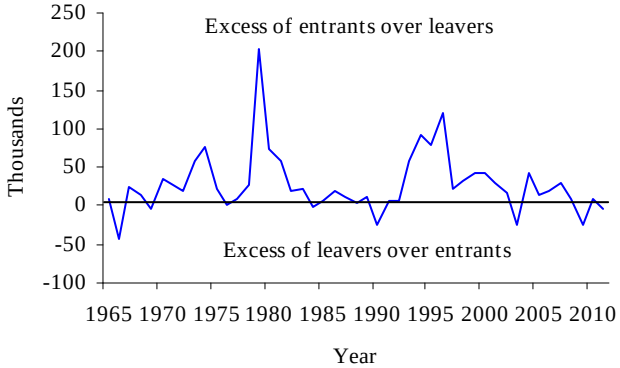
[= (b) Natural increase + (c) Net movement]



(b) Natural increase



(c) Net movement



Note :

(3) Population figures from 1996 onwards are compiled based on the "resident population" method adopted since August 2000 and are broadly comparable with figures for 1995 and before. Notwithstanding this, the population figure for 1996 compiled based on the old method (i.e. 6 311 000) has been used in calculating the population growth and net movement from 1995 to 1996.

Stock-flow relationship

To understand the equation shown in the above section, take the size of population at a point of time as a pool of water (i.e. the population "stock"), its level being raised by adding water from the tap (i.e. inflows of births and entrants) and lowered by draining water away through the plughole (i.e. outflows of deaths and leavers). Thus, the population size measures the stock at a particular point of time, whereas changes in all the components, or "flows", contribute to changes in the stock size during an interval of time.

Crude birth rate

The childbearing trend in the population is measured by the birth rate, which is expressed in terms of the number of births per 1 000 population.

The crude birth rate is calculated by dividing the number of known "live birth(4) " born in a calendar year by the average population size during the calendar year. Usually the mid-year population is taken as the average population size.

Thus, in 2011, with 95 400# live births and a mid-year population of 7 071 600, the crude birth rate is equal to

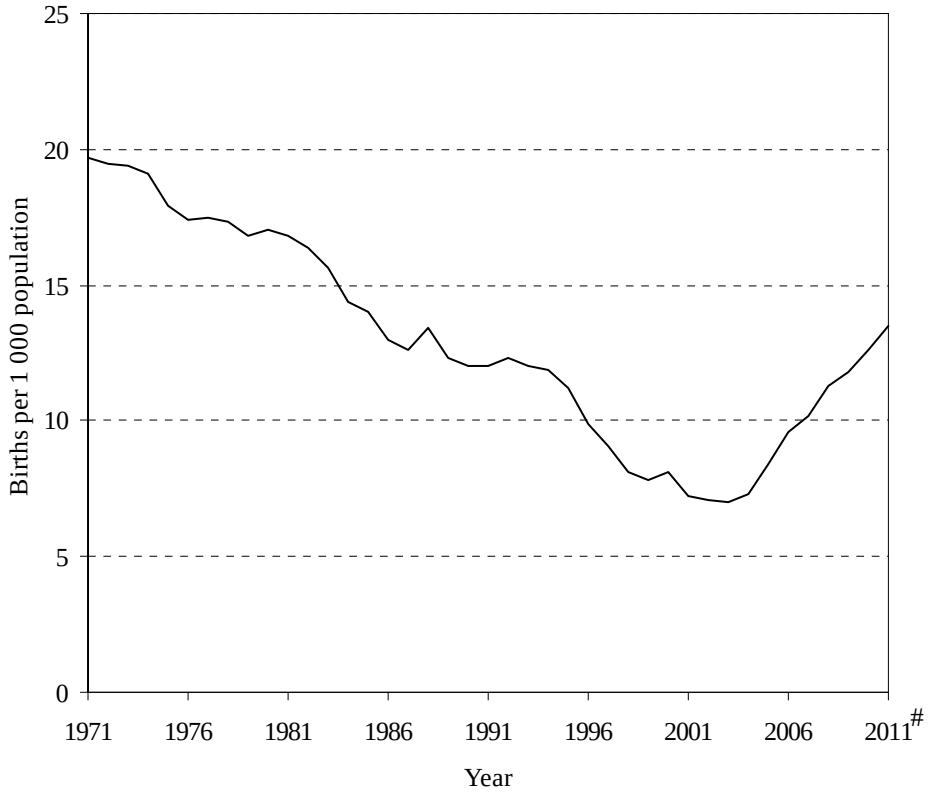
$$\frac{95\,400^{\#}}{7\,071\,600} \times 1\,000 = \underline{\underline{13.5^{\#}}}$$

or 13.5# per 1 000 population.

Note :

(4) Live births are defined as babies born with evidence of life, such as respiration, movement of voluntary muscles or heartbeat after complete expulsion or extraction.

Chart 1.3 Crude birth rate, 1971–2011



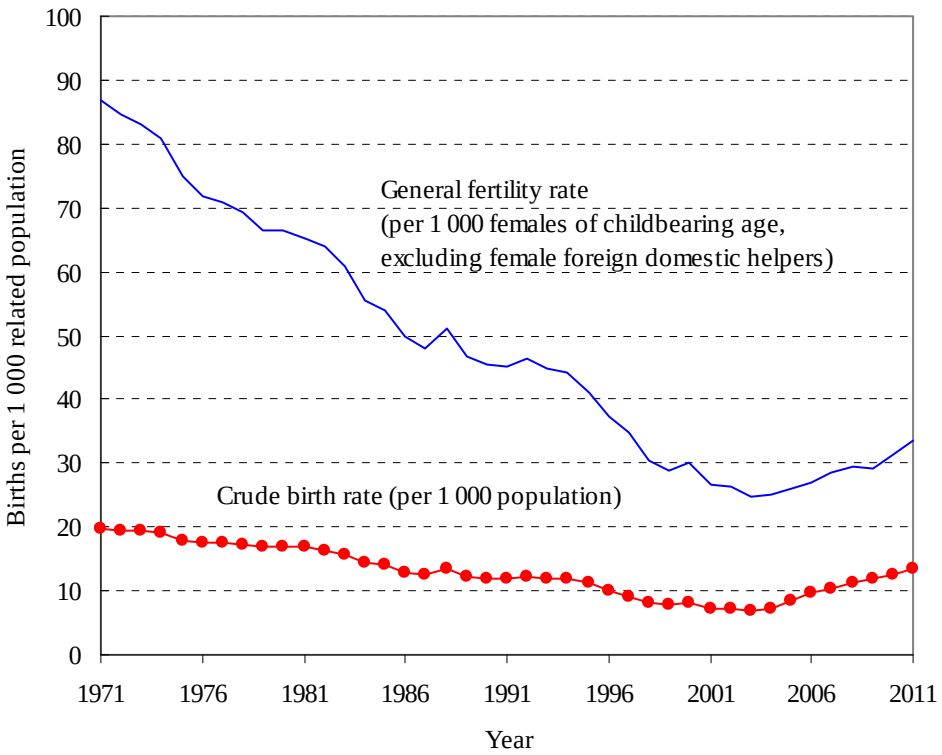
General fertility rate

Sometimes, the crude birth rate does not reflect correctly the birth trend. For instance, an inflow of male immigrants increases the population size, leading to a lowered birth rate, while actually the birth trend may not have changed.

This pitfall can be avoided by using the general fertility rate, which relates the number of births to the number of women of childbearing age (i.e. those aged between 15 and 49) in the population. The rate is usually expressed in terms of births per 1 000 females of childbearing age.

There is a special point to note for Hong Kong. There are a large number of female foreign domestic helpers working here, the great majority of whom would not give birth to children. To better reflect the fertility situation in Hong Kong, female foreign domestic helpers are excluded from the number of women of childbearing age in computing the general fertility rate.

Chart 1.4 General fertility rate and crude birth rate, 1971–2011



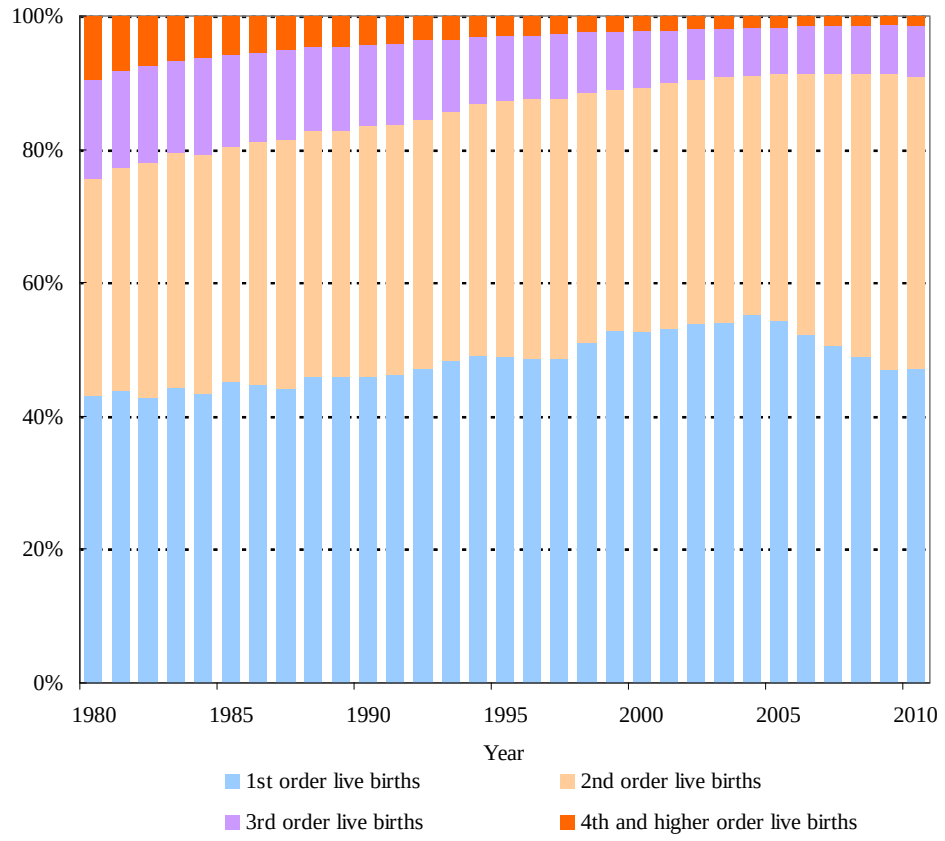
Decline in birth trend

Since the 1970's, the birth rate in Hong Kong has been falling rapidly. This is mainly because of late marriage. The median age of women who got married for the first time(5) in 1971 was 22.9, whilst that in 2011 was 28.8#.

Over the same period, there is also a significant increase in the prevalence of spinsterhood. In 1971, the percentage of never married women in the age group 40-44 was 3%. In 2011, this figures reached 17%.

Furthermore, people tend to have children later after marriage and have fewer children than before. Amongst all the births recorded in 2010, 90.8% were the first or second births to their mothers as compared with 75.6% in 1980. Correspondingly, the average size of domestic households dropped from 3.9 persons per household in 1982 to 2.9 in 2011, indicating a tendency towards the formation of smaller households.

Chart 1.5 Live births by order, 1980–2010



Note :

(5) Median age at first marriage is an indicator of the average age of persons at their first marriage such that 50% of these persons are above this age while the other 50% are below it.

Crude death rate

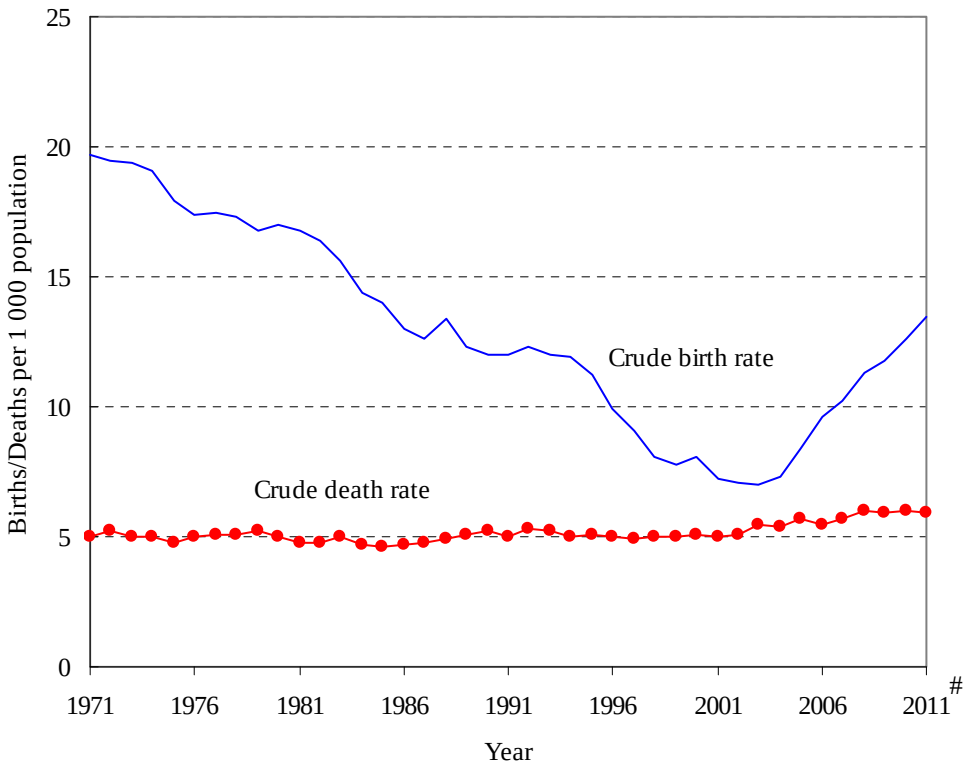
To measure the death trend, the crude death rate is calculated by dividing the number of known deaths in a calendar year by the average population size during the calendar year. Usually the mid-year population is taken as the average population size and the crude death rate is expressed in terms of deaths per 1 000 population.

In 2011, with 41 400# deaths and a mid-year population of 7 071 600, the crude death rate is equal to

$$\frac{41\ 400^{\#}}{7\ 071\ 600} \times 1\ 000 = \underline{\underline{5.9^{\#}}}$$

or 5.9# per 1 000 population.

Chart 1.6 Crude birth rate and crude death rate, 1971–2011



Rate of natural increase

The excess of known live births over known deaths occurring in a year is called the natural increase in population.

The rate of natural increase is given as the excess of known live births over known deaths occurring in a year per 1 000 mid-year population of that year. This indicator shows the rate at which the population growth is due to vital events (viz. births and deaths) only. The gain or loss in population owing to net movement is not taken into account.

During the year 2011, there were 95 400# known live births and 41 400# known deaths. With a population of 7 071 600, the rate of natural increase during the year is equal to

$$\frac{95\ 400^{\#} - 41\ 400^{\#}}{7\ 071\ 600} \times 1\ 000 = \underline{\underline{7.6^{\#}}}$$

or 7.6# per 1 000 population.

Alternatively, the rate of natural increase may be found by computing the difference between the crude birth rate and the crude death rate. For the above example, the rate of natural increase may also be calculated as crude birth rate for 2011 minus crude death rate for 2011 (i.e. 13.5# – 5.9# = 7.6#) or 7.6# per 1 000 population.

Effect of age structure on birth and death rates

For a population with more women of childbearing age, the crude birth rate will normally be higher. For a population with more old people, the crude death rate can be expected to be higher. It is therefore necessary to study the age composition of a population before definitive statements can be made about its birth and death trends. In more in-depth studies of the population, there are methods to isolate the age composition effects in order that birth and death trends at different points of time or amongst different countries can be accurately compared.

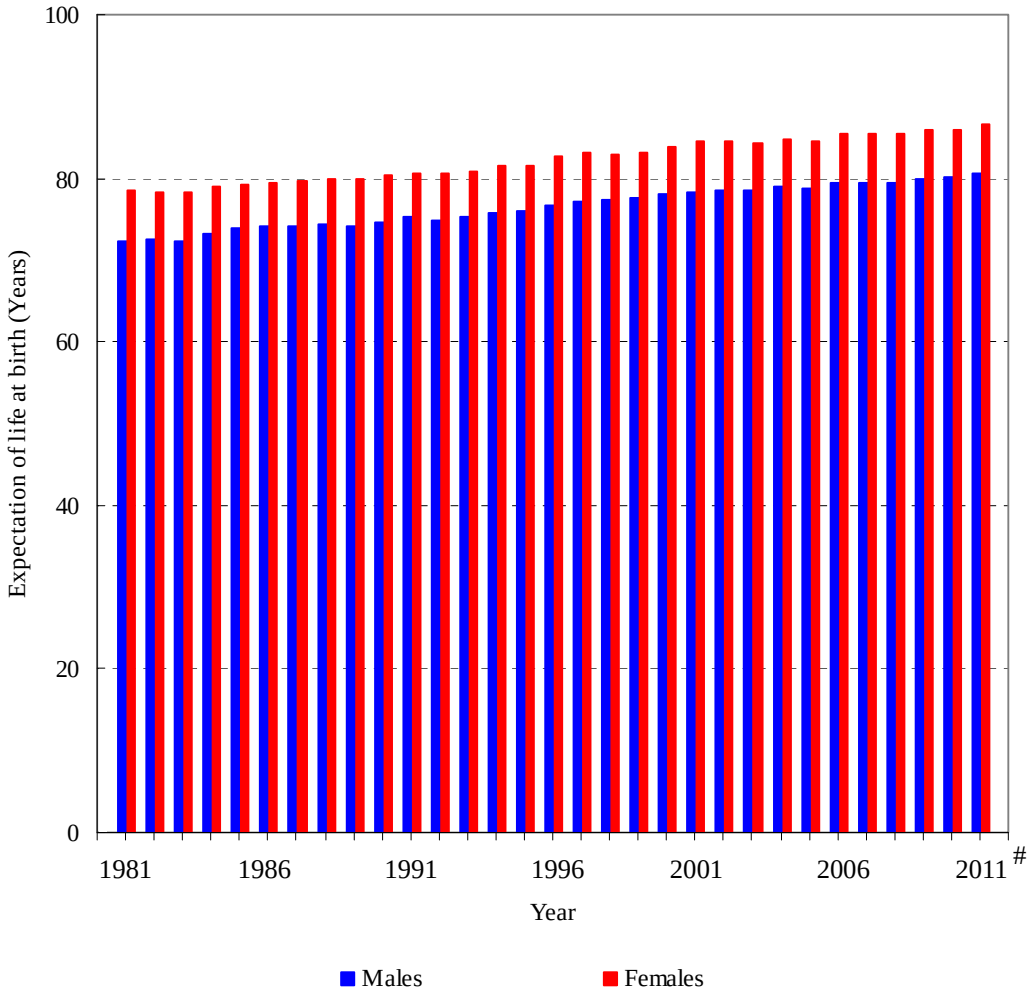
Expectation of life at birth

Since the 1950's, the population of Hong Kong has experienced a great reduction in death risks. The crude death rate fell from 10.2 per 1 000 population in 1951 to 5.0 per 1 000 population in 1971. The rate has remained rather stable since then. It has to be noted, however, that at a time when the population of old persons is increasing, a stable crude death rate actually suggests an overall improvement in the death trend.

A lower death trend would mean that people enjoy a longer life. Thus, the death trend may also be measured by the "expectation of life at birth". This indicates how long a child born in a year, say 2011, would expect to live, assuming that upon reaching various ages in his/her life, he/she would be subject to the same death risks as those faced by people of the respective ages in 2011.

The expectation of life at birth in 2011 was 80.5[#] years for males and 86.7[#] years for females, representing respectively a substantial increase of 8.2[#] years for both since 1981.

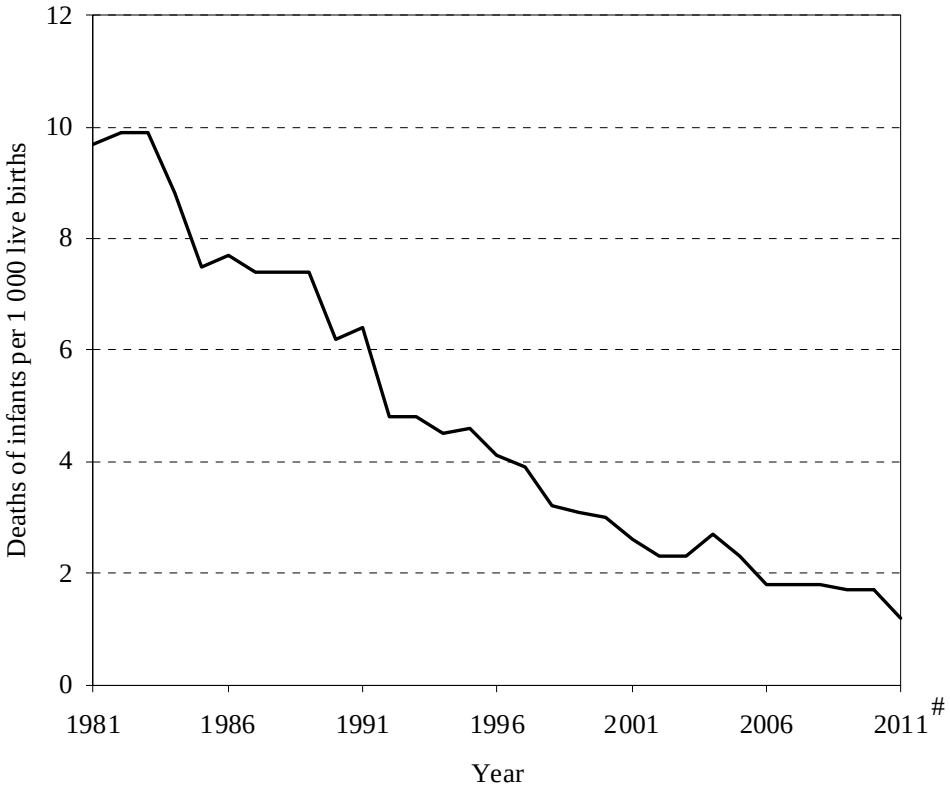
Chart 1.7 Expectation of life at birth, 1981–2011



Infant mortality rate

The decline in death rate in the past three decades was mainly due to improvements in birth care and child health, which reduced deaths of infants. The relevant indicator to study the phenomenon is the infant mortality rate, which is obtained by dividing the number of deaths of infants aged under one year by the total number of live births. The rate is usually expressed in terms of deaths per 1 000 live births.

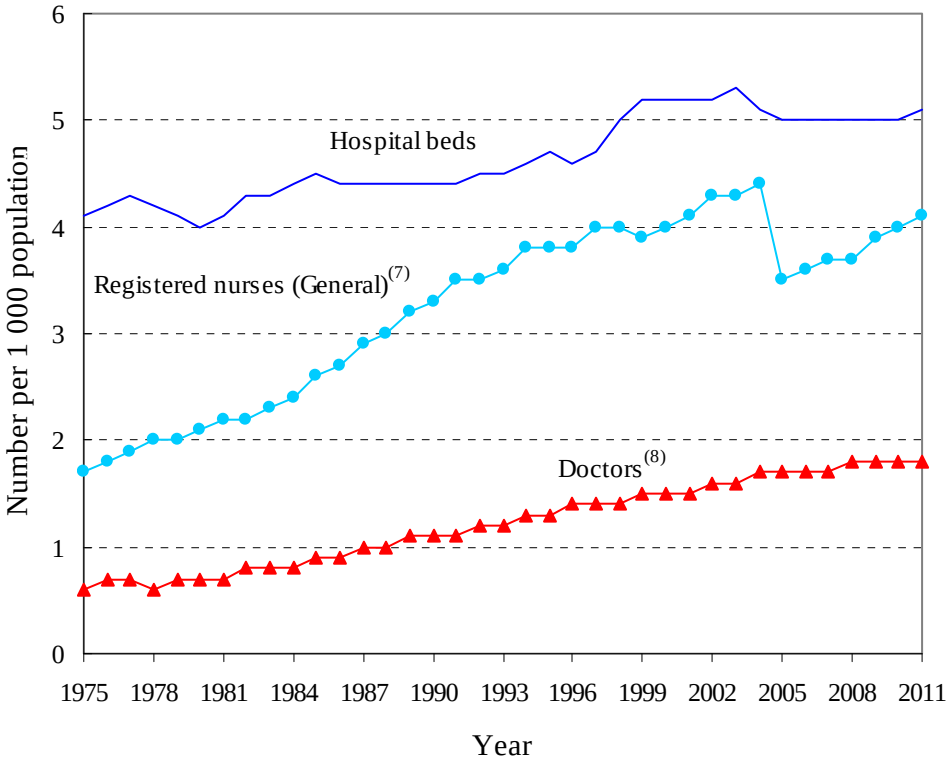
Chart 1.8 Infant mortality rate, 1981–2011



Better medical and health services

The general decline in death trend can be associated with better medical and health services available nowadays. People also have better knowledge about health and are more health conscious.

Chart 1.9 Rate of hospital beds and selected types of registered healthcare professionals, 1975–2011[6]



Notes :

(6) Figures are as at end of the year.

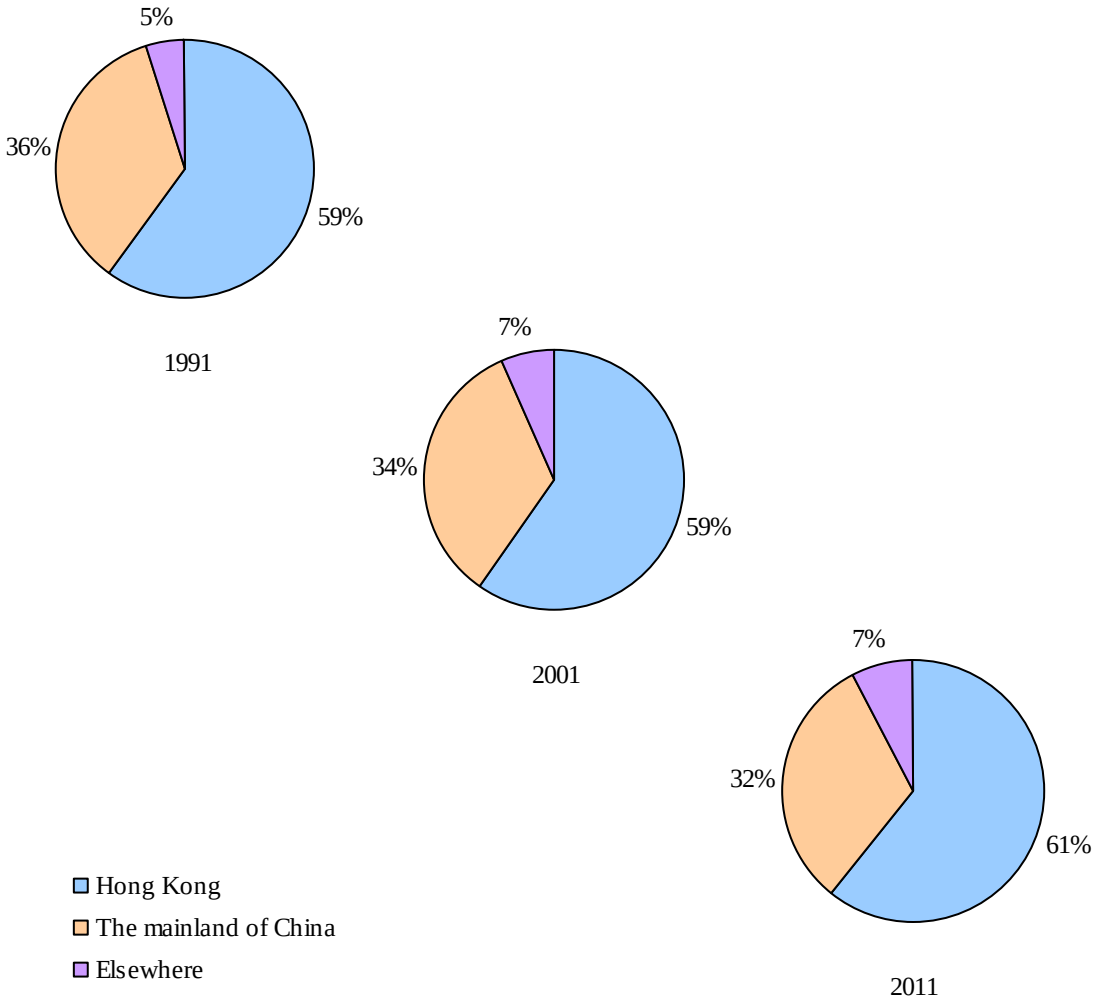
(7) The drop in 2005 is due to the removal of names of more than 6 000 registered nurses (General) from the register/roll in accordance with Section 7(3)(e) and Section 13(3)(e) of the Nurses Registration Ordinance, Cap. 164, Laws of Hong Kong.

(8) Figures refer to doctors with full registration on the local and overseas lists.

Movement of population

A substantial part of the Hong Kong population are entrants from the mainland of China. However, the locally born population always form the largest group. This can be seen by comparing the proportion of the population by their place of birth. In recent years, there has been an increasing number of residents who were born outside Hong Kong and some of them are foreign domestic helpers.

Chart 1.10 Distribution of population by place of birth, 1991, 2001 and 2011(9)



Note :

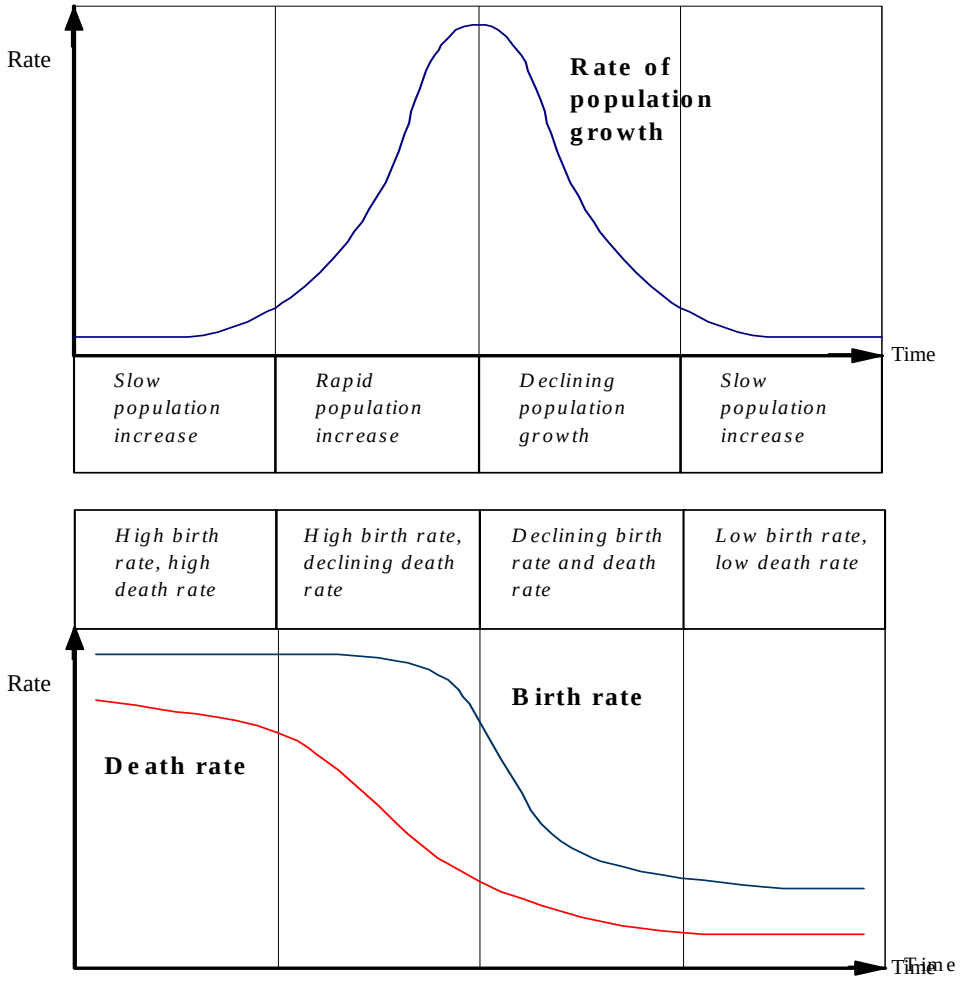
(9) Statistics shown in the charts above refer to information collected via population censuses. The area of pies is proportional to population size.

Demographic transition model as a general pattern of population growth

The demographic transition model describes changes in birth and death rates as a population passes from a traditional society to an urbanised and industrialised one. Generally, birth and death rates are high in traditional societies and low in modern societies. According to different rates of birth and death, changes in population are thought to occur in four stages :

(1) high birth and death rates;
(2) high birth rate and declining death rate;
(3) declining birth and death rates; and
(4) low birth and death rates.

Chart 1.11 Demographic transition model



As the population grows naturally by excess of births over deaths, it would increase at varying speeds in these different stages. This, of course, assumes that flows of movement are not significant. Otherwise, rather different pictures can emerge.

Further information

The above contents present only part of the information produced by C&SD on the topics concerned. For further information regarding the topics discussed in this chapter (e.g latest statistics, statistical reports, concepts and methods), please visit the following sections of the C&SD Website :

Population estimates
Population censuses and by-censuses
Demographics
Health

Exercise

Growth pattern of different countries

Births, deaths and net movement are the three factors affecting population growth. Annual growth rates, rates of natural increase, crude birth rates and crude death rates are useful indicators of the growth pattern.

Readers can try to calculate these rates from the exhibited table where data on population size, births and deaths are given for different countries.

The aim of this exercise is to enable readers to get familiarised with the fundamental concepts in population growth.

Generally speaking, the rate of natural increase may be taken as the difference between the crude birth rate and the crude death rate.

To further test one's understanding of the transition model as a general pattern of population growth, the last part of this exercise is to classify those countries into different transitional stages based on the birth and death rates they obtain. There are no universal rules to calibrate the rates as high or low. However, as a practical guide, a crude birth rate of 25 per 1000 population or above may be regarded as high; and for a crude death rate, over 15 per 1000 population or above is high.

(1) From the data given below, complete the table.
(Figures given in thousands; rates expressed in “per 1 000 population”.)

	(1)	(2)	(3)	(4)	(5)	(6)
Country	Population at year t	No. of births in year t	No. of deaths in year t	Crude birth rate	Crude death rate	Rate of natural increase
Country A	29 863	1 322	525			
Country B	15 941	723	332			
Country C	20 155	249	132			
Country D	32 268	332	226			
Country E	9 749	433	181			
Country F	1 315 844	17 558	8 795			
Country G	74 033	1 860	421			
Country H	1 517	51	17			
Country I	82 689	702	853			
Country J	10 098	96	132			

(2) From the rates obtained in part (1), classify the countries in various stages of the demographic transition model.

Stage (1) High birth rate High death rate	Stage (2) or (3) High birth rate or declining birth rate Declining death rate	Stage (4) Low birth rate Low death rate

2 Population structure

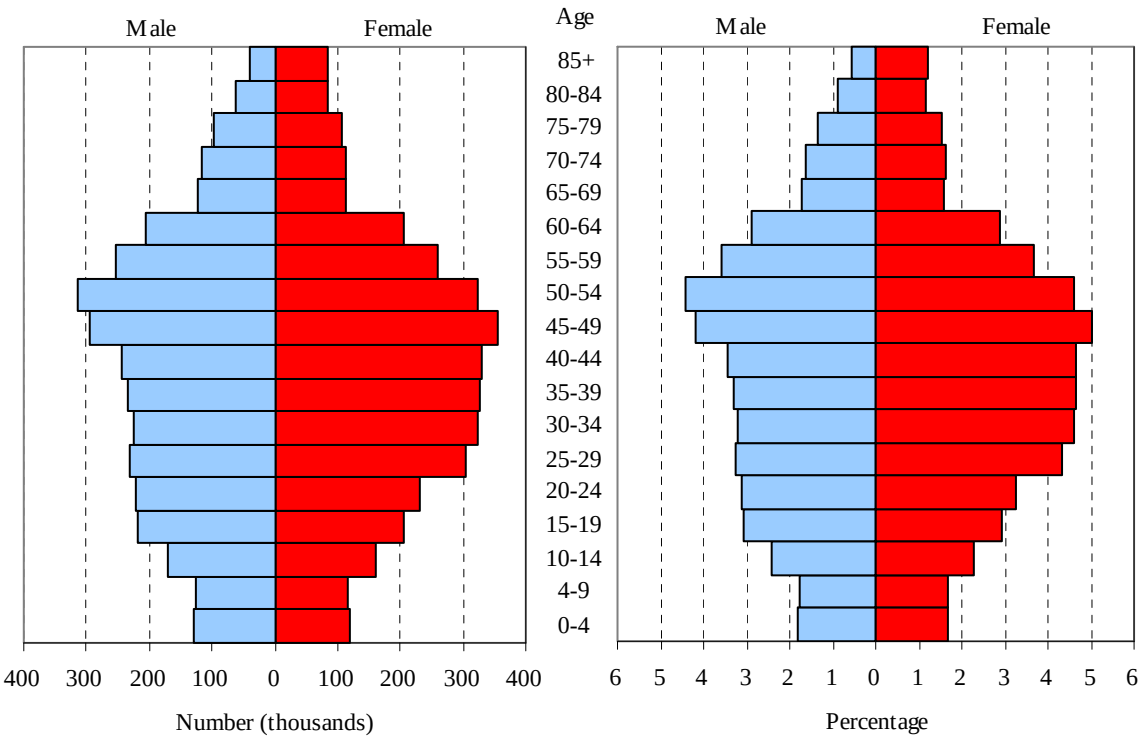
Introduction

Individuals differ in terms of their sex, age, employment, and so on. A population, made up of individuals, is characterised by the distribution of individuals in regard to these features. This chapter introduces various measures for describing population characteristics and their changes over time.

Population pyramid

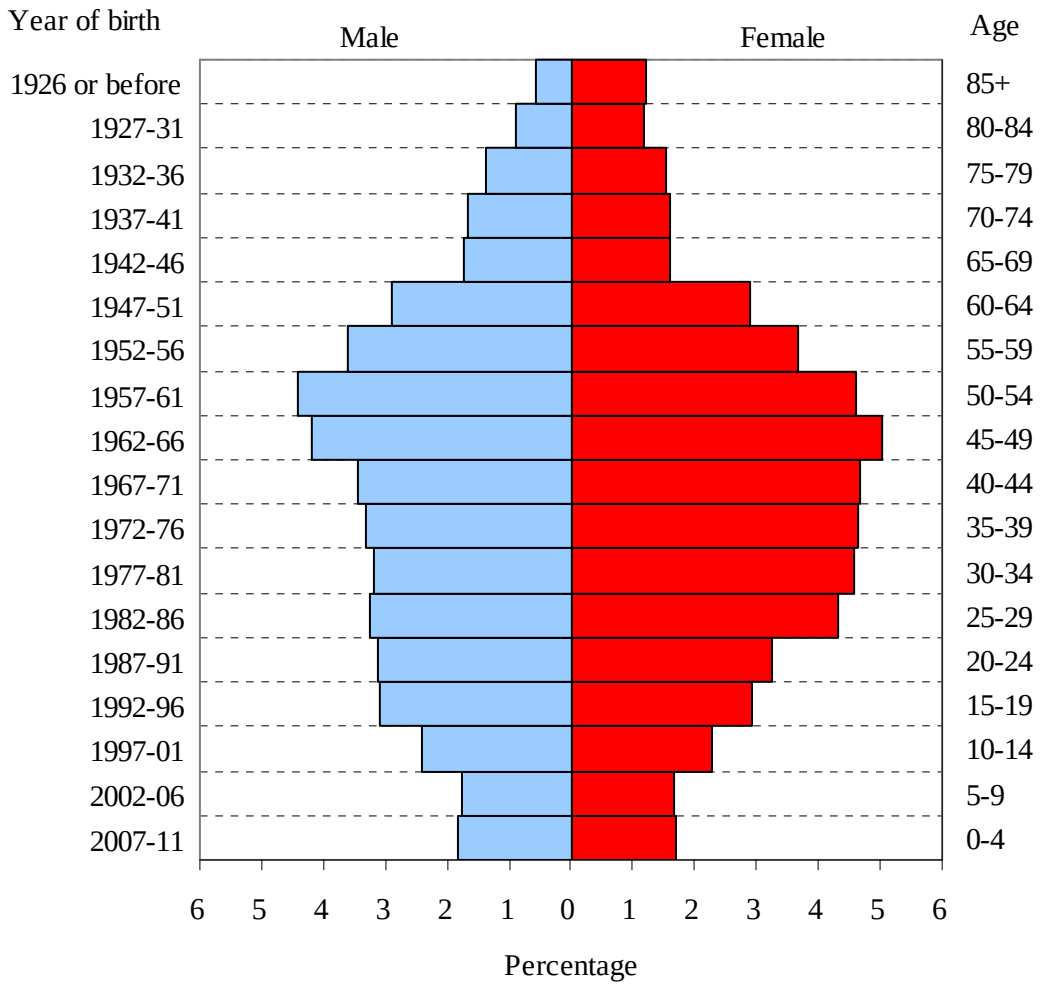
The age-sex distribution of a population is most clearly presented in the graphical form of a "population pyramid". The pyramid can be scaled in either population numbers or percentages. The percentage distribution is more appropriate when two or more populations are compared.

Chart 2.1 Population of Hong Kong by age and sex, mid-2011



The age-sex distribution reflects the combined effects of the past and recent trends in births, deaths and net movement. It is a rather persistent feature, and will not change drastically in a short span of time. Irregularities, like age or sex groups of exceptional size, reflect unusual numbers of births, deaths or net movement at some points of time in the past. Normally, births and net movement have larger impacts on the age-sex distribution than deaths.

