Statistics Canada Statistique Canada

Methodology Branch

Household Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes des ménages

Canada

WORKING PAPER

METHODOLOGY BRANCH

ESTIMATION OF ENROLMENT BY SINGLE YEAR OF AGE FROM TOTAL ENROLMENT

HSMD-98-005E

Jack Singleton

Education, Culture & Tourism Methods Section

Household Survey Methods Division

Statistics Canada

June 1998

ESTIMATION OF ENROLMENT BY SINGLE YEAR OF AGE FROM TOTAL ENROLMENT

Jack Singleton[1]

ABSTRACT

Annual counts of numbers of individuals enrolled in various levels of education (e.g. elementary, secondary, university, college, trade/vocational) are available for most countries. For some countries these counts are available for each age (i.e. by single year of age or by age) whereas for others counts are produced for all ages combined. This document gives several methods for estimation of enrolment by single year of age from total enrolment. The choice among methods depends on the historical enrolment data available. Four methods are presented for estimation of elementary/secondary enrolment by age, and three methods for estimation of post-secondary (university, college and trade/vocational) enrolment by age. For all methods it is assumed annual age-specific population estimates are available, as are total enrolment counts for each level of education. The estimates produced are reconciled with known total enrolments to produce estimates consistent with the known totals. The methods are designed for decomposing a known total enrolment into enrolments by age, rather than for prediction of enrolment by age. The main body of the document describes each method in non-technical terms. Specific formulae for each method are given in the appendices.

[1]Jack Singleton, Education, Culture and Tourism Methods Section, Household Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6

ESTIMATION DE L'EFFECTIF PAR ANNÉE D'ÂGE À PARTIR DE L'EFFECTIF TOTAL

Jack Singleton[2]

RÉSUMÉ

Il existe des données annuelles sur le nombre de personnes inscrites aux divers ordres d'enseignement (p. ex. élémentaire, secondaire, universitaire, collégial, et professionnel et technique) dans la plupart des pays. Dans certains pays, ces données sont présentées pour chaque âge (c.-à-d. par année d'âge ou par âge), alors que dans d'autres, les chiffres sont produits pour l'ensemble des âges. Ce document porte sur plusieurs méthodes qui servent à estimer l'effectif par année d'âge à partir de l'effectif total. La méthode à utiliser dépend des données chronologiques qui existent sur l'effectif. On présente quatre méthodes pour estimer l'effectif par âge aux niveaux élémentaire et secondaire, et trois méthodes pour estimer l'effectif par âge au niveau postsecondaire (universitaire, collégial, et professionnel et technique). Dans tous les cas, on suppose qu'il existe des estimations annuelles de la population pour chaque âge, ainsi que des données sur l'effectif total de chaque ordre d'enseignement. On fait le rapprochement des estimations produites et des données connues sur les effectifs totaux pour produire des estimations compatibles avec les totaux connus. Les méthodes ont été conçues pour décomposer un effectif total connu en plusieurs effectifs par âge, plutôt que pour faire des prévisions de l'effectif par âge. Chacune des méthodes est décrite dans le corps du document, dans un langage non spécialisé. Les formules propres à chaque méthode sont données en annexe.

[2]Jack Singleton, Section des méthodes de l'éducation, de la culture et du tourisme, Division des méthodes d'enquêtes des ménages, Statistique Canada, Ottawa (Ontario), K1A 0T6

1 Introduction

1.1 Purpose

Often it is desired to have estimates of enrolment separately for each age (i.e. "by single year of age" or "by age") although only total enrolment (all ages combined) is available. The purpose of this document is to give several methods for estimation of enrolment by single year of age from total enrolment. The choice of method depends on the set of historical and current data available. These methods produce estimates of enrolment by single year of age, as a function of population estimates by single year of age, historical enrolment counts (possibly by single year of age or by grade/year of study), total enrolment and other inputs. These estimates are reconciled to produce estimates consistent with known total enrolment counts.

It should be pointed out that these methods are not intended to produce enrolment projections, but to "break down" or decompose a known total enrolment into enrolments by single year of age. Although the methods are designed for estimation of enrolment by age for the current year or a previous year for which total enrolment is known, they may be used to break down an estimate of future total enrolment into enrolment by age. In this document the methods are presented in the context of estimation of enrolment by single year of age for the current year.

Estimates of enrolment by age can be calculated separately for each level of education for which total enrolment is available. Several methods are provided for each of elementary and university. It is assumed secondary enrolments can be calculated similarly to elementary, and other forms of post-secondary (e.g. college, trade/vocational) similarly to university.

1.2 Inputs

It is assumed that estimates of the population by single year of age are available for all years, past, current and future. The estimate may take the form of i) a census population count, possibly adjusted; ii) an inter-censal population estimate; or iii) a population projection. For age a in year t, the population estimate will be denoted P_a^t throughout this document.

It is also assumed that total enrolment counts for elementary and university are available for all years, past and current. Unlike population, it is assumed that there are no projections of future enrolments available. Total elementary and university enrolments in year t will be denoted $ELEM^t$ and $UNIV^t$, respectively, in this document.

One possible source of error is that the date on which the population estimate and enrolment count are based may differ, as well as the target population of each. Also the date relative to which age is calculated may differ between enrolment counts and population estimates. This document will not go into detail regarding adjustment for these biases, although they must be taken into account.

2 Elementary and secondary enrolments

It is assumed that there is a national age of entry to school (assumed to be 5 years of age in this note) and that children enter school at kindergarten (grade 0) and progress through to grade 12. If the age of entry, grade of entry or highest grade differ from the above, the formulae can still be used, with the proper substitutions. This formulation assumes all children enter school at the national age of entry to school; if this is not the case, this methodology is not appropriate for national enrolments.

It is assumed students progress from grade to grade (i.e. successful completion of one school year or grade leads to entry into the next), and that each of elementary and secondary is defined nationally in terms of certain grades (e. g. elementary as grades 0 to 8, secondary as grades 9 to 12). It is assumed there are three outcomes to a year of schooling:

- successful completion and enrolment in the next grade the following year ("pass");
- unsuccessful completion and re-enrolment in the same grade the following year ("fail");
- departure from the school system ("drop out").

For simplicity in calculation, generally it is assumed an individual can repeat a grade ("fail") at most once in a lifetime, and that all departures occur at the end of the school year. It is also assumed the number of re-entrants to the school system is negligible; if this is not the case, the definitions of the repeater and dropout rate must be extended to include re-entrants.

In most nations there is a range of ages of compulsory school enrolment, and it is assumed that enrolment is close to 100 per cent for those ages. At some points in the analysis it is necessary to assume there are no dropouts, although in other instances the calculations will allow specification of a dropout rate.

Consistent with the National Centre for Education Statistics (NCES) publication "Projections of Education Statistics to 2007", we include only ages 5-18 for elementary, and only ages 12-34 for secondary. Enrolments are negligible for remaining ages (p.123).

As mentioned earlier, the methods will be presented for elementary enrolment. Four scenarios will be considered, in order from least to most information available. In Section 2.1 it is assumed no historical enrolment data is available, although in Section 2.1.0 it is assumed estimates of either the number or proportion of individuals not in school is available, and in Section 2.1.1 that grade-specific estimates of the dropout and repeater rates are available. In Section 2.2 it is assumed historical enrolments by grade are available, and in Section 2.3 that historical enrolments by age are available for certain years. Section 2.4 discusses secondary enrolments.

2.1 No historical enrolment data available

Let us assume only population estimates by age are available. As most children are enrolled in school, the population estimate for a certain age can be taken as an estimate of elementary enrolment of that age, provided elementary and secondary can be defined in terms of nonoverlapping and contiguous age ranges (else there will be children of a certain age assigned to both levels or neither level). However this approach may lead to systematic overestimation of enrolment if there are significant numbers of children home-schooled or otherwise not in school. In addition this approach may lead to enrolment estimates which conform to population estimation concepts. For example, children of diplomats will not be counted in the enrolment estimate if they are not included in the population estimate, or population estimates may be defined in terms of age on July 1st, whereas enrolment may be based on age on January 1st.

2.1.0 Number or proportion of individuals not in school available

If numbers of children known not to be in school are available from alternate sources, such as a count of children home-schooled, enrolment can be estimated as the population estimate less the number of children known to be not in school.

One could estimate enrolment as a specified fraction of the population. The fraction can vary among ages and over time, and can be computed from auxiliary information such as:

- administrative records;
- census data;
- labour force data;
- sample-survey studies;
- enrolment and population data of other regions or countries in the world considered "similar".

For example, suppose a recent survey of parents indicated 2 per cent of children aged 5 to 12 were not enrolled in school. Then elementary enrolment could be estimated as 98 per cent of the population estimate for ages 5 to 12. It is possible that the fraction enrolled could change over time, or could vary among ages.

2.1.1 Estimates of repeater rate and dropout rate available

Let us assume only population estimates by age are available but the method of Section 2.1.0 is not deemed appropriate. Without further assumptions it is difficult to proceed in estimation of enrolment by age. To aid in modeling progression, it is assumed students progress from grade to grade, as often enrolment is discussed in terms of grades rather than ages. If it is possible to specify, for each grade, the proportion of students who drop out of school while completing this grade and the proportion who repeat this grade, then enrolment by age can be estimated from population estimates by age using the methodology described below. Note that due to the assumption of three outcomes to any year of schooling, the proportion who pass a grade is given by the proportion who neither fail nor repeat. For ease in calculation we assume an individual can repeat a grade at most once and do not allow for individuals "skipping" one or more grades nor for re-entrants into the school system. If these assumptions are not reasonable the method can be modified accordingly.