Chart 2.2 Age-sex distribution of Hong Kong population, mid-2011

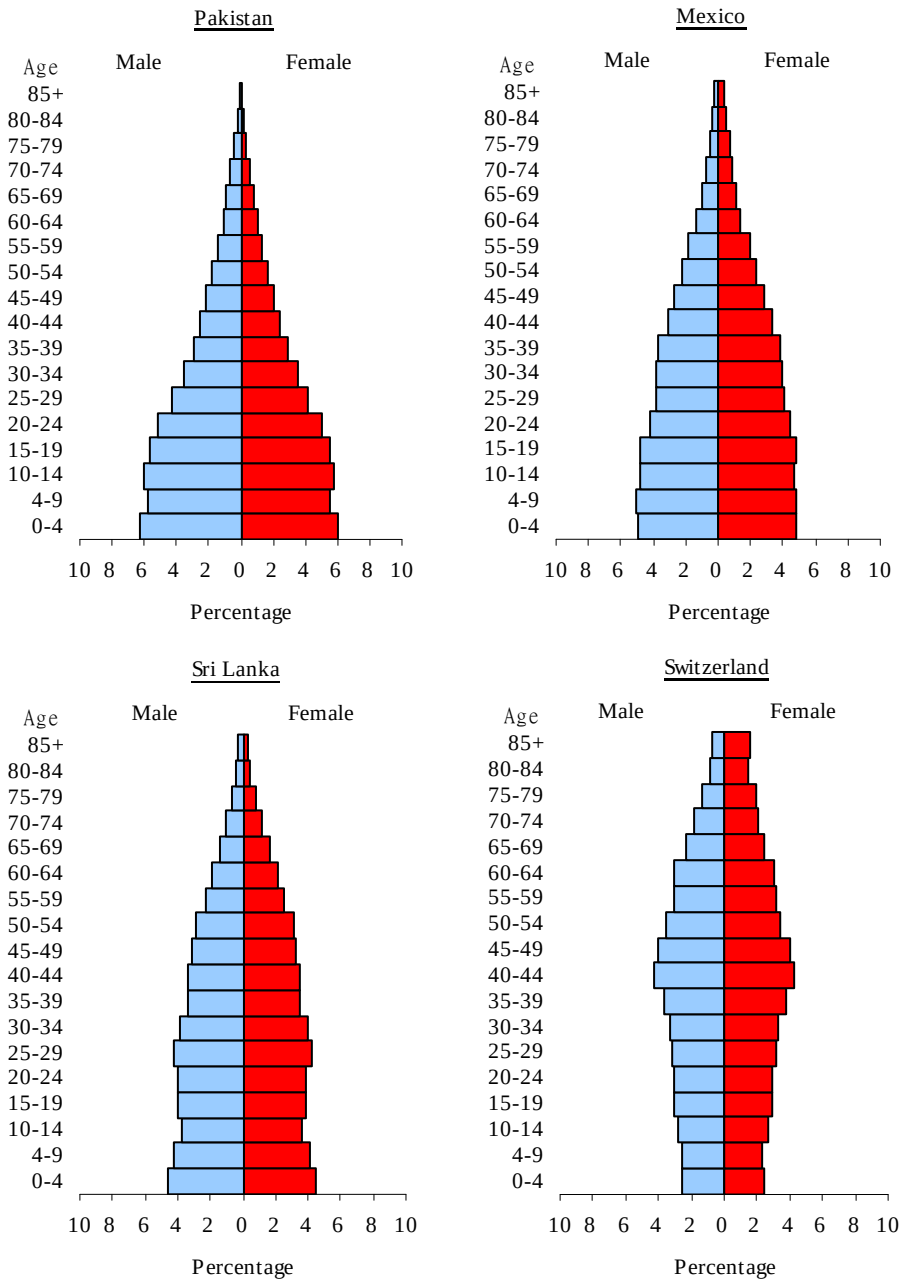


Shape of population pyramid in relation to the growth pattern of a population

In a growing population with high birth rates, the numbers of people in younger age groups are large and the age structure would take on the shape of a typical pyramid. If death rates are also high, the pyramid, which has a broad base, would narrow rapidly towards the top. Conversely, as the birth trend declines and the death trend improves, the pyramid would have a relatively narrow base and a middle section of nearly the same width; it does not begin to converge to the vertex until after very old age groups.

The shape of a population pyramid is thus closely related to the growth pattern of a population. Countries in different transitional stages may take on varying shapes of the population pyramid for their age-sex structures. For instance, the pyramid for Pakistan (2010) has a very broad base that narrows upwards very rapidly. It has almost a triangular shape. It results from a very high crude birth rate of 28.1 per 1 000 population and relatively low expectation of life at birth, 63.8 years for males and 65.4 years for females (2005–2010).

Chart 2.3 Population pyramids (in percentage) of selected countries, 2010



Source : World Population Prospects : The Population Database, Population Division of the of the United Nations Department of Economic and Social Affairs

In contrast, the pyramid for the Switzerland (2010) has a much narrower base, a little bit wider middle section and a slowly converging tip. The shape is almost semi-elliptical. It illustrates the effects of a low crude birth rate of 9.9 per 1 000 population and high expectation of life at birth, 79.3 years for males and 84.1 years for females (2005–2010).

The pyramids for Mexico (2010) and Sri Lanka (2010) illustrate configurations intermediate between those for Pakistan and Switzerland; their crude birth rates were 20.6 and 19 per 1 000 population respectively and their respective expectations of life at birth were 73.7 and 71.2 years for males and 78.6 and 77.4 years for females (2005–2010).

Expectation of life at birth is preferred here to crude death rate as an indicator of death trend because the crude death rates for the four countries are found highly distorted by the different age structures of their populations.

Apart from the natural effect of births and deaths, it should be noted that these four pyramids are affected by other irregular variations and therefore may deviate from the theoretical triangular and semi-elliptical shapes that the demographic transition model prescribes.

Besides the detailed description of age-sex structure provided by the population pyramid, the age-sex distribution of a population can also be given by other summary measures.

Sex ratio

To compare the relative size of the male group and female group in the population, the sex ratio is calculated. The ratio is obtained by dividing the number of males in the population by the number of females. The sex ratio is usually expressed as the number of males per 1 000 females.

At mid-2011, there were 3 303 000 males and 3 768 600 females in Hong Kong. Thus the sex ratio was

$$\frac{3\,303\,000}{3\,768\,600} \times 1\,000 = \underline{\underline{876}}$$

or 876 males per 1 000 females.

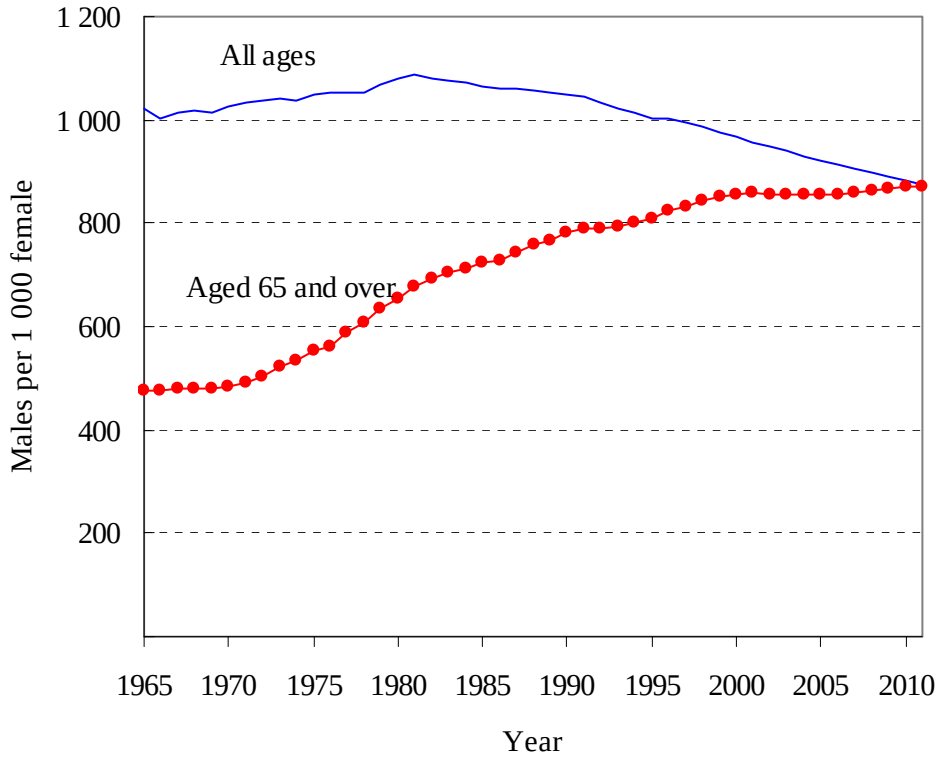
Sex ratios by age group

Sex ratios may also be calculated separately for various age groups. For instance, the sex ratio at birth compares the number of male births to female births. Usually, as a biological norm, more male babies are born than female babies.

Sex ratio (number of males per 1 000 females)

Age group	Mid-2001	Mid-2006	Mid-2011
0 – 14	1 072	1 063	1 073
15 – 19	1 058	1 039	1 052
20 – 24	927	914	965
25 – 29	848	803	753
30 – 34	784	771	696
35 – 39	844	749	715
40 – 44	977	834	735
45 – 49	1 005	964	829
50 – 54	1 078	987	963
55 – 59	1 165	1 033	980
60 – 64	1 143	1 097	1 004
65 and over	858	856	871
All ages	956	912	876

Chart 2.4 Sex ratio of Hong Kong population by age group, mid-1965 to mid-2011



As revealed by the above chart, the sex ratio for the older ages was on the rise over the years. This reflects a greater improvement in the death trend for males than for females in recent years.

Increasing proportion of females in Hong Kong population

In Hong Kong, the proportion of females in the total Hong Kong population increased continuously in the past ten years. The sex ratio of the population fell below parity. The ratio dropped from 956 in mid-2001 to 876 in mid-2011.

The increase in the proportion of females in the population was brought about by three major factors.

Effect of foreign domestic helpers

Firstly, there was a significant increase in the number of foreign domestic helpers, who were mostly young and middle-aged females, in the past decade. Owing to this factor, the number of females in the age group of 20-39 was relatively large, with the corresponding sex ratio at mid-2011 being 768 males per 1 000 females.

Sex ratio (number of males per 1 000 females)

Age group	Mid-2001	Mid-2006	Mid-2011
0 – 19	1 068	1 055	1 066
20 – 39	**845**	**803**	**768**
40 – 64	1 043	955	888
65 and over	858	856	871
All ages	956	912	876

After excluding foreign domestic helpers, the sex ratio for the age group of 20-39 at mid-2011 rose to 936. The sex ratio for the entire population was 948.

Sex ratio, excluding foreign domestic helpers
(number of males per 1 000 females)

Age group	Mid-2001	Mid-2006	Mid-2011
0 – 19	1 069	1 056	1 066
20 – 39	**985**	**941**	**936**
40 – 64	1 089	993	932
65 and over	858	857	872
All ages	1 021	971	948

Effect of one-way permit holders from the mainland of China

The second factor contributing to the increasing proportion of females in the population is that a large proportion of the one-way permit holders from the mainland of China were women who came to Hong Kong to join their husbands. In 2011, there were 43 379 one-way permit holders from the mainland of China and about 69% of these new entrants were females.

Effect of higher expectation of life for females

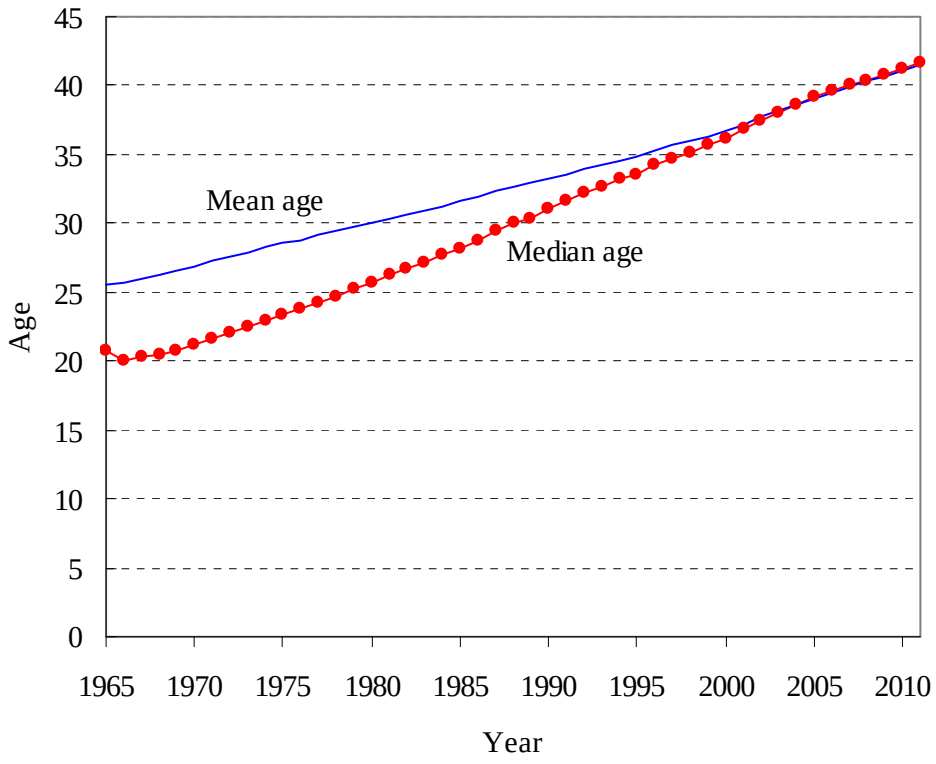
The third factor was that females generally live longer than males. The expectation of life at birth of females was 86.7# years while that for males was only 80.5# years in 2011.

The fact that females generally live longer than males will result in a lower sex ratio for the older ages than that for the entire population (this phenomenon can be seen from Chart 2.4 on page 2:6).

Mean and median ages

To describe the age structure of a population, its average age can be used. A simple average, or the "mean", can be obtained by dividing the sum of ages of the total population by the population size. Alternatively, a measure is provided by the "median" age of the population. Median age of the population is an indicator of the average age of the population such that 50% of the total population are above this age while the other 50% are below it.

Chart 2.5 Mean and median ages of Hong Kong population, mid-1965 to mid-2011



In a set of data, the "median" is normally a better measure of its average value if there are extreme values, since the "mean" will be much affected by those extreme values. In the case of age, the general experience is that the median age is considered more preferable.

"Young" and "old" populations

A population may be described as "young" or "old".

In general, populations with median ages under 20 may be described as "young" and those with median ages 30 and over as "old". Those with median ages in between 20 and 29 are referred to as "populations of intermediate age". A population may be said to be "ageing" when its median age is on the rise.

Dependency ratios

It is interesting and useful to divide the population into three groups by age, namely the young (i.e. those aged between 0 and 14), those in the working age (i.e. aged between 15 and 64) and the elderly (i.e. aged 65 and over) and study their relative sizes.

The overall dependency ratio is computed by dividing the number of the young and the elderly in the population by those in the working age. The components of this measure may be calculated separately as the "child dependency ratio" and "elderly dependency ratio".

Given the following population data for Hong Kong in mid-2011, the various dependency ratios are computed accordingly :

Age group	Population at mid-2011
0 – 14	823 500
15 – 64	5 306 700
65 and over	941 400

Child dependency ratio

$$= \frac{823\,500}{5\ 306\,700} \times 1\ 000 = \underline{\underline{155}}$$

or 155 persons aged under 15 per 1 000 population of working age

Elderly dependency ratio

$$= \frac{941\,400}{5\,306\,700} \times 1\ 000 = \underline{\underline{177}}$$

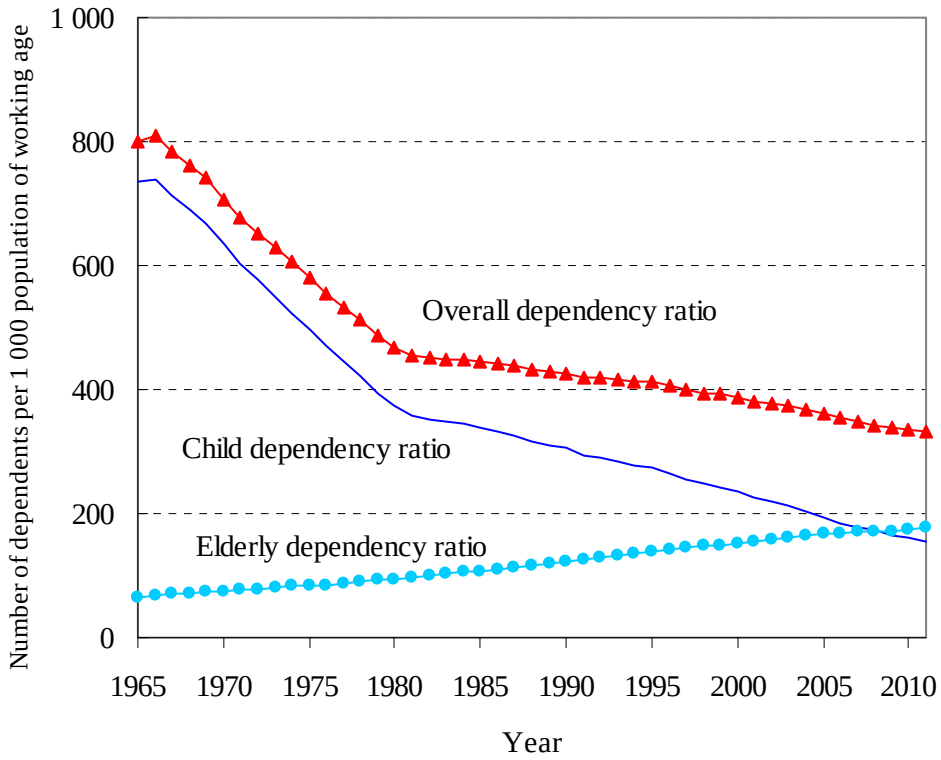
or 177 persons aged 65 and above per 1 000 population of working age

Overall dependency ratio

$$= \frac{(823\,500) + (941\,400)}{5\,306\,700} \times 1\ 000 = \underline{\underline{333}}$$

or 333 persons aged under 15 or aged 65 and above per 1 000 population of working age

Chart 2.6 Dependency ratios of Hong Kong population, mid-1965 to mid-2011



Geographical distribution of the population

Hong Kong is delineated into 18 geographical districts : 4 on Hong Kong Island; 5 in Kowloon and 9 in the New Territories. These geographical districts are called District Council districts. The numbers of people living in different districts vary considerably.

In mid-2011, 630 273 persons lived in the Sha Tin district. It was the highest populated district. There were only 141 327 persons living in the Islands district, where was the lowest populated district.

The proportion of the population living in Hong Kong Island decreased from 19.9% in 2001 to 18.0% in 2011 while the proportion living in the New Territories increased from 49.8% to 52.2%. This change in the geographical distribution of the population reflects the effectiveness of the government's policy in regard to the development of new towns in the New Territories.

Population by District Council district, mid-2011

District Council district	Population(1)
Central and Western	251 519
Wan Chai	152 608
Eastern	588 094
Southern	278 655
Yau Tsim Mong	307 878
Sham Shui Po	380 855
Kowloon City	377 351
Wong Tai Sin	420 183
Kwun Tong	622 152
Kwai Tsing	511 167
Tsuen Wan	304 637
Tuen Mun	487 546
Yuen Long	578 529
North	304 134
Tai Po	296 853
Sha Tin	630 273
Sai Kung	436 627
Islands	141 327

Note :

(1) Figures exclude marine population.

Employment structure of the population

"Economically active" persons are those aged 15 and over who are either working or seeking jobs. These people make up the "labour force". Children at school, the retired and home-makers are economically inactive persons.

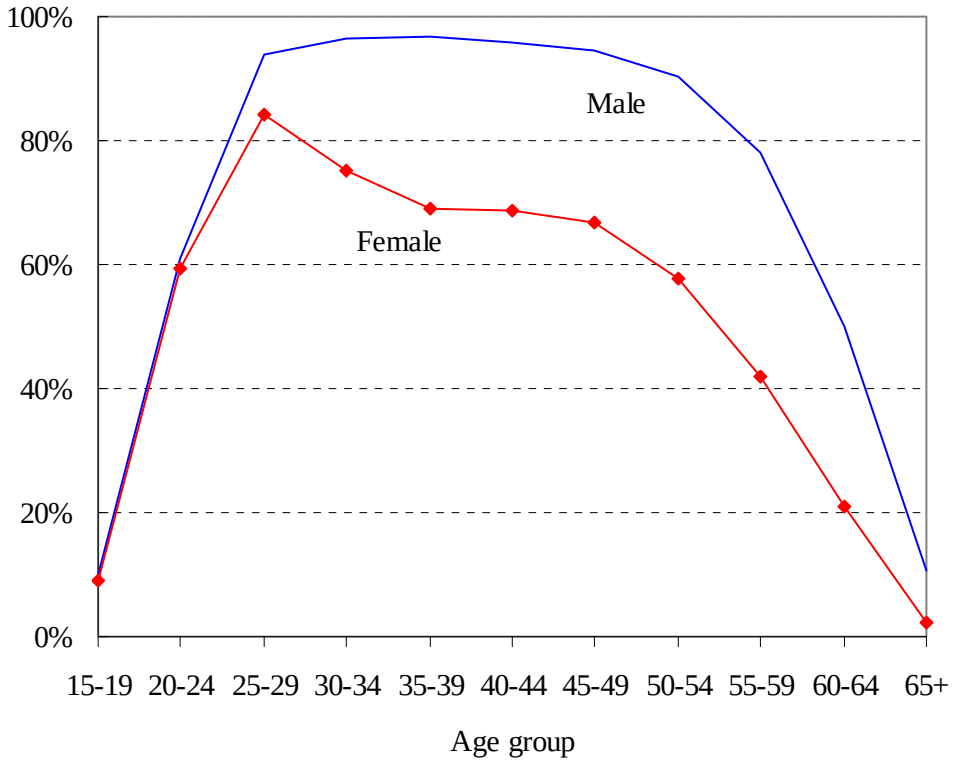
The labour force of Hong Kong increased steadily from 3.37 million in 2000 to 3.70 million in 2011.

"Labour force participation rates" are used to measure the proportion of economically active persons in the population aged 15 and over. These rates are usually calculated separately for males and females in different age groups. Because more women concentrate on household duties, their labour force participation rates are generally lower than males. The age-specific labour force participation rates for males are highest at their prime working ages (i.e. between 25 and 54) while the rates for females reach the peak at the age group 25-29 and decline gradually in the older age groups.

The labour force participation rates in the age groups 15-19 and 60 and over for both sexes were substantially lower than those in other age groups, as most members of the former group were still at school while a fair proportion of those in the latter group were retired people.

The characteristics of foreign domestic helpers differ from those of the entire population, particularly the labour force participation rate. Because of this, the age-sex specific labour force participation rates presented in Chart 2.7 have excluded foreign domestic helpers in order to better reflect the profile of the local population.

Chart 2.7 Labour force participation rates of Hong Kong population (excluding foreign domestic helpers), 2011

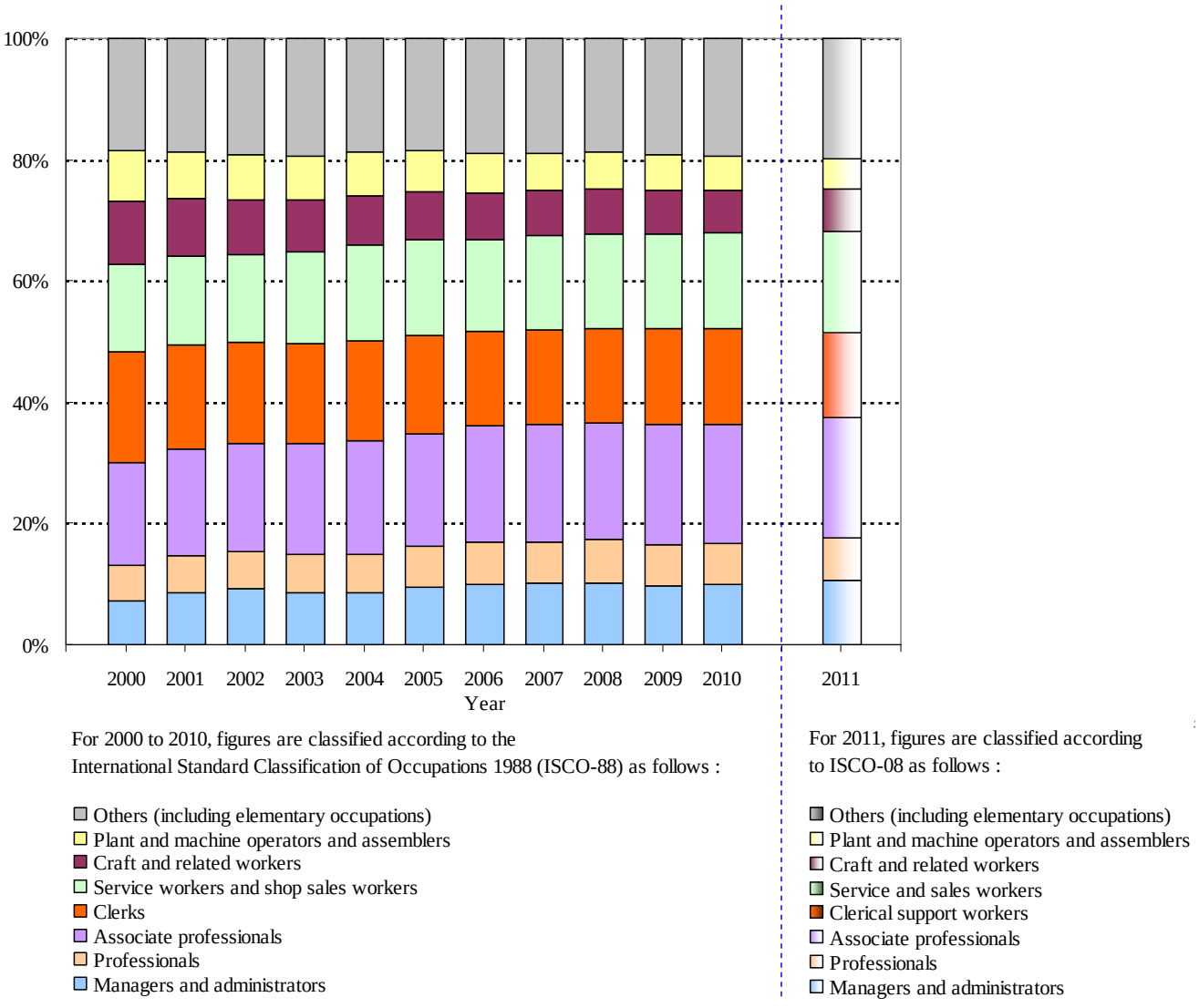


For employed persons who are working, their characteristics can be described in terms of the occupations and industries in which they are engaged.

Distribution of employed population by occupation

Occupation refers to the kind of work that a person does. Examples are teachers, doctors, shop sales workers, engineers, knitting machine operators and farmers.

Chart 2.8 Distribution of employed population by occupation(2), 2000–2011



Note :

(2) Statistics for 2001-2010 are compiled based on International Standard Classification of Occupations 1988 (ISCO-88). Statistics for 2011 are compiled based on International Standard Classification of Occupations 2008 (ISCO-08). Hence, figures for 2011 are not directly comparable to earlier figures.

Distribution of employed population by industry sector

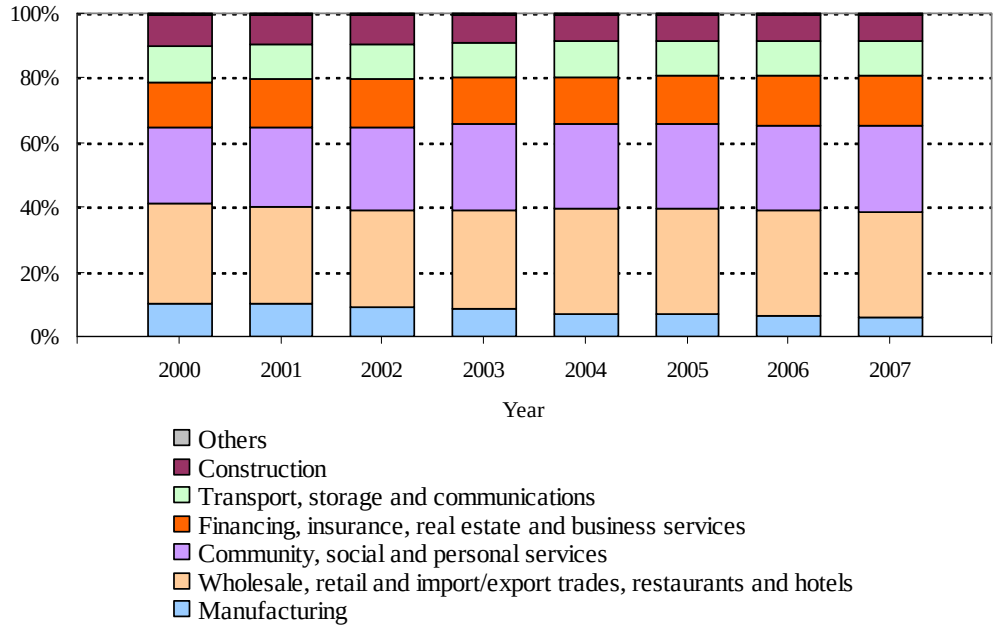
Industry, on the other hand, refers to the kind of activity carried out by the establishment or enterprise in which a person works. Examples of some establishments are schools, hospitals, shops, construction firms, factories and farms. Examples of "industries" are business services, retail, construction, manufacturing and agriculture.

The economic characteristics of a population reflect the stage of its economic development. In general, the labour force tends to shift away from agriculture to manufacturing industries, and then to services industries.

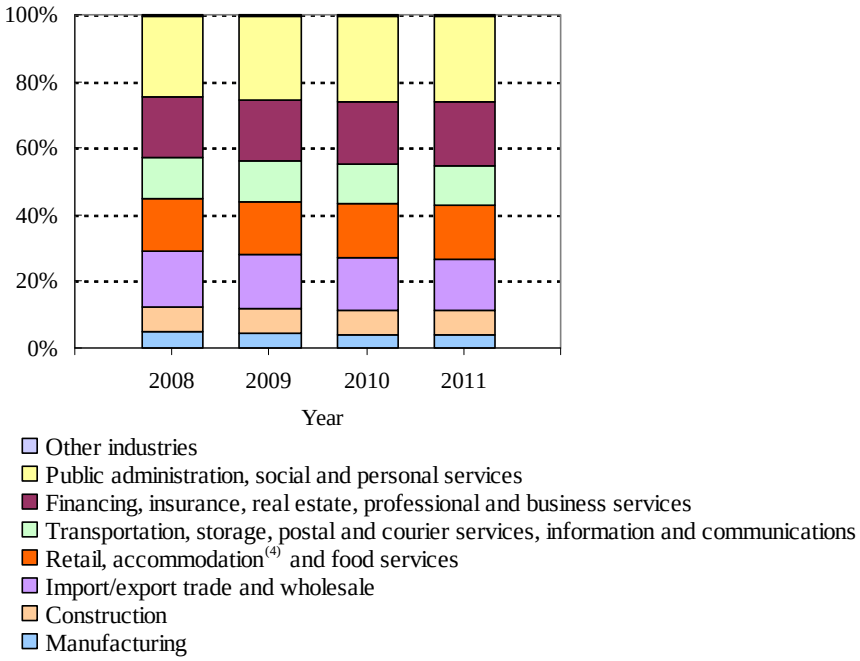
In Hong Kong, the economic characteristics of the population follow this general pattern in economic development. Over the last 20 years, the importance of services industries has gradually grown, at the expense of manufacturing industries. Consequently, more and more people work in clerical and services jobs, rather than as production workers.

Chart 2.9 Distribution of employed population by industry [3], 2000–2011

(a) Based on Hong Kong Standard Industrial Classification (HSIC) Version 1.1



(b) Based on HSIC Version 2.0



Notes:

(3) Statistics for 2001-2007 are compiled based on the "Hong Kong Standard Industrial Classification (HSIC) Version 1.1". Statistics for 2008 and onwards are compiled based on "HSIC Version 2.0". Hence, figures for 2008 and onwards are not directly comparable to earlier figures.

(4) Accommodation services cover hotels, guesthouses, boarding houses and other establishments providing short term accommodation.

Further information

The above contents present only part of the information produced by C&SD on the topics concerned. For further information regarding the topics discussed in this chapter (e.g. latest statistics, statistical reports, concepts and methods), please visit the following sections of the C&SD Website :

Population estimates
Population censuses and by-censuses
Demographics
Labour force

Exercise

Age-sex structure of the 18 districts

Population pyramid is a useful tool to depict the age-sex structure of a population.

This exercise demonstrates how to present age-sex data by means of population pyramids. It further shows how age-sex structure differs in different districts of Hong Kong, which has implications for the requirement for social services and facilities in these districts.

Given the age-sex data of the 18 districts in pages 2:22 to 2:24, please prepare a population pyramid for each of these districts on graph papers according to the formats given in pages 2:25 to 2:27.

Some of the 18 districts may be called “old” districts and others called “young” districts. Please rank these districts according to their aging extent, by referring to the median age of the district populations.

(1) With reference to the statistics obtained from the 2011 Population Census for the 18 districts shown on pages 2:22 to 2:24, prepare the population pyramids.

(2) Rank the aging extent of these districts by comparing their median ages.

District Council district	Median age	Ranking for aging extent (Note) ("1" for the most aged district)
Central and Western	41.3	
Wan Chai	43.0	
Eastern	44.1	
Southern	42.9	
Yau Tsim Mong	41.1	
Sham Shui Po	43.2	
Kowloon City	42.5	
Wong Tai Sin	44.5	
Kwun Tong	42.8	
Kwai Tsing	41.9	
Tsuen Wan	41.3	
Tuen Mun	40.4	
Yuen Long	38.6	
North	40.2	
Tai Po	41.4	
Shatin	41.5	
Sai Kung	39.3	
Islands	39.1	

(3) With reference to the statistics shown on pages 2:22 to 2:24, please analyse how various districts may differ in their requirements for public services and facilities.

Note : Where there are ties in rank, the tied observations are assigned the mean of the ranks which they jointly occupy.

Percentage distribution of population by age and sex, June 2011

	Central and Western			Wan Chai			Eastern		
Age group	Male	Female	Total	Male	Female	Total	Male	Female	Total
0-4	2.2	2.0	4.2	2.0	1.9	3.9	1.7	1.6	3.4
5-9	1.6	1.6	3.2	1.6	1.5	3.0	1.5	1.4	3.0
10-14	2.0	1.9	3.9	1.7	1.6	3.3	2.2	2.0	4.2
15-19	2.3	2.2	4.5	2.2	1.9	4.1	2.8	2.7	5.4
20-24	2.9	3.0	5.9	2.4	2.7	5.1	2.7	2.8	5.5
25-29	3.7	5.1	8.8	3.3	4.8	8.2	2.9	4.0	7.0
30-34	3.4	5.0	8.4	3.4	5.1	8.4	3.0	4.5	7.5
35-39	3.6	5.1	8.7	3.6	5.3	9.0	3.1	4.4	7.5
40-44	3.5	5.0	8.5	3.4	4.8	8.2	3.3	4.6	7.9
45-49	3.7	4.7	8.4	3.7	4.8	8.5	4.1	5.0	9.1
50-54	4.0	4.3	8.2	4.1	4.4	8.6	4.4	4.7	9.1
55-59	3.5	3.6	7.1	3.6	3.9	7.5	3.8	4.0	7.8
60-64	3.1	3.0	6.1	3.2	3.3	6.5	3.4	3.5	6.9
65-69	1.5	1.5	3.0	1.6	1.6	3.2	1.8	1.7	3.6
70-74	1.8	1.8	3.6	1.8	2.1	3.9	1.9	2.0	3.9
75-79	1.5	1.7	3.1	1.7	1.9	3.6	1.7	1.8	3.5
80-84	1.0	1.3	2.2	1.1	1.5	2.5	1.0	1.4	2.5
85+	0.6	1.3	2.0	0.8	1.6	2.4	0.7	1.4	2.2
All ages	46.0	54.0	100.0	45.3	54.7	100.0	46.2	53.8	100.0
Median age			41.3			43.0			44.1

	Southern			Yau Tsim Mong			Sham Shui Po		
Age group	Male	Female	Total	Male	Female	Total	Male	Female	Total
0-4	1.6	1.5	3.1	2.4	2.3	4.7	1.9	1.7	3.6
5-9	1.7	1.6	3.4	1.8	1.6	3.4	1.7	1.6	3.3
10-14	2.4	2.2	4.6	2.2	2.1	4.3	2.4	2.2	4.6
15-19	2.9	2.7	5.6	2.5	2.3	4.8	3.0	2.8	5.8
20-24	3.0	3.2	6.2	2.3	2.5	4.8	2.9	3.0	5.9
25-29	3.3	4.6	7.9	3.1	4.3	7.4	3.0	4.1	7.1
30-34	2.9	4.3	7.2	3.9	5.6	9.5	2.9	4.2	7.1
35-39	3.1	4.4	7.5	3.9	5.4	9.3	3.1	4.5	7.6
40-44	3.4	4.7	8.1	3.5	4.7	8.2	3.4	4.5	7.9
45-49	4.5	5.5	9.9	3.8	4.6	8.4	3.9	4.8	8.7
50-54	4.5	4.8	9.3	3.7	3.8	7.5	4.0	4.1	8.1
55-59	3.6	3.8	7.4	3.3	3.4	6.7	3.7	3.7	7.4
60-64	2.9	3.0	5.9	3.3	3.3	6.6	2.9	2.9	5.8
65-69	1.6	1.5	3.1	1.7	1.6	3.4	1.9	1.8	3.7
70-74	1.7	1.7	3.4	1.8	1.8	3.6	2.2	2.1	4.3
75-79	1.4	1.6	3.1	1.5	1.7	3.2	1.8	2.1	3.9
80-84	1.0	1.3	2.3	1.0	1.3	2.4	1.2	1.6	2.8
85+	0.7	1.4	2.1	0.7	1.3	2.0	0.8	1.6	2.3
All ages	46.2	53.8	100.0	46.4	53.6	100.0	46.7	53.3	100.0
Median age			42.9			41.1			43.2

Percentage distribution of population by age and sex, June 2011 (cont'd)

	Kowloon City			Wong Tai Sin			Kwun Tong		
Age group	Male	Female	Total	Male	Female	Total	Male	Female	Total
0-4	1.9	1.7	3.7	1.4	1.3	2.8	1.8	1.6	3.4
5-9	1.7	1.6	3.3	1.5	1.4	3.0	1.8	1.7	3.5
10-14	2.4	2.2	4.6	2.3	2.2	4.6	2.5	2.4	4.8
15-19	2.9	2.7	5.5	3.2	3.1	6.2	3.1	3.0	6.1
20-24	2.7	2.9	5.6	3.2	3.4	6.6	3.0	3.1	6.0
25-29	3.1	4.1	7.2	2.9	3.7	6.6	2.9	3.8	6.7
30-34	3.3	4.8	8.1	2.7	3.8	6.4	2.9	4.1	7.0
35-39	3.3	4.7	8.0	2.9	4.0	6.9	3.2	4.4	7.6
40-44	3.4	4.7	8.2	3.3	4.4	7.7	3.5	4.8	8.3
45-49	4.0	4.8	8.8	4.5	5.4	9.9	4.0	4.8	8.9
50-54	4.0	4.1	8.1	4.6	4.7	9.3	4.2	4.4	8.6
55-59	3.4	3.5	6.9	3.4	3.5	6.9	3.4	3.5	6.9
60-64	3.1	3.1	6.1	2.7	2.8	5.5	2.8	3.0	5.8
65-69	1.8	1.7	3.6	2.0	1.8	3.8	2.0	1.9	4.0
70-74	2.0	2.0	4.0	2.1	2.1	4.2	2.0	2.0	4.1
75-79	1.7	1.9	3.5	1.9	2.3	4.2	1.8	1.9	3.7
80-84	1.1	1.5	2.6	1.3	1.8	3.1	1.1	1.4	2.6
85+	0.7	1.6	2.3	0.8	1.6	2.4	0.7	1.3	2.0
All ages	46.4	53.6	100.0	46.7	53.3	100.0	46.8	53.2	100.0
Median age			42.5			44.5			42.8