Assuming children begin school in kindergarten (grade 0 in this document) at age 5, the assumption of at most one repeated grade implies that kindergarten students will be age 5 or 6, grade 1 age 6 or 7, grade 2 age 7 or 8, etc.; i.e. grade g students will be age g+5 or g+6. Also children age 5 will be enrolled only in kindergarten, age 6 in kindergarten or grade 1, age 7 in grade 1 or grade 2, etc.; i.e. children age a are enrolled in either grade a-6 or grade a-5. Assuming (for the moment) that there is no immigration into the school system, children in grade 1 either entered school the previous year and passed kindergarten or entered two years prior and either repeated kindergarten or are repeating grade 1. Similarly children in grade 2 either entered two years previously and passed kindergarten and grade 1, or entered three years ago and repeated either kindergarten or grade 1, or are repeating grade 2. The same reasoning holds for higher grades; individuals in grade g either entered school g years ago and passed every grade or entered (g+1) years ago and either repeated a grade or are repeating grade g. Thus under the assumption of no immigration into the school system the number of students in grade g in a certain year can be calculated from the numbers that entered school g and (g+1) years ago.

Let us now consider estimation the components mentioned above. Let us assume that the population estimate of individuals age 5 can be taken as a reasonable estimate of the number of children who entered school that year. The number of children in grade 1 in the current year is the sum of three components, specifically the numbers who:

i) entered school and passed kindergarten last year;
ii) entered school and failed kindergarten two years ago and passed kindergarten last year;
iii) entered school and passed kindergarten two years ago and failed grade 1 last year.

These three terms are estimated as, respectively:

- the population estimate for age 5 last year multiplied by the fraction of kindergarten students who pass.
- the population estimate for age 5 two years ago multiplied by the fraction of kindergarten students who repeat multiplied by the fraction of kindergarten students who do not drop out (note the assumption of one failure per lifetime precludes failing kindergarten a second time).
- the population estimate for age 5 two years ago multiplied by the fraction of kindergarten students who pass multiplied by the fraction of grade 1 students who repeat.

The first term gives an estimate of the number of grade 1 students age 6 computed from the number of children age 5 one year ago, and the second and third terms together give an estimate of the number of grade 1 students age 7 computed from the number of children age 5 two years ago.

A similar formulation can be applied to other grades, although there will be (g+2) terms summed, specifically the number who passed every grade up to and including g-1, the number who repeated each of grades 0, 1, 2, ..., g-1 and the number repeating grade g. The first term gives an estimate of the number of grade g students age g+5 computed from the number of children age 5 g years ago, and the subsequent terms together give an estimate of the number of grade g students age g+6 computed from the number of children age 5 (g+1) years ago. As students age a are enrolled in only grades a-5 and a-6, the estimate of enrolment age a is the sum of the estimate of the number of grade a-5 students age a, and the estimate of the number of grade a-6 students age a.

Formulae for calculation of the above are given in Appendix A. A few points to be noted are:

- the formulae are defined in terms of the repeater and dropout rates[3].
- the fraction passing a grade differs between students who never repeated a grade and those who have repeated a grade; the latter either drop out or pass, whereas the former may pass, repeat or drop out.
- the formulae also provide for net immigration by including population growth factors (see Appendix A for definition). It is assumed the progression behaviour (repeater and drop-out rates) of immigrants into the system is similar to non-immigrants.

Reconciliation of estimated elementary enrolment by age with known total elementary enrolment

As the estimates of enrolment by age of Sections 2.1, 2.1.0 and 2.1.1 are calculated independently of known total enrolment, it may be that the sum of the age-specific estimates differs from the known total enrolment. In order to preclude this inconsistency, the estimates are adjusted, as described in Appendix A, to give estimates reconciled with the known total enrolment. The same reconciliation will be done for estimates of elementary enrolment by age produced by many of the methods below.

2.2 Historical enrolments by grade available

Let us now assume enrolments by grade are available for all years. We can proceed to calculate enrolments by age from enrolments by grade, although we will modify the assumptions of Section 2.1.1 to ease calculations. In particular:

- we will no longer assume an individual can repeat a grade at most once; thus the repeater rate is applied to all children, rather than just those who have never repeated a grade previously.
- we will assume the population growth factor for individuals age g+5 is applied to individuals enrolled in grade g, although the latter includes children older than age g+5.

For any given year, the enrolment in a grade can be decomposed into those who successfully completed the previous grade the year before and those who are repeating the grade. Thus the enrolment in grade g is the sum of:

- the repeater rate for grade g multiplied by the enrolment in grade g the year prior and;
- the successful completion rate for grade (g-1) multiplied by the enrolment in grade (g-1) the year prior.

---

[3]Once two of the repeater, dropout and successful completion rates are specified, the third is determined as the three sum to one by the assumption of three outcomes to a year of schooling.

To account for immigration into the school system, the first component is multiplied by the population growth factor for age (g+5) for the year prior, and the second is multiplied by the population growth factor for age (g+4) the year prior.

The above relation involves both known quantities (enrolments by grade and population growth factors) and unknown quantities (repeater rate for grade g and successful completion rate for grade g-1). Since the equation corresponding to the above relation contains two unknowns it cannot be solved unless one of the two unknowns is specified. However we note the successful completion rate can be specified in terms of the repeater and dropout rates, which implies that the repeater rate for grade g appears in two equations, namely the specification of the enrolment in grade g and in grade g+1. Consequently if we specify values for the dropout rate and repeater rate for grade 0, the repeater rate for grade 1 can be obtained by solving the above equation when g=1. Using this estimate of the repeater rate for grade 1 together with a specified value for the dropout rate for grade 1 gives the repeater rate for grade 2 (solve the above when g=2). It follows that if the dropout rate for all grades and the repeater rate for kindergarten (grade 0) are specified, then repeater rates for all grades can be obtained (see Appendix A).

Using these repeater and dropout rates and the relation given above, the enrolment in grade g can be decomposed into estimates of the number repeating the grade and the number who successfully completed the previous grade the year before (see Appendix A for details). Assuming the former are of age (g+6) and the latter of age (g+5), the above gives estimates of enrolment by age and grade, from which estimates of enrolment by age follow as individuals age a are assumed enrolled only in grades (a-5) and (a-6). As in Sections 2.1, 2.1.0 and 2.1.1, the estimated enrolments by age are reconciled with known total enrolment to give age-specific enrolment estimates consistent with the known total. See Appendix A for more details.

2.3 Historical enrolments by age available

Let us suppose that historical elementary enrolments by age are available for one or more years, and that we are estimating elementary enrolment by age for the current year. The approach we will take is to use the historical data to estimate the current elementary enrolment rate (i.e. the ratio of enrolment to population) for each age, and apply the estimated enrolment rate to the population estimate. That is, elementary enrolment age a is estimated as the product of the estimated elementary enrolment rate for age a and the population estimate for age a for the current year. Once again the estimated elementary enrolments by age are reconciled with the known total elementary enrolment.

This approach assumes the historical enrolment rates can be used to predict the current enrolment rate. If this assumption is not valid, the approach is not appropriate and may give misleading estimates.

If there is only one year of historical enrolments by age available, the enrolment rate for that year can be used as an estimate of the current enrolment rate. If more than one year of historical enrolments by age are available, then the current enrolment rate can be estimated as a function, such as a weighted average, of the historical enrolment rates; see Appendix A for some examples.

Let us now suppose that for one or more years the percentage breakdown by age of elementary enrolment is known, although the enrolment counts by age are not known. That is, for each age the percentage of total elementary enrolment is known, although the number of individuals enrolled is not. From the historical breakdowns by age of elementary enrolment an estimate of the current breakdown of elementary enrolment by age can be obtained, similarly to the estimation of current enrolment rate from historical enrolment rates. The estimate of the current breakdown by age is applied to known current total elementary enrolment to produce estimates of elementary enrolment by age. That is, for each age the estimate of the proportion of current elementary enrolment that age is multiplied by current total elementary enrolment to give an estimate of current elementary enrolment that age. See Appendix A for details. This approach is appropriate if the historical breakdown by age of elementary enrolment is reasonable for the current year. If not, the estimated enrolments by age may be misleading.

2.4 Secondary enrolments

For secondary grades the estimates of enrolment by age will be reconciled with total secondary enrolment rather than total elementary enrolment (consistent with "Projections of Education Statistics to 2007", secondary enrolments include only ages 12-34). See Appendix A for details.

Assuming elementary and secondary are each defined in terms of grades or years of schooling and that the first year or grade of secondary school follows the last year or grade of elementary school, the methods given in Sections 2.1.1 and 2.2 for estimation of enrolments by age and grade can equally well be applied to secondary grades as to elementary grades. Some of the assumptions made for elementary enrolments may not be appropriate for secondary enrolments, such as only one repeated grade per lifetime and the absence of re-entrants to the school system. In addition for the "teenage" years assumption of a zero dropout rate may not be tenable. It is important to note that these methods assume individuals enrolled in grade g are either age g+5 or g+6. Consequently if the highest grade is 12, no enrolments will be calculated for ages beyond 18. Therefore it is advisable to use a different method of estimation for secondary enrolments for ages 19-34, such as estimation of an enrolment rate (Section 2.3) or estimation based on auxiliary information such as census or labour force data.