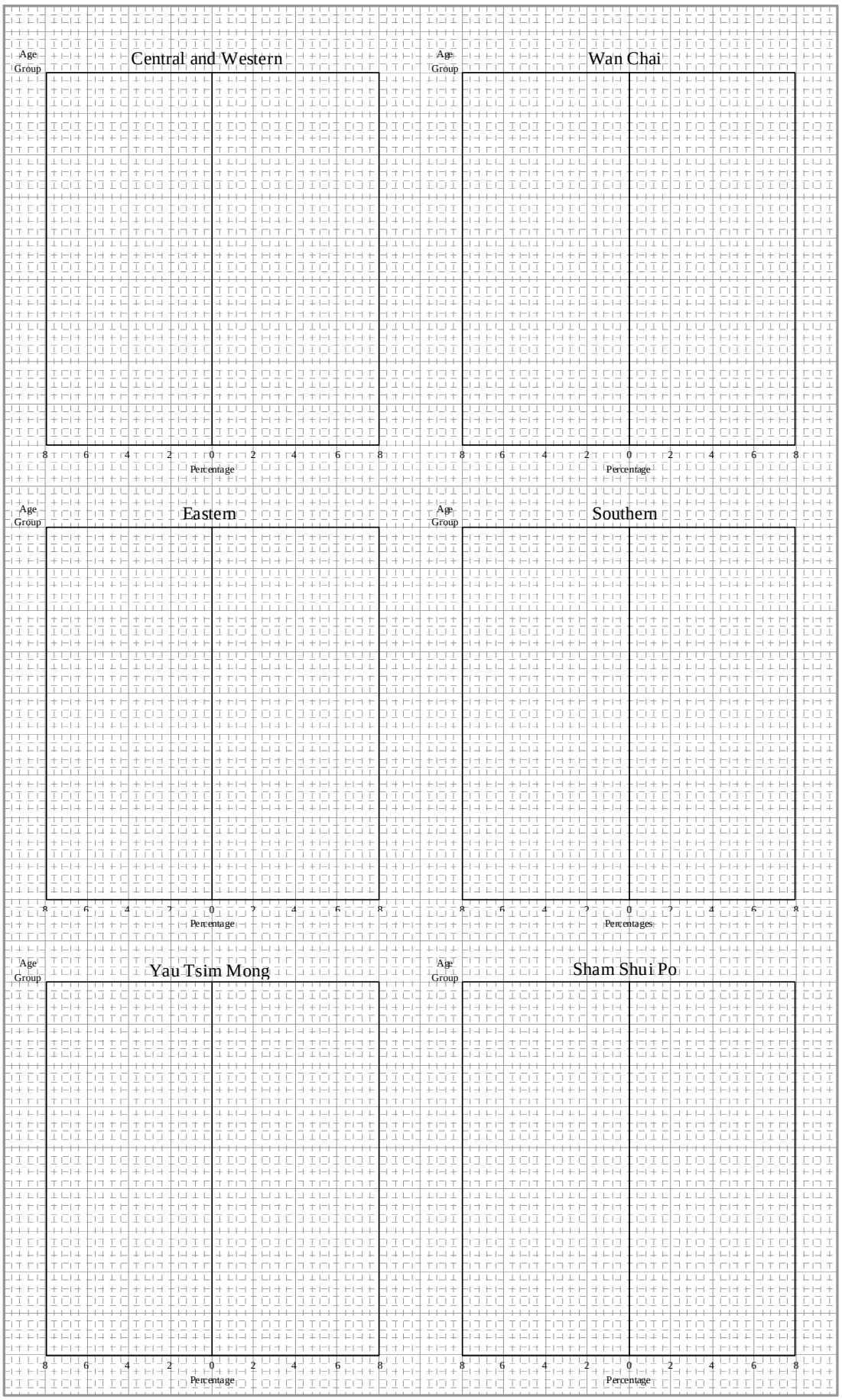
	Kwai Tsing			Tsuen Wan			Tuen Mun		
Age group	Male	Female	Total	Male	Female	Total	Male	Female	Total
0-4	1.6	1.5	3.1	2.2	1.9	4.2	1.8	1.6	3.4
5-9	1.9	1.7	3.7	2.0	1.9	3.9	1.6	1.5	3.1
10-14	2.4	2.2	4.6	2.5	2.3	4.8	2.4	2.2	4.6
15-19	3.2	3.0	6.1	2.9	2.8	5.7	3.3	3.1	6.4
20-24	3.4	3.5	6.9	2.5	2.6	5.1	3.5	3.5	6.9
25-29	3.3	4.2	7.4	2.8	3.8	6.6	3.6	4.6	8.2
30-34	3.0	4.0	7.0	3.5	5.1	8.6	3.8	5.1	8.9
35-39	3.4	4.7	8.1	3.6	5.2	8.8	3.3	4.5	7.8
40-44	3.6	4.6	8.1	3.7	5.1	8.8	3.0	4.1	7.2
45-49	3.8	4.6	8.5	4.4	5.3	9.7	4.2	5.0	9.3
50-54	4.3	4.5	8.7	4.4	4.5	8.9	4.9	5.0	9.9
55-59	3.5	3.7	7.1	3.2	3.3	6.5	4.1	4.2	8.2
60-64	2.9	3.1	6.0	2.8	2.7	5.5	3.2	3.1	6.3
65-69	2.2	2.0	4.2	1.7	1.7	3.4	1.7	1.4	3.1
70-74	1.9	1.7	3.6	1.6	1.6	3.2	1.2	1.0	2.3
75-79	1.5	1.5	3.0	1.3	1.4	2.7	0.9	1.0	1.9
80-84	0.9	1.1	2.1	0.9	1.1	2.0	0.5	0.7	1.3
85+	0.6	1.2	1.8	0.5	1.2	1.7	0.4	0.8	1.1
All ages	47.3	52.7	100.0	46.6	53.4	100.0	47.4	52.6	100.0
Median age			41.9			41.3			40.4

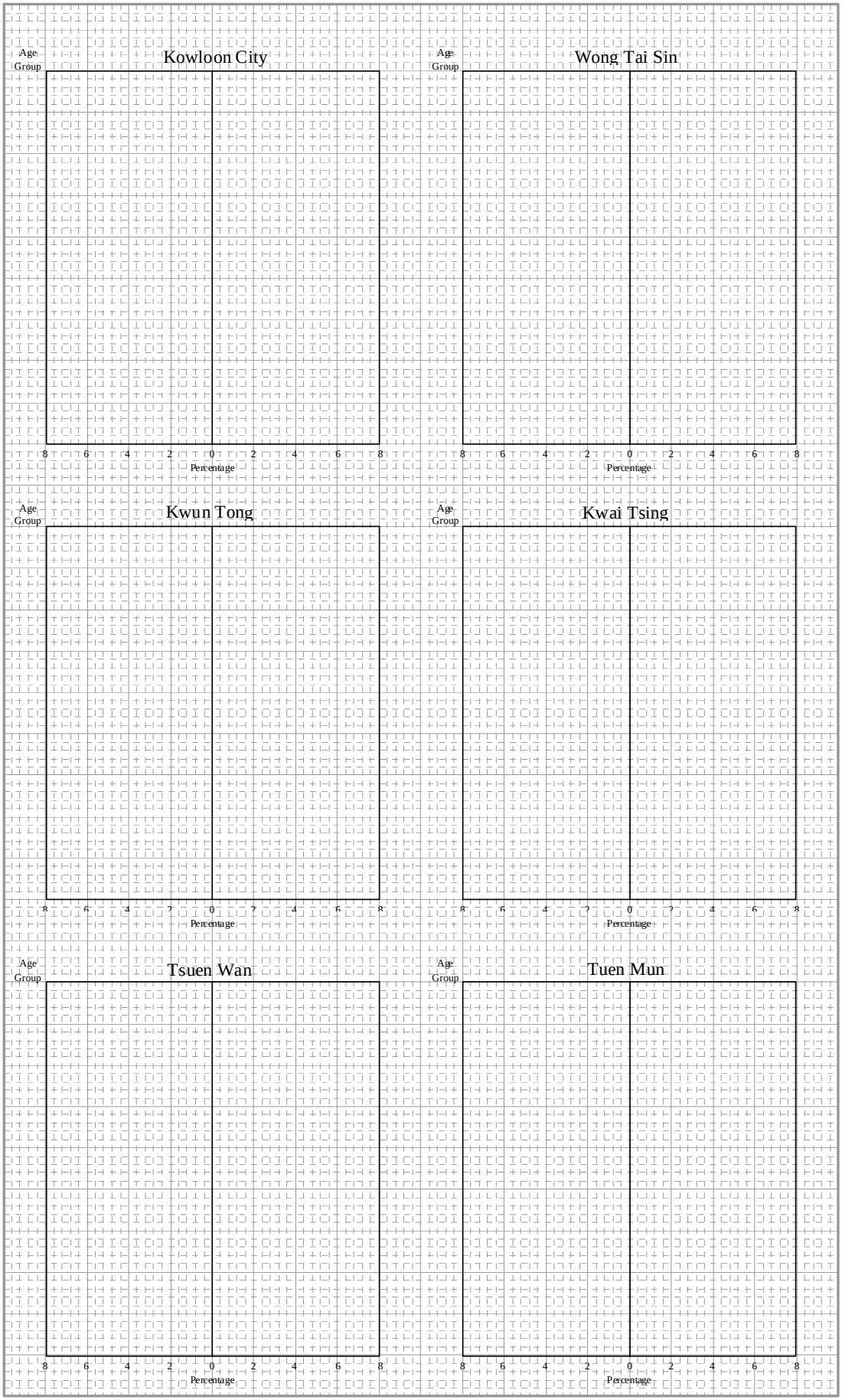
Percentage distribution of population by age and sex, June 2011 (cont'd)

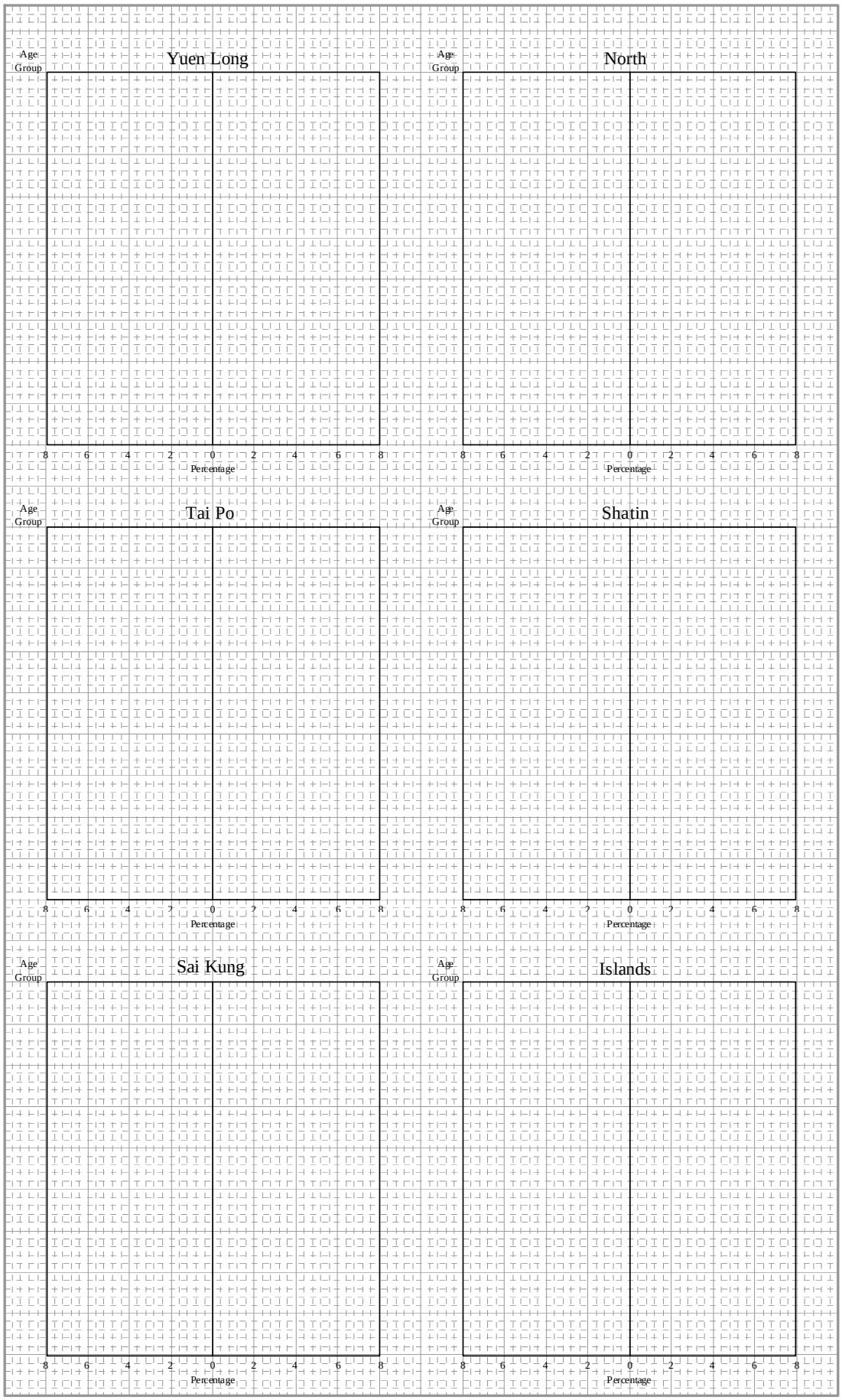
	Yuen Long			North			Tai Po		
Age group	Male	Female	Total	Male	Female	Total	Male	Female	Total
0-4	1.8	1.7	3.6	1.8	1.7	3.5	1.6	1.5	3.1
5-9	2.1	1.9	4.0	2.0	1.8	3.8	1.5	1.5	3.0
10-14	3.1	2.9	6.0	2.6	2.5	5.1	2.1	2.0	4.1
15-19	3.8	3.7	7.5	3.5	3.3	6.8	3.3	3.1	6.3
20-24	3.5	3.7	7.2	3.8	4.0	7.8	3.9	4.0	7.9
25-29	3.4	4.6	7.9	3.5	4.6	8.1	3.8	5.1	8.9
30-34	3.2	4.6	7.8	3.0	4.4	7.4	3.3	4.8	8.0
35-39	3.5	4.8	8.3	3.0	4.1	7.0	2.8	3.9	6.7
40-44	3.6	4.9	8.5	3.4	4.6	8.0	2.8	3.9	6.7
45-49	4.4	5.3	9.7	4.4	5.2	9.6	4.3	5.2	9.5
50-54	4.6	4.6	9.1	4.9	5.0	10.0	5.4	5.5	10.9
55-59	3.2	3.2	6.4	3.6	3.5	7.1	4.3	4.3	8.6
60-64	2.3	2.1	4.4	2.6	2.5	5.0	2.8	2.6	5.5
65-69	1.3	1.2	2.4	1.4	1.2	2.6	1.6	1.3	2.9
70-74	1.2	1.1	2.3	1.3	1.3	2.6	1.2	1.2	2.4
75-79	1.0	1.0	2.0	1.1	1.3	2.3	1.1	1.2	2.3
80-84	0.6	0.8	1.5	0.7	1.0	1.7	0.7	0.9	1.6
85+	0.4	0.9	1.3	0.5	1.0	1.5	0.5	1.0	1.5
All ages	46.9	53.1	100.0	47.2	52.8	100.0	46.9	53.1	100.0
Median age			38.6			40.2			41.4

	Shatin			Sai Kung			Islands		
Age group	Male	Female	Total	Male	Female	Total	Male	Female	Total
0-4	1.7	1.5	3.2	2.0	2.0	4.0	2.2	2.1	4.3
5-9	1.7	1.6	3.3	2.0	1.9	3.9	2.4	2.2	4.6
10-14	2.3	2.1	4.5	2.6	2.4	5.0	3.1	2.8	5.9
15-19	3.0	2.9	5.9	3.2	3.1	6.2	3.3	3.2	6.4
20-24	3.3	3.4	6.7	3.4	3.6	7.0	3.1	3.1	6.3
25-29	3.5	4.7	8.2	3.2	4.4	7.6	2.9	3.9	6.8
30-34	3.4	4.9	8.3	3.2	4.7	7.8	3.2	4.7	7.9
35-39	3.2	4.4	7.6	4.0	5.6	9.7	3.8	5.4	9.2
40-44	3.2	4.4	7.5	4.0	5.5	9.5	4.0	5.5	9.6
45-49	4.2	5.0	9.2	4.6	5.5	10.2	4.3	5.2	9.6
50-54	4.7	4.9	9.6	4.5	4.6	9.1	4.0	4.2	8.2
55-59	4.1	4.2	8.3	3.2	3.2	6.4	3.3	3.1	6.5
60-64	3.2	3.1	6.4	2.3	2.2	4.6	2.5	2.4	4.9
65-69	1.8	1.6	3.4	1.3	1.3	2.7	1.3	1.0	2.3
70-74	1.4	1.3	2.7	1.1	1.1	2.2	1.2	1.2	2.4
75-79	1.1	1.3	2.4	0.8	1.0	1.8	1.0	1.1	2.1
80-84	0.7	0.9	1.6	0.5	0.8	1.3	0.6	0.8	1.4
85+	0.4	1.0	1.4	0.3	0.7	1.1	0.5	1.0	1.5
All ages	46.8	53.2	100.0	46.6	53.4	100.0	46.9	53.1	100.0
Median age			41.5			39.3			39.1

Note : Individual percentage may not add up to 100 due to rounding.


Central and Western
Wan Chai
Eastern
Southern
Yau Tsim Mong
Sham Shui Po
Age Group
Percentage
8 6 4 2 0 2 4 6 8


Kowloon City
Wong Tai Sin
Kwun Tong
Kwai Tsing
Tsuen Wan
Tuen Mun
Age Group
Percentage
8 6 4 2 0 0 2 4 6 8


Yuen Long
North
Tai Po
Shatin
Sai Kung
Islands
Age Group
Percentage
8 6 4 2 0 2 4 6 8

3 Impact of population growth on society

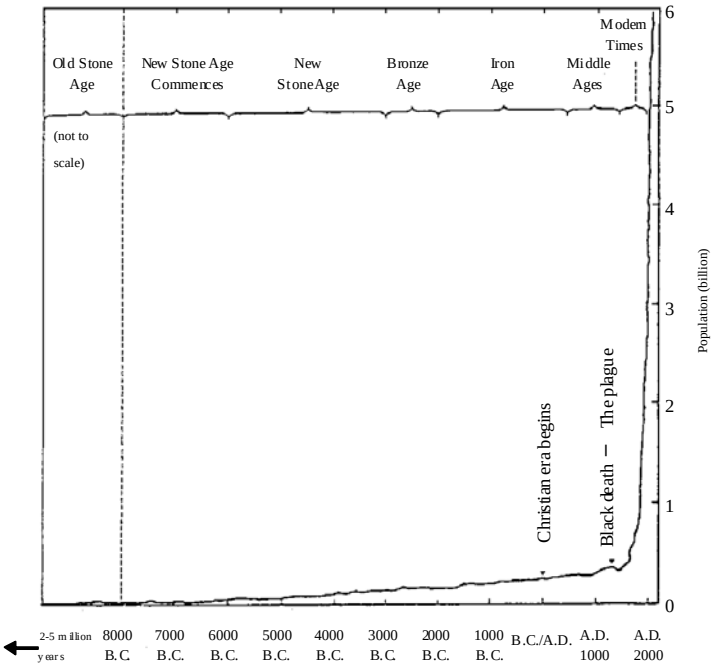
Introduction

This chapter discusses the impact of population growth on society. It shows how statistical data can be used to identify the problems caused. Population control and forward planning of service requirements are solutions to the problems. To carry out such work, population data of good quality are required.

The current population of the world is estimated to be around 6.9 billion by the United Nations Population Division, being mainly a result of an unprecedented growth of population in the twentieth century. This very large population size has caused much concern about the availability of resources to meet the demand of all the people for food, clothing, housing, education, medical services, etc.

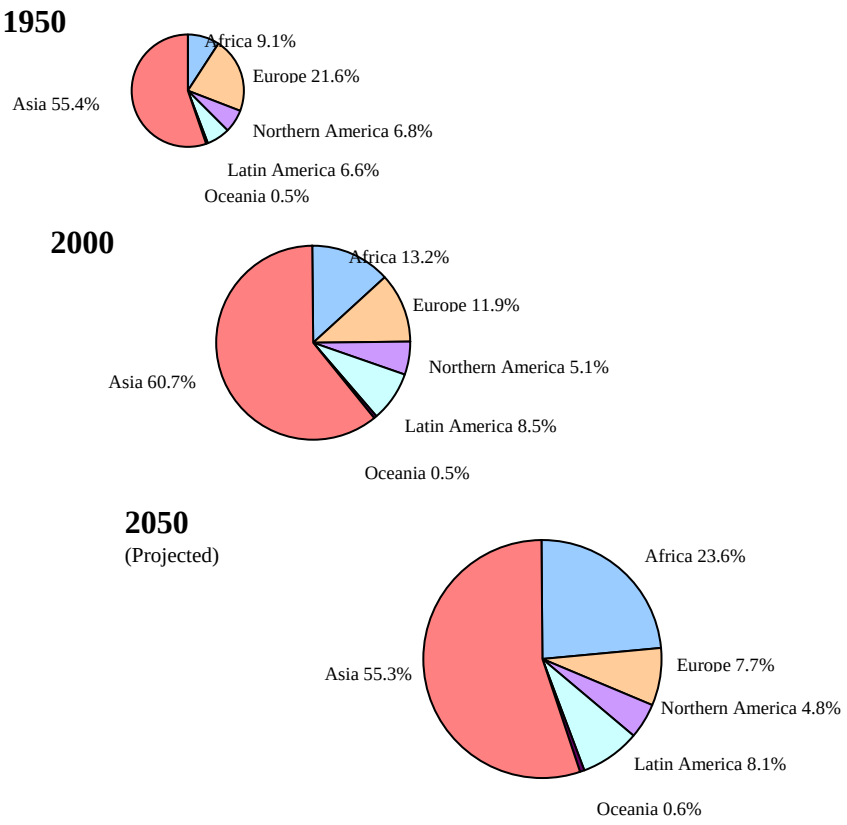
The great burden from rapid population growth is likely to persist in the near future. Latest population projections indicate that by 2050, the world population will increase to 9.1 billion. In particular, population problems will be more serious in some poorer areas or countries where populations grow faster than in more developed regions.

Chart 3.1 World population history



Source : Jean van der Tak, Carl Haub, and Elaine Murphy, "Our Population Predicament : A New Look," Population Bulletin, Vol. 34, No. 5 (Population Reference Bureau, Inc., Washington, D.C., 1979), p.2.

Chart 3.2 Distribution of world population by major area, 1950, 2000 and 2050[(1), (2)]



Notes :

(1) Latin America includes Central America, South America and the Caribbean.

(2) Area of pies is proportional to population size.

Source : World Population Prospects : The Population Database, Population Division of the of the United Nations Department of Economic and Social Affairs

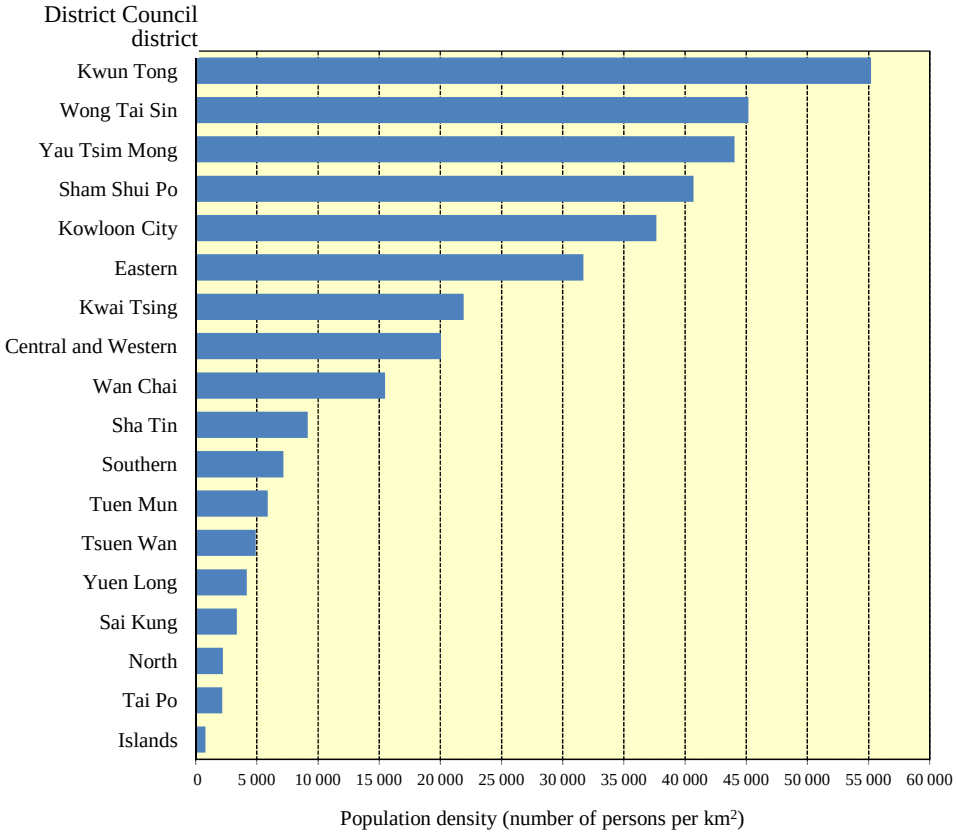
Impact of population growth and changes in population structure on Hong Kong

In Hong Kong, there are a number of concerns arising from the continuous growth of population and changes in its structure.

Increasing population density

With 7.07 million people (excluding marine population) living on its land area of only 1 104 square kilometres, Hong Kong is one of the most densely populated places in the world. In 2011, the territory as a whole had an average of 6 544 persons per square kilometre, compared to 6 237 people per square kilometre in 2001. For some areas, the population density was even higher. For instance, Kwun Tong, with 55 204 persons per square kilometre, was the most densely populated district in 2011.

Chart 3.3 Population density by District Council district, 2011



To reduce urban congestion and to meet the new housing needs of the growing population, systematic and coordinated development of new towns began in the early 70's. Since then, there has been rapid development of new towns in the more remote areas. At present, there are twelve new towns, namely, Tuen Mun, Sha Tin, Kwai Chung, Tai Po, Tsuen Wan, Fanling/Sheung Shui, Tsing Yi, Tseung Kwan O, Ma On Shan, Yuen Long, Tin Shui Wai and North Lantau.

The development of new towns has led to a marked re-distribution of the population, from the older urban areas on Hong Kong Island and in Kowloon to the new towns. Based on the results of the 2011 Population Census, about 171 853 persons had moved their homes from Hong Kong Island and Kowloon to the new towns during the period from 2006 to 2011. Similar movements of 275 309 persons during the period from 1996 to 2001 and 228 279 persons during the period from 2001 to 2006 were respectively estimated in the 2001 Population Census and the 2006 Population By-census. Based on the results of respective Population Census, people living in the new towns accounted for about 47% of the Hong Kong land population in 2011, against 45% in 2001.

Traffic congestion

The large population of Hong Kong generates a huge volume of traffic which, coupled with a lack of space, places a great strain on the internal transport systems.

To obtain an indication of traffic congestion, the number of motor vehicles can be compared with the total length of roads. As at end-2011, there were 302 motor vehicles per kilometre of road, compared with 275 as at end-2001. The increase, was mainly contributed by the increase in the number of private cars.

As in places all over the world, people of Hong Kong rely much on public transport services in their daily living. In 2011, public transport carried 4.3 billion passengers(3), compared to 3.9 billion in 2001.

In the past years, railways remained to be a major carrier among the various public transport modes. They comprise the Mass Transit Railway (MTR) heavy rail systems (which include nine local lines and the Airport Express Line), the Light Rail and the Hongkong Tramways. In 2011, the daily patronage of railways was 4.7 million passenger journeys, as compared with 3.5 million in 2001. Over the same period, the share for railways in total public transport journeys increased from 32.4% in 2001 to 39.7% in 2011.

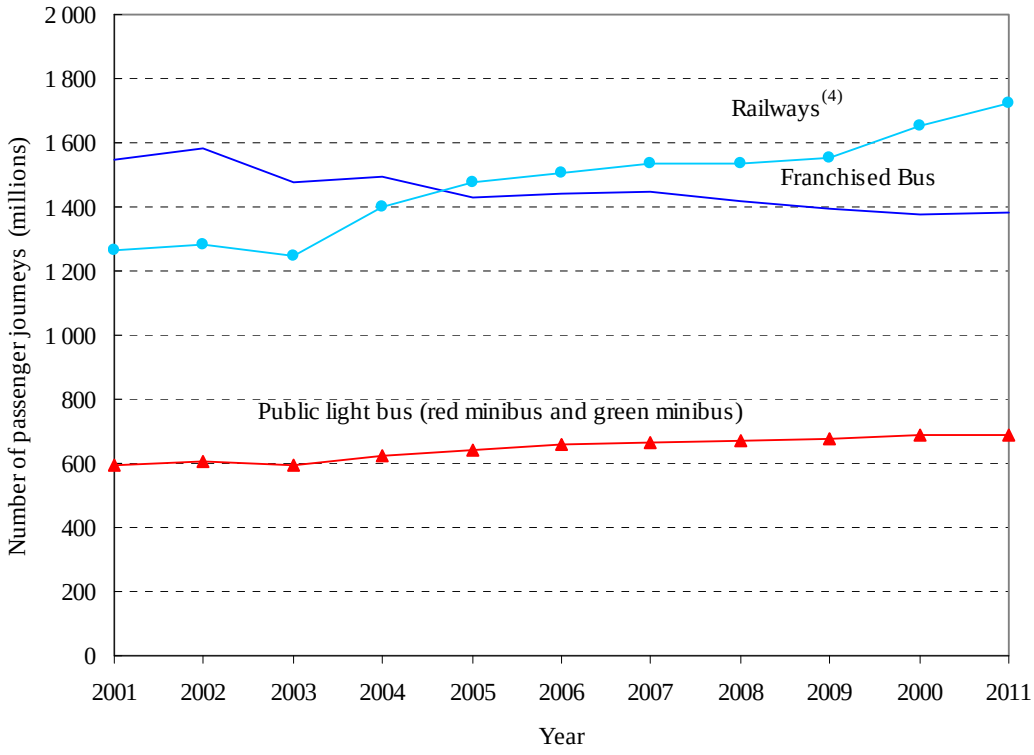
Another important part of the public transport system in Hong Kong is franchised buses. In 2011, they carried nearly 3.8 million passenger journeys or 31.8% of all public transport journeys a day. Compared with ten years ago, the daily patronage was 4.2 million passenger journeys, or 39.7% of the total public transport journeys.

To ensure the smooth and efficient movement of people and goods, careful co-ordination and management are needed. This involves a programme to improve the road network, expansion of public transport services and measures to achieve more economic use of the limited road capacity.

Note :

(3) The number of passengers cited in transport statistics actually refers to the number of passenger journeys. Since a person may make more than one journey during a period of time, the figure is usually larger than the number of people who have ever made journeys during the period.

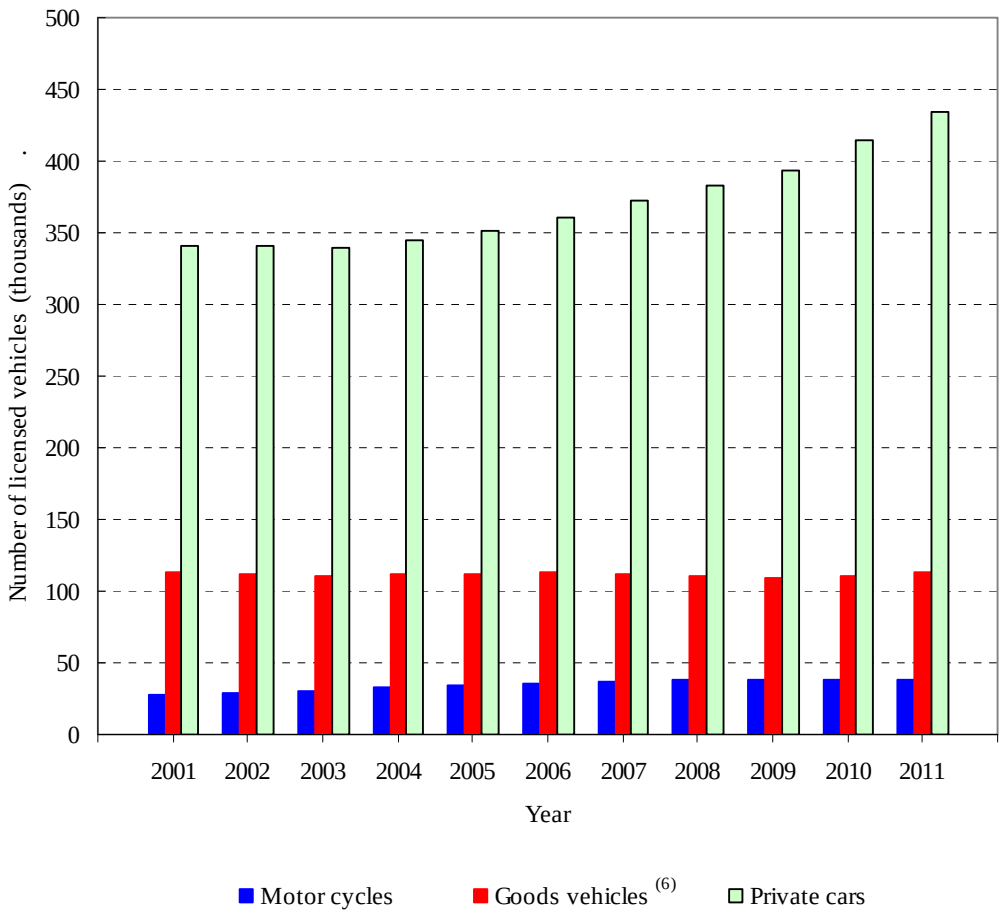
Chart 3.4 Passenger journeys by major mode of public transport, 2001–2011



Note :

(4) Railways comprise the Mass Transit Railway (MTR) heavy rail systems (which include nine local lines and the Airport Express Line), the Light Rail and the Hongkong Tramways.

Chart 3.5 Motor vehicles licensed, 2001–2011[5]



Notes :

(5) Figures refer to end of the year.

(6) Goods vehicles include Special Purpose Vehicles.

Increasing demand for housing

As Hong Kong is scarce of land, the government has to ensure that there is steady and sufficient supply of serviced land for meeting the demand for housing brought about by the continuous growth of the population. To achieve this, the government needs to develop strategic growth areas (including new and existing land for housing), renew urban areas, redevelop old districts, reclaim more land, perform better town planning and rezone agricultural and industrial land for housing development where appropriate.

During 2001 to 2011, about 424 900 residential flats were completed to meet the housing needs of people of Hong Kong. As at end of September 2011, the total stock of permanent quarters in Hong Kong was 2 603 000.

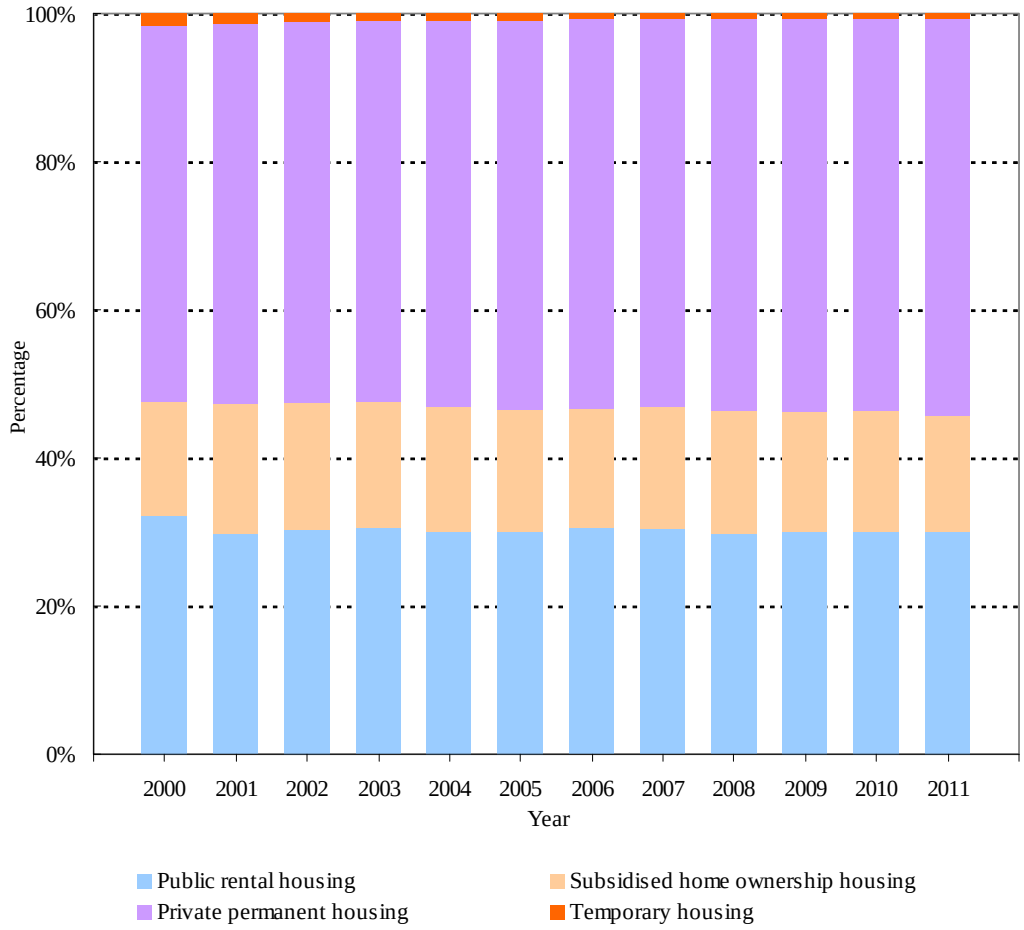
Of the 2.36 million domestic households in 2011, 30% resided in public rental housing and 53.4% in private permanent housing(7). While the proportion of domestic households living in subsidised home ownership housing(8) increased from 15.5% in 2000 to 15.8% in 2011, the proportion of domestic households in temporary housing declined from 1.7% to 0.8%.

Notes :

(7) Figures include private housing blocks, flats built under the Urban Improvement Scheme of the Hong Kong Housing Society, villas/bungalows/modern village houses, simple stone structures and quarters in non-residential buildings. As from 2002, subsidised sale flats that can be traded in open market are also put under this category.

(8) Subsidised home ownership housing includes flats built under the Home Ownership Scheme, Middle Income Housing Scheme, Private Sector Participation Scheme, Buy or Rent Option Scheme and Mortgage Subsidy Scheme, and flats sold under the Tenants Purchase Scheme of the Hong Kong Housing Authority. Also includes flats built under the Flat for Sale Scheme and Sandwich Class Housing Scheme of the Hong Kong Housing Society. As from 2002, subsidised sale flats that can be traded in open market are excluded.

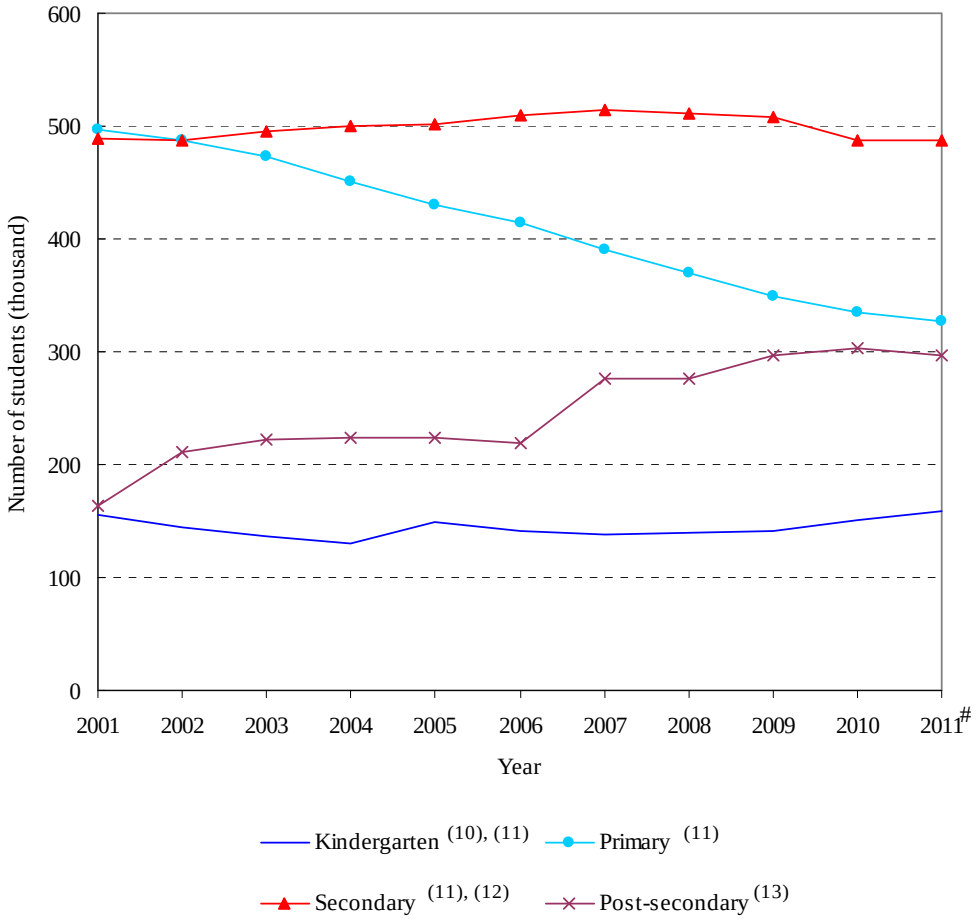
Chart 3.6 Distribution of domestic households by type of housing, 2000–2011



Increasing demand for post-secondary education

Despite the continuous growth of population size, the number of people aged below 15 was on the decline. Owing to such changes in the age structure of the population, the number of students studying in primary education showed a great decrease in the past ten years. On the other hand, there was a rapid rise in the number of post-secondary education students as a result of improvement in education opportunities. The overall post-secondary participation rate for the 17-20 age group has increased to over 60%, more than double the level of about 30% a decade ago.

Chart 3.7 Student enrolment by level of education, 2001–2011[9]



Notes :

(9) Figures include both full-time and part-time students attending long programmes lasting for at least one school/academic year.

(10) Figures from 2005 onwards include pupils attending kindergarten classes (i.e. nursery, lower and upper classes) in kindergarten-cum-child care centres upon harmonisation of pre-primary services in September 2005. Figures from 2008 onwards also include special child care centres registered under the Social Welfare Department.

(11) Figures include students in special schools.

(12) Apart from day schools and special schools, figures also include students attending evening schools, craft level courses and programmes of the Project Yi Jin.

(13) Figures include students attending universities and colleges offering post-secondary courses including certificate/diploma, associate degree or equivalent and bachelor degree or above; and also non-local registered or exempted courses leading to non-local higher academic qualifications and jointly operated with non-local institutions from 2002 onwards. Starting from 2007, figures also include all students attending self-financing programmes offered by the University Grants Committee-funded institutions and their extension arms.

Increasing elderly dependency ratio

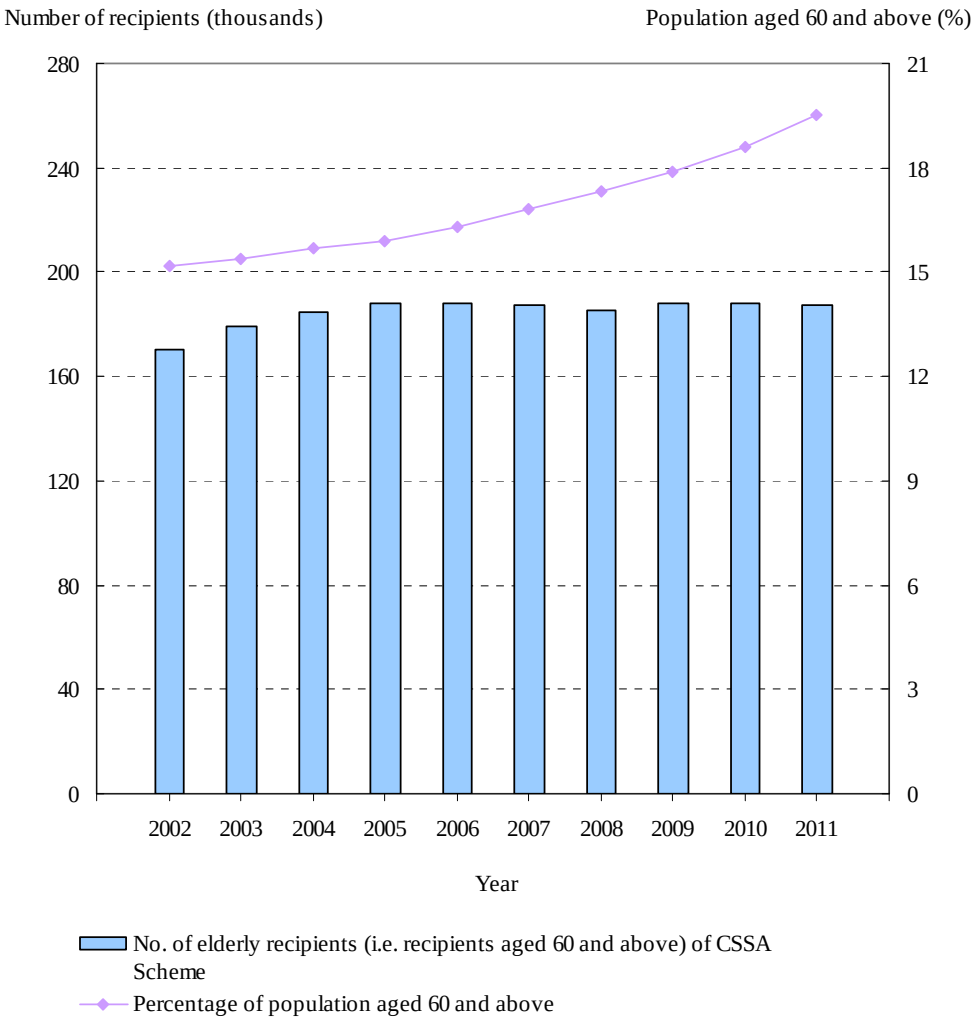
The population of Hong Kong is ageing, as can be seen from the median age of the population increasing from 33.6 in 1995 to 41.7 in 2011. As a result, the elderly dependency ratio increased from 138 per 1 000 population of working age in 1995 to 177 per 1 000 population of working age in 2011. It is common to consider the elderly as dependent on those aged between 15 and 64. The rise in the elderly dependency ratio reflects that there is an increase in the demand for social services, in particular social welfare and medical and health services.

Increasing demand for social welfare services

The percentage of population aged 60 and above increased gradually, resulting in increasing demand for social welfare services. The number of elderly recipients (i.e. recipients aged 60 and above) of the Comprehensive Social Security Assistance (CSSA) Scheme increased from 170 500 in end-2002 to 187 100 in end-2011. The CSSA Scheme is designed to provide members of the community who cannot support themselves financially with monetary assistance to meet their basic and special needs.

The government also provides a wide range of elderly and community services. Community services include the operation of social centres for the elderly, district elderly community centres, neighbourhood elderly centres, day care centres/units for the elderly, enhanced home and community care services, and integrated home care services.

Chart 3.8 Number of elderly recipients of the comprehensive social security assistance (CSSA) scheme and percentage of population aged 60 and above, 2002–2011(14)



Note :

(14) Figures refer to December of the year.

Population projections

For the purpose of forward planning of services, the likely size, structure and distribution of the population in future years must be projected. A reliable projection requires careful examination of past trends of fertility, mortality and movement to determine their possible changes in the future. The assumed changes are then added onto the base population, which is derived from a population census or by-census, year after year to obtain the projection results.

Population projections of Hong Kong are compiled by the Census and Statistics Department at intervals of two to three years to provide a common basis for government planning in various programme areas such as education, housing, transport, social services, medical and health services and infrastructural facilities. They are also widely applied in business development and planning in the private sector. The latest set of population projections was released in July 2010, with a 30-year projection period from mid-2010 to mid-2039.

"Component method" for population projections

The "component method", which is commonly used internationally, is adopted for making the population projections. Under this method, the size and age-sex structure of the projected population in each of the projection years are obtained by applying different assumptions made on the fertility, mortality and movement conditions for the projection year to the population situation in the preceding year. Therefore, in compiling the population projections, the population situation in a base year (i.e. mid-2009 in the latest set of projections) will be used as the starting point. It is then brought forward under separate projections of fertility, mortality and movement, year after year until the end of the projection period. While fertility, mortality and movement assumptions are formulated mainly on the basis of the past trends and recent developments pertaining to the socio-economic conditions in Hong Kong, government policies, if involved, are taken as that the existing policies will continue to apply.

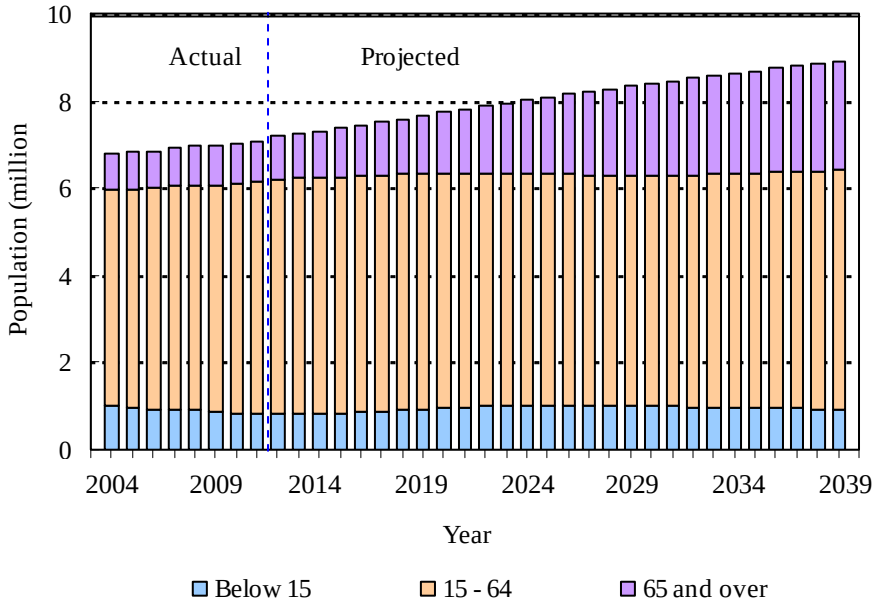
Major projection results

In the latest set of projections, the Hong Kong population is projected to increase at an average annual rate of 0.8%, from 7.00 million in mid-2009 to 8.89 million in mid-2039. Between 2010 and 2017, the annual growth rate of the population would accelerate somewhat, from 0.9% in 2010 to 1.0% in 2017. Thereafter, the annual growth rate is expected to slacken gradually, to 0.5% towards the end of the projection period.

Population on an ageing trend

The population is expected to remain on an ageing trend. The proportion of the population aged 65 and over is projected to rise markedly, from 13% in 2009 to 28% in 2039, though the rise would be gradual up to around 2019 (when the proportion would reach 17%) and would be at a much faster pace thereafter. Meanwhile, the proportion of the population aged under 15 would decrease gradually from 12% in 2009 to 11% by the end of the projection period.

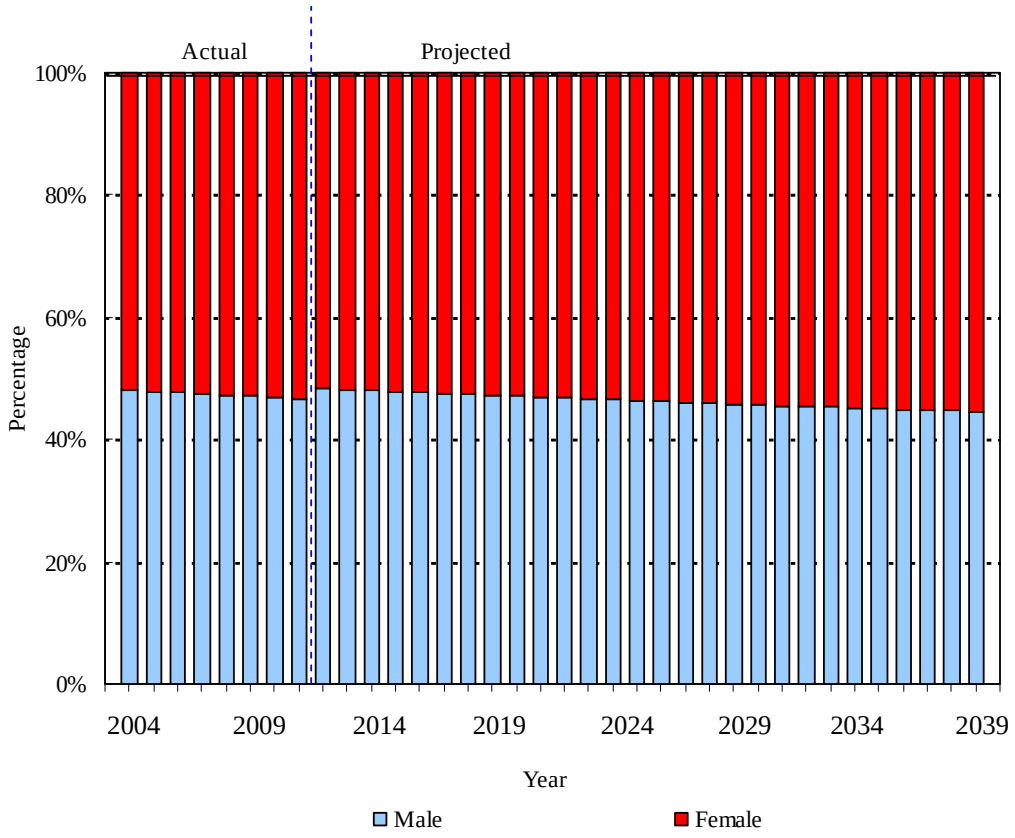
Chart 3.9 Hong Kong population by age group, 2004–2039



Decreasing sex ratio

Another noteworthy point is that the proportion of males in the population would keep on decreasing. In 2009, the sex ratio (i.e. the number of males per 1 000 females) of the population was 891. Due to the continued entry of One-way Permit Holders (many being Hong Kong men's wives in the Mainland) and foreign domestic helpers (mostly being younger females) into Hong Kong in the coming years and the fact that females in general live longer than males, the sex ratio is projected to fall noticeably to 744 in 2039. Even with the effects of foreign domestic helpers excluded, the sex ratio would still come down from the level of 957 in 2009 to 804 in 2039.

Chart 3.10 Proportion of male and female populations, 2004–2039 (including foreign domestic helpers)



Further information

The above contents present only part of the information produced by C&SD on the topics concerned. For further information regarding the topics discussed in this chapter (e.g. latest statistics, statistical reports, concepts and methods), please visit the following sections of the C&SD Website :

Population estimates
Population censuses and by-censuses
Population projections
Demographics
Labour force
Transport, communications and tourism
www.aaa.comHousing and property
Education
Social welfare
Health

Exercise

Identifying impact of population growth on society

To arouse the awareness of various population and social problems, readers can try to identify such problems in the districts they live, in the whole territory of Hong Kong, or in other countries.

For example, readers may consider the problem of traffic congestion in Hong Kong by looking up statistics compiled by the Census and Statistics Department or other government departments (e.g. Transport Department) in order to quantify the magnitude of traffic congestion in broad terms, and to consider how various problems can be solved. Examples of relevant statistics include number of motor vehicles licensed, length of roads and number of passengers using different kinds of public transport.

4 Survey methods

Introduction

Statistics on births, deaths and net movement used in making population projections are compiled from birth and death registration records and immigration clearance records. In addition, information on the latest size and age-sex structure of the population and on various other aspects of persons in it has to be obtained. This can be achieved by conducting a population census or by-census.

Administrative records, such as the registration records mentioned above, and survey results, such as those obtained through a population census/by-census, are the typical sources of data from which statistics are compiled.

Administrative records provide an economical and accurate data source. However, since they are primarily intended for certain specific purposes, they may not meet the statistical requirements, unless special provision has been made to cater for such requirements. Naturally, there are situations where meeting such requirements would mean distorting the original purposes of the records too much. In such circumstances we would have to give up the idea of generating the required statistics as a by-product of the administrative system, and resort to other means of data collection.

Surveys are another important means of data collection whereby the objects of our study are interviewed, observed or given questionnaires to fill in. If we design the survey such that all objects of study will be contacted, one way or another, the survey is specifically referred to as a "census". If only a sample of the objects is taken for contacting later on, then the survey is more clearly described as a "sample survey". The language in this aspect is, however, a little loose. When the word "survey" is used, quite often it is implied that only a sample is taken.

Census versus sample survey

In deciding on whether a complete census or sample survey should be carried out, a number of factors need to be considered. In studies in which the "population" (i.e. the group of persons or other objects we wish to study) is large, a proper sample survey, as compared to a census, is very useful for cutting down the resources required, including money and manpower, and for reducing the overall burden imposed on all the objects to be studied. We can also collect, summarise and analyse data more quickly when the information is urgently needed.

On the other hand, a complete census can be considered if very accurate information is required on small sub-divisions of the population under study, and when comprehensive data on a population are needed for forming a benchmark to support further statistical studies, such as projections.

Population census and by-census

In a population census, basic information is collected from all persons in the population of a country/territory while in a population by-census, only a sample of persons is covered. Usually a population census is conducted once every ten years. A by-census may be conducted in between two censuses to update the information obtained from the last census.

Because of its long history world-wide, and when there is not likely to be confusion, "population census" is abbreviated as "census" for convenience. Thus, it should be noted that there can actually be census of agriculture, census of industry and so on apart from population census.

Population censuses and by-censuses taken in Hong Kong

The history of population census in Hong Kong can be traced back to 1841. Earlier censuses were simple, with headcounts taken only. The 1961 Census was the first population census in which detailed information was collected on a broad spectrum of demographic, social and economic characteristics of the population. Since then, it has become a well-established practice in Hong Kong to conduct population censuses and by-censuses at regular intervals. Censuses were conducted in 1961, 1971, 1981, 1991, 2001 and 2011. By-censuses were conducted in 1966, 1976, 1986, 1996 and 2006.

Details of last population census are available in the section "The 2011 Population Census" in Chapter 1.

Objectives of conducting population censuses and by-censuses

In a dynamic city like Hong Kong, many changes in major characteristics of the population occur in a short span of time. Up-to-date and reliable information on the size, composition and geographical distribution of the population is necessary for forward planning by the government. Moreover, as Hong Kong develops into an international business centre, census and by-census data are widely used in assessing market conditions and making investment plans. Such data will also help researchers in studying important social issues.

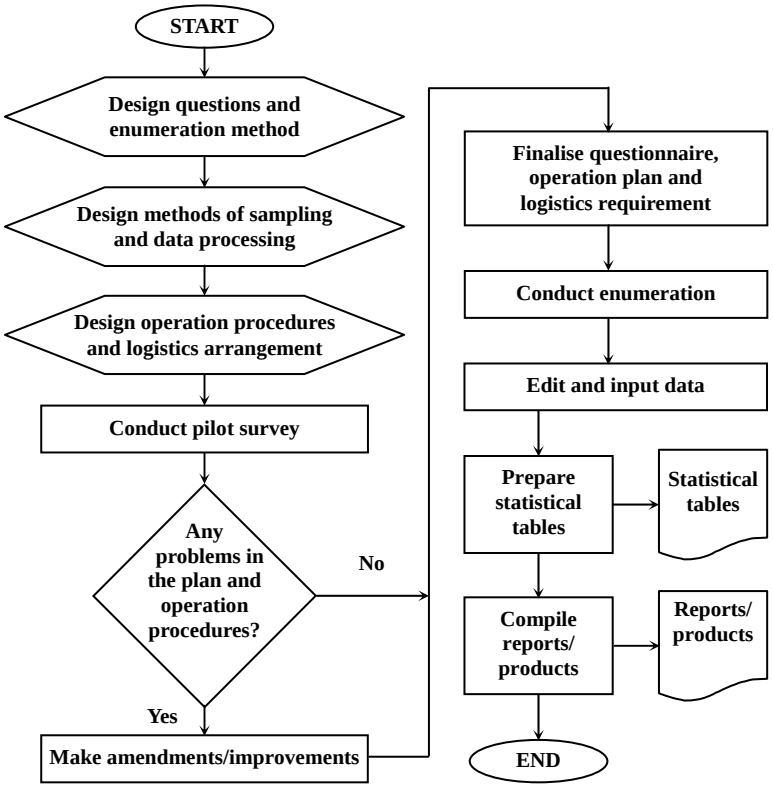
Preparation for conducting a census or by-census

Conducting a population census or by-census involves a vast amount of preparatory work. These preparatory activities include :

(1) geographical work, such as preparing maps and lists of places;
(2) determining the data needs of the government, business and other sectors;
(3) choosing the questions to be asked and the statistical tables to be compiled;
(4) designing the sampling methodology and method of estimation;
(5) deciding on the method of enumeration, e.g. whether field workers are to conduct household interviews or household members to fill in questionnaires by themselves; and
(6) planning the data processing procedures.

To ensure the effectiveness of the census/by-census operation plan, a pilot survey is usually conducted during the preparatory stage of every population census/by-census so as to test the questionnaire and operation procedures as well as to confirm the logistics requirement. Based on the results of the pilot survey, the operation plan is then amended/improved as appropriate for actual implementation. The pilot survey is particularly important for ensuring the smooth conduct of large scale surveys (such as population censuses/by-censuses) which involve the mobilisation of a large number of workers and other resources.

Chart 4.1 Main steps in planning and conducting a population census/ by-census



Census order

Furthermore, legislation is passed to provide a legal basis for conducting the census or by-census. Under the Census and Statistics Ordinance, the Census and Statistics Department (C&SD) is vested with the authority to conduct the population census or by-census. The legislation requires all persons to give particulars for themselves and their dwelling places in the data collection process. It also lays down penalties for those who refuse to do so. On the other hand, it contains strict provisions for safeguarding the confidentiality of the information supplied by individual households and persons. Such information will only be used for compiling statistics and will not be released to any parties, including other government departments.

Public support for the census/by-census

Although there are legal provisions, success of the census/by-census operation depends very much on the public's co-operation in providing adequate and accurate information. The public should realise that data collected from the census/by-census are essential to the future planning of Hong Kong and it is their civic duty to assist.

For the 2011 Population Census, a considerable number of students and teachers were recruited as temporary field workers to undertake various tasks. They thus had a chance to participate actively in this important project. Training on interviewing and other necessary skills was provided. Some school premises were used as field centres for the field operation.

Processing of census/by-census data

After raw data are collected from households and individuals in the population, they must undergo a number of processes before statistical outputs can be produced. The raw data are first edited to correct for inconsistent answers. They are then converted into codes for input into the computer. Finally, various compilations and computations are performed by the computer to produce statistical tables in appropriate formats.

Dissemination of census and by-census results

Results of the more recent rounds of census or by-census are published in the forms of summary reports and detailed reports. Users can refer to them easily and make use of the data for various purposes. In particular, data users can use the free-of-charge online Interactive Data Dissemination System to build their own statistical tables. Results of the 2011 Population Census are obtainable from the website of the 2011 Population Census.

Other surveys

Apart from population censuses and by-censuses, C&SD and other government departments also carry out other surveys to collect various types of social and economic data.

Some examples of surveys carried out by C&SD are :

- Monthly Survey of Retail Sales;
- Quarterly Survey of Employment and Vacancies;
- Annual Survey of Economic Activities;
- General Household Survey; and
- Household Expenditure Survey.

Examples of surveys carried out by other government departments are :

- Enrolment Survey (by Education Bureau);
- Survey of Drug Use among Students (by Security Bureau); and
- Health Manpower Survey (by Department of Health).

Important points to note in conducting surveys

In planning and conducting surveys, there are a number of major points which need to be carefully considered :

(A) Overall planning

(1) The objectives of the survey should be clearly specified.

(2) The population covered by the survey should be clearly defined with reference to the survey objectives. "Population", in the context of a survey, refers to the totality of objects under study. The objects may be human beings or otherwise.

(3) It should be ensured that respondents of the survey are able to provide the information required.

(4) Adequate resources such as time, money and manpower should be available for the survey.

(5) In conducting a survey, professional advice is often required and preferably should be sought at an early stage to ensure that the various steps of a survey are properly planned. This is because rectification is normally very difficult to make when the survey is completed and its results are found invalid.

(B) Design and selection of sample (in case a sample survey instead of a census is to be conducted)

(1) Probability sampling methods should be used, as far as practicable, rather than non-probability sampling methods. In probability sampling methods, every unit in the population has a known and non-zero probability of being selected. This makes it scientifically valid to draw inferences from the sample results about the entire population which the sample represents. Examples of probability sampling methods include simple random sampling, stratified random sampling and cluster sampling.

(2) Non-probability sampling methods such as haphazard sampling (i.e. picking the sampling units as they come along or as convenient), quota sampling (i.e. selecting sampling units according to some criteria, such as 50% being male and 50% being female, but yet the selection of units is at the discretion of the interviewers) and self-initiated telephone/online polling (i.e. interested individuals dial a telephone number/visit a website at their own initiative to give views) should be avoided because such methods normally induce bias to the survey results.

(3) The sample size depends on the characteristics of the population, the choice of method of sampling and the desired level of precision of estimates. It can be worked out using established statistical methodology. In general, a larger sample size is needed if estimates of higher precision are required.

(4) Substitution of sampling units or respondents should not be done once they have been selected according to a sample design.

(C) Design of questionnaire

(1) The questionnaire used for the survey should contain questions which are relevant, precise, unambiguous, well-sequenced and appropriately phrased. If the questions involve expression of opinions, care should be taken to avoid questions leading respondents towards the direction of a certain answer or questions being loaded in favour of a particular response.

(2) Requiring respondents to provide an answer while they are unable to do so or genuinely have no opinion will result in invalid answers. Answers like "don't know" or "no opinion" should be allowed in the questionnaire, as appropriate.

(3) The questionnaire should not be too long.

(4) To ensure the effectiveness of the questionnaire, it should be tested on potential respondents.

(D) Collection of raw data

(1) The mode of data collection, such as the use of personal interviewing, telephone interviewing, self-administered questionnaires or postal questionnaires, should be carefully selected by considering the respondents' willingness to co-operate, the degree of complexity of the subject of enquiry and other relevant factors.

(2) Various arrangements related to fieldwork should be carefully planned in order to ensure smooth and efficient operation. Fieldwork procedures should be thoroughly tested before implementation.

(3) The interviewers should be carefully briefed on the concepts and definitions of terms used in the survey and properly trained in procedures before they start working.

(4) Interviewers should ensure that respondents understand the questions. Probing should be done where necessary. This, however, should not be overdone otherwise unnecessary influence will be exerted on respondents.

(5) Every effort should be made to minimise non-contacts with respondents or refusals to respond because a low response rate may induce a serious bias in the survey results. Adequate publicity of the survey, proper introduction of the surveying organisation, clear identification of interviewers, and giving advance notice to sampled respondents should help. Arrangements should also be made to follow up on non-contact respondents and to persuade uncooperative respondents to participate.

(6) The identity and information supplied by individual respondents should be kept confidential. With such assurance, respondents would feel at ease and would be more willing to provide individual or organisational information and opinions. It should be emphasised that sample survey is a kind of statistical work and not an investigation on individuals.

(E) Processing of raw data and compilation and analysis of statistics

(1) Raw data should be carefully and thoroughly checked before compilation of statistics.

(2) Appropriate statistical methodology should be adopted in compiling and analysing the data. In particular, in case a sample design has been used where the probability of selection of every individual respondent is not equal, it is necessary that proper weighting methods are applied to survey results to give unbiased population estimates.

(3) Due deliberation should be made to decide on an appropriate sub-grouping scheme in analysing survey results at some disaggregate levels, taking into account the precision of estimates derived from the sample.

(F) Dissemination of survey results

(1) It is the usual practice to publish the survey results if the subject of the survey is related to a topic of public interest.

(2) The survey results should be presented in the form of aggregate statistics so that information supplied by individual respondents would not be disclosed.

(3) A complete survey report should contain details about different aspects of the survey, in particular details on population coverage, sample design, sample size, sampling error, response rate and likely sources of non-sampling error. Where appropriate, specimen of the questionnaire should be attached.

(4) When the survey results are released to the media, sufficient details on the survey methodology should be provided in addition to survey findings in order that the media may report both. Otherwise, members of the public will have no basis to assess the reliability of those findings.

Assessing quality of survey results

There are many factors that are crucial to the quality of data obtained from a sample survey. In assessing the data quality of a survey, at least the following questions should be considered :

(1) Whether the sample of the survey is selected in accordance with a probability sampling method;
(2) Whether the sample size is sufficiently large;
(3) Whether the questionnaire is well designed; and
(4) Whether the response rate of the survey is sufficiently high.

A well-designed and properly selected sample is crucial to the quality of results derived from a survey. Nevertheless, some survey-takers may not appreciate the need for conducting surveys scientifically. "Street-corner surveys" and "self-initiated telephone/online polling" that we frequently come across are examples of non-scientific surveys.

In "street-corner surveys", interviewers pick on passers-by haphazardly. Even though sometimes there are stipulations such as 50% of interviewees being male and 50% being female (such selection procedures are known as "quota sampling"), there is still much discretion for the interviewer to pick the interviewees within such stipulations. Thus, interviewees who seem to be more approachable will have a much greater tendency to be selected. Moreover, out of convenience, busy street corners are often chosen for interviewing and thus people who like to avoid busy spots will have little chance of being selected. Therefore, samples taken are not representative of the whole population.

Also, respondents of "self-initiated telephone/online polling" are not selected through an objective method. Because of this, we can expect that people who tend to be more concerned about the subject and have stronger views will be more willing to call/go online while others may not. Obviously, data so obtained cannot be taken for granted to represent the target population.

In a scientific survey, the sample should be selected from a well-defined population by probability sampling methods. Only when probability sampling methods are employed, statistical inference based on the data obtained from the sample can be made.

Apart from using a scientific sampling method, the size of the sample (i.e. the number of selected respondents) of a survey should be sufficiently large. Otherwise, the results of the survey may be subject to substantial error.

In surveys, data are very often collected by way of a questionnaire, though sometimes by observations. Hence, the design of the questionnaire will affect directly the quality of data to be collected. As an example, consider the question : "Many people are against smoking. Do you smoke?" This is obviously a leading question. In fear of counteracting the social norm, a respondent may have a tendency of giving a negative response even though he or she is a habitual smoker. The accuracy of data obtained from such a question is dubious.

Last but not least, the response rate achieved in a survey is also an important indicator that helps assess the quality of survey results. Response rate is the percentage of selected respondents who are successfully interviewed. Low response rate will introduce non-response bias to the data because the characteristics of the non-respondents may be quite different from those of respondents. Therefore, the results of surveys with a low response rate are likely to be biased and misleading.

References

Croxton, Frederick E., Applied General Statistics, Chapter 2, Prentice-Hall, 1975.

Ilersic, A.R. and Pluck, R.A., Statistics, Chapter XVI, HEL Ltd., 1979.

Moser, C.A. and Kalton, G., Survey Methods in Social Investigation, Chapter 2, Heinemann Educational Books Ltd.,1979.

Hong Kong Statistical Society, A Practical Guide to Sample Surveys, 1992.

Exercise

Conducting surveys

Readers can design and conduct a survey by drawing reference to the following illustrative example on a school survey.

(1) Listing and sampling

There are several possible methods :

(i) A list of all students in the school is prepared. All students are going to be surveyed. This is a “census”.

(ii) The above list is prepared. By drawing lots, a pre-set number of students is selected. (A more advanced treatment is to use “random” numbers but lot-drawing can achieve the same purpose.) This is a sampling method called “simple random sampling”.

(iii) Another convenient, yet scientific method is to make a list of all classes, randomly select a pre-set number of classes by lot-drawing, and survey all students in the selected classes. This method is called “cluster sampling”.

Methods (ii) and (iii) above involve “random” selection, which should be distinguished from haphazard choice of students or classes. The students selected from the methods would form a representative sample for the whole school.

(2) Questionnaire design

The type of information to be collected has to be decided first. Subsequently questions are designed to collect the required information. The information would usually include basic data about the students, such as their age, sex, class, etc. The primary questions may focus on some of their characteristics or about their opinions in certain respects. A specimen questionnaire is provided herewith, collecting information on students’ habit of watching TV. More items may be added at wish.

(3) Data collection and editing

To collect the information, students may interview respondents or distribute the questionnaires for self-completion by the respondents themselves. If the interviewing method is used, the interviewers should be trained to conduct interviews properly, ensuring that the questions are understood by respondents and complete answers are supplied. If self-completion method is used, questionnaires normally have to be simple and fully self-explanatory. The questionnaires are then edited. If there are errors, they have to be rectified.

(4) Data processing and analysis

Data in the questionnaires are then processed, with the aid of the computer if necessary. Statistical tables are produced. Graphs may also be prepared for further discussion and analysis. Some simple statistical indicators may also be calculated from the data collected.

Below are more hints in organising the school survey：

(i) To set the questionnaire, try to limit the number of questions to no more than some 15 questions. Keep it concise yet precise. Avoid open-ended questions in the simpler survey.

(ii) Before carrying out the actual school survey, try to test the questionnaire with some potential respondents to see whether it is clear and easy to understand, and whether respondents are able to provide the information.

(iii) It is necessary to make certain that every selected student has completed one and only one questionnaire. Follow up with omissions and rectify duplications.

(iv) It is necessary to instill in the interviewers a sense of respect for private information. They should be aware of the need not to reveal to any unauthorised person the information with regard to individual persons obtained in the interviews. (A small ceremony may be held in which interviewers will take an oath on this obligation. This may also boost respondents' confidence in giving true information.)

(v) To process the data, students can work in groups for coding, editing, counting, computing, etc.

(vi) Editing is a process to correct inconsistencies and to eliminate omissions found in answers on the returned questionnaires.

(vii) There are many ways to present the findings. For example, obtain simple frequency count of each variable and present the frequency distributions in the form of bar graphs, line graphs or pie charts.

(viii) To find out relationship amongst different variables, it may be interesting to cross-tabulate some of them as follows：

- number of students by type of housing by monthly family income;
- number of students by type of housing by number of TV sets;
- number of students by type of housing by number of hours spent on watching TV; and
- number of students by age by the most favourite kind of TV programmes.

Points to note

The “population” in the school survey (whether a census of the students or a sample survey of them) is comprised of students of the school in question. The findings are applicable to students of the school (if the survey is conducted properly). One should be careful not to generalise the findings to, say, students of the district where the school is situated or even to all students of Hong Kong. If we do wish to find out facts about all students of Hong Kong, the design of the survey will have to be more complex and obviously the operation will be much bigger in scale.

Survey questionnaire

Serial No. ：_____

Tick the appropriate answers.

(1) Age：_____ (in completed years)

(2) Sex： 1 ☐ Male 2 ☐ Female

(3) No. of family members (including yourself)：_____

(4) Type of housing
- 1 ☐ Public rental housing
- 2 ☐ Subsidised home ownership housing
- 3 ☐ Private permanent housing
- 4 ☐ Non-domestic housing
- 5 ☐ Temporary housing

(5) Monthly family income：
- 1 ☐ Under $6,000
- 2 ☐ $6,000 - $14,999
- 3 ☐ $15,000 - $29,999
- 4 ☐ $30,000 - $59,999
- 5 ☐ $60,000 and over

(6) How many TV sets do you have at home?
- 1 ☐ One
- 2 ☐ Two
- 3 ☐ Three and more
- 4 ☐ None

(7) How many hours did you spend on watching TV during last week?
(note：include videotapes and discs (including VCD, DVD, etc.) as well; at home or elsewhere)
- 1 ☐ Less than 5 hours
- 2 ☐ 5 – less than 10 hours
- 3 ☐ 10 – less than 20 hours
- 4 ☐ 20 – less than 30 hours
- 5 ☐ 30 hours and more
- 6 ☐ Not at all

(8) What kind of TV programmes do you like most?

1 ☐ Cartoon
2 ☐ Drama series
3 ☐ Musical show
4 ☐ Variety show
5 ☐ Sports
6 ☐ News
7 ☐ Documentary
8 ☐ TV movie
9 ☐ Others (please specify) _____

5 Uses and misuses of statistics

Introduction

"Statistics" has two aspects : First, as information in quantitative form relating to an aggregate group, i.e. the numbers themselves; second, as the science of dealing with such information, i.e. the principles and methods which have been developed for handling data. Information pertaining to an individual element – be it a person, firm or other subject of interest – is, on its own, normally not referred to as "statistics". In the context of statistical work, such information is called raw data, which become statistics only after going through some compilation or processing steps such as summation or calculation of average. Hence, the two aspects of statistics, as information and methodology, have interlocking relationships.

Statistics are playing an increasingly important role in various fronts. They are widely used in business, education, science, engineering and economic and social research. It is also a useful tool to facilitate better understanding of the society and discussion of socio-economic issues in a scientific manner.

Although statistics are useful, it is important to use them properly. Statistical data have to be of adequate quality before they are put to use and this relies on proper methods of collection and compilation. When interpreting and analysing statistics, appropriate techniques should be applied, or else wrong conclusions may be arrived at.

Uses of statistics

(1) Description of situations

People are always looking for relevant information in a form which can help understand the circumstances at work or in their daily lives. For example, a headmaster observes that students of the S1 class seem to be taller than those of the same class five years ago. (This is a piece of qualitative information as it represents the perception of a phenomenon.)

To ascertain this, information on the height of each student in the S1 class and those in the same class five years ago are studied. However, there are hundreds of students and going over the records alone may still give only an impression of the students' heights in general, leading to merely a broad confirmation of the earlier impression.

The most common way to tackle this kind of situation is to calculate the "mean" heights for the two groups. The summarisation of individual students' data to an aggregate form, i.e. "statistics", is a much more precise, quantitative description of the situation.

(2) Comparison

Comparison is to assess whether two or more groups of items have different properties, e.g. comparison between districts, between countries, or between different points of time.

In the above example of students' heights, the headmaster can compare the average heights of the two groups of students and then make statements on which group of students are as a whole actually taller, and by how much.

(3) Detection of relationship

e.g. relationship between smoking and having lung cancer, relationship between income and expenditure

(4) Evaluation

Evaluation of the effectiveness of action programmes, e.g. Is the "Water Conservation Campaign" successful?

(5) Prediction

Prediction about the future based on existing trends and knowledge of relevant developments, e.g. sales forecast and population projection.

(6) Support for policy formulation

e.g. data on current and future road traffic are used in a study to decide whether a highway should be constructed

(7) Operational and process control

e.g. quality control of manufactured goods

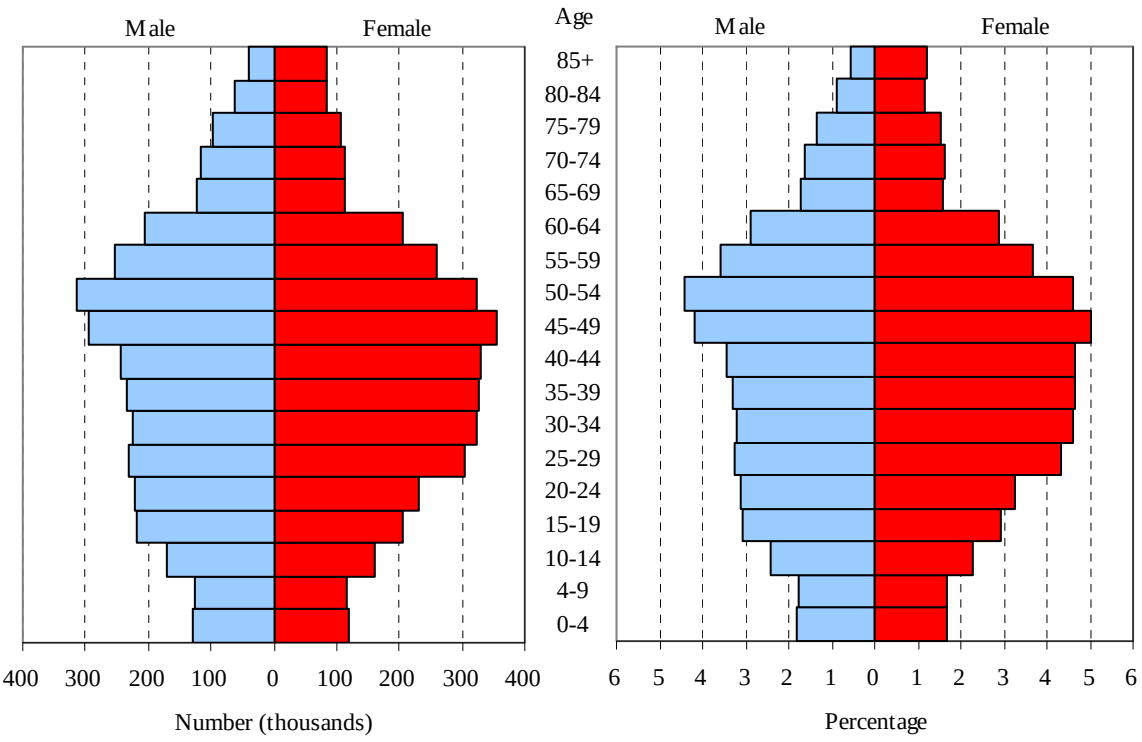
Social statistics

Social statistics are concerned with such subjects as the social environment, demographic characteristics of people, their activities and their opinions and attitudes.