The methods of Section 2.3 for estimation of the current elementary enrolment rate from historical elementary enrolments by age can be used analogously for secondary enrolments. However for higher ages (especially those beyond age 17 or 18) it may be desirable to include auxiliary variables in the estimation of the secondary enrolment rate, similarly as is done in Section 3.3 for the university participation rate. Particular attention must be paid to estimation of the secondary enrolment rate for ages beyond 17 or 18, as for these ages only a minority of the population is enrolled in secondary school. As mentioned previously, for these ages estimation of the enrolment rate or of the number enrolled based on labour force or other sources of activity information should be considered whenever possible and may be preferable to estimation based only on population estimates and historical enrolment counts.

3 Post-secondary enrolments

The underlying model assumed for post-secondary enrolments differs from that for elementary and secondary enrolments. For the latter, as described in Section 2, it is assumed individuals enter school at a uniform age, and progress from grade to grade until either graduation or abandonment of studies ("dropping out"). It is assumed elementary/secondary enrolment is highly correlated with population for individuals below a certain age (usually age 18 in North America), as few individuals in this age range are able to pursue options other than education, such as employment. The high correlation between enrolment and population may not exist for secondary enrolment for ages beyond 17, for which different estimation methods may be appropriate (see Section 2.4). Although this does not factor into estimation of enrolment by age, it is assumed that elementary and secondary education is provided by the state, funded by levies such as property and other taxes, rather than by fees directly charged to enrolled students, although there may exist privately-funded institutions providing elementary and/or secondary education. This characterization of elementary and secondary enrolments is quite consistent with the system in place in the Western world.

Let us now consider the assumptions underlying estimation of post-secondary enrolment. Once elementary and secondary studies are complete, it is assumed there are several different paths an individual may follow. While some individuals enter or attempt to enter the work force without further education, others may decide to pursue post-secondary education either on a full-time or part-time basis, possibly combined with full- or part-time employment. In addition the individual may alternate periods of education and employment. In summary, unlike elementary and secondary education, there are many paths through post-secondary education. Furthermore post-secondary education is available in many forms, such as theoretical university degree studies, applied college diploma programs and job-specific trade/vocational certification programs. Post-secondary programs are available from both state- and privately-funded institutions, although in both cases it is assumed the individual pays a "tuition" fee for enrolment. The decision to pursue post-secondary education entails both a direct cost and indirect or "opportunity" cost of the income forgone while pursuing studies. Consequently the decision to pursue post-secondary education is a function of factors which have little influence on elementary and secondary enrolment, such as the availability of funding for post-secondary education (scholarships, bursaries, student loans), the direct cost of attending (tuition and other fees, possibly cost of living "away from home"), the availability and desirability of other alternatives (measured through the unemployment rate, average income and other indicators of economic prosperity) and "value" placed on post-secondary education by potential employers. Moreover these factors also contribute to the decision of which form of post-secondary education should be pursued. The fraction of the population enrolled in a form of post-secondary education is characterized by a participation rate, rather than an enrolment rate.

Consistent with "Projections of Education Statistics to 2007", we include only ages 16 and over for calculation of post-secondary enrolments, as enrolments are negligible for remaining ages (p.123). Enrolments for ages 25-29, 30-34 and 35 and over are grouped together, and the cohort of students aged 35 and over will be related to the population aged 35-44.

Progression through post-secondary programs is defined in terms of years of study, rather than grades. There are two interpretations of a year of study:

i) the year of study is the effective number of years since the individual began the program;

ii) the program of study is defined in terms of a sequence of years of study, and completion of the program occurs when all years of study are completed.

Thus in i) there is a one-to-one correspondence between calendar or academic years and years of study, whereas in ii) an individual may complete a year of study over one or more calendar or academic years.

Several methods of estimating enrolment by age will be presented in the context of university enrolment, although the same methods may be applied for other types of post-secondary education such as college and trade/vocational. Some methods may be applied to individual programs of study within university education. In Sections 3.1 and 3.2 it is assumed age-specific counts of the numbers of new entrants are available. Section 3.1 further assumes age-specific counts of the number of graduates are available, and Section 3.2 that counts of enrolment by year of study are available for two consecutive years. Section 3.3 assumes historical enrolments by age are available for certain years, and Section 3.4 discusses college and trade/vocational enrolments.

3.1 Numbers of new entrants and graduates available

Let us assume that for each year and for each program of study the number of graduates and the number of new entrants are each available by age.