Some important social statistics are :

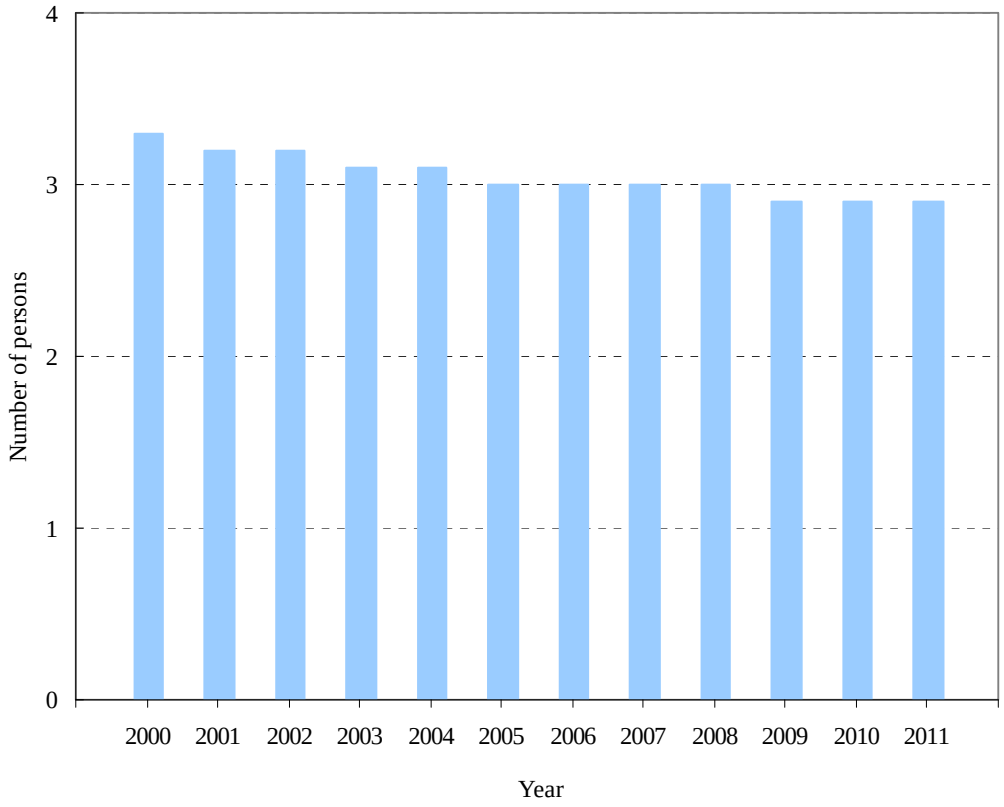
(1) Demographic characteristics
- population size
- population composition
- births
- deaths
- net movement
- marriages

Chart 5.1 Population of Hong Kong by age and sex, mid-2011



(2) Domestic households[1]
- household size
- household composition
- household income

Chart 5.2 Average domestic household size, 2000–2011

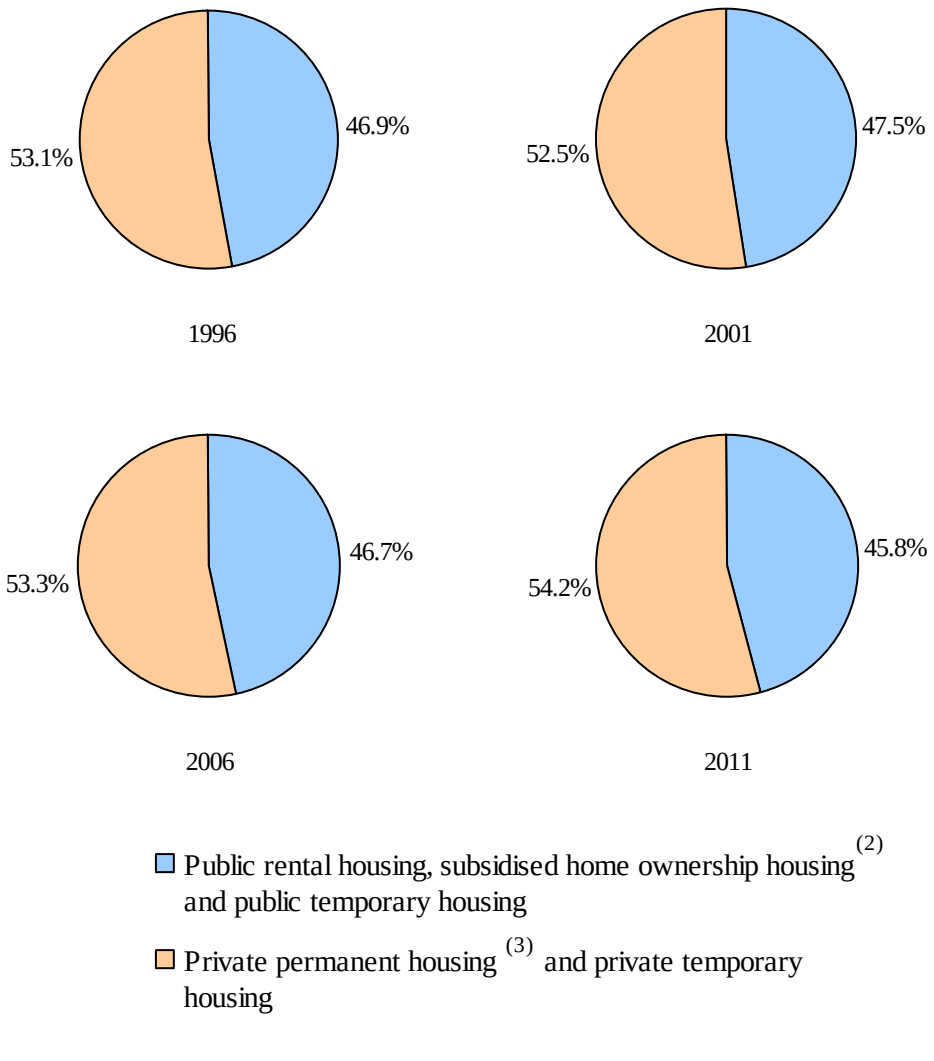


(3) Housing
- stock of permanent quarters
- housing facilities
- housing type
- degree of sharing
- tenure of accommodation
- rent

Note :

(1) A domestic household consists of a group of persons who live together and make common provision for essentials for living (e.g. sharing food and utensils). These persons need not be related. If a person makes provision for essentials for living without sharing with other persons, he/she is also regarded as a household. In this case, the household is a “one-person household”.

Chart 5.3 Percentage distribution of domestic households by type of housing, 1996–2011



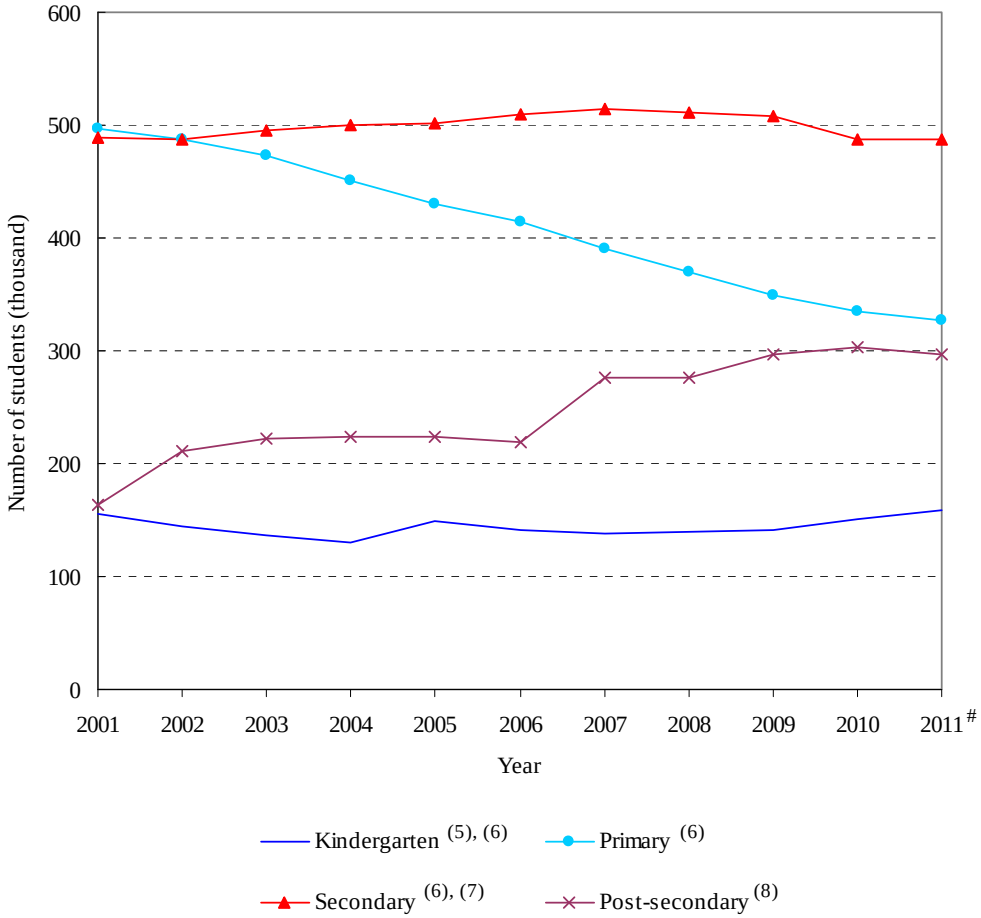
Notes :

(2) Subsidised home ownership housing includes flats built under the Home Ownership Scheme, Middle Income Housing Scheme, Private Sector Participation Scheme, Buy or Rent Option Scheme and Mortgage Subsidy Scheme, and flats sold under the Tenants Purchase Scheme of the Hong Kong Housing Authority. Also includes flats built under the Flat for Sale Scheme and Sandwich Class Housing Scheme of the Hong Kong Housing Society. As from 2002, subsidised sale flats that can be traded in open market are excluded.

(3) Figures include private housing blocks, flats built under the Urban Improvement Scheme of the Hong Kong Housing Society, villas/bungalows/modern village houses, simple stone structures and quarters in non-residential buildings. As from 2002, subsidised sale flats that can be traded in open market are also put under this category.

(4) Education
- educational attainment of population
- schools
- teachers
- student enrolment

Chart 5.4 Student enrolment by level of education, 2001–2011[(4)]



Notes :

(4) Figures include both full-time and part-time students attending long programmes lasting for at least one school/academic year.

(5) Figures from 2005 onwards include pupils attending kindergarten classes (i.e. nursery, lower and upper classes) in kindergarten-cum-child care centres upon harmonisation of pre-primary services in September 2005. Figures from 2008 onwards also include special child care centres registered under the Social Welfare Department.

(6) Figures include students in special schools.

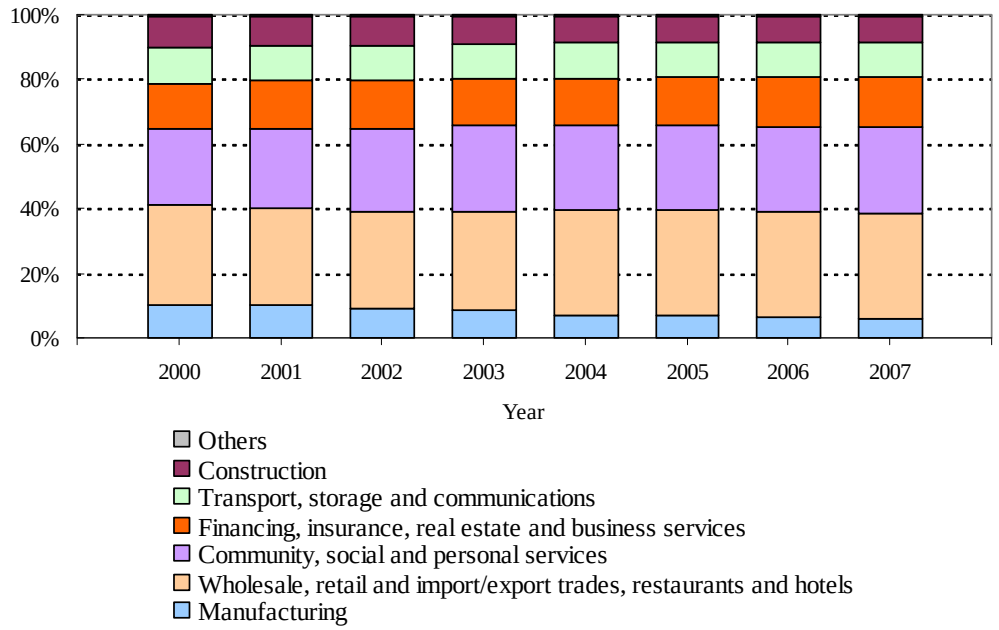
(7) Apart from day schools and special schools, figures also include students attending evening schools, craft level courses and programmes of the Project Yi Jin.

(8) Figures include students attending universities and colleges offering post-secondary courses including certificate/diploma, associate degree or equivalent and bachelor degree or above; and also non-local registered or exempted courses leading to non-local higher academic qualifications and jointly operated with non-local institutions from 2002 onwards. Starting from 2007, figures also include all students attending self-financing programmes offered by the University Grants Committee-funded institutions and their extension arms.

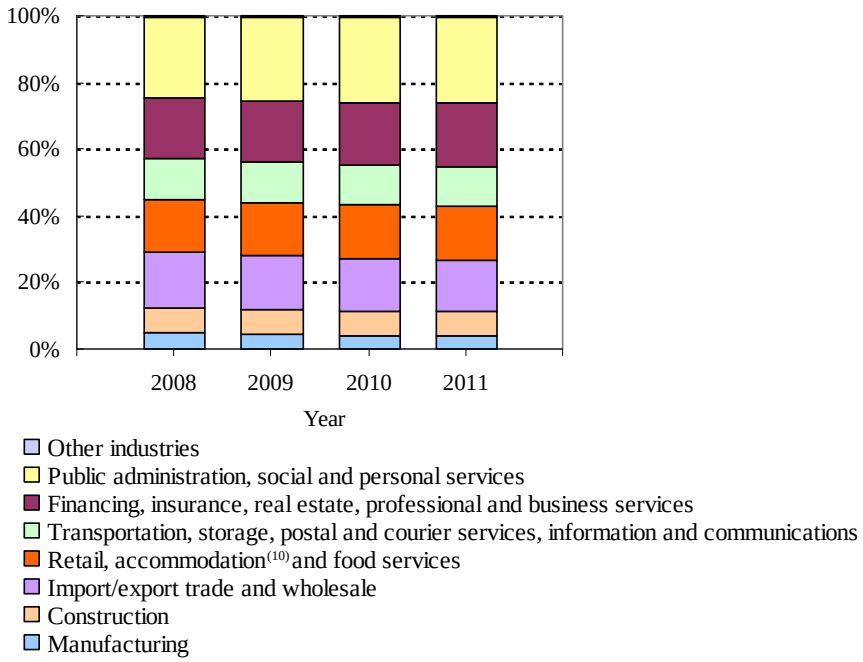
(5) Employment
- labour supply
- unemployment and underemployment
- distribution of employed persons by industry sector
- distribution of employed persons by occupation
- hours worked
- earnings from employment

Chart 5.5 Distribution of employed population by industry[9], **2000–2011**

(a) Based on Hong Kong Standard Industrial Classification (HSIC) Version 1.1



(b) Based on HSIC Version 2.0



Notes:

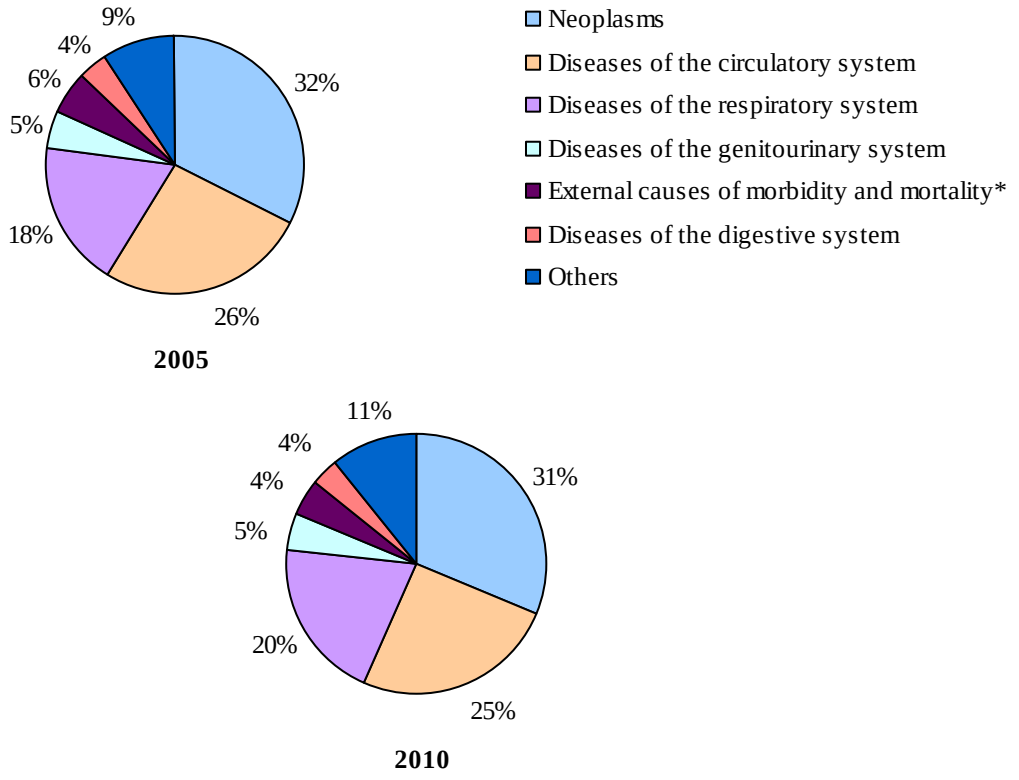
(9) Statistics for 2001-2007 are compiled based on the "Hong Kong Standard Industrial Classification (HSIC) Version 1.1". Statistics for 2008 and onwards are compiled based on "HSIC Version 2.0". Hence, figures for 2008 and onwards are not directly comparable to earlier figures.

(10) Accommodation services cover hotels, guesthouses, boarding houses and other establishments providing short term accommodation.

(10)

(6) Medical and health
- healthcare facilities and services
- health manpower
- patients
- diseases
- causes of death

Chart 5.6 Registered deaths by cause, 2005 and 2010[(11), (12)]



Notes :

(11) Classification of diseases and causes of death is based on the International Statistical Classification of Diseases and Related Health Problems (ICD) 10th Revision from 2001 onwards. Figures from 2001 onwards may not be comparable with figures for previous years which were compiled based on the ICD 9th Revision.

(12) The percentages may not add up to 100% due to rounding.

* According to the ICD 10th Revision, when the morbid condition is classifiable under Chapter XIX as "injury, poisoning and certain other consequences of external causes", the codes under Chapter XX for "external causes of morbidity and mortality" should be used as the primary cause of death.

(7) Social welfare
- beneficiaries and benefits
- services for the elderly
- family and child welfare services
- services for young people
- rehabilitation and medical social services
- institutional services

(8) Law and order
- crimes
- victims
- judicial services
- prisoners/inmates and correctional services

(9) Culture, entertainment and recreation
- cultural and recreational facilities
- leisure, entertainment and sports presentations

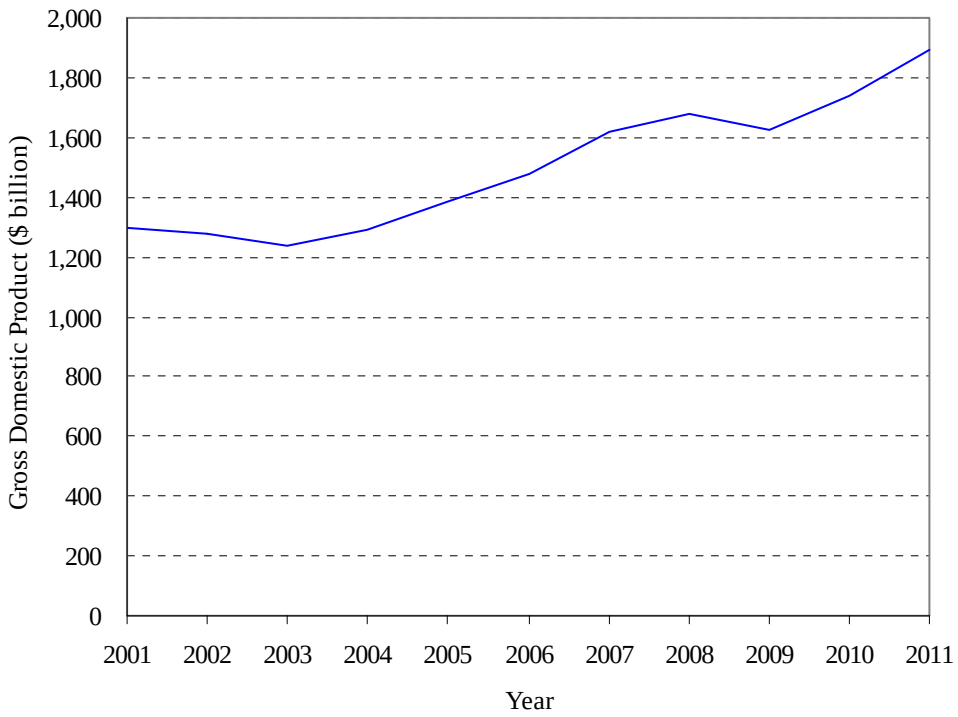
Economic statistics

Economic statistics are statistics related to the structure and performance of an economy and the operating characteristics of different industries. Some statistics can be treated as either economic or social statistics under different analytical frameworks (e.g. wages, employment).

Some important economic statistics are :

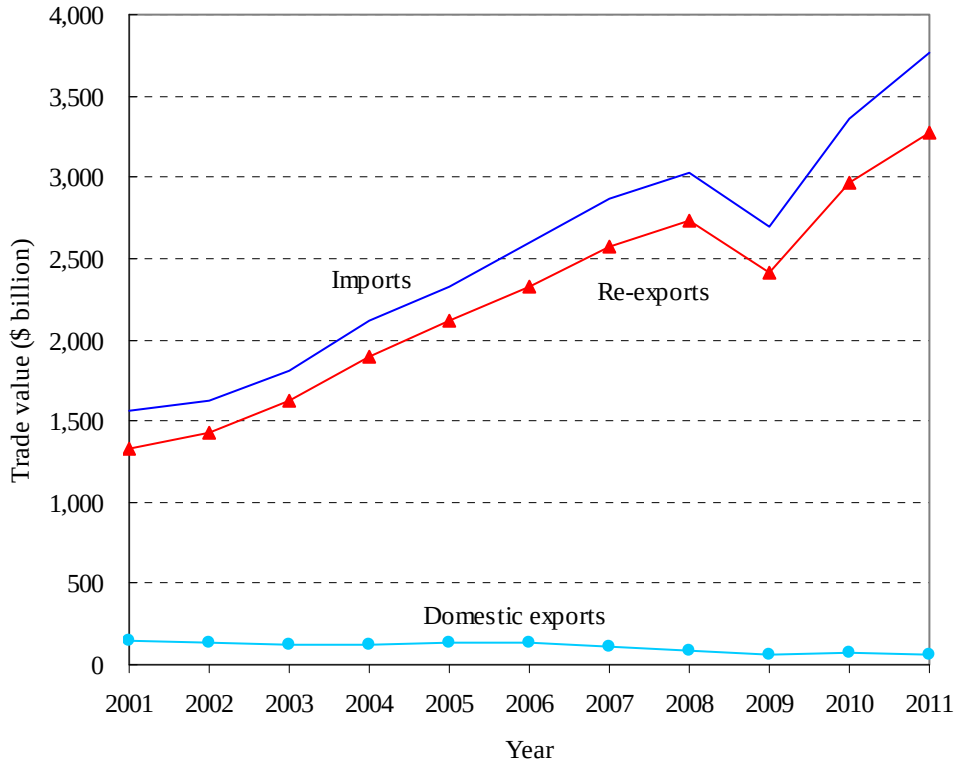
(1) National Income and Balance of Payments statistics
- private consumption expenditure
- Gross Domestic Product
- investment income
- Gross National Product
- Balance of Payments account
- International Investment Position
- External Debt

Chart 5.7 Gross Domestic Product of Hong Kong (at current market prices), 2001–2011



(2) External trade
- imports and exports of goods
- trade involving outward processing in the mainland of China
- trade in services
- offshore trade

Chart 5.8 Values of domestic exports[13], re-exports[14] and imports[15], 2001–2011



Notes :

(13) "Imports" are goods which have been produced or manufactured in places outside the jurisdiction of Hong Kong and brought into Hong Kong for domestic use or for subsequent re-exportation as well as Hong Kong products re-imported.

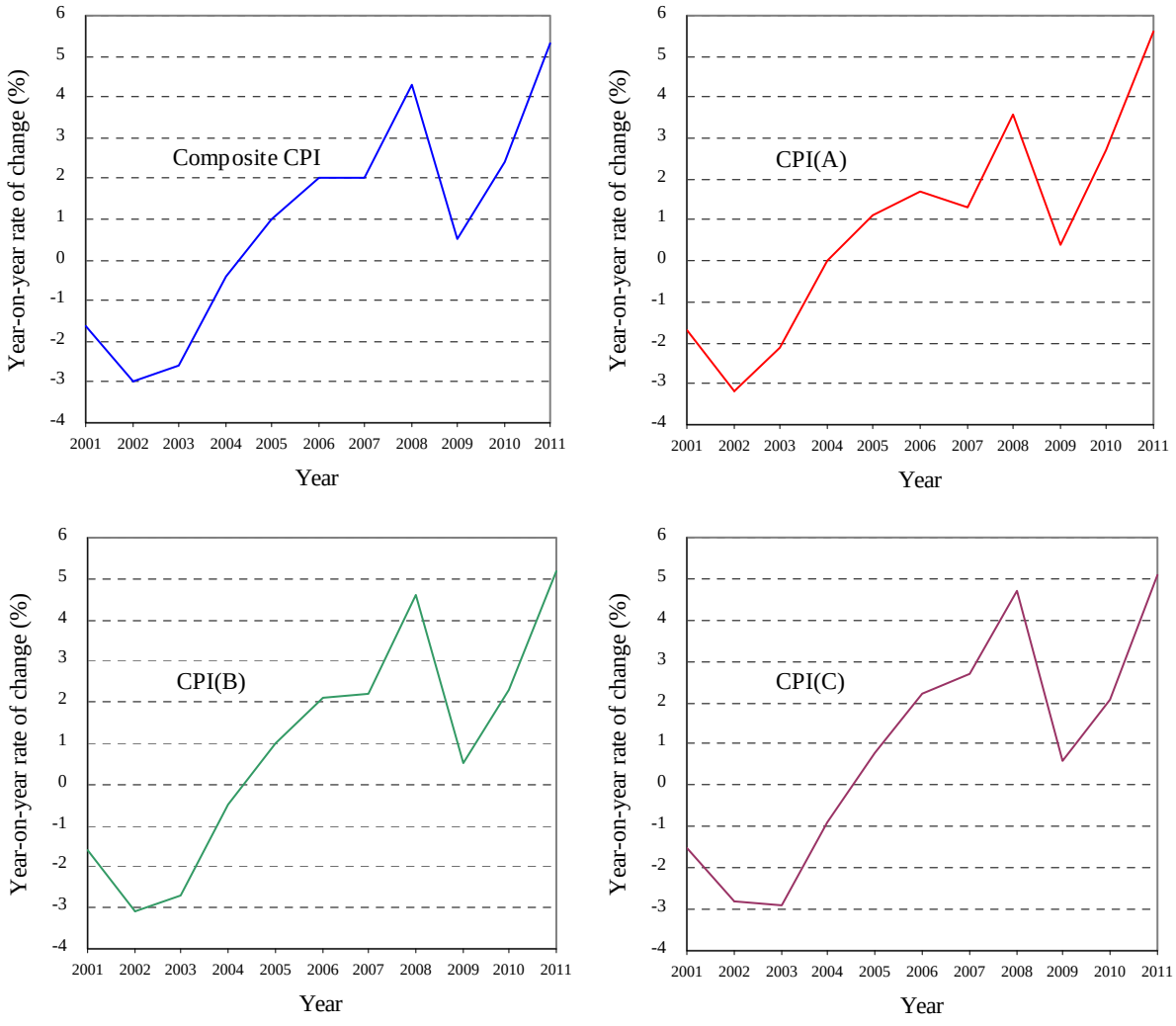
(14) "Domestic exports" are the natural produce of Hong Kong or the products of a manufacturing process in Hong Kong which has changed permanently the shape, nature, form or utility of the basic materials used in manufacture.

(15) "Re-exports" are products which have previously been imported into Hong Kong for subsequent re-exportation without having undergone in Hong Kong a manufacturing process which has changed permanently their shapes, nature, forms or utilities.

(3) Labour
- number by persons engaged in different industries
- wages and earnings

(4) Prices
- Consumer Price Indices (CPIs)
- import and export prices
- wholesale and retail prices

Chart 5.9 Year-on-year rates of change in the Consumer Price Indices[16], 2001–2011 (October 2009 – September 2010 = 100)[17]



Notes :

(16) CPIs measure change in the total cost of a given basket of goods and services representative of the purchases made by households in a particular expenditure group in a specified time period.

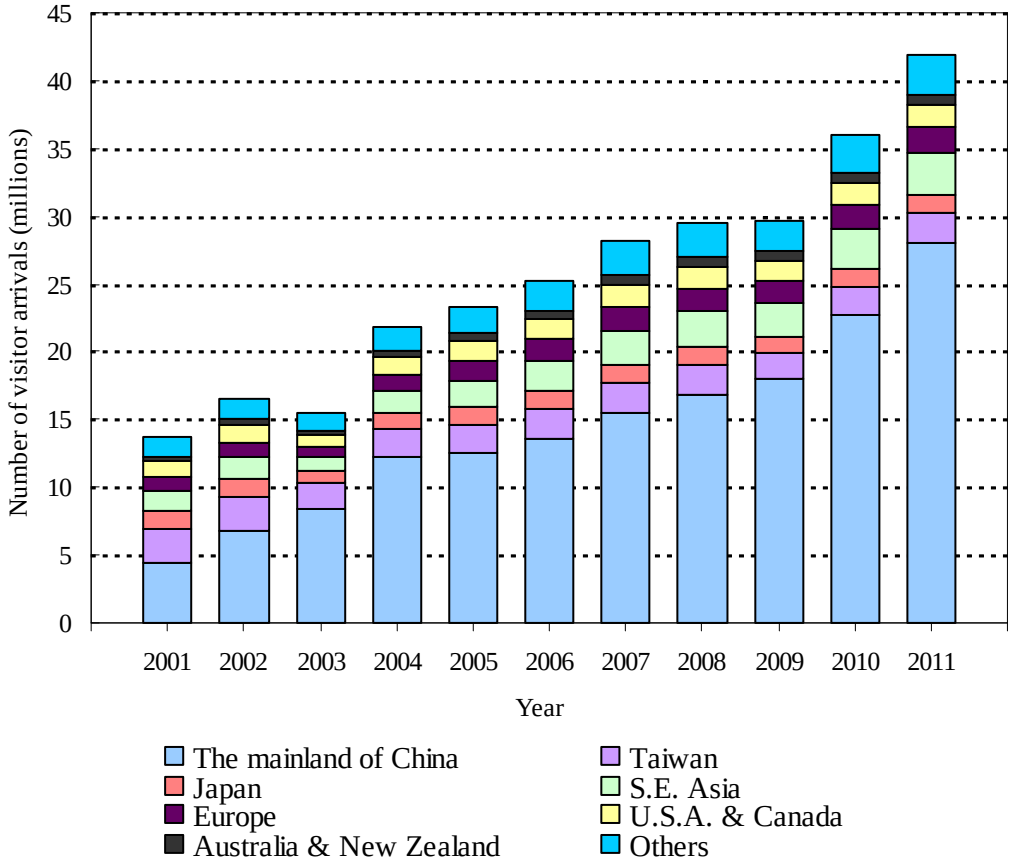
(17) From October 2010 onwards, the year-on-year rates of change are derived from the 2009/10-based CPIs. The year-on-year rates of change before October 2010 were derived using the index series in the base periods at that time (for instance the 2004/05-based index series), compared with the index a year earlier in the same base period.

(5) Operating characteristics and business performance of different economic sectors

- manufacturing
- electricity and gas supply
- sewerage, waste management and remediation activities
- building, construction and real estate
- import/export, wholesale and retail trades
- accommodation and food services
- transportation, storage and courier services
- information and communications, financing and insurance, professional and business services

(6) Tourism - visitor arrivals and their spending in Hong Kong

Chart 5.10 Visitor[(18)] arrivals by country/territory of residence, 2001–2011



Note :

(18) The term “visitor” is used, and not “tourist”, because there are people who come for business or other purposes as well as those who come for leisure purposes.

(7) Finance
- deposits and loans
- foreign exchange
- stock exchange turnover and index of share prices
- public accounts
- asset management

Misuses of statistics

Nowadays, people become increasingly more scientific-minded. Arguments that are without the support of relevant figures can hardly convince people. Thus, we can see how important statistics are. However, if the data are not used properly, one will make a blunder in trying to be smart. Not only will the argument be refuted and queried, the decision-maker may also be misled to make wrong decisions, hence causing serious consequences.

To fully understand the proper ways of using statistics, one has to acquire the knowledge through in-depth learning and accumulation of experience over a long period. The following are some common mistakes which should be avoided.

Reference periods of statistics unknown

When making statements about comparisons of statistics over time, it is important to specify the reference periods in question. A statement like “the number of students has grown 30%” is not meaningful. It is necessary to specify that the growth has taken place over a certain interval of time, be it 1 year, 5 years or 10 years, etc.

Confusion in regard to base of comparison

A restaurant claims in its advertisement that its prices will be cut by 100%. It is, of course, clear that prices cannot be cut 100%, else the servings formerly sold would now be given away free! Food formerly sold at $60 per order will be sold at $30 after the price deduction. The restaurant has calculated the 100% from ($60 - $30) / $30 × 100% by using the reduced price as the base of comparison. This is incorrect. The proper way of calculation is to use the original price as the base, hence the reduction is only 50%, i.e. ($60 - $30) / $60 × 100%.

Improper averaging of percentages and ratios

When individual percentages and ratios are calculated on different bases, they should not be directly added together to get the overall average.

Example :

Population[19] , 2011 Population Census

Area	Male	Female	Sex ratio (No. of males per 1 000 females)
Hong Kong Island	585 092	685 784	853
Kowloon	983 423	1 124 996	874
New Territories	1 733 779	1 957 314	886
Total	3 302 294	3 768 094	876

It is incorrect to derive the sex ratio for Hong Kong by adding the individual ratios for different areas and dividing by 3. This would give a wrong value of 871 [(853 + 874 + 886) / 3]. The true value of 876 may be calculated only from the total number of males and females [(3 302 294 / 3 768 094) × 1 000].

Note :

(19) Excluding the marine population.

Misinterpretation of changes

Example (1)

John's monthly salary is $8,000. When we say that his monthly salary is going to be ***increased by*** two times next month, what will be his new salary?

The answer is $24,000, i.e. ($8,000 + $8,000 × 2). However, if his salary is going to be ***increased to*** two times instead, then the revised salary will be $16,000 = $8,000 x 2.

Example (2)

The sales volume of a company dropped from $1,000,000 in 2010 to $500,000 in 2011. Could we say "The sales volume dropped by two times"?

The answer is no. Sales volume cannot drop by two times (200%). By dropping one time (100%), it already becomes zero! The correct description should be "Sales dropped by half or dropped by 50% [($1,000,000 – $500,000) / $1,000,000 = 50%]".

Example (3)

One of the two members of a household was sick. Though it is not wrong to say "50% of the members were sick", a statement like this is not meaningful when the base figure is too small.

Look at another example. The Customer Services Department of a company could not reply to a complaint according to the scheduled time. It is not meaningful to say that the customer service of the company was bad based solely on the incident mentioned above, especially if the total number of complaints received by the company in that particular year was only one.

Mis-presentation of data

The same set of statistical data may be presented in different ways to give quite different impressions to readers. Hence, if data are not presented properly, illusions would easily occur and readers would then be misled. Some common examples of mis-presentation of data are going to be discussed here.

Charts using an improper vertical scale

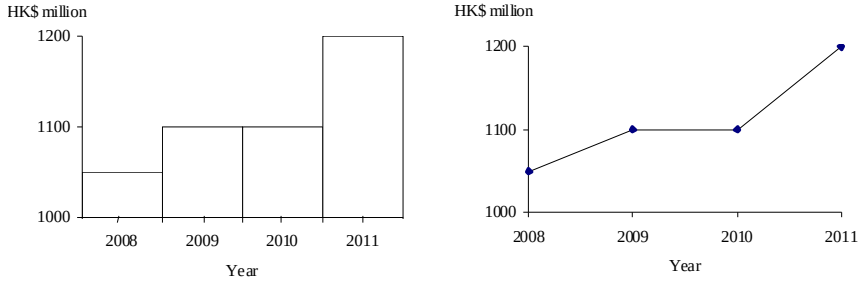
It is essential that charts with an arithmetic scale should begin at the zero base line in order to show the true picture. However, when the data to be presented graphically are of very large values but with small variations, people would tend to adopt a larger scale and plot the chart from a selected initial point for easy observation and comparison purposes. Watch out for this, as the true relationship between the data is bound to be exaggerated and a distorted impression is conveyed. Differences between the data now appear greater, and the slopes indicating the trends seem steeper (Chart 5.11A).

In order to convey a true picture of the relationship while still highlighting the variations, a broken scale chart is often used (Chart 5.11B). It enables the data variations to be observed easily but still reminds readers to watch out for the possible distorted impression that may be conveyed by the chart.

Chart 5.11

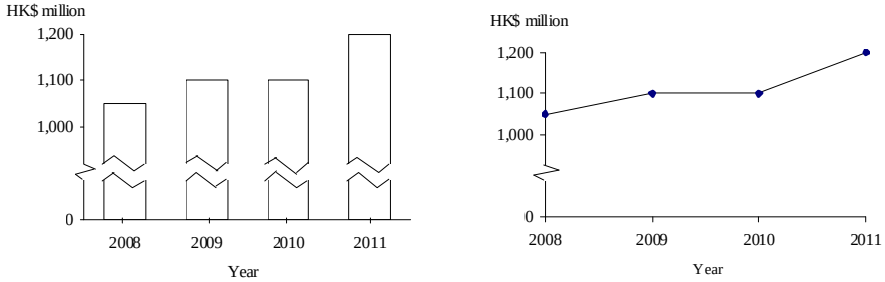
(A) Charts not using the broken scale

Sales turnover of Company ABC, 2008–2011



(B) Charts using the broken scale

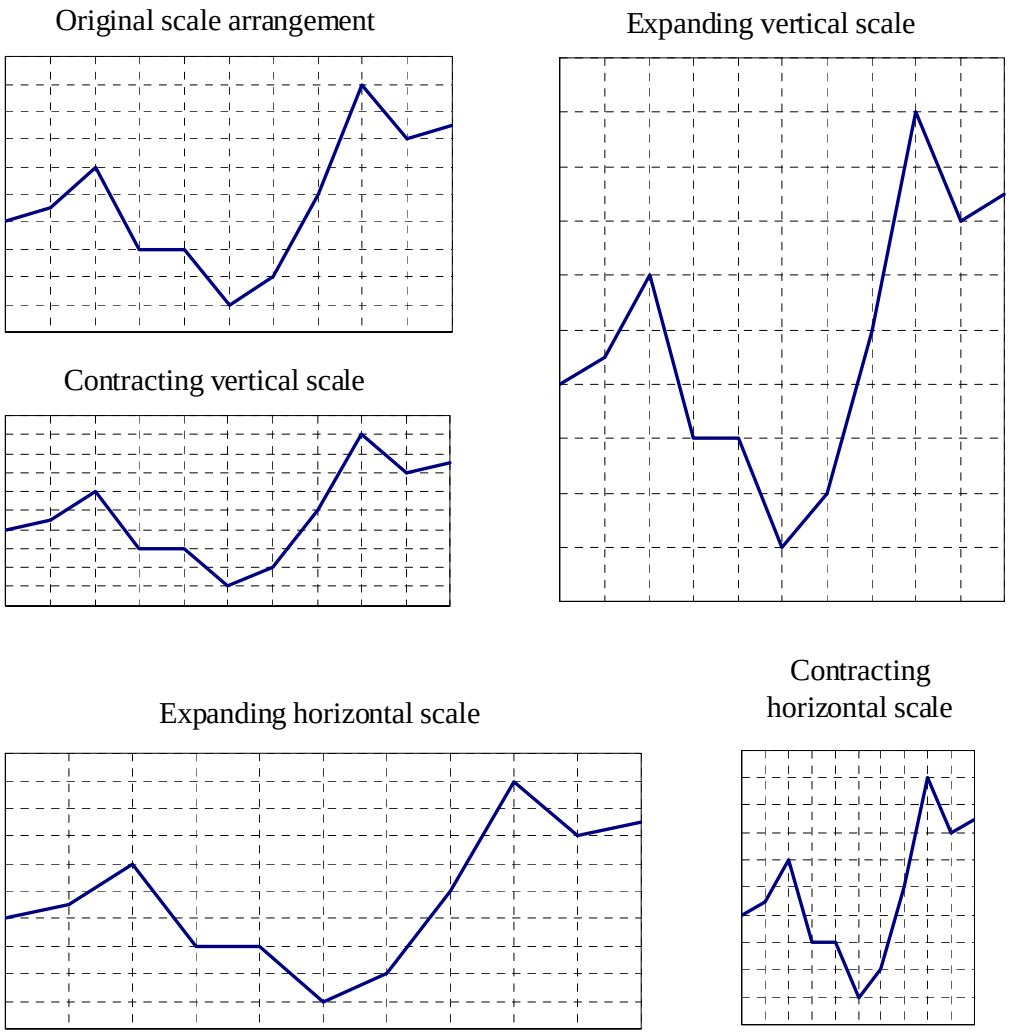
Sales turnover of Company ABC, 2008–2011



Illusions caused by expanding/contracting the grid

Seemingly, large movement in the trend depicted in a chart may result through expanding the vertical axis or contracting the horizontal axis of the chart. On the other hand, movement in the trend will appear less active than the actual situation if the horizontal axis is expanded or the vertical scale contracted to a large extent.

Chart 5.12 Changing the visual image : Contracting or expanding vertical (amount) scale or horizontal (time) scale tends to change the visual picture

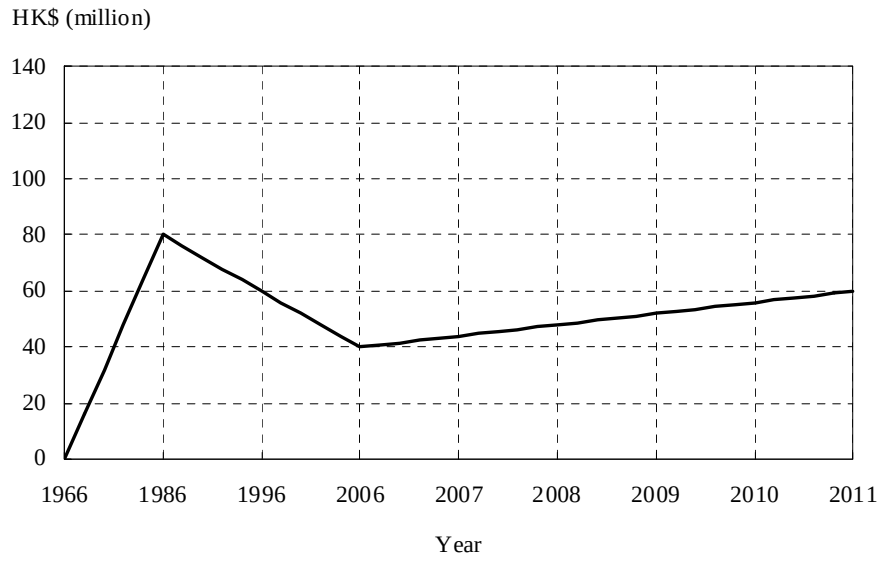


Illusions caused by skipping the grid

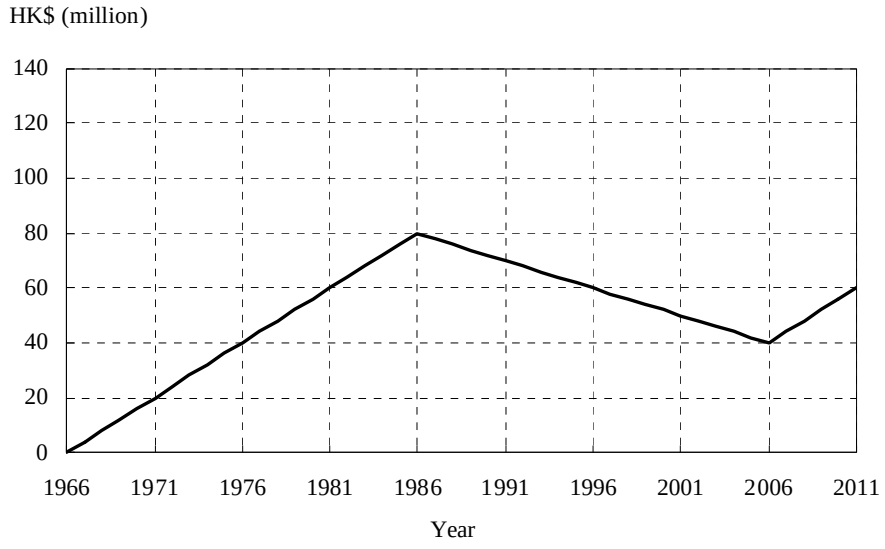
A familiar layout in reports and advertisements is seen in Chart 5.13A. In order to dramatise the story, a little trick is done with the time scale. It is not noticeable at a casual glance. Chart 5.13B shows what the actual trend looks like when the correct grid spacing is used for each year.

Chart 5.13

(A)



(B)



References

Croxton, Frederick E., Applied General Statistics, Chapter 1, Prentice-Hall, 1975.

Ilersic, A.R. and Pluck, R.A., Statistics, Chapter XVII, XVIII and XIX, HFL Ltd., 1979.

Moroney, M.J., Facts from Figures, Penguin Books Ltd., 1956.

Reichmann, W.J., Use and Abuse of Statistics, Penguin Books Ltd., 1961.

Exercise

A. Confusion in regard to base

Bank Y charges very high interest rates.

A man borrowed \$100,000 from Bank Y for 15 days. He received only \$80,000 in cash. The remaining \$20,000 was deducted as interest for the 15-day period.

Bank Y claimed that the interest rate was 20% ($= \$20{,}000 / \$100{,}000 \times 100\%$) per 15 days. Is this correct?

B. Misuses of statistics in daily endeavours

Reader can collect materials from newspaper cuttings, articles in magazines and commercials on television which may appear to involve misuses of statistics. Readers can then examine various aspects of such examples and analyse whether they are indeed misuses of statistics. Subsequently, readers may make suggestions on how errors can be rectified, or how methods of presentation can be improved.

6 Rate, ratio, proportion and percentage

Introduction

Very often, a figure is not very meaningful unless it is compared with another one. For instance, there are 100 cinemas in a certain city. This may look a large figure. However, when compared with the city's population (say, 10 million persons), this represents merely 1 cinema to every 100 000 persons. On the other hand, suppose there are only 10 cinemas in another city. Although the number of cinemas may look small, actually 1 cinema is just serving 10 000 persons there if the population of the city comprises only 100 000 persons.

Look at another example. Suppose there is an increase of 1 000 participants in a certain medical scheme. This may appear very impressive at first sight. However, if the scheme originally has 100 000 participants, the increase is merely 1%. On the contrary, an increase of 100 participants in a scheme originally with 500 participants represents a substantial increase of 20%.

"Relative numbers", which are derived by comparing two related figures(1), are generally more useful for analysis than "absolute figures". Rate, ratio, proportion and percentage are examples of some commonly used relative numbers.

Rate

"Rate" is obtained by expressing the measurement of something as per unit of something else. The quantities compared are not necessarily of the same kind and hence rate always bears a unit. An example of a rate is car speed, say 50 km per hour. Another example is population density, which is obtained by dividing the size of population in a district by the area of the district. The rate of 100 000 persons per cinema mentioned above is another example describing the degree of availability of movie entertainment facilities.

In compilation of "rates", the choice of correct bases is very important. Otherwise, the results would be misleading. For accurate comparison of "rates", a base closely related to the figures to be measured should be used. Take for an example, to calculate the voting rate of a city, the use of the total number of citizens having the right to vote as the base is considered more appropriate than the total number of citizens in the city.

Note :

(1) Comparisons are meaningful when the base figures are not too small. Thus, it would not be meaningful to make a statement that 25% of the members of a household are sick when 1 out of its 4 members falls ill.

Having different rates compiled from smaller groups of a population, we often want to get an overall picture for the whole population concerned. However, one should be careful that individual rates calculated from different bases should not be added together for getting the overall rate.

Ratio

"Ratio" is another way of expressing the relationship between figures. It is normally derived by comparing quantities of the same kind and therefore unlike rate, it does not bear a unit. Sometimes, the relationship is expressed as a ratio to one. For example, the ratio between boys and girls in a class is 3 : 1. However, if it suits our purpose better, we can express the relationship as a ratio to any other number (i.e. other than 1). For the previous example, the ratio between boys and girls can also be written as 30 : 10.

Proportion and percentage

"Proportion" and "percentage" are special cases of ratio. "Proportion" relates a "part" to the "whole"; while "percentage" is the case where the relationship between figures is presented as a ratio to one hundred.

Some frequently used relative numbers

Crude death rate

The crude death rate for a given calendar year is obtained by dividing the number of known deaths occurring in a community during that calendar year by the mid-year population of that community, and expressing the result in terms of deaths per 1 000 population.

In 2011, the number of known deaths in Hong Kong was 41 400[#] and the mid-year population was 7 071 600. The crude death rate for 2011 was therefore 5.9[#] per 1 000 population.

Age-sex specific death rate

Mortality varies significantly with age and sex. To study in greater depth the pattern of mortality of people in different age groups and sex, the so-called age-sex specific death rates (or age-sex specific mortality rates) may be referred to. This rate is expressed as the number of known deaths per 1 000 population in the same age group and same sex. For example :

$$\text{Age-sex specific death rate for males in the age group 60-64} = \frac{\text{Number of known deaths of males in the age group 60-64}}{\text{Number of males in the age group 60-64}} \times 1\,000$$

Crude birth rate

Similar to crude death rate, crude birth rate is calculated by dividing the number of known live births during a calendar year by the mid-year population for that year, and expressing the result in terms of births per 1 000 population.

In 2011, the number of known live births in Hong Kong was 95 400[#] and the mid-year population was 7 071 600. The crude birth rate for 2011 was therefore 13.5[#] per 1 000 population.

Unemployment rate

The unemployment rate is obtained by dividing the total number of unemployed persons by the labour force in the community. The labour force or economically active population refers to the land-based non-institutional population aged 15 and over who satisfy the criteria for inclusion in the employed population or unemployed population. The unemployment rate of Hong Kong was 3.4% in 2011.

Industrial accident rate in the manufacturing sector

The industrial accident rate in the manufacturing sector is expressed as the number of reported casualties per 1 000 manufacturing workers. The structure of this relative number, with the number of manufacturing workers as the base, highlights the fact that the base figure should relate to the subject matter in question as much as possible. (In this case, manufacturing workers are those “at risk” of industrial accidents and the number of such workers is to be used as the base.) In 2010, the industrial accident rate in the manufacturing sector was 16.8 per 1 000 manufacturing workers.

Crime rate

Crime rate is usually expressed as the number of cases of reported crime per 100 000 population. In 2011, the crime rate for Hong Kong was 1 074 per 100 000 population.

Age specific prosecution rates

To study in greater depth the tendency of committing crimes by people in different age groups, the so-called age specific prosecution rates may be referred to. This rate is usually expressed as the number of persons prosecuted per 100 000 population in the same age group. For example :

$$\text{Age specific prosecution rate for the age group 16-20} = \frac{\text{Number of persons in the age group 16-20 who are prosecuted}}{\text{Number of persons in the age group 16-20}} \times 100\,000$$

Sex ratio

The relationship of the number of males to the number of females in the population is given by the sex ratio, which states the number of males per 1 000 females.

As in mid-2011, there were 3 303 000 males and 3 768 600 females in Hong Kong. Thus, the sex ratio was 876 males per 1 000 females.

Proportion of domestic households resided in subsidised home ownership housing

To relate the number of domestic households resided in subsidised home ownership housing(2) to the total number of domestic households in Hong Kong, we can obtain the proportion of domestic households resided in subsidised home ownership housing.

The proportion of domestic households resided in subsidised home ownership housing increased slightly from 15.5% in 2000 to 16.3% in 2010.

Note :

(2) Subsidised home ownership housing includes flats built under the Home Ownership Scheme, Middle Income Housing Scheme, Private Sector Participation Scheme, Buy or Rent Option Scheme and Mortgage Subsidy Scheme, and flats sold under the Tenants Purchase Scheme of the Hong Kong Housing Authority. Also includes flats built under the Flat for Sale Scheme and Sandwich Class Housing Scheme of the Hong Kong Housing Society. As from 2002, subsidised sale flats that can be traded in open market are excluded.

Percentage of population by age group

The percentage of population by age group is obtained by dividing the total number of people in a particular age group by the total population size and then multiplying by 100 percent.

As in mid-2011, 11.6% of people in the Hong Kong Population were aged under 15, 75.0% aged 15-64 and 13.3% aged 65 and over.

Proper use of relative numbers

Care must be taken when handling or interpreting relative numbers; otherwise we may arrive at incorrect conclusions or mislead other people. The following are two examples to illustrate this.

Example 1 Visitor[(3)] arrivals in Hong Kong

Visitor's country/ territory of residence	2006 Thousand	2006 (Percentage of total)	2011 Thousand	2011 (Percentage of total)
The mainland of China	13 591	(53.8%)	28 100	(67.0%)
Taiwan	2 177	(8.6%)	2 149	(5.1%)
South East Asian countries	2 274	(9.0%)	3 202	(7.6%)
Japan	1 311	(5.2%)	1 284	(3.1%)
U.S.A. and Canada	1 495	(5.9%)	1 623	(3.9%)
Europe	1 548	(6.1%)	1 801	(4.3%)
Australia and New Zealand	662	(2.6%)	748	(1.8%)
Others	2 193	(8.7%)	3 015	(7.2%)
Total	25 251	(100%)	41 921	(100%)

Note :

(3) The term "visitor" is used, and not "tourist", because there are people who come for business or other purposes as well as those who come for leisure purposes.

Comparing the percentage figures for 2006 and 2011, one may notice that figures for all countries, except the mainland of China, have dropped. For instance, the drops of Taiwan, Japan and Europe are respectively 3.5 "percentage points(4)", 2.1 "percentage points" and 1.8 "percentage points"

If not careful, one may judge from the percentage drops only and make a wrong conclusion that visitors from all these areas have become "less likely" to visit Hong Kong. In fact, a decrease in the share does not necessarily imply there is a decrease in the actual number. Take "South East Asian countries" as an example. Despite a drop of 1.4 percentage point in share, the number of visitors from "South East Asian countries" actually increased by 928 000 persons, or 40.8% (i.e. (3 202 000 – 2 274 000) / 2 274 000 × 100%).

The situation can be attributed mainly to the significant increase in the number of visitors from the mainland of China. The percentage share of the total number of visitors from the mainland of China has gone up from 53.8% in 2006 to 67.0% in 2011 while the shares of visitors from other areas have decreased relatively.

Merely mentioning the reduction in the proportion of visitors, though correct, may not create a right impression in the less statistically-minded people. Referring to both the numbers and the proportions of visitors will enable a more complete and accurate picture to be presented.

Note :

(4) "Percentage point" refers to the absolute sum or difference between two rates or percentages, which is obtained by direct subtraction of the figures concerned. As a matter of fact, an increase of *x* percentage points and an increase of *x* per cent are not the same. Suppose the passing rates of two different rounds of examination for a class of students were 80% and 84% respectively, the passing rate increased by 4 percentage points (84% – 80% = 4 percentage points). Alternatively we can also say that the passing rate recorded an increase of 5% (i.e. (84 – 80) / 80 × 100%). However, the second way of presentation is generally not recommended.

Example 2 Registered deaths from external causes of morbidity and mortality in Hong Kong by age group, 2010

Age group	No. of registered deaths		Population at mid-2010	No. of registered deaths per 100 000 persons in the age group
0 – 14	27	(1.4%)	844 200	3.2
15 – 49	770	(41.3%)	3 770 000	20.4
50 – 64	405	(21.7%)	1 491 500	27.2
65 and above	659	(35.4%)	918 500	71.7
Total	1 864[(5)]	(100.0%)	7 024 200	

Based on the above table, someone may make a single statement : "The majority of people who died of external causes of morbidity and mortality in 2010 was those aged 15-49 (41.3%), followed by those aged 65 and above (35.4%) and 50-64 (21.7%)." This may create a false impression that the death situation in the year was more serious for people aged 15-49. Looking at the table more closely, it is observed that the death rate of those aged 15-49 was 20.4 per 100 000 persons. The death rate increased to 27.2 per 100 000 persons for those aged 50-64 and further to 71.7 per 100 000 persons for those aged 65 and above. The death rate for the age group 65 and above was in fact the highest.

It should be noted that the "statement" itself is not wrong. However, the proper way of presenting a situation is to mention the other relevant facts as well so that a more complete picture of the whole situation is provided to enable a clearer understanding of the actual situation.

Reference

Croxton, Frederick E., Applied General Statistics, Chapter 7, Prentice-Hall, 1975.

Note :

(5) Including registered deaths of unknown age.

Exercise

Calculating rate

Consider the following hypothetical example on studying the passing rate of an open examination :

District		Population[(Note)] in 2011	No. of students sitting for the examination in 2011	No. of students passing the examination in 2011
Hong Kong Island	Central and Western	251 500	60 293	44 014
	Wan Chai	152 600	41 110	28 366
	Eastern	588 100	144 528	93 943
	Southern	278 700	63 142	44 831
Kowloon	Yau Tsim Mong	307 900	57 773	36 397
	Wong Tai Sin	420 200	85 917	51 550
	....	....	....	....
New Territories	Shatin	630 300	138 281	76 055
	Tai Po	296 900	156 552	60 494
	....	....	....	....
Total		7 071 600	1 403 211	841 927

(i) Comment on the statement :
"The overall passing rate of the examination in 2011 was 11.9%,
i.e. (841 927 / 7 071 600) × 100%."

(ii) Calculate the passing rates for the 4 districts on Hong Kong Island and the overall passing rate for Hong Kong Island.

(iii) Suppose that the number of students passing the examination in the Wan Chai district in 2006 was 23 479. Can it be concluded that since there was an increase in the number of students passing the examination in 2011, the academic results of students in the Wan Chai district had improved during the period from 2006 to 2011?

Note : Individual figures do not add up to the total population of Hong Kong because the marine population are not included.

7 Measures of central tendency

Introduction

Given a set of data, it is often required to describe them in terms of a single value around which the data are distributed. This value is a measure of "central tendency", indicating the central position of the given set of data. This central value is called an average.

Several types of average can be defined. The most common ones are arithmetic mean (or briefly "the mean"), median and mode. These measures of central tendency are applied in different situations depending on the data and the intended purpose.

Mean

The arithmetic mean of a set of data is defined as the sum of the values of the given set of data divided by the number of data points.

$$\overline{X} = \frac{X_1 + X_2 + \ \ldots \ + X_N}{N}$$

where $\overline{X}$ is the mean;
N is the total number of data points; and
X_i is the value of the i^{th} data point, for i = 1, 2, ..., N.

Examples :

(1) The mean height of the students in a class is 1.52 m.

(2) The mean daily maximum air temperature for the year 2011 was 25.8^{o}C. (The method is to find out the maximum air temperature recorded for each of the 365 days in the year 2011 and then calculate the arithmetic mean of the 365 data points.)

Median

The median of a set of data is the middle value of the given set of data when they are arranged in order, ascending or descending, of magnitude. In case there are two middle values (i.e. when the data set has an even number of data points), then the median is the arithmetic mean of these two middle values.

Examples :

(1) The median of the set of numbers {4, 6, 7, 9, 10} is 7.

(2) The median of the set of numbers {1, 3, 4, 4, 7, 9, 10, 11, 13, 14} is 8, i.e. (7 + 9)/2.

(3) The median monthly household income of domestic households in Hong Kong in 2011 was $20,000. In other words, 50% of the domestic households in Hong Kong earned more than $20,000 household income per month in 2011 while the other 50% earned less than $20,000 per month.

(4) The median age of men who married for the first time in Hong Kong in 2011 was 31.1[#].

Mode

The mode of a set of data is the value which occurs with the largest frequency, i.e. the most common value in the given set of data. The mode may not exist, and even if it does exist it may not be unique.

Examples :

(1) The set {1, 3, 3, 5, 7, 8, 8, 8, 9, 10, 10, 11, 12} has a mode 8.

(2) The set {2, 3, 4, 6, 8, 9} does not have any mode.

(3) The set {1, 3, 4, 4, 4, 6, 7, 7, 9, 9, 9, 10, 12} has two modes, 4 and 9.

(4) The modal household size in Hong Kong was 2 in the fourth quarter of 2011.

Grouped data

When presenting a large data set, it is often useful to classify the data into different classes and to find out the number of data points belonging to each class (i.e. the class frequency). Data organised in this way are called grouped data whereas a tabular arrangement of the data by class showing the corresponding class frequency is called a frequency distribution or frequency table.

Example :

The weights (in kg) of individual students in a class are as follows :

61, 62, 63, 63, 63, 64, 64, 64, 64, 65, 65, 65, 66, 66, 66,
66, 66, 67, 67, 67, 67, 68, 68, 68, 69, 69, 69, 69, 70, 70.

To help give a clear overall picture of the weights of the students, the above 30 data points can be re-organised into several classes and a frequency distribution can be constructed.

Step 1 Determine the largest and smallest values in the data set and find the range (i.e. the difference between the largest and smallest values).

Smallest value	=	61
Largest value	=	70
Range	=	9

Step 2 Divide the range into a convenient number of classes having the same size (i.e. same class interval[(1)]). The number of classes is usually taken between 5 and 20, depending on the values of the data.

Given that there are only 30 data points in the data set, it is not appropriate to divide the range into a large number of classes (otherwise there will be too few observations within each class). For this example, 5 is a sensible number of classes. The class intervals to be used are thus :

61 – 62;
63 – 64;
65 – 66;
67 – 68; and
69 – 70.

Note :

(1) The smaller value of each class interval shown above is called the lower class limit and the larger value the upper class limit. The mid-point of the class interval is known as the class mark.

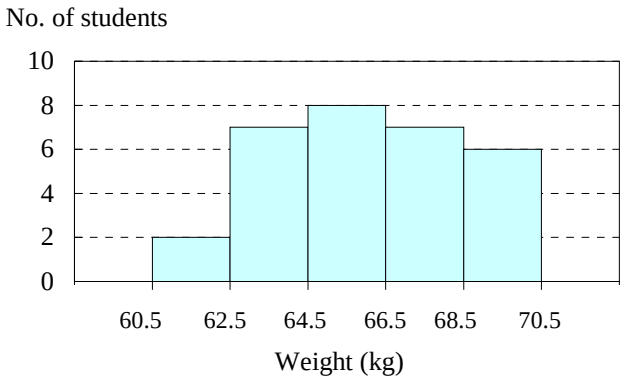
Step 3 Find the class frequency for each class and then construct the frequency distribution.

Weight[(2)] (kg)	Number of students
61 – 62	2
63 – 64	7
65 – 66	8
67 – 68	7
69 – 70	6
Total	30

Histogram

A histogram is a vertical bar graph, with no space between the bars. It is commonly used for presenting frequency distributions of grouped data with continuous class boundaries, with class boundaries shown as points on the horizontal axis and the frequencies as units on the vertical axis. The frequency corresponding to a class is then represented by the area of a bar with the base as the class interval.

Chart 7.1 Histogram of the number of students



The population pyramid may be visualised as formed from two horizontally placed histograms, one being the age distribution of males and the other being the age distribution of females.

Note :

(2) If the weights are rounded to the nearest kg, the class interval, say, 61 kg – 62 kg, theoretically includes all measurements from 60.5000…kg to 62.4999…kg. These two class limits, indicated briefly by the exact numbers 60.5 kg and 62.5 kg, are called true class limits or class boundaries.

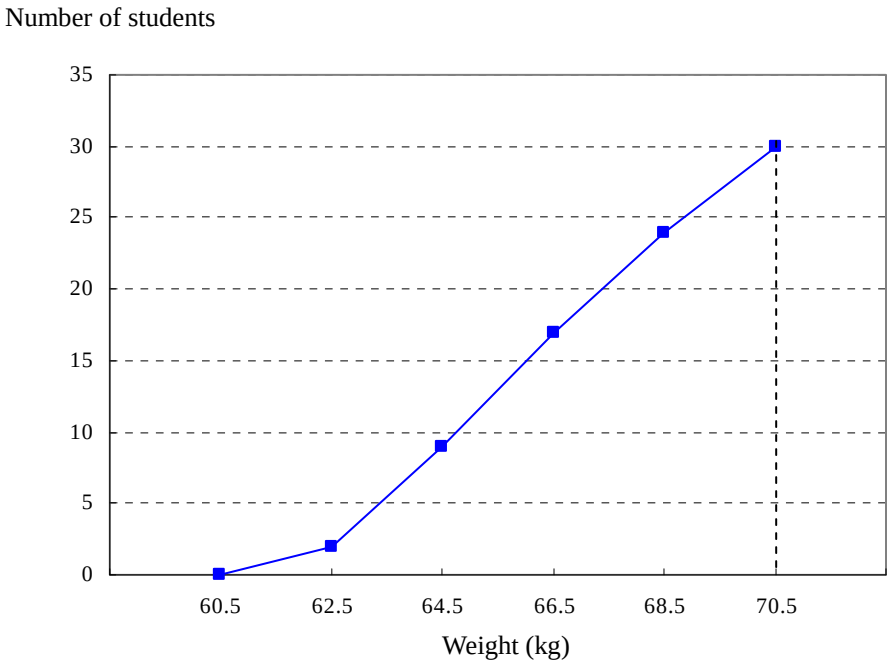
Cumulative frequency polygon

For some purposes, it is desirable to consider the total frequency of all values greater than or equal to the lower class boundary of each class (i.e. the "more than" basis for counting total frequency) or the total frequency of all values less than the upper class boundary of each class (i.e. the "less than" basis for counting total frequency). In such circumstances, it is necessary to construct the cumulative distribution (in tabular form) or the cumulative frequency polygon (in graphical form).

To construct a cumulative frequency polygon, frequencies accumulated for each class are plotted at the lower class limits when the "more than" basis is used. When the "less than" basis is used, cumulated frequencies are plotted at the upper class limits.

Regardless of the basis used, the Y-axis should be drawn long enough to accommodate the total frequency. The points, each representing the cumulative frequency of a particular class, are then joined by a series of straight lines and the S-curve so obtained may be extended to touch the X-axis by dropping a perpendicular line from its highest point. The polygon thus formed is a cumulative frequency polygon.

Chart 7.2 Cumulative frequency polygon of the number of students



The cumulative frequency polygon is particularly useful for interpolation. However, it is important to note that the use of straight lines to connect the series of known points is only an approximation to the frequency distribution of the items within any given interval. Nevertheless, this approximation is generally considered to be sufficiently accurate under most circumstances.

Mean of grouped data

The formula to be used to calculate the mean of a set of grouped data is :

$$\overline{X} = \frac{f_1 X_1 + f_2 X_2 + \ ... \ + f_k X_k}{f_1 + f_2 + \ ... \ + f_k}$$

where $\overline{X}$ is the mean;

X_i is the class mark (mid-point of the class interval) of the i^{th} class, for i = 1, 2, ..., k; and

f_i is the frequency of the i^{th} class, for i = 1, 2, ..., k.

The mean thus obtained depends on the way the data are grouped and is only an approximation. However, it reduces the burden of numerical calculation.

Finding the median of grouped data from histogram

Making use of the histogram, an approximation of the median of a set of grouped data can be found.

Step 1 Construct the histogram for the data set.

Step 2 Identify the median class, i.e. the class which contains the "middle value".

Step 3 Calculate the sum of frequencies of classes lower than the median class, $(\sum f)_l$.

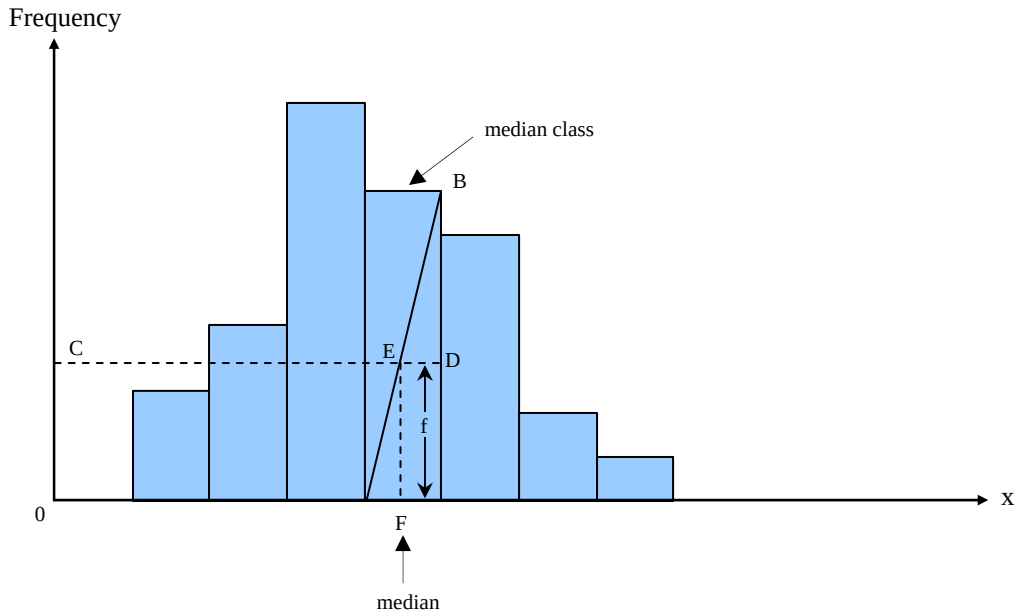
Step 4 Find f, using the expression $f = N/2 - (\sum f)_l$
where N is the total frequency; and
$(\sum f)_l$ is the sum of frequencies of classes lower than the median class.

Step 5 Construct a diagonal in the median class. The line should be drawn from the lower left corner to the upper right corner (AB).

Step 6 At the height corresponding to " f " found in step 4, draw a horizontal line (CD) to cut the median class. The line will intersect with the diagonal.

Step 7 Drop a perpendicular line from this point of intersection (E) to the X-axis. The point thus located on the X-axis (F) will give the value of the median.

Chart 7.3



Finding the median of grouped data from cumulative frequency polygon

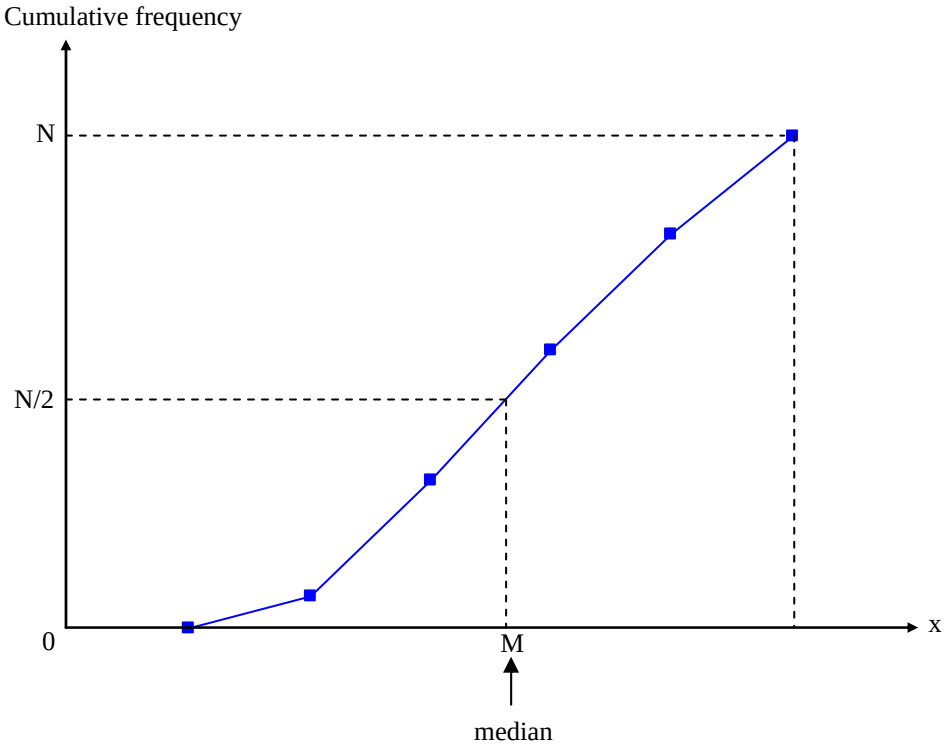
Another graphical method of finding the median is to draw a cumulative frequency polygon.

Step 1 Construct the cumulative frequency polygon for the data set.

Step 2 At the cumulative frequency height N/2 (N is the number of data points), draw a horizontal line until it touches the cumulative frequency polygon.

Step 3 Draw a perpendicular line from this point of intersection to the X-axis. The point thus located on the X-axis (M) will give the value of the median.

Chart 7.4



Finding the mode of grouped data from histogram

A rough estimate of the mode can also be found with the use of the histogram.

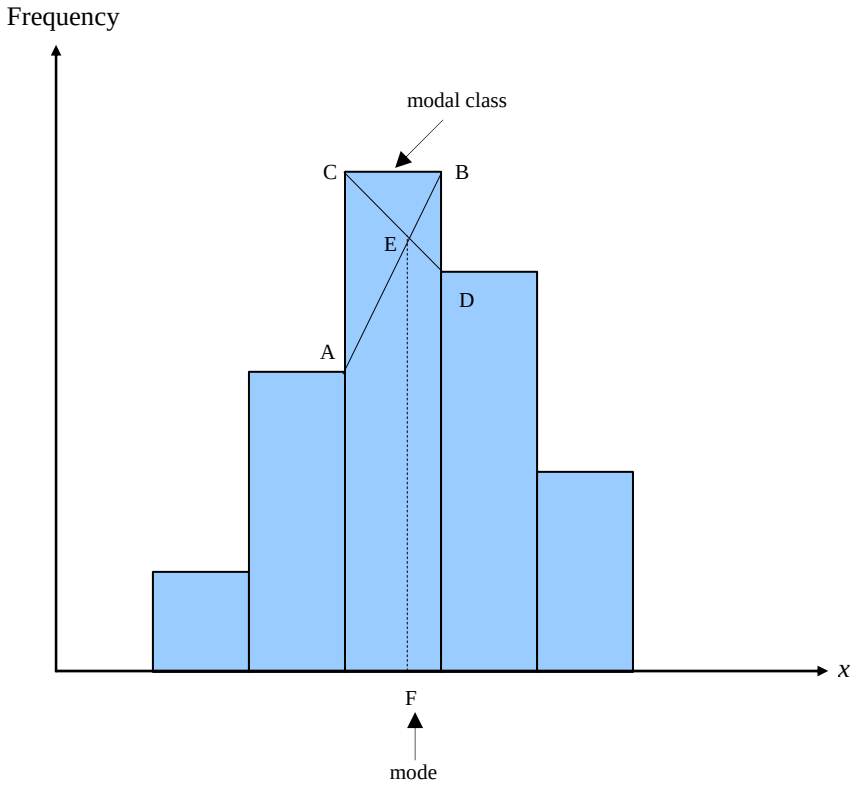
Step 1 Construct a histogram for the data set.

Step 2 Identify the modal class, i.e. the class with the largest frequency.

Step 3 Draw 2 diagonals as shown in Chart 7.5 (AB and CD).

Step 4 Drop a vertical line from the point of intersection of these 2 diagonals (E) to the X-axis. The point thus located on the X-axis (F) will give the value of the mode.

Chart 7.5



Use of mean and median

The mean is the most widely used of all the measures of central tendency.

For an approximately symmetric set of data such as {13, 14, 15, 17, 18.5}, the mean 15.5 gives the central location effectively.

The mean, however, is affected by the value of every item in the set of data, and the presence of a few extremely large (or extremely small) items may result in a mean which could be misleading. For example, for the set {1 000, 2 000, 3 000, 3 500, 4 500, 5 000, 30 000}, the mean is 7 000 which is not a good measure of the central tendency. However, the median, 3 500, is more satisfactory because the existence of a few extreme values generally would not cause the median to fluctuate much.

In daily life, the mean monthly household income is easily affected by extreme values. Hence, the median is more suitable than the mean in measuring the central tendency of household income. In 2011, the median monthly household income of domestic households in Hong Kong was $20,000.

Supplementary notes - weighted mean

Sometimes, we associate the data (X_1, X_2, ..., X_k) with certain weighting factors or weights (W_1, W_2, ..., W_k), depending on the significance or importance attached to the data points. In this case,

$$\overline{X} = \frac{W_1X_1 + W_2X_2 + \ ... \ + W_kX_k}{W_1 + W_2 + \ ... \ + W_k}$$

The mean of a set of grouped data is in fact a mean of the class marks weighted by the class frequencies.

Normally, the use of the weighted mean leads to a different conclusion from that by the use of the unweighted mean. In fact, the concept of weighted mean is commonly used in daily life.

Example :

The following is a summary of the grade points of the final examination and the mid-term test scored by two students A and B, with the final examination grades weighted 3 times as much as the mid-term test grades :

	Grade points		Weights
	A	**B**	
Final examination	82	90	3
Mid-term test	78	66	1

If the weights are disregarded,

the mean score of A $= \dfrac{82+78}{2} = \underline{\underline{80}}$,

and

the mean score of B $= \dfrac{90+66}{2} = \underline{\underline{78}}$.

We would say that A did better than B.

On the other hand, if the weights are taken into consideration,

the weighted mean score of A $= \dfrac{82\times 3+78\times 1}{3+1} = \underline{\underline{81}}$,

and

the weighted mean score of B $= \dfrac{90\times 3+66\times 1}{3+1} = \underline{\underline{84}}$.

We would then say that B did better than A.

References

Croxton, Frederick E., Applied General Statistics, Chapter 9, Prentice-Hall, 1975.

Spiegel, Murray R., Theory and Problems of Statistics, Chapter 3, Schaum Publishing Co., 1961.

Exercise

Calculating mean, median and mode

Consider a set of figures：

{1, 9, 3, 7, 8, 12, 9}

(i) What is the sum of these seven values?

(ii) What is the mean?

(iii) What is the median?

(iv) What is the mode?

(v) Is the median equal to the mean of the same seven values?

(vi) Are the mean and median the same when the set of figures is {1, 3, 7, 11, 13}?

(vii) What does this suggest about the conditions when the mean and median are the same?

8 Measures of dispersion

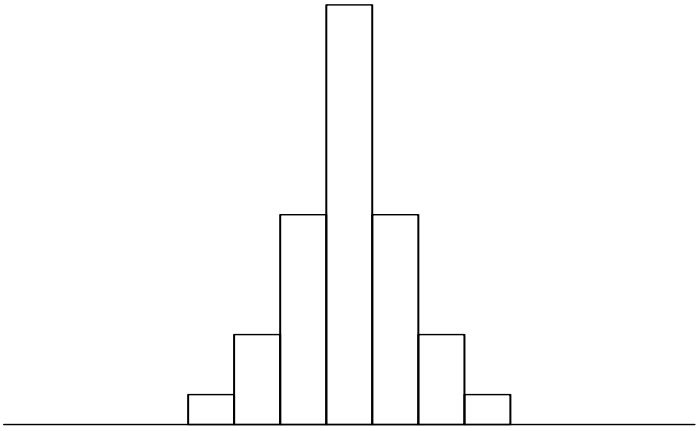
Introduction

Information on the central tendency of a set of data alone is not sufficient to describe its distribution. For instance, two groups of data may have the same average, yet having different degrees of spreading amongst the data points. To have a better knowledge of a set of data, it is necessary to measure the extent to which the individual values vary within the set, or equivalently, the extent to which individual values vary around the average.

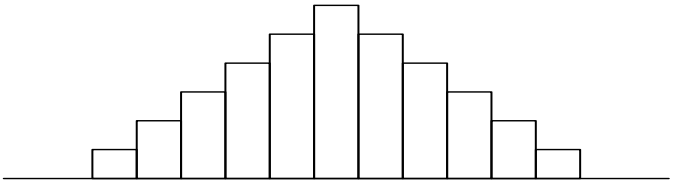
The degree to which numerical data tend to spread around an average value is called the dispersion of the data. Various measures of dispersion are available. The more common ones are range, mean deviation, variance and standard deviation.

Chart 8.1 Various degrees of variation

Small variation



Large variation

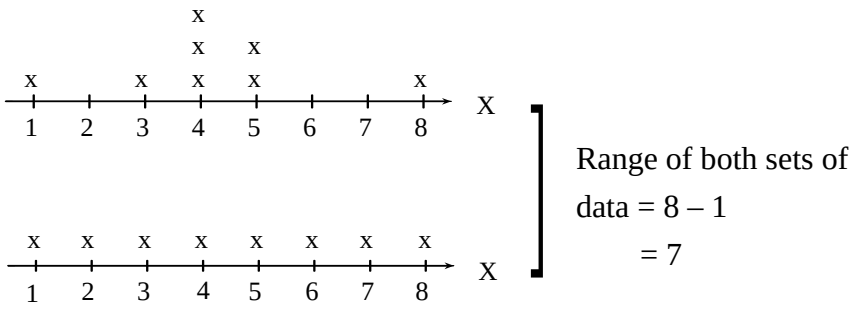


Range

The range of a set of numbers is the difference between the maximum and the minimum values. This measure is greatly affected by extreme values. The advantage of this measure is simplicity. The disadvantage is that it does not take the intermediate values into account at all.

Chart 8.2 Range

The following two sets of data have the same range but different dispersions :



Mean deviation (Mean absolute deviation)

Mean deviation, also known as mean absolute deviation, is a measure summarising how large the extent a data set deviates from its mean, irrespective of the direction of spread of individual data points in the set. For a given set of numerical data, the mean deviation is defined as the sum of the absolute deviations (i.e. the absolute values of the deviations from the mean) of individual data points divided by the total number of data points.

$$\text{md} = \frac{\left|X_1 - \overline{X}\right| + \left|X_2 - \overline{X}\right| + \cdots + \left|X_N - \overline{X}\right|}{N}$$

$$= \frac{\sum_{i=1}^{N} \left|X_i - \overline{X}\right|}{N}$$

where md is the mean deviation;

$\overline{X}$ is the mean;

N is the total number of data points;

X_i is the value of the i^{th} data point, for i = 1, 2, …,N; and

$\left|X_i - \overline{X}\right|$ is the absolute value of the deviation from the mean of the i^{th} data point, for i = 1, 2, …,N.

For example, consider a set of data {13, 3, 5, 1, 8} which has a mean of 6. By subtracting the mean from each of the data points in the set, the deviations(1) from the mean can be found, which are 7, -3, -1, -5 and 2 respectively.

The mean deviation of the set of data can then be computed by finding the simple average of the absolute deviations :

$$\frac{7+3+1+5+2}{5} = \underline{\underline{3.6}}$$

This tells that the data points within the given set lie on average 3.6 units from their mean of 6.

From the above example, it can be seen that if we simply add up the individual deviations, a zero sum will be obtained despite the dispersion of the data points around the mean. This is because the negative deviations just happen to balance the positive deviations. To avoid such problem, the absolute values of the deviations, or absolute deviations, are therefore used instead in the computation of the mean deviation. With the negative signs of those negative deviations removed, only the "distances" but not the "directions" in which the individual data points spread away from the mean are indeed considered.

Another point worth noting is that the sum of absolute deviations is highly sensitive to the total number of data points in the given set of data. For two sets of data with same degree of dispersion, the set which has more data points is bound to have a greater sum of absolute deviations. To enable comparisons of data sets of different sizes to be made, it is essential to discount the size effect by dividing the sum of absolute deviations by the total number of data points.

Note :

(1) Some of the deviations so derived are negative. This shows that the values of the data points concerned are smaller than the mean. Conversely, the deviations will be positive when the data points have a value larger than the mean.

Variance

The average of the squared deviations from the mean is called the variance.

$$\sigma^2 = \frac{\left(X_1 - \overline{X}\right)^2 + \left(X_2 - \overline{X}\right)^2 + \cdots + \left(X_N - \overline{X}\right)^2}{N}$$

$$= \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N}$$

where σ^2 is the variance;
$\overline{X}$ is the mean;
N is the total number of data points;
X_i is the value of the i^{th} data point, for i = 1, 2, …, N; and
$\left(X_i - \overline{X}\right)^2$ is the squared deviations from the mean, for i = 1, 2, …, N.

In our example above, the variance of the data set is 17.6
(i.e. $\frac{7^2 + (-3)^2 + (-1)^2 + (-5)^2 + 2^2}{5}$).

This is another way of measuring the size of the deviations irrespective of their signs, since squares of negative numbers are always positive.