First let us impose some definitions:

- <u>enrolment in year t</u> is defined as enrolment at the beginning of academic year t/t+1 (e.g. 1997 corresponds to academic year 1997/98).
- <u>new entrant in year t</u> is defined as an individual who begins the program of study in calendar year t.
- <u>graduate in year t</u> is defined as an individual who graduates from the program of study in calendar year t.
- <u>dropout in year of study i</u> is defined as an individual enrolled in year of study i but not in year of study i+1; in other words, the individual abandons his/her studies either during or after the ith year of study. If i corresponds to the final year of study, graduation plays the role of enrolment in year of study i+1.

It is necessary to assume the following:

- all new entrants are in the first year of the program and no students enter beyond first year.
- students who drop out of or leave the program do not return.
- all students study full-time.
- the program is D years in duration and all students who complete the program do so in D years of study and thus graduate D years after entering.
- dropouts are uniformly distributed over the D years of study. This assumption holds independently for each age.

The progression of students under these conditions is quite regimented. Students enter in first year and either continue on to the next year of study the following year or abandon their studies (by choice or by choice of the institution). Once studies are discontinued the individual is not able to return to the program. If the assumption of no resumption of studies is too restrictive, it can be assumed that individuals who drop out return in first year.

Let us focus on a particular class of students in this program, who enter in year t and graduate D years later. For each age a, we know the number of new entrants (first-year students) age a in year t and the number of graduates age a+D in year t+D. Let us illustrate with a numerical example with a=18, D=4.

Suppose 1,000 students aged 18 enter in year t, of whom 600 graduate four years later. The number of individuals from this group who do not complete the program (i.e. dropouts) is given by the difference in these two numbers (1000-600=400). We assume one-quarter of this number (400/4=100) drop out during each of the first, second, third and fourth years of study. Thus the number enrolled in the second, third and fourth years of study, in years t+1, t+2 and t+3, respectively, is given by the number of new entrants in year t less the number who dropped out thus far. As these individuals were age 18 in year t, those enrolled in second year are age 19 in year t+1, those in third year age 20 in year t+2 and those in fourth year age 21 in year t+3. In our example, since 100 dropped out during the first year of study, the number enrolled in second year is 1000-100=900. A further 100 dropped out during the second year, so the number enrolled in third year is 900-100=800. Similarly the number enrolled in fourth year is 700, of whom 100 drop out during the fourth year, leaving us with 600 graduates.

To estimate enrolment age a in year t in this particular program, use the above to estimate the enrolment age a in year t in each year of study (1, 2, ..., D), and sum the values. Individuals age a in year of study i+1 in year t were new entrants of age (a-i) in year (t-i) for i=0, 1, 2, ..., D-1. Thus to estimate enrolment age a in year t, it is necessary to apply the above analysis to D different classes of new entrants.

Summing estimated enrolment by age over all university programs gives estimates of university enrolment by age. Similarly to elementary enrolments, these estimated enrolments by age are reconciled with known total university enrolment. See Appendix B for details.

3.2 Enrolments by year of study available for two consecutive years and number of new entrants available

Suppose enrolments by year of study are available for two consecutive reference years. A methodology for calculating enrolments by year of study for all years is described in the Indicators of Education System (INES) Ninth Technical Group Meeting publication "Quick survey form on unit costs" (TGM-97-4). In this approach "Ideally the year of study should be the effective number of years spent by a student since the beginning of...studies" (TGM-97-4, p.2). However it is recognized that in some cases proxies such as the course year may be used. Also it is assumed the number of individuals enrolled in the first year of the program equals the number of new entrants (i.e. no entries beyond first year, and all first-year students are new entrants). The calculation methodology assumes that all cohorts (i.e. classes of new entrants) follow the same transition behaviour as the individuals observed between the two periods for which data is available. This will be explained with an example. Suppose data for a three-year program is observed for years 1990 and 1991. Thus transition behaviour between the second and third years is observed for the 1989 entry cohort, and between first and second years for the 1990 entry cohort. Under the assumption of similar transition behaviour for all cohorts, the percentage passing from first to second year for all entry cohorts will be assumed to be the same as that for the 1990 entry cohort, and from second to third as that for the 1989 entry cohort.