Standard deviation

The variance measures the scatter of the data points from their mean, but it is expressed in "squared units". To return to the original unit of measurement, one takes the square root of the variance. This gives a measure known as standard deviation.

$$\sigma = \sqrt{\frac{\left(X_1 - \overline{X}\right)^2 + \left(X_2 - \overline{X}\right)^2 + \cdots + \left(X_N - \overline{X}\right)^2}{N}}$$

$$= \sqrt{\frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N}}$$

where σ is the standard deviation;
$\overline{X}$ is the mean;
N is the total number of data points;
X_i is the value of the i^{th} data point, for i = 1, 2, …, N; and
$\left(X_i - \overline{X}\right)^2$ is the squared deviations from the mean, for i = 1, 2, …, N.

In our example, the standard deviation of the data set is 4.2 (i.e. $\sqrt{17.6}$).

Both variance and standard deviation are more powerful measures of the spread of data, as they can be used for manipulating other sophisticated statistical calculations.

In general, for two sets of data with equal (or more or less similar) values for their means, the set with a higher variance and standard deviation has a distribution with bigger dispersion.

Application of some measures of dispersion

Example 1 : Standard scores **(** Z **)**

$$Z = \frac{X_i - \overline{X}}{\sigma}$$

The standard score is a transformation of raw scores for comparison purposes. The standard score is commonly used in examinations for assessing a student's standing amongst a group of students.

Consider the marks in Mathematics and English of a class of ten students :

	Scores	
Student	Mathematics	English
(1)	95	60
(2)	90	50
(3)	80	55
(4)	87	69
(5)	79	61
(6)	55	68
(7)	70	70
(8)	97	59
(9)	75	71
(10)	55	72

Student (4) scores 87 in Mathematics and 69 in English. In order to understand the student's standing amongst his fellow classmates, we can separately calculate his standard scores in the two subjects.

	Mathematics	English
Class mean ($\overline{X}$)	78.3	63.5
Class standard deviation (σ)	14.2	7.2
Student (4)'s standard score (Z)	$\frac{87 - 78.3}{14.2}$	$\frac{69 - 63.5}{7.2}$
	= 0.61	= 0.76

Comparing his original score in each subject with the respective mean score, we may already observe that he is doing better than the average in both subjects. If supplemented by the standard scores, we may further observe that he stands relatively better in English amongst his classmates than he does in regard to Mathematics.

Example 2 : Standard deviation as an indication of precision

Two balances, A and B, are used for measuring the weight of a piece of baggage of about 100 kg, each for 20 times.

There may be the thinking that a balance should always give the same reading in respect of the weight of the baggage no matter how many times weighing is done. In fact, this is not the case as illustrated by the experiment below :

	Balance readings (kg)	
Balance measurement	Balance A	Balance B
1	100.3	100.3
2	100.3	100.2
3	100.0	99.7
4	99.8	100.1
5	99.7	99.8
6	100.3	100.2
7	100.1	100.2
8	99.6	99.9
9	100.2	100.4
10	100.3	100.3
11	100.2	100.0
12	99.6	100.3
13	100.4	100.2
14	99.6	99.8
15	99.7	99.9
16	100.3	100.1
17	100.0	100.4
18	99.6	100.1
19	100.4	100.2
20	99.6	99.9
Mean	100.0	100.1
Standard deviation	0.3	0.2

Balance A gives a standard deviation of 0.3 kg while Balance B gives a standard deviation of 0.2 kg. Which balance is more “precise”?

The answer is Balance B. It should, however, be noted that we are actually not sure whether Balance B or Balance A measures weight more accurately (i.e. we do not know whether the weight of the baggage is actually closer to 100.0 kg or 100.1 kg).

References

A.S. C. Ehrenberg, A Primer in Data Reduction, Chapter 2, John Wiley & Sons Ltd.

Croxton, Frederick E., Applied General Statistics, P. 189-198, Prentice-Hall, 1975.

Spiegel, Murray R., Theory and Problems of Statistics, Chapter 4, Schaum Publishing Co., 1961.

The English Universities Press Limited, Introduction to Management Statistics, 1967.

Exercise

Application of standard score

The following are the examination marks of a class of ten students. Comparing their total standard scores, which student in the class would you consider to have attained the best overall result?

	Examination scores		
Student	English	Chinese	Mathematics
(1)	40	62	80
(2)	57	60	60
(3)	60	70	95
(4)	48	64	92
(5)	53	66	75
(6)	66	72	78
(7)	72	70	86
(8)	49	67	87
(9)	69	60	99
(10)	70	74	83

	Examination scores			Standard examination scores			
Student	English	Chinese	Mathematics	English	Chinese	Mathematics	Total
(1)	40	62	80				
(2)	57	60	60				
(3)	60	70	95				
(4)	48	64	92				
(5)	53	66	75				
(6)	66	72	78				
(7)	72	70	86				
(8)	49	67	87				
(9)	69	60	99				
(10)	70	74	83				
Mean							
Standard deviation							

The student in the class who has attained the best result is _____ .

9 Official statistics

Introduction

Official statistics are generally statistics compiled by the government based on raw data from such sources as administrative records, special returns and surveys. Where surveys are conducted, there are often statistical legislations requiring persons or companies to supply the required raw data.

The Census and Statistics Department (C&SD) is the central statistical office of Hong Kong. C&SD together with statistical units established in various government departments and bureaux form the Government Statistical Service. Broadly speaking, most general-purpose statistics come under the responsibility of C&SD. The statistical units in various government departments and bureaux are responsible for specific-purpose statistics (for dedicated use in their respective work) and provide necessary support in the application of statistics.

In this knowledge-based era, more and more people are making use of statistics in their work and daily lives. From macro-economic analysis to daily administration and long term planning of the government, to decision-making of individuals and business firms, and to enhancing people's understanding of the society and their participation in discussions about social and economic policies, official statistics play an important role in a wide range of endeavours.

Some of the more common economic and social statistics are introduced in the following sections.

(A) Economic statistics

Economic statistics comprise, among others, the following areas : trade statistics, national income statistics, Balance of Payments statistics, labour statistics, price statistics and operating statistics of businesses.

Just as doctors need to use medical equipment when they make diagnosis, economic analysts need to make reference to statistics in forming their views from theoretical considerations. Economic statistics serve as measuring instruments which enable the analysts to judge whether different economic sectors in the economy are in harmony, whether the economy is stagnant or overheated, and whether performance will improve or decline. Although different analysts may have very different views on economic prospects or on the role of the government, there is no doubt about the importance of good measuring instruments to them.

Apart from facilitating macro-economic analysis, economic statistics are also very useful in business planning. Businessmen can make use of the statistics in deciding what businesses to invest in, what products to produce, where to locate shops, at what wages they should employ workers, and so on.

Trade statistics

Hong Kong is an externally oriented economy. Its economic performance is highly related to the performance of the external sector. Hong Kong external trade comprises trade in goods (i.e. merchandise trade) and trade in services.

Trade in goods covers movements of merchandise between Hong Kong and her trading partners, by land, air, water and by post. It comprises imports, domestic exports and re-exports.

Trade in services includes exports and imports of services. Exports of services are the sales of services to the rest of the world, whereas imports of services are the purchases of services from the rest of the world.

Merchandise trade

Merchandise trade statistics provide important indicators of the performance of the external sector of the economy in respect of merchandise trade. They can be analysed by the three components of external merchandise trade, namely, domestic exports, re-exports and imports. In addition to aggregate statistics of trade in respect of each of the components, detailed statistics are also available on Hong Kong’s trade with different countries/territories and for different commodities.

Besides providing very useful information on the performance of an economy, trade statistics can also be used by business firms in decision-making on such areas as forecasting business prospects, measuring market shares, planning investment and formulating marketing, production and development strategies.

As a free port, Hong Kong has minimal import and export licensing requirements. Most products do not need licences to enter or leave Hong Kong. Unlike many other economies, there is no import tariff. Yet, persons who import or export goods (except for exempted goods) are required to submit import/export declarations to the Customs and Excise Department within 14 days after the importation/exportation of goods. The import/export declarations are subsequently forwarded to C&SD for further processing and compilation of Hong Kong's external merchandise trade statistics. The import/export declaration is a typical example of special returns for collecting raw data essentially used for trade statistics compilation. Electronic submission of trade declarations has been made mandatory as from April 2000. In 2011, around 18.4 million import/export declarations were processed by C&SD.

As stated in "The Basic Law of the Hong Kong Special Administrative Region of the People's Republic of China", the Hong Kong Special Administrative Region is a separate customs territory. Import/export declaration is also required of Hong Kong's trade with the mainland of China, and statistics relating to this are thus included in Hong Kong's trade statistics.

In 2011, the value of Hong Kong's total merchandise trade was $7,101.8 billion, higher than that in 2010 by 11%. Of this, $3,337.3 billion were attributable to total exports and $3,764.6 billion to imports.

Between 2001 and 2011, the values of total exports and imports increased respectively at an average annual growth rate of 8.5% and 9.2%.

Total exports comprise domestic exports and re-exports. Domestic exports are the natural produce of Hong Kong or the products of a manufacturing process in Hong Kong which has changed permanently the shape, nature, form or utility of the basic materials used in manufacture. Re-exports are products which have previously been imported into Hong Kong for subsequent re-exportation. Unlike domestic exports, they have not undergone in Hong Kong any manufacturing process which has changed permanently their shapes, nature, forms or utilities. The values of both domestic exports and re-exports are recorded on f.o.b. (free-on-board) basis, i.e. on the basis of the value of the commodities when they leave the customs boundary of Hong Kong.

Imports are goods which have been produced or manufactured in places outside the jurisdiction of Hong Kong and brought into Hong Kong for domestic use or for subsequent re-export as well as Hong Kong products re-imported. Their values are recorded on c.i.f. (cost, insurance and freight) basis, i.e. on the basis of the value of the imported commodities when they enter into Hong Kong. C.i.f. value of an imported commodity is usually higher than its f.o.b. value, as the former includes also the insurance and transportation fees for delivering the commodity from the exporting territory to the importing territory.

The mainland of China and the U.S.A. were the two largest destinations of Hong Kong's domestic exports in 2011. They accounted for 46.8% and 10.9% respectively of the value of domestic exports. Over the past decade, the relative importance of the mainland of China as a destination of Hong Kong's domestic exports grew substantially. This was related to the economic growth in the mainland, with increased demand for imports. At the same time, trade involving outward processing was also a contributing factor, as raw materials and semi-manufactures were exported to the mainland for processing.

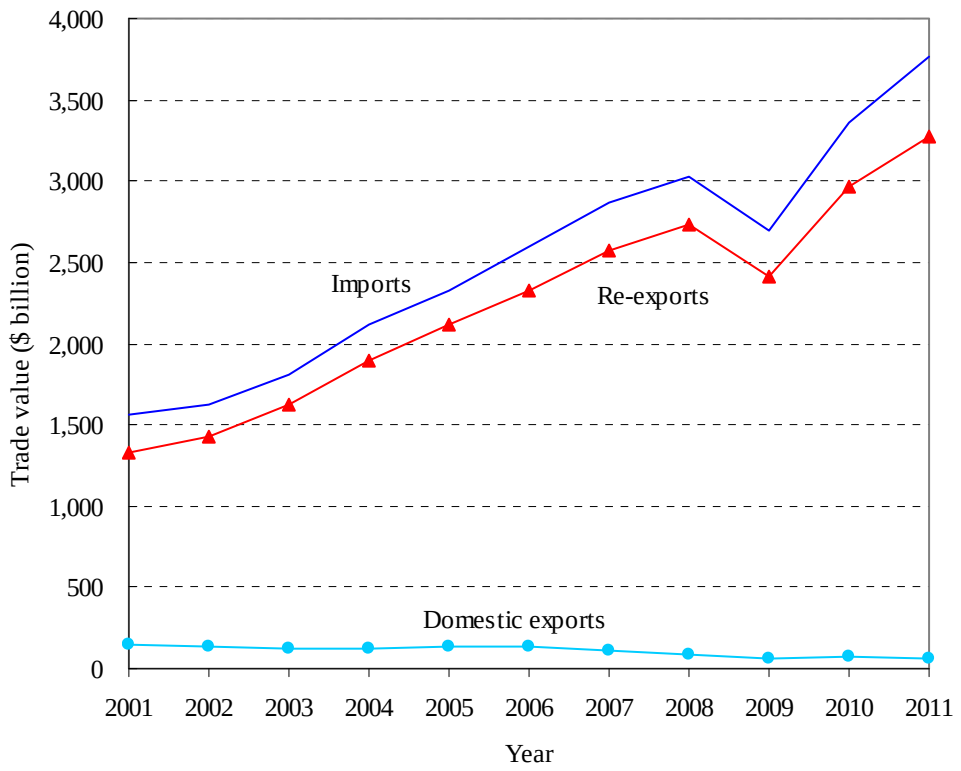
Hong Kong's domestic exports mainly include jewellery, goldsmiths' and silversmiths' wares, and other articles of precious or semi-precious materials; plastics in primary forms; metalliferous ores and metal scrap; non-ferrous metals and electrical machinery, apparatus and appliances, and electrical parts thereof.

The mainland of China and the U.S.A. were also the two largest destinations of Hong Kong's re-exports in 2011. They accounted for 52.5% and 9.9% respectively of the total re-export value. Benefited from the open door policy of the mainland of China, the importance of re-export trade in Hong Kong's external merchandise trade had become more pronounced in the past ten years or so. The share of re-exports among total exports increased from 89.6% in 2001 to 98.0% in 2011.

The major commodities in Hong Kong's re-exports include electrical machinery, apparatus and appliances, and electrical parts thereof; telecommunications equipment; office machines and automatic data processing machines; clothing and non-metallic mineral manufactures.

Most of Hong Kong's imports come from the mainland of China, Japan, Singapore, Taiwan and the U.S.A. In 2011, the mainland of China accounted for 45.1% of the total import value, compared to 43.5% in 2001. The increase in imports from the mainland of China was partly attributable to outward processing activities because imports from the mainland of China included those goods which had undergone outward processing there and subsequently imported into Hong Kong.

Chart 9.1 Values of domestic exports, re-exports and imports, 2001–2011



Trade in services

With the structural change in the Hong Kong economy, the importance of services industries has been on the rise. Being an externally oriented economy, Hong Kong is an important centre not only for trade in goods, but also for trade in services (TIS). TIS in Hong Kong has been maintaining observable growth in recent years.

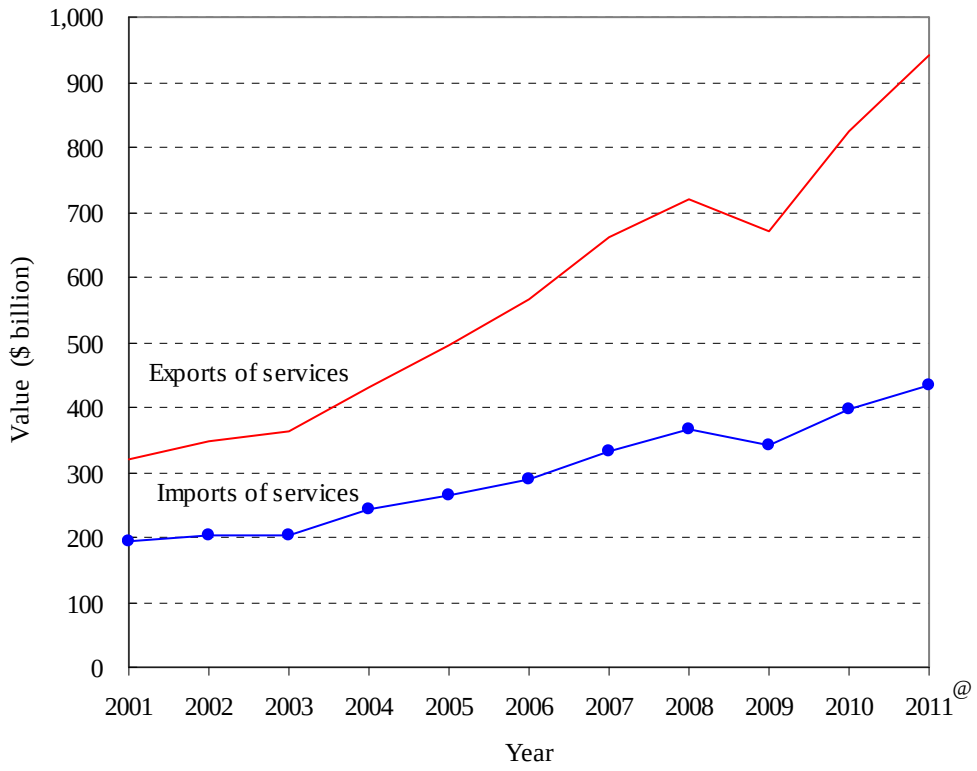
Hong Kong's TIS statistics can be traced back to the reference year 1980. Owing to data limitation, the classification of services at that time was only based on major industries in which the Hong Kong businesses were engaged. In response to the growing demand for detailed statistics on TIS in recent years, Hong Kong has started releasing detailed TIS statistics by type of services and country/territory on an annual basis since 2002. TIS of Hong Kong are categorised into six major service groups, viz. transportation services, travel services, insurance services, financial services, merchanting and other trade-related services (MOTS), and other services.

During the period 2001 to 2011, the value of total exports of services rose significantly, from $320.8 billion in 2001 to $941.3 billion in 2011, at an average annual growth of 11.4%. Total imports of services also increased significantly, from $194.2 billion in 2001 to $434.9 billion in 2011, at an average annual growth of 8.4%. The more important exports of services included MOTS (at $274.6 billion in 2011) and transportation services (at $254.0 billion in 2011), which also brought about a substantial surplus in the overall TIS after netting out the total imports of services.

Analysing exports of services of Hong Kong by geographical breakdown(1) in 2010, the mainland of China (28.3%) and the United States of America (21.2%) were the most important destinations, followed by the United Kingdom (7.8%), Japan (5.9%) and Taiwan (4.2%).

For imports of services(1) of Hong Kong, the mainland of China (26.6%) and the United States of America (14.6%) were the most important sources, followed by Japan (8.2%), the United Kingdom (6.8%) and Singapore (6.0%).

Chart 9.2 Values of exports and imports of services, 2001–2011



Note :

(1) Since geographical breakdown figures of financial intermediation services are not available, figures by country/territory breakdown do not include financial intermediation services.

National income statistics

The Gross Domestic Product (GDP) and Gross National Product (GNP) are core statistics in the field of national income. They are both important indicators of the economic performance of an economy.

GDP is an aggregate measure of the total value of production of all resident producing units of a country or territory in a specified period (e.g. a year) before deducting allowance for consumption of fixed capital. By providing statistics on the production of various sectors of the economy, GDP statistics are particularly relevant for analysis related to production activities within the economy (e.g. employment, productivity, industrial output and investment in equipment) and changes in the economic structure.

GDP of Hong Kong is compiled using the production approach and the expenditure approach. Under the production approach, GDP is the sum of **value added** of all resident producing units (e.g. factories, shops, service organisations) of Hong Kong. Simply speaking, the value added of a producing unit is the value of its gross output less the value of intermediate consumption (the value of goods and services used up in the course of production).

It is essential to apply the concept of value added in GDP compilation in order to avoid double counting the value of production. Because various producing units are indeed closely linked in the course of production, the goods or services produced by a producing unit may be used as the intermediate consumption of another producing unit. If the gross output of individual producing units are added up directly, the values of certain goods and services will be counted more than once and the GDP so derived will exaggerate the true value of production.

From another angle, goods and services produced will eventually be consumed by final users which will be captured in the total expenditure on final demand for goods and services. Thus, the expenditure approach can also be applied for compilation of GDP.

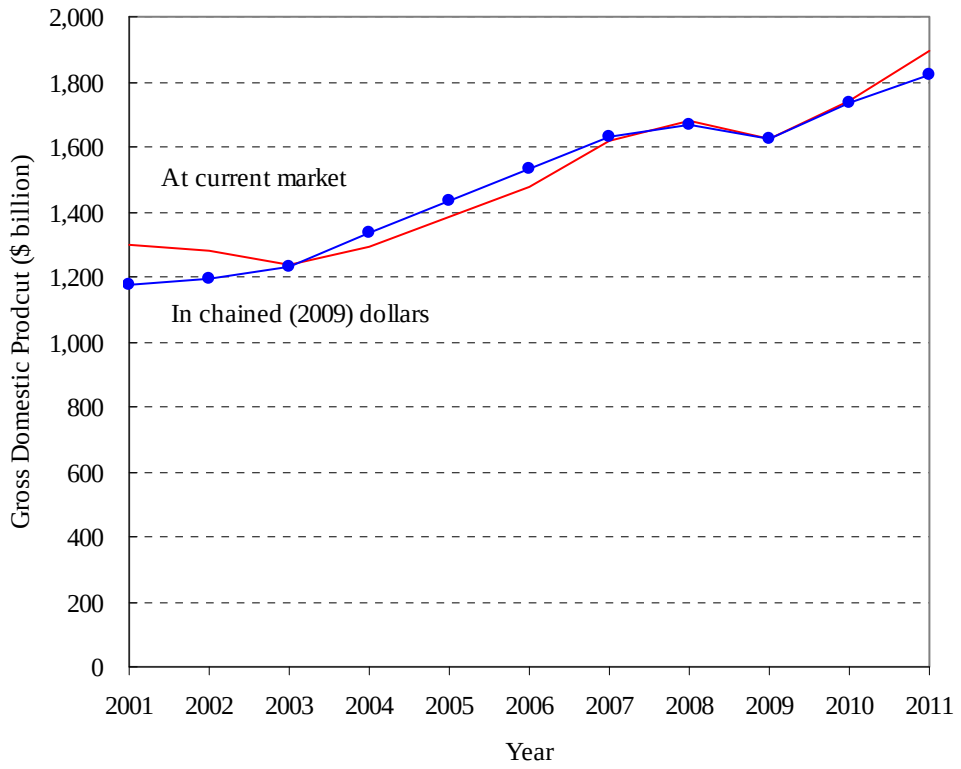
Under the expenditure approach, GDP is compiled as the total value of goods and services produced for final use, net of their import content which does not attribute to domestic production. Note that goods and services used as intermediate input for production are not included to avoid double-counting.

The production and expenditure approaches represent how components within the GDP can be viewed from two different perspectives. In theory, there is only one GDP figure. However, as the compilation of GDP requires a large volume of data gathered from various sources and the data sources used in different approaches are not the same, some minor statistical discrepancy may exist between the GDP statistics compiled under the two different approaches.

Over any period of time, changes in GDP valued at current market prices reflect changes in both the volume and the price of goods and services produced. To measure the growth of an economy in real terms, the effect of price changes has to be eliminated. To remove the price change element, current price estimates are converted to "volume" estimates. To compile the volume measures of GDP and its components, the annually re-weighted chain linking approach is adopted. Under this calculation method, the volume measure of GDP for each year is compiled by aggregating the volume estimates of its components using their price structures in the preceding year as weights. For a particular year, the volume estimates of major components revalued at preceding year prices are first produced by "deflating" the current price values of sub-components by the relevant price indices at the most disaggregated levels. The volume estimate of GDP is then obtained by aggregating the volume estimates of GDP major components revalued at preceding year prices.

Hong Kong's GDP for 2011 at current market prices amounted to $1,893.9 billion, about 1.5 times the corresponding figure of $1,299.2 billion for 2001. GDP for 2011 in chained (2009) dollars was $1,823.1 billion, or 1.6 times the corresponding figure of $1,174.3 billion for 2001. This represented an average annual growth rate of 4.5% in real terms between 2001 and 2011.

Chart 9.3 Gross Domestic Product of Hong Kong, 2001–2011



Apart from being a measure of the overall economic performance of an economy, GDP is often used in international comparisons of state of economic well-being. In the light of the variations in the scale of different economies, "per capita GDP" is a

more relevant indicator for comparing the economic well-being in different places. Per capita GDP is computed by dividing the total GDP in a year by the population size of the economy concerned in the same year.

At current market prices, per capita GDP for Hong Kong increased from $193,500 in 2001 to $267,812 in 2011. This represented an average annual growth rate of 3.3%. To eliminate the effect of price changes over the ten-year period, the average annual growth rate of per capita GDP in real terms can be referred to, which was 4.0% during the period.

The Gross National Product (GNP) is another useful measure of economic performance. It measures the total income earned by residents of a country or territory(2) from engaging in various economic activities, irrespective of whether the economic activities are carried out within the economic territory or outside. Therefore, GNP is particularly useful for analysing economic situations relating to income of residents, investment behaviour, domestic demand and inflation.

GNP is derived from GDP by adding factor income earned by residents from outside the economic territory and subtracting factor income earned by non-residents from within the economic territory. Factor income includes income from direct investment, portfolio investment and other investment(3), and compensation of employees. Similar to GDP, GNP can be compiled on the basis of current market prices and in chained dollars.

The chained dollar estimate of GNP for a particular year is obtained by multiplying the chain volume index for that year by the current price value in the reference year. In compiling the continuous time series of the chain volume indices of GNP, the annually re-weighted chain linking approach is adopted.

In Hong Kong, the first set of GNP statistics was produced in 1995 in respect of reference year 1993.

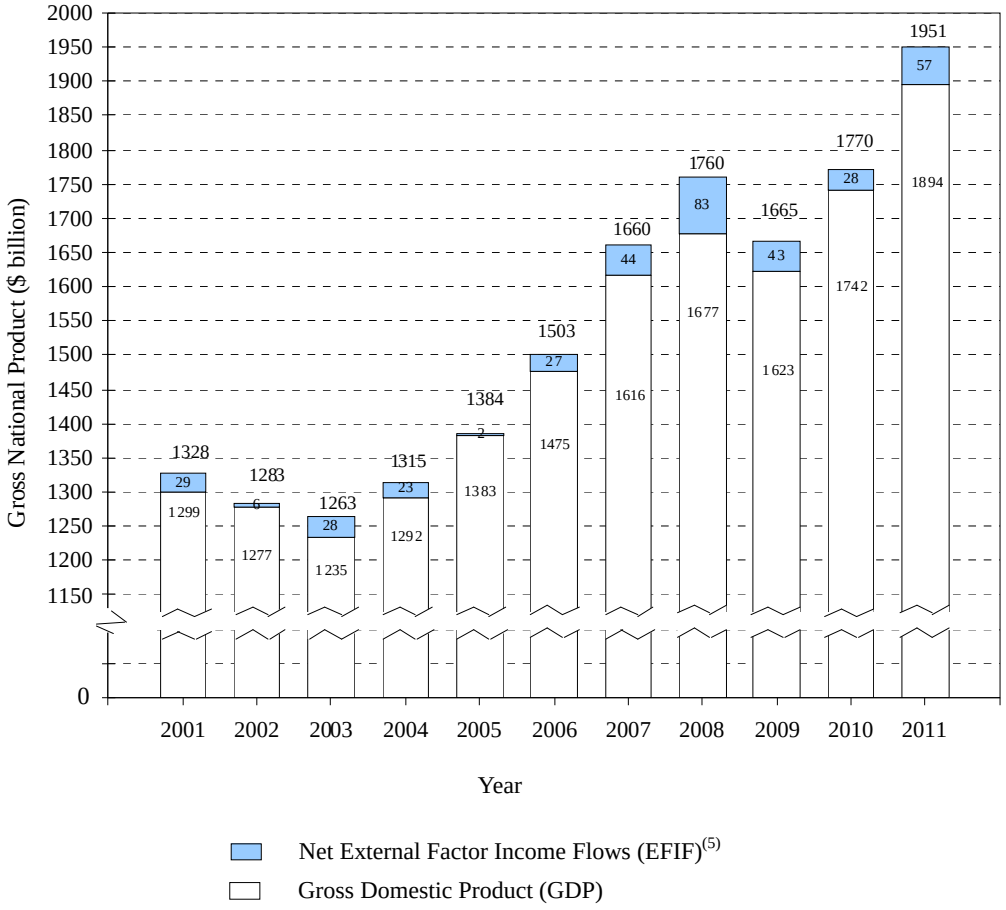
GNP for 2011 in chained (2009) dollars, obtained by discounting the effect of price changes from the GNP at current market prices, was $1,876.7 billion, or about 1.6 times the corresponding figure for 2001.

Notes :

(2) Residents of a country or territory include both individuals and organisations. According to statistical international standards, resident individuals refer to those who normally stay in the economic territory of the country or territory for at least 12 months or longer, or intend to do so, irrespective of their nationality. For organisations, residents refer to those which ordinarily operate in the economic territory of the country or territory. The residence status of individuals and organisations thus depends on their centre of economic interests.

(3) Please refer to the section on “Balance of Payments Statistics” in page 9:11 for more about direct investment, portfolio investment and other investment.

Chart 9.4 Gross National Product of Hong Kong (at current market prices), 2001–2011[4]



Notes :

(4) Individual figures may not add up exactly to the total due to rounding.
(5) Net EFIF equals to external factor income inflow net of external factor income outflow.

Balance of Payments statistics

Just as an individual person would keep an account for himself in respect of his current income and expenditure and his assets and liabilities, an economy would do the same using the method of "national accounting". This would trace the economy all the way through production, distribution of income and capital accumulation. In particular, the economy would keep an account on all its transactions with the rest of the world. Such an account is called the Balance of Payments (BoP) account.

A BoP account is a statistical statement that systematically summarises, for a specific time period, the economic transactions of an economy with the rest of the world. A complete BoP account comprises the following two broad accounts : (a) the current account; and (b) the capital and financial account.

The current account largely measures the flows of real resources including exports and imports of goods and services, income receivable and payable abroad, and current transfers from and to abroad. Current account transactions reflect the provision and acquisition of real resources by an economy to and from other economies.

Current transfers are transactions in which residents of an economy provide/receive real or financial resources that are likely to be consumed immediately or shortly, to/from non-residents without the receipt/provision of equivalent economic values in return. Current transfers are unilateral in nature and are offsetting entries in the BoP account for one-sided transactions. They include workers' remittances, donations, official assistance, pensions and so on.

Within the capital and financial account, the capital account measures external transactions in capital transfers, and in the acquisition or disposal of non-produced, non-financial assets. Capital transfers are the transfers of ownership of a fixed asset or the forgiveness of a liability without receiving any economic value in return. Cash transfers that are linked to, or conditional on, the acquisition or disposal of a fixed asset are also capital transfers. Capital transfers consist mainly of debt forgiveness and migrants' transfers. Migrants' transfers are contra-entries to the flows of goods or changes in financial items that arise from the migration of individuals between two economies. The financial account records transactions in financial assets and liabilities between residents and non-residents. It shows how an economy's external transactions are financed. Transactions in the financial account are classified into direct investment, portfolio investment, financial derivatives, other investment and reserve assets.

Direct investment refers to external investment which allows an investor of an economy to have a lasting interest and a significant degree of influence or an effective voice in the management of an enterprise located in another economy. For statistical purpose, an effective voice is taken as equivalent to a holding of 10% or more of the voting power in an enterprise.

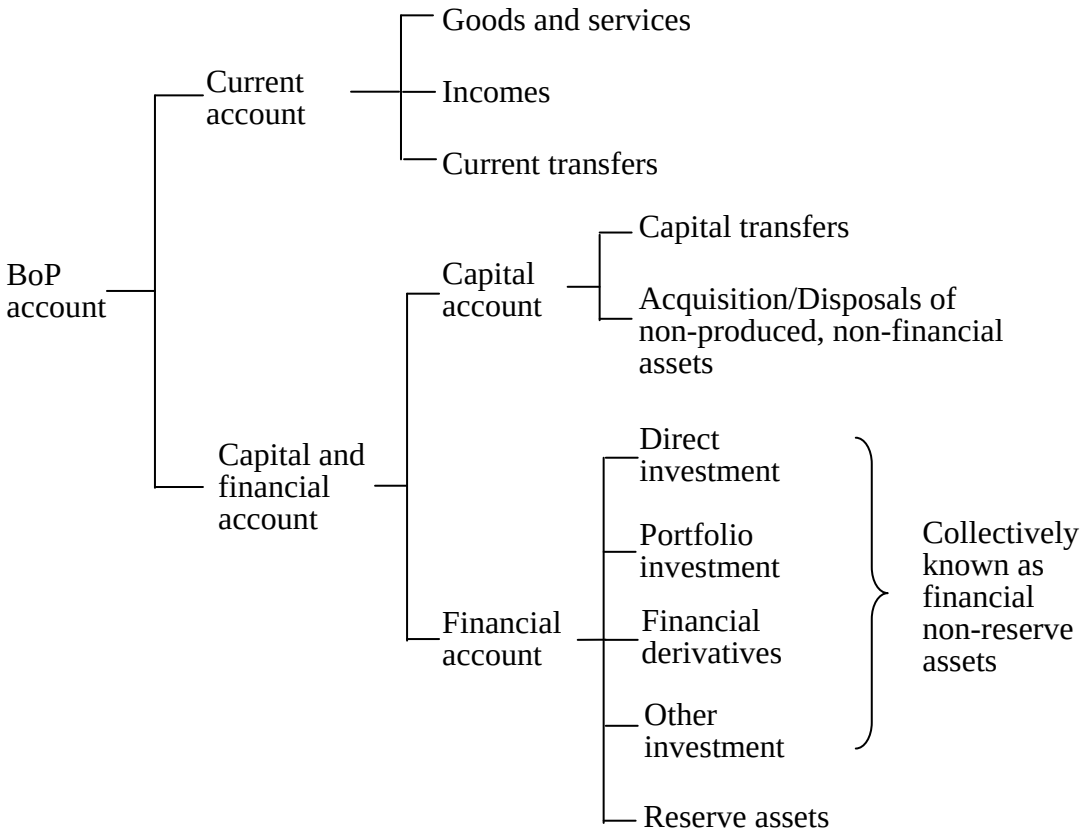
Portfolio investment refers to investment in non-resident equity securities and debt securities (e.g. bonds and notes, money market instruments). Compared with direct investors, portfolio investors in equity securities and debt securities of non-resident enterprises have no lasting interest or influence in the management of the companies they invest. A holding of less than 10% voting power in an enterprise is regarded as portfolio investment.

Financial derivatives are financial instruments that are linked to a specific financial instrument or indicator or commodity, and through which specific financial risks can be traded in financial markets (including on exchanges and over-the-counter) in their own right.

Other investment refers to other financial claims on and liabilities to non-residents that are not classified as direct investment, portfolio investment, financial derivatives or reserve assets. This includes non-marketable loans, currency and deposits, trade credits, financial leases and so on.

Reserve assets consist of foreign currency assets that are readily available to and controlled by the monetary authority of an economy (in the case of Hong Kong, the Hong Kong Monetary Authority) for directly financing payment imbalances and for indirectly regulating the magnitude of such imbalances through intervention in foreign exchange markets to affect the currency exchange rate of that economy.

The following diagram shows the individual major components of the BoP account :



In principle, the sum of the current account balance, the capital account balance and the net change in financial non-reserve assets, i.e. the overall balance of payments, is equal to the net change in reserve assets.

For a reference period, if an economy receives more foreign currencies than it pays in external transactions in goods, services, income and investment, as well as in external transfers and remittances, then it is said to have a BoP surplus, which is equal to its overall net inflow of funds from the rest of the world. Conversely, there will be a BoP deficit which is equal to its net outflow of funds.

The first set of complete BoP account for Hong Kong for the reference year of 1997 was published by C&SD in April 1999. Currently, annual and quarterly BoP account with detailed breakdowns of capital and financial flows are compiled. The availability of a complete BoP account is a major statistical development for Hong Kong. It is also one of the requirements stipulated in the Special Data Dissemination Standard (SDDS) of the International Monetary Fund (IMF), which Hong Kong has subscribed to and which underpins Hong Kong's commitment to the international financial community for providing data in compliance with international standards.

In 2010, Hong Kong recorded a BoP surplus amounting to $71.1 billion (as a ratio of 4.1% to GDP). This was lower than the BoP in 2009, which was a surplus of $549.3 billion (as a ratio of 33.9% to GDP).

Of the major BoP components, the current account surplus narrowed to $96.2 billion (as a ratio of 5.5% to GDP) in 2010, from $139.2 billion (as a ratio of 8.6% to GDP) in 2009. For financial non-reserve assets, there was a net outflow of $122.3 billion (as a ratio of 7.0% to GDP) in 2010 against the net inflow of $357.7 billion (as a ratio of 22.0% to GDP) in 2009.

Besides the BoP account, the International Investment Position (IIP) statement is also an important component under the framework of BoP statistics.

While the BoP financial account measures transactions in external financial assets and liabilities during a period, the IIP statement reflects the level of such assets and liabilities at the end of the period. The former is a flow concept, whilst the latter is a stock concept. Transactions are economic flows during a specific time period. On the other hand, stocks are positions in, or holdings of, assets and liabilities at a particular time point. Transactions in assets and liabilities, among other factors, will affect the stock of these assets and liabilities. In addition to transactions, price changes and exchange rate variations as well as some other adjustments also have an impact on the value of stock of an economy's external financial assets and liabilities when expressed in the currency of the host economy.

C&SD started to release the annual IIP statistics for Hong Kong in mid-2002. Currently, time series on annual IIP statistics since the end of 2000 and quarterly IIP statistics since the first quarter of 2010 are available.

At the end of 2010, Hong Kong's external financial assets and liabilities amounted to $22,277.4 billion and $17,141.9 billion respectively. After netting out the external financial liabilities from the external financial assets, Hong Kong stood as a net creditor, with net external financial assets amounting to $5,135.5 billion at the end of 2010. This corresponded to 295% of the GDP in that year.

BoP statistics are very useful for monetary and financial monitoring purposes and also for formulation of related policies in both the territorial and international contexts. The statistics provide an objective basis for assessing the financial soundness and the macro-economic and monetary situation of Hong Kong, which is a crucial element for many business decisions. The availability of BoP statistics also helps discern the various forces in the foreign exchange market affecting the exchange rate of the Hong Kong dollar, and is useful for analysing the situation in regard to money supply and demand, domestic demand and inflation, and movements in the interest rates.

Labour statistics

Labour statistics are very useful economic statistics regarding the labour force and production of Hong Kong. They reflect the characteristics of the economically active population and their economic activities. Since working is also an important aspect of one's life, labour statistics are considered as social statistics as well.

The labour force participation rate (LFPR) is obtained by expressing the size of the labour force as a percentage of the population of working age, i.e. population aged 15 and over. This is a measure of the working age population's propensity to be in the labour force. The size of the labour force of Hong Kong in 2011 was 3.70 million and the LFPR was 60.1%.

The unemployment rate, in particular, is one of the most important indicators describing the labour market situation. It refers to the proportion of "unemployed population" in the "labour force". The unemployment rate was 4.9% in 2000. Because of economic downturn, the unemployment rate rose to the high level of 7.9% in 2003 as result of the outbreak of Severe Acute Respiratory Syndrome during March to June of the year. Yet, the labour market improved drastically towards the end of 2003 and this trend sustained thereafter. Consequently, the unemployment rate dropped significantly to 3.5% in 2008. Impacted by the outbreak of the global financial tsunami, the unemployment rate rose to 5.3% in 2009. With the improving economy, the unemployment rate dropped to 3.4% in 2011.

As the unemployment rate is subject to effects of seasonal variations (e.g. there are generally more job-seeking graduates entering the labour market in late summer, leading to a higher unemployment rate), short-term fluctuations, say between adjacent months or quarters, in the unemployment rate do not necessarily signal an actual change in the labour market conditions. Hence, when studying the underlying trend of unemployment, it is common to use the seasonally adjusted unemployment rate, which is obtained by using statistical methods that estimate and

remove the seasonal variations from the original figures.

The definition used for measuring unemployment follows closely that recommended by the International Labour Organisation (ILO). According to the definition, a person aged 15 or over is classified as unemployed if he/she satisfies all of the three following criteria :

(a) has not had a job and has not performed any work for pay or profit during the seven days before enumeration;

(b) has been available for work during the seven days before enumeration; and

(c) has sought work during the thirty days before enumeration.

However, if a person aged 15 or over fulfils conditions (a) and (b) above only and he/she has not sought work during the thirty days before enumeration because he/she believed that work was not available, he/she is still classified as unemployed, and is regarded as a so-called “discouraged worker”.

It should be emphasised that “whether available for work” is an important condition for defining unemployment. Those persons who are not available for work because of one reason or another (e.g. being fully engaged in household duties) are not classified as unemployed even though they are without work, as they do not fulfil condition (b) of the above definition of unemployment, i.e. not available for work during the seven days before enumeration.

When looking at the unemployment figures, it should be noted that “unemployment” is a state of affairs, which is not equivalent to the event of “losing a job”. For persons who have been dismissed from their jobs or who have left their jobs for certain reasons, but have managed to take up another job within a short period of time, they may not be counted in the unemployed population by definition, even though they have experienced the event of “losing a job” during the period.

Chart 9.5 Unemployment rate in Hong Kong, 2000–2011



Statistics in respect of "underemployed persons" also provide another useful piece of information about the labour force. The underemployed persons are defined as those employed persons who have involuntarily worked less than 35 hours during the seven days before enumeration and have sought additional work during the thirty days before enumeration, or have not sought but have been available for additional work during the seven days before enumeration. Those employed persons who worked less than 35 hours per week but have not been available for additional work and have not sought additional work (e.g. full-time teachers in bi-sessional schools) are not classified as underemployed.

Statistics on labour force, employment, unemployment and underemployment are compiled by C&SD based on the data collected from a continuous General Household Survey (GHS) conducted on households (and their members). The sample size of the GHS in a three-month period is about 24 000 households and the survey methodology adopted is in line with international standards and practices.

Apart from the aggregate figures mentioned above, a large amount of statistics analysed by various groupings (e.g. industrial and occupational distributions of the employed population) are also available from the GHS. These statistics are provided to the government departments/bureaux concerned for reference and are also published in the survey reports which are for sale to the public.