Assume enrolments by year of study are available for years t-1 and t. As described in TGM-97-4 (p. 6-7), the transition ratio (for year t) from the (i-1)th year of study to the ith year of study is defined as the ratio of enrolment in year of study i in year t to the enrolment in year of study (i-1) in year (t-1). The transition rate can be calculated for year t for all years of study, as enrolments by year of study are available for years t-1 and t. The transition rate gives the proportion of students enrolled in year of study i-1 in year t-1 who proceed to the next year of study the following year. As we assume equivalence between year of study and year of enrolment, the transition rate gives the proportion of students enrolled in year of study i-1 in year t-1 who continue in the program the following year. As described in Appendix B, the proportion of new entrants in year t-(i-1) who are enrolled in year of study i in year t can be obtained as the product of the transition ratios for years of study 1, 2, ..., i in year t, due to the assumption of similar transition behaviour between cohorts. This proportion is called the conditional probability to ith year of study in year t, and in Appendix B it is shown that this proportion is independent of year t, so we may speak of the conditional probability to ith year of study. Applying this probability to a cohort of new entrants gives the number who will be enrolled in the ith year of study i-1 years later. Thus if counts of new entrants by age are available, estimates of enrolment by age and year of study can be produced. Aggregating over all years of study produces estimates of enrolment by age. Once again these estimates of enrolment by age are reconciled with known total university enrolment to produce consistent estimates.

This methodology can also be used to calculate the average duration of studies. Suppose a cohort begins studies as new entrants in some year. The conditional probabilities can be used to obtain the number from this cohort who complete exactly i years of study, and from this the average duration of studies can be calculated. Details are given in Appendix B. The average duration of studies is the average number of years of study completed by both graduates and dropouts, and should not be confused with the duration of program of Section 3.1. Also completion of a year of study is defined as enrolment in that year of study, and does not imply successful completion of that year of study.

3.3 Historical enrolments by age available

If historical university enrolments and participation rates by age are available for one or more years, they can be used to estimate the current university participation rate in a manner similar to estimation of the current elementary enrolment rate described earlier. However, unlike the elementary enrolment rate, the university participation rate is also a function of external factors, such as the economic climate, availability of employment, government policy regarding post-secondary education, tuition fees, availability of student funding, importance of university education among employers, etc. NCES uses the unemployment rate and a "four-period weighted average of per-capita real disposable income" in "Projections of Education Statistics to 2007" (p.131). For NCES, "The regression method used to estimate ... was pooled least squares with first-order autocorrelation correction" (p. 131 NOTE).

Similarly to elementary, university enrolment by age is estimated by multiplying the estimated participation rate by the population estimate, provided the estimated participation rate is appropriate for the current year. As usual, these estimates are reconciled with the known total university enrolment.

As was described for elementary, if the percentage breakdown (rather than enrolment counts) by age of university enrolment is known for one or more years, then enrolment by age can be estimated as the historical share multiplied by current total enrolment. This approach assumes the historical breakdown by age is appropriate for the current year. See Appendix B for details.

University enrolment can be divided into two components, which for purposes of this document will be denoted non-permanent residents (npr) and permanent residents (pr). The former includes students from abroad who possess a so-called "student visa", and the latter includes citizens of the country. Often differential fees apply for the two groups (npr considerably higher) and the number of npr students may be dictated by a quota. Thus enrolment in the two groups is modeled quite differently. The npr enrolment normally equals the quota, and as such is a function of policy rather than participation rates. If policy in unchanging, the npr enrolment can be taken as constant. The pr enrolment responds to external factors as well as to demographic trends, and thus is modeled as a function of several inputs, as described above.

Let us assume an estimate of the npr enrolment by age is available each year. If npr are included in population estimates (as in Canada), it is necessary to extract them if the npr component of enrolment is estimated separately. The estimated participation rate then applies to the pr component of the population, and enrolment by age is estimated as the sum of the estimate of npr enrolment and the product of the non-npr population estimate and the estimate of the (non-npr) participation rate.

3.4 College and trade/vocational enrolments

Estimates of college and trade/vocational enrolment by age will be reconciled with total college and trade/vocational enrolment, respectively. The formula is similar to that for university enrolments, as again ages 16-34 are included.

In general college programs are of shorter duration than university programs, and trade/vocational of shorter duration still. In addition often trade/vocational programs involve alternating periods (several weeks or months) of work and schooling. Consequently year-to-year transition methods like those of Sections 3.1 and 3.2 may not be appropriate. For college and trade/vocational programs enrolment may be estimated based only on the number of individuals enrolled in past years, as often in these programs enrolment is at capacity.

Appendix A
Estimation of elementary enrolment by age

Π- and Σ-operators

Throughout these appendices the Π-operator (product) and Σ-operator (summation) will be used. Use of these operators is illustrated below.

$$\prod_{i=1}^{n} a_i = a_1 * a_2 * \ldots * a_n$$
$$\sum_{j=1}^{m} b_j = b_1 + b_2 + \ldots + b_m$$

For b<a,

$$\prod_{i=a}^{b} a_i = 1$$
$$\sum_{j=a}^{b} b_j = 0$$

Estimates and estimators

Suppose we wish to estimate the quantity X using a function $\hat{X}$. In proper terminology the function $\hat{X}$ is known as an estimator of X and values produced by or taken on by $\hat{X}$ are termed estimates. However for ease we will abuse terminology and refer to $\hat{X}$ as an estimate of X. A "hat" above a variable indicates it is an estimator/estimate of the corresponding quantity.