Furthermore, statistics on employment (i.e. number of employees and vacancies in different industries) and wage and payroll statistics are also compiled by the department. They are commonly referred to by enterpreneurs and human resources management personnel, since they provide information which is very useful for formulation of personnel management strategies. For example, firms can compare their staffing and vacancy situations with those of the industry taken as a whole or with those of other industries. They can also compare wage levels of their employees with others in the same industry. The relevant data are collected from surveys conducted on business establishments.

Price statistics

The Consumer Price Index (CPI) is another widely quoted item of official statistics. It is an important economic indicator which summarises changes over time in the price levels of consumer goods and services generally purchased by households. The year-on-year rate of change in the CPI is commonly used to reflect the inflation or deflation affecting households. Specifically, it measures changes in the total cost of a given basket of consumer goods and services representative of the purchases made by households in a specified time period.

The CPI has been widely accepted and used as a reference for salary adjustment purposes so as to maintain the purchasing power of employees in the face of changing prices.

The prices of some consumer goods and services may be subject to seasonal variations causing the CPI to rise or fall. Therefore, when studying changes in the CPI, it is common to refer to year-on-year changes (e.g. comparing the CPI of a month with that of the same month in the preceding year) so as to eliminate the effects of seasonal factors. Another alternative is to use the deseasonalised CPI. The latter is obtained by using statistical methods that estimate and remove the seasonal variations from the original index.

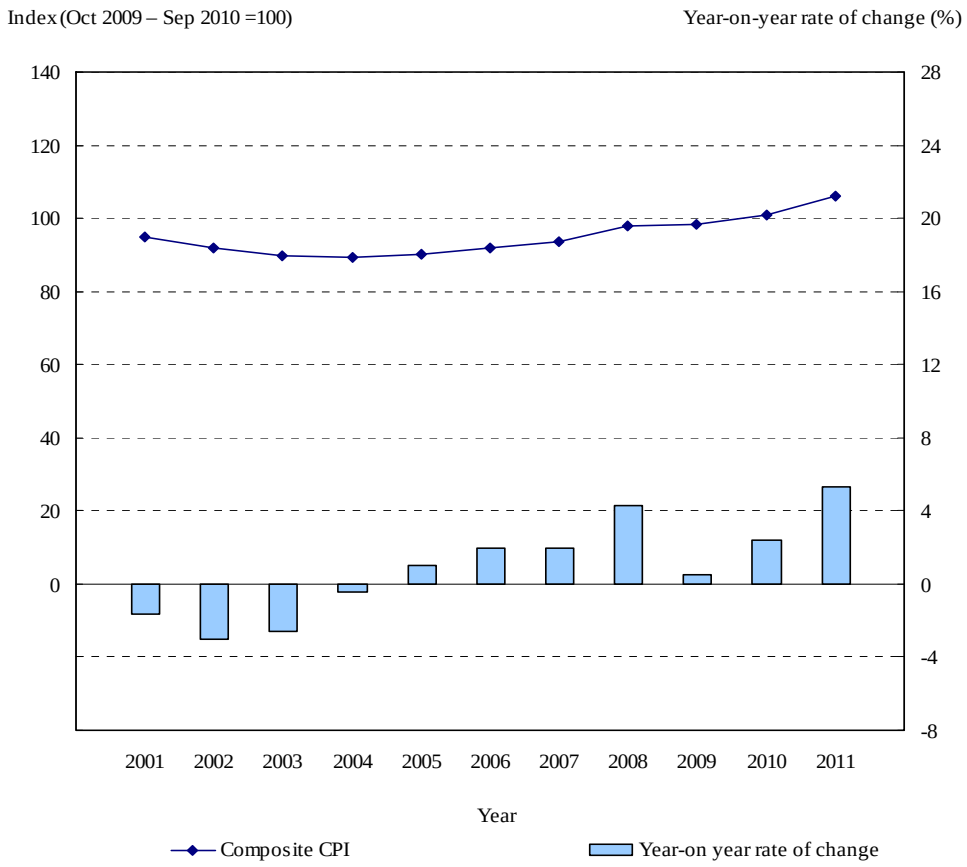
Households in different expenditure ranges have different expenditure patterns. Different series of CPIs are therefore compiled by C&SD to reflect the impact of consumer price changes on households in different expenditure ranges. The CPI(A), CPI(B) and CPI(C) are compiled based on the expenditure patterns of households in the relatively low, medium and relatively high expenditure ranges respectively. The Composite CPI, which is compiled based on the overall expenditure pattern of all these households taken together, reflects the impact of consumer price changes on the household sector as a whole.

Two types of data are required for compiling the CPI, namely, price movements of consumer goods and services and their expenditure weights. Given the wide variety of consumer goods and services, it can be imagined that a lot of price data must be collected for the compilation of the CPI. Indeed, C&SD collected some 47 000 price quotations each month from around 4 000 retail outlets and service providers for the purpose of compiling the index.

As for the expenditure weights, they are derived from the results of a Household Expenditure Survey which is conducted once every five years. As households spend more on some items and less on the others, similar price movements in different items may have different impacts on the households and effects on the overall price change. Therefore, a weighting system which represents the relative importance, in terms of expenditure, of individual items in the basket of consumer goods and services bought by households is required for the compilation of the CPI. The weight of each item represents the share of the item in the total expenditure of households.

From 2002 to 2004, the CPI registered annual decreases due mainly to the weakness of the global economy. Besides, the increasing economic integration between Hong Kong and the Mainland might have accelerated the price convergence between two places in those years. Along with the improved economic conditions, the CPI reverted to annual increases since 2005. Since the mid-2007, consumer prices began to increase significantly amidst the upsurge in food and energy prices in international markets and appreciation of renminbi. The increase started to moderate towards the end of 2008 due to the outbreak of the financial tsunami, and registered a much lower rate of increase in 2009. The CPI went back to the increasing trend again since mid-2010, mainly driven by the increases in food prices and private housing rents.

Chart 9.6 Composite Consumer Price Index and its year-on-year rate of change, 2001–2011[6], [7]



In making reference to the CPI or its change over time, it should be noted that the information thus conveyed reflects the collective experience of all households about inflation/deflation and that it does not always correspond to the experience of individual households. As each household has its own expenditure pattern and prices of different goods and services increase or decrease at varying rates, inflation/deflation does not affect all households to the same extent. If a household spends a lot on goods and services with rapid price increases/decreases, the household will feel a greater impact of inflation/deflation.

Besides, people are generally psychologically more aware of isolated cases of drastic price changes, especially those which affect them personally. Thus, one would tend to notice and remember only those large changes and do not realise that there are other cases of moderate or small changes. Even individual members of the same

Notes :

(6) The CPIs from October 2009 onwards are compiled based on expenditure weights obtained from the 2009/10 Household Expenditure Survey. The CPIs for earlier periods are compiled based on old weights and have been re-scaled to the new base period for linking with the new index series.

(7) From October 2010 onwards, the year-on-year rates of change are derived from the 2009/10-based CPIs. The year-on-year rates of change before October 2010 were derived using the index series in the base periods at that time (for instance, the 2004/05-based index series), compared with the index a year earlier in the same base period.

household may feel different impacts of price changes. Therefore, while personal experience may be subjective or biased, the CPI provides an objective assessment of price changes affecting households generally.

Another point to note is that the CPI measures changes in the price level of consumer goods and services purchased by households. If we buy more goods or consume goods of higher quality than before, then the increase in expenditure is not caused by price increase.

Some people may also think that using a fixed expenditure pattern in the computation of price index within a period of five years would affect the reliability of the CPI. This is in fact not true because household expenditure patterns (i.e. the relative expenditure among various items of goods and services) only change very gradually over time. Updating the expenditure pattern at five-yearly intervals conforms to international practices and is adequate for maintaining accuracy of the CPI.

Operating statistics of businesses

Statistics from economic surveys conducted periodically by C&SD provide information about business trend, cost structure and profit situation in various lines of businesses. Firms can make use of the statistics to compare their own performance against the general situation within their specific lines of business and adjust their business strategies accordingly.

Other economic statistics

Shipping statistics are compiled by C&SD in collaboration with other parties.

(B) Social statistics

Social statistics comprise, among others, the following areas : population statistics, education statistics, transport statistics, housing statistics, health statistics, social welfare statistics and crime statistics.

Population statistics

Population statistics have already been discussed in Chapters 1 to 3. It may be emphasised again here that population statistics, which depict the size, growth and socio-economic characteristics of the Hong Kong Population, are among the most useful items of statistical information for government administrators. They are of prime importance in the formulation and administration of government policies in such areas as education, housing, transport, health and social welfare services. For businesses, population statistics are useful in identifying potential target markets more accurately and in supporting better investment decision-making.

Other social statistics

Education statistics, transport statistics, housing statistics, health statistics, social welfare statistics and crime statistics are basically compiled by the respective government departments to meet specific planning and administrative needs. They are also useful to the government as a whole in formulating broad-range policies. Co-ordination is necessary to ensure that the statistics relating to different areas can be used together for planning activities which cut across many facets of the community.

On education, there are three main areas of statistics, viz. pupils, teachers and school provision. The Education and Manpower Bureau compiles a lot of statistics which are very useful for formulating policies and planning facilities.

Regular public transport operating statistics and related statistics on transport are compiled. These provide a basis for understanding the transport and traffic situation of Hong Kong and for transport planning.

Housing statistics cover housing stock and housing production in both the public and private sectors. They also relate to the situation in various housing estates. They are very useful in reflecting the housing condition of the people of Hong Kong and are important references when planning for housing development is undertaken.

Health statistics provide information about illness, health care and health facilities. Social welfare statistics attempt to establish the number and characteristics of individuals in the community who require assistance, in particular the aged, the unemployed, the disabled, the orphaned or deserted children, the delinquent and families with special problems. These statistics are useful in planning, review and monitoring of the provision of social welfare services.

Crime statistics are compiled on the basis of crimes reported to the police, which partially reflect the state of law and order in society. Some crimes are not reported and statistics on unreported crimes are collected through the conduct of a survey on crime victimisation.

To support policy analysis and forward planning, various statistical techniques can be applied to available data for evaluating the current situation and forecasting future demand.

From a general user’s perspective, social statistics provide an objective, in-depth view of the respective facets of society, as they can effectively reflect developments and shortfalls in various areas of living conditions and provision of social services.

Ways of obtaining official statistics

In meeting the divergent needs of data users, C&SD disseminates official statistics through different media, emphasising the service principles of friendly access and prompt delivery.

For statistics of common interest, C&SD issues press releases to the mass media. The date schedule for issuing around 115 press releases of regular nature in a given year is made known to the mass media in September of the preceding year. The schedule is also available to the general public for reference on C&SD Website.

C&SD also produces different types of publications and statistical products (including statistical tables and CD-ROM products). Member of the public can download publications and statistical tables free-of-charge from the C&SD Website. Print versions of some publications and CD-ROM products are available for purchase and collection on the spot at the Publications Unit of C&SD (Address : 19/F Wanchai Tower, 12 Harbour Road, Wan Chai, Hong Kong). For those publications with print versions, they are also available for sale online at the Government Bookstore of the Information Services Department.

There may be times when users require statistics more detailed than those contained in the press releases and publications of C&SD. They are welcome to approach us for assistance by telephone, facsimile, letter, electronic mail or personal visit. A list of telephone enquiry numbers and email addresses is given at C&SD Website (www.censtatd.gov.hk/service_desk/enquiries/index.jsp).

To cater for the changing technological environment and service needs, continuous efforts have been made in enhancing the contents and functionality of the C&SD Website with a view to further promoting better understanding and application of statistics in the community. A new mobile version of the C&SD Website has been launched recently to facilitate navigation of the latest Hong Kong statistics on handheld devices.

Proper understanding of the concepts and definitions of different kinds of statistics is imperative for their correct interpretation and application. Besides providing statistical information to the public, C&SD is also prepared to give advice on the interpretation and application of statistics so as to maximise their utility to users.

Further information

The above contents present only part of the economic and social statistics produced by C&SD. For further information regarding the official statistics covering various social and economic aspects of Hong Kong (e.g. latest statistics, statistical reports, concepts and methods), please visit the following sections of the C&SD Website :

Labour
External trade
National Income and Balance of Payments
Prices
Business performance
The Four Key Industries and the Six Industries
Energy
Housing and property
Public accounts, money and finance
Science and technology
Transport, communications and tourism
Education
Health
Social welfare
Law and order
Culture, entertainment and recreation
Environment, climate and geography

Exercise

The following is a quiz to test your understanding of the concepts and key points covered in the previous chapters. Please give a ✓ in the appropriate box (☐) for the correct answer, or fill in the blanks (__) for the correct rankings.

1. The population of Hong Kong in mid-2011 was
   - ☐ 5.7 million
   - ☐ 6.5 million
   - ☐ 7.1 million
   - ☐ 7.3 million

2.

Year	Crude birth rate in Hong Kong (Number of births per 1 000 population)
2005	8.2
2011	*x*

   x =
   - ☐ 6.1
   - ☐ 8.2
   - ☐ 9.3
   - ☐ 13.5

3. If the crude death rate in a given year was 5.0 deaths per 1 000 population and the crude birth rate was 7.2 births per 1 000 population, the rate of natural increase per 1 000 population in the period was
   - ☐ -2.4
   - ☐ 2.2
   - ☐ -12.0
   - ☐ 12.0

4. The population in mid-2011 = a; and
   the population in mid-2006 = b.
   The compound average growth rate per annum of the population in this period was

   ☐ $\left(\sqrt[4]{\frac{b}{a}}-1\right)\times 100\%$

   ☐ $\frac{\left(\frac{b}{a}-1\right)}{4}\times 100\%$

   ☐ $\left(\sqrt[5]{\frac{b}{a}}-1\right)\times 100\%$

   ☐ $\frac{\left(\frac{b}{a}-1\right)}{5}\times 100\%$

5. In 2011, the size of the labour force in Hong Kong was
   ☐ 3.0 million
   ☐ 3.5 million
   ☐ 3.7 million
   ☐ 3.9 million

6. Which category of persons is regarded as economically inactive?
   ☐ Unpaid family workers
   ☐ Employers
   ☐ The unemployed
   ☐ Full-time home-makers

7. Which of the following is not an approach in the compilation of the Gross Domestic Product?
   ☐ Expenditure approach
   ☐ Production approach
   ☐ Industrial approach

8. Most of the visitors visiting Hong Kong in 2011 came from
   - ☐ U.S.A.
   - ☐ Japan
   - ☐ The mainland of China
   - ☐ Canada

9. The arithmetic mean of [10, 20, 30, 40, 50] is
   - ☐ 30
   - ☐ 35
   - ☐ 40
   - ☐ 45

10. The mode of [3, 10, 14, 23, 29, 30, 31] is
    - ☐ 20
    - ☐ 23
    - ☐ 31
    - ☐ none

11. When half of the elements of a population are smaller than a value X and half are greater than it, X is the
    - ☐ Mean
    - ☐ Median
    - ☐ Mode

12. The mean deviation of [2, 5, 13, 28] is
    - ☐ 0
    - ☐ 8.5
    - ☐ 12
    - ☐ 34

13. Among mean, median and mode, which one is the most easily affected by extreme values?
    - ☐ Mean
    - ☐ Median
    - ☐ Mode

14. The price of an article A was $100 in 2010 and $50 in 2011. This means that the price of A in the period has decreased by
    - ☐ 50%
    - ☐ 100%
    - ☐ 150%
    - ☐ 200%

15. Which one of the following charts is visually distorted and should not be used?
    - ☐ Chart A
    - ☐ Chart B

Chart A

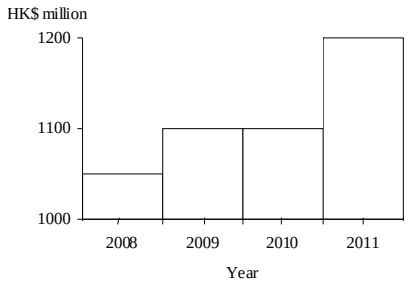
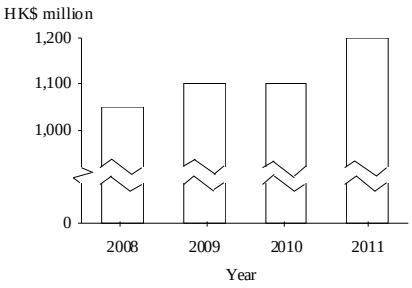


Chart B



Solutions to exercises

Chapter 1 Population size and growth

(1) From the data given below, complete the table.
(Figures given in thousands; rates expressed in "per 1 000 population".)

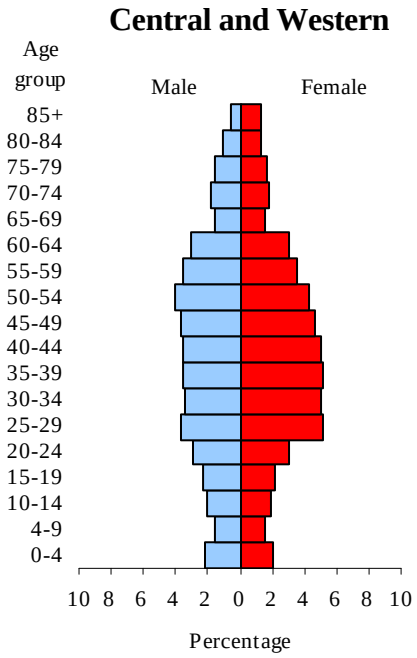
	(1)	(2)	(3)	(4)	(5)	(6)
Country	Population at year T	No. of births in year T	No. of deaths in year T	Crude birth rate	Crude death rate	Rate of natural increase
Country A	29 863	1 322	525	***44***	***18***	***26***
Country B	15 941	723	332	***45***	***21***	***24***
Country C	20 155	249	132	***12***	***7***	***5***
Country D	32 268	332	226	***10***	***7***	***3***
Country E	9 749	433	181	***44***	***19***	***25***
Country F	1 315 844	17 558	8 795	***13***	***7***	***6***
Country G	74 033	1 860	421	***25***	***6***	***19***
Country H	1 517	51	17	***34***	***11***	***23***
Country I	82 689	702	853	***8***	***10***	***-2***
Country J	10 098	96	132	***10***	***13***	***-3***

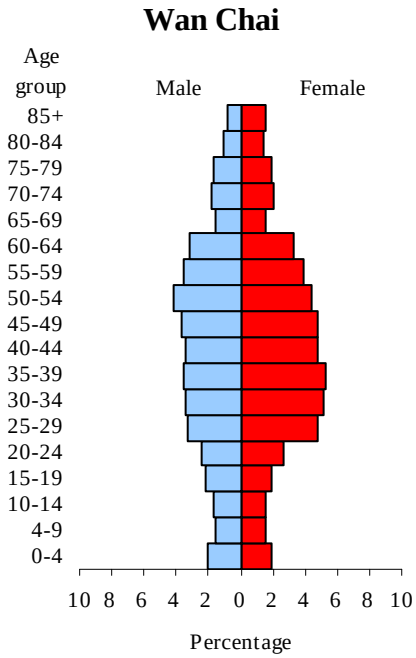
(2) From the rates obtained in part (1), classify the countries in various stages of the demographic transition model.

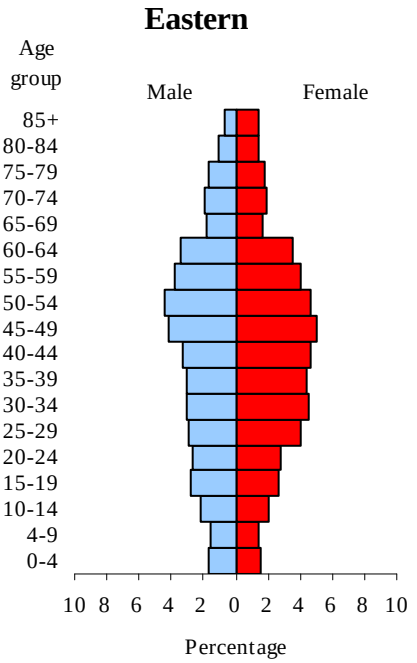
Stage (1) High birth rate High death rate	Stage (2) or (3) High birth rate or declining birth rate Declining death rate	Stage (4) Low birth rate Low death rate
Country A	***Country G***	***Country C***
Country B	***Country H***	***Country D***
Country E		***Country F***
		Country I
		Country J

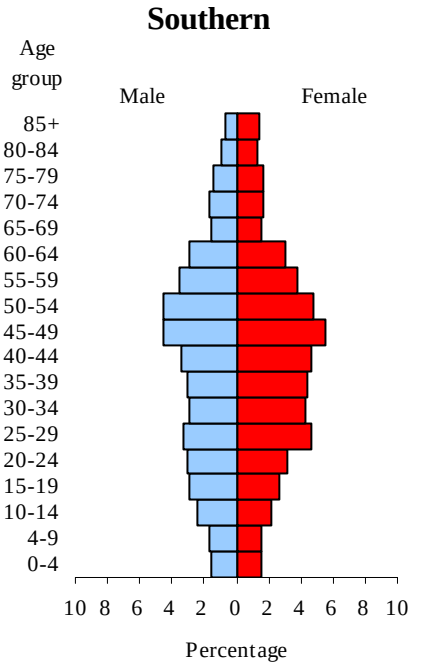
Chapter 2 Population structure

(1) Population pyramids (in percentage) of the 18 districts in Hong Kong, June 2011



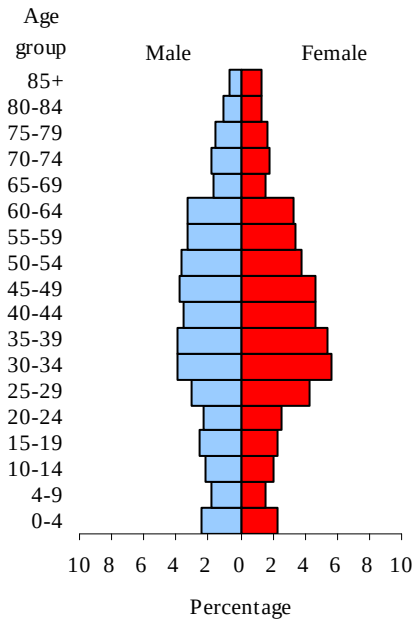




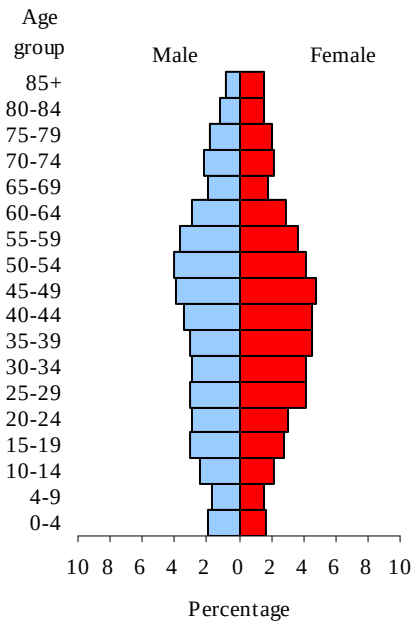


(1) Population pyramids (in percentage) of the 18 districts in Hong Kong, June 2011 (cont'd)

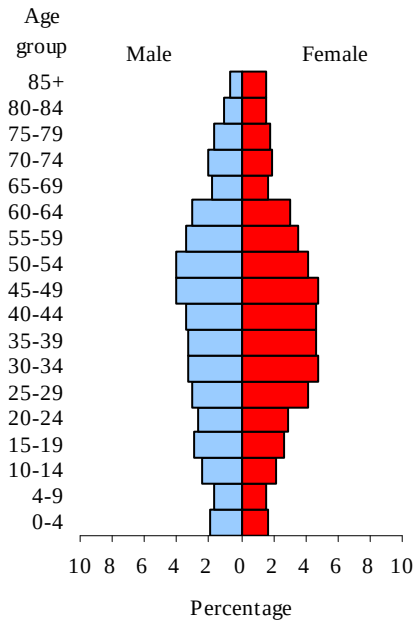
Yau Tsim Mong



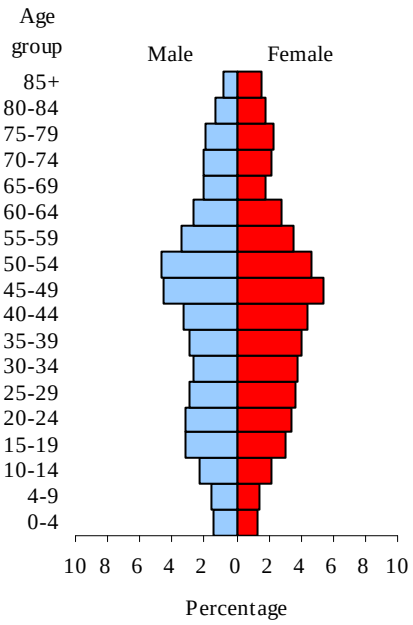
Sham Shui Po



Kowloon City

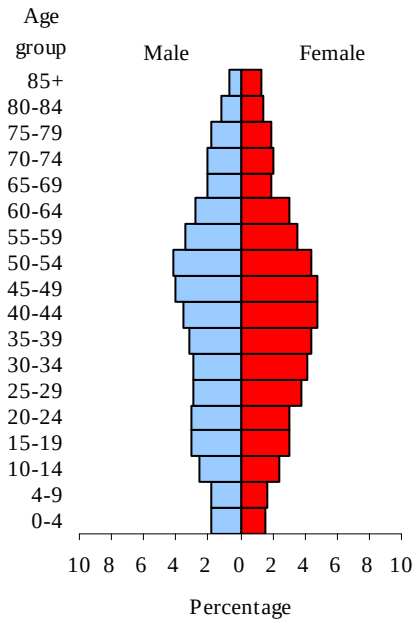


Wong Tai Sin

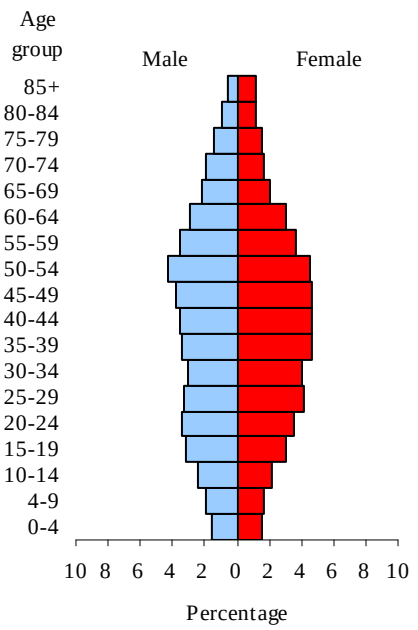


(1) Population pyramids (in percentage) of the 18 districts in Hong Kong, June 2011 (cont'd)

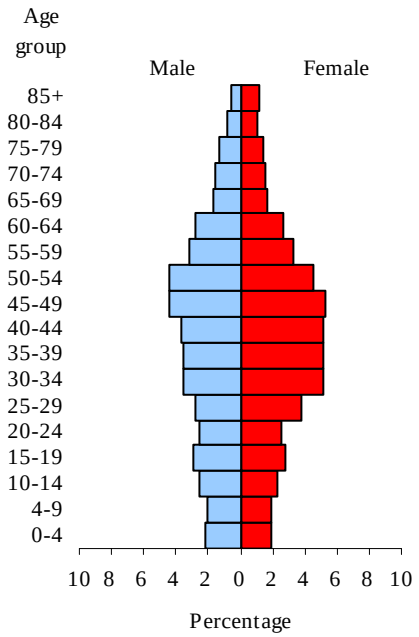
Kwun Tong



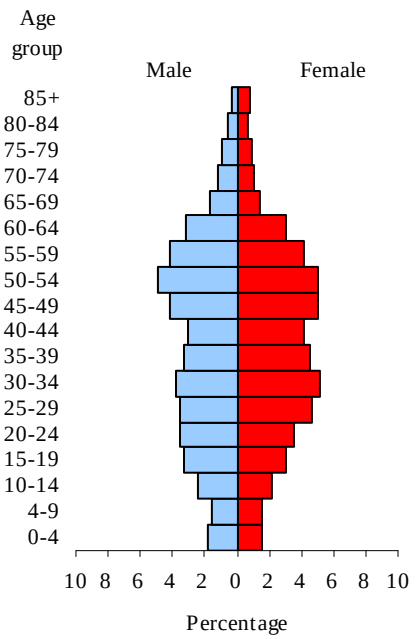
Kwai Tsing



Tsuen Wan

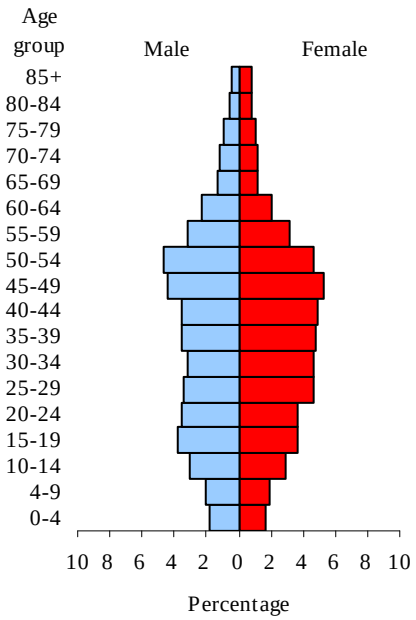


Tuen Mun

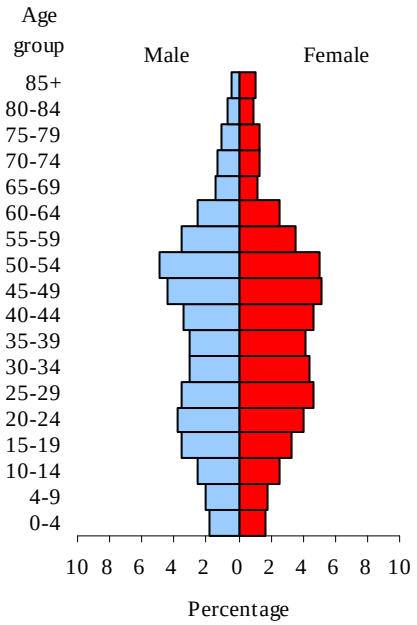


(1) Population pyramids (in percentages) of the 18 districts in Hong Kong, June 2011 (cont'd)

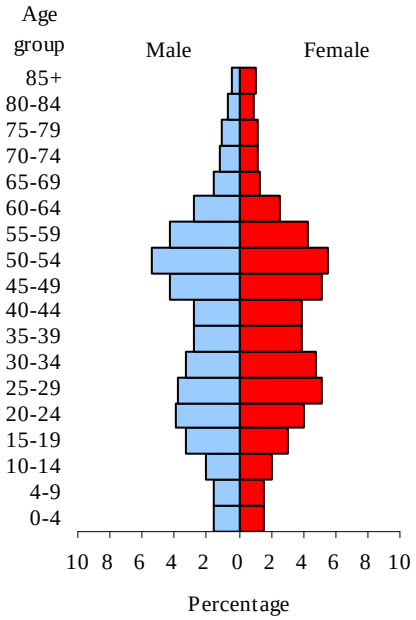
Yuen Long



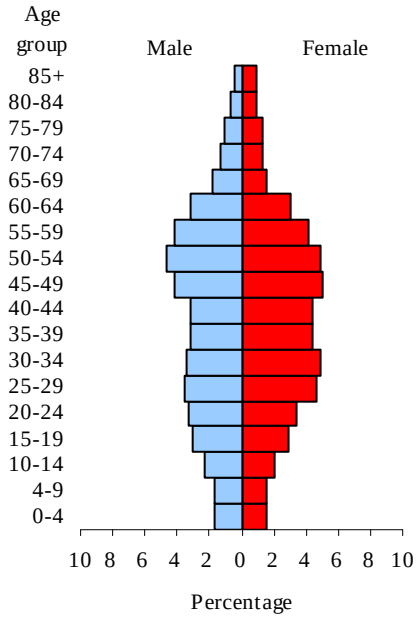
North



Tai Po

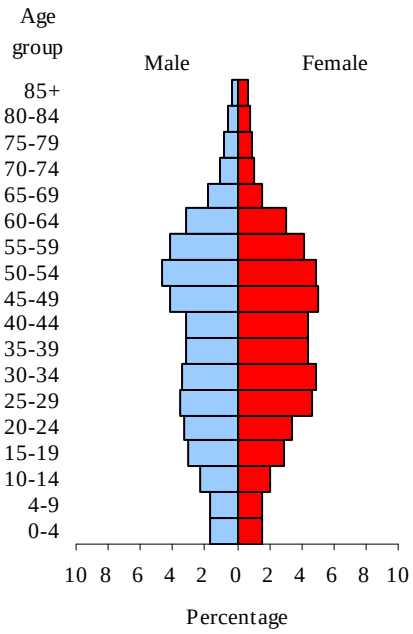


Shatin

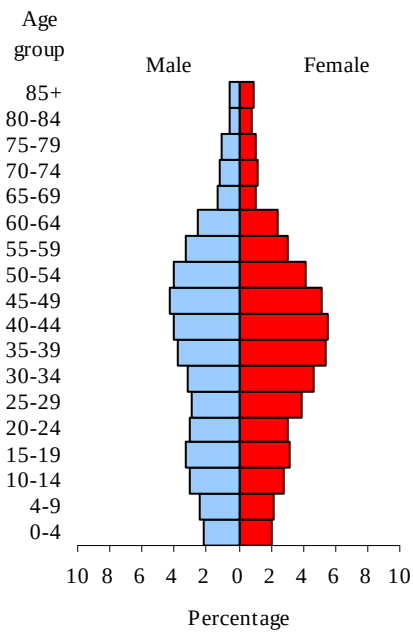


(1) Population pyramids (in percentage) of the 18 districts in Hong Kong, June 2011 (cont'd)

Sai Kung



Islands



(2) Rank the aging extent of these districts by comparing their median ages.

District Council District	Median age	Ranking for aging extent (Note) ("1" for the most aged district)
Central and Western	41.3	11.5
Wan Chai	43.0	4
Eastern	44.1	2
Southern	42.9	5
Yau Tsim Mong	41.1	13
Sham Shui Po	43.2	3
Kowloon City	42.5	7
Wong Tai Sin	44.5	1
Kwun Tong	42.8	6
Kwai Tsing	41.9	8
Tsuen Wan	41.3	11.5
Tuen Mun	40.4	14
Yuen Long	38.6	18
North	40.2	15
Tai Po	41.4	10
Shatin	41.5	9
Sai Kung	39.3	16
Islands	39.1	17

Note : Where there are ties in rank, the tied observations are assigned the mean of the ranks which they jointly occupy.

Chapter 5 Uses and misuses of statistics

The man actually received $80,000 in cash from Bank Y and he had to repay $100,000 on the 16th day; therefore the interest rate should be :

$$\frac{\$20,000}{\$80,000} \times 100\% = 25\%$$

or 25% per 15 days.

Chapter 6 Rate, ratio, proportion and percentage

(i) (841 927 / 7 071 600) is just the proportion of the number of students passing the examination to the total population, irrespective of whether they have taken the examination. The accurate measure of the overall passing rate is (841 927 / 1 403 211), i.e. only those students who have taken the examination should be counted in the base used for comparison.

(ii) The passing rates for the four districts on Hong Kong Island should be：

Central and Western district	=	(44 014 / 60 293) × 100%	=	73%
Wan Chai district	=	(28 366 / 41 110) × 100%	=	69%
Eastern district	=	(93 943 / 144 528) × 100%	=	65%
Southern district	=	(44 831 / 63 142) × 100%	=	71%

The overall passing rate for Hong Kong Island should be：

$$\frac{(44\ 014 + 28\ 366 + 93\ 943 + 44\ 831)}{(60\ 293 + 41\ 110 + 144\ 528 + 63\ 142)} \times 100\% = \underline{\underline{68\%}}$$

(iii) As the increase in the number of students passing the examination might be due to the fact that there were more students in the Wan Chai district sitting for the examination in 2011, the comparison for the two years should therefore be based on the passing rate for the district in the two respective years.

As we only have limited information (data for two years only, viz. 2006 and 2011), it is difficult for us to draw the conclusion that the academic results of students in the Wan Chai district improved during the period from 2006 and 2011. Maybe, the results in years 2006 and 2011 are outliers in the data series.

Chapter 7 Measures of central tendency

Consider a set of figures：

{1, 9, 3, 7, 8, 12, 9}

(i) What is the sum of these seven values? **[49]**

(ii) What is the mean? **[7]**

Note：
- **Mean = Sum of the figures ÷ Number of data points**
- **Sum of the figures = Mean × Number of data points**

(iii) What is the median? **[8]**

(iv) What is the mode? **[9]**

(v) Is the median equal to the mean of the same seven values? **[No]**

(vi) Are the mean and median the same when the set of figures is {1, 3, 7, 11, 13}?
[Yes; Mean = Median = 7]

(vii) What does this suggest about the conditions when the mean and median are the same?
[When the set of data is symmetrical, the mean and median will be the same. However, having the same mean and median, the set of data is not necessarily symmetrical. For example, in the set of figures {1, 7, 9, 13, 15}, although both the mean and median are equal (= 9), the set of data is not symmetrical.]

Chapter 8 Measures of dispersion

	Examination scores			Standard examination scores			
Student	English	Chinese	Mathematics	English	Chinese	Mathematics	Total
(1)	40	62	80	***-1.8***	***-1.0***	***-0.3***	***-3.1***
(2)	57	60	60	***-0.1***	***-1.4***	***-2.2***	***-3.7***
(3)	60	70	95	***0.2***	***0.7***	***1.1***	***2.0***
(4)	48	64	92	***-1.0***	***-0.5***	***0.8***	***-0.7***
(5)	53	66	75	***-0.5***	***-0.1***	***-0.8***	***-1.4***
(6)	66	72	78	***0.7***	***1.2***	***-0.5***	***1.4***
(7)	72	70	86	***1.3***	***0.7***	***0.2***	***2.3***
(8)	49	67	87	***-0.9***	***0.1***	***0.3***	***-0.5***
(9)	69	60	99	***1.0***	***-1.4***	***1.5***	***1.1***
(10)	70	74	83	***1.1***	***1.6***	***0.0***	***2.7***
Mean	***58.4***	***66.5***	***83.5***				
Standard deviation	***10.3***	***4.7***	***10.6***				

The student in the class who has attained the best result is **(10)**.
(Reason: Student (10) got the highest total standard score.)

Chapter 9 Official statistics

The following is a quiz to test your understanding of the concepts and key points covered in the previous chapters. Please give a ✓ in the appropriate box (☐) for the correct answer, or fill in the blanks (__) for the correct rankings.

1. The population of Hong Kong in mid-2011 was
   - ☐ 5.7 million
   - ☐ 6.5 million
   - ☑ 7.1 million
   - ☐ 7.3 million

2. Crude birth rate in Hong Kong

Year	(Number of births per 1 000 population)
2005	8.2
2011	*x*

x =

   - ☐ 6.1
   - ☐ 8.2
   - ☐ 9.3
   - ☑ 13.5

3. If the crude death rate in a given year was 5.0 deaths per 1 000 population and the crude birth rate was 7.2 births per 1 000 population, the rate of natural increase per 1 000 population in the period was
   - ☐ -2.4
   - ☑ 2.2
   - ☐ -12.0
   - ☐ 12.0

4. The population in mid-2006 = a; and
   the population in mid-2011 = b.
   The compound average growth rate per annum of the population in this period was

   ☐ $\left(\sqrt[4]{\frac{b}{a}}-1\right)\times 100\%$

   ☐ $\frac{\left(\frac{b}{a}-1\right)}{4}\times 100\%$

   ☑ $\left(\sqrt[5]{\frac{b}{a}}-1\right)\times 100\%$

   ☐ $\frac{\left(\frac{b}{a}-1\right)}{5}\times 100\%$

5. In 2011, the size of the labour force in Hong Kong was
   ☐ 3.0 million
   ☐ 3.5 million
   ☑ 3.7 million
   ☐ 3.9 million

6. Which category of persons is regarded as economically inactive?
   ☐ Unpaid family workers
   ☐ Employers
   ☐ The unemployed
   ☑ Full-time home-makers

7. Which of the following is not an approach in the compilation of the Gross Domestic Product?
   ☐ Expenditure approach
   ☐ Production approach
   ☑ Industrial approach

8. Most of the visitors visiting Hong Kong in 2011 came from
   - ☐ U.S.A.
   - ☐ Japan
   - ☑ The mainland of China
   - ☐ Canada

9. The arithmetic mean of [10, 20, 30, 40, 50] is
   - ☐ 30
   - ☑ 35
   - ☐ 40
   - ☐ 45

10. The mode of [3, 10, 14, 23, 29, 30, 31] is
    - ☐ 20
    - ☐ 23
    - ☐ 31
    - ☑ none

11. When half of the elements of a population are smaller than a value X and half are greater than it, X is the
    - ☐ Mean
    - ☑ Median
    - ☐ Mode

12. The mean deviation of [2, 5, 13, 28] is
    - ☐ 0
    - ☑ 8.5
    - ☐ 12
    - ☐ 34

13. Among mean, median and mode, which one is the most easily affected by extreme values?
    - ☑ Mean
    - ☐ Median
    - ☐ Mode

14. The price of an article A was \$100 in 2010 and \$50 in 2011. This means that the price of A in the period has decreased by

- ☑ 50%
- ☐ 100%
- ☐ 150%
- ☐ 200%

15. Which one of the following charts is visually distorted and should not be used?

- ☑ Chart A
- ☐ Chart B

Chart A

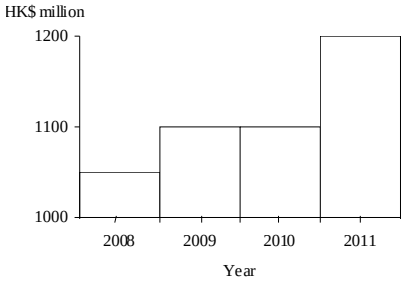


Chart B