Notation and definitions

P_a^t: population estimate of individuals age a in year t

The quantity $(1+p_a^t)=P_{a+1}^{t+1}/P_a^t$ is denoted the population growth factor for age a in year t, and gives the rate of increase in the population of age a in year t between year t and year t+1.

$ELEM^t$: total elementary enrolment in year t

E_a^t: elementary enrolment in year t of individuals age a
${}_gE_a^t$: elementary enrolment in year t in grade g of individuals age a

${}_gE^t$: elementary enrolment in year t in grade g

The quantity $e_a^t=E_a^t/P_a^t$ is denoted the elementary enrolment rate for age a in year t.

r_g: repeater rate for grade g, applicable in all years
d_g: dropout rate for grade g, applicable in all years

The successful completion rate for grade g is given by $(1-d_g-r_g)$, and gives the proportion of individuals enrolled in grade g who will pass on to grade g+1 the following year.

Inputs and outputs

For each scenario below a list of inputs and outputs is given. Values must be specified for each quantity in the list of inputs, for all grades g, ages a and years t (unless otherwise noted). From the inputs the estimation method of the scenario produces values (estimates) for the variables (estimators) found in the list of outputs. As explained earlier, we will abuse terminology and refer to the variables in the list of outputs as estimates.

Reconciliation of estimated elementary enrolment by age with known total elementary enrolment

Many of the scenarios below produce estimates $\hat{E}_a^t$ of E_a^t independently of the known total elementary enrolment $ELEM^t$. As a result, there is no reason that the age-specific estimates of elementary enrolment will sum (over all ages) to total elementary enrolment. As the goal of this estimation methodology is to break down known total elementary enrolment into elementary enrolment by age, it is important that the age-specific estimates sum to the known total elementary enrolment. To this end, an adjusted estimate $\widehat{ELEM}_a^t$ given by:

$$\widehat{ELEM}_a^t = \frac{ELEM^t}{\sum_{a=5}^{18} \hat{E}_a^t} * \hat{E}_a^t$$

is produced for ages a=5, 6, 7, ...,18. The choice of notation is motivated by the fact that the adjusted estimate can be viewed as an estimate of the portion of total elementary enrolment which is age a, the logical notation for which would be $ELEM_a^t$. Note that:

$$\sum_{a=5}^{18} \widehat{ELEM}_a^t = ELEM^t$$

Sections in these appendices are numbered in correspondence to sections in the main body of the text.

2.1 No historical enrolment data available

Inputs:

P_a^t: population estimate of individuals age a in year t

Outputs:

$\hat{E}_a^t$: estimate of elementary enrolment age a in year t

Enrolment by age can be estimated as the population estimate, i.e.:

$$\hat{E}_a^t = P_a^t$$

2.1.0 Number or proportion of individuals not in school available

Inputs:

P_a^t: population estimate of individuals age a in year t
N_a^t: number of individuals age a not enrolled in school in year t
n_a^t: proportion of individuals age a not enrolled in school in year t

Outputs:

$\hat{E}_a^t$: estimate of elementary enrolment age a in year t

If the number of individuals age a not enrolled in school is known, the estimate of enrolment is:

$$\hat{E}_a^t = P_a^t - N_a^t$$

If the proportion of individuals age a not enrolled in school is known, the estimate of enrolment is:

$$\hat{E}_a^t = n_a^t * P_a^t$$

2.1.1 Estimates of repeater rate and dropout rate available

Inputs:

P_a^t: population estimate of individuals age a in year t
r_g: repeater rate for grade g, applicable in all years
d_g: dropout rate for grade g, applicable in all years

Outputs:

${}_g\hat{E}_a^t$: estimate of ${}_gE_a^t$ for g=a-5 and g=a-6 (i.e. a=g+5 and a=g+6)

$\hat{E}_a^t$: estimate of elementary enrolment age a in year t

As discussed in the text, assuming at most one repeated grade and assuming individuals begin school at age 5 in grade 0 (kindergarten), the estimated elementary enrolment is zero unless a=g+5 (no repeated grades) or a=g+6 (one repeated grade).

The estimate of elementary enrolment in grade g of individuals age g+5 in year t is given by:

$${}_g\hat{E}_{g+5}^t = P_5^{t-g} \prod_{i=0}^{g-1} \left[(1-d_i-r_i)\,(1+p_{5+i}^{t-g+i}) \right]$$

The estimate of elementary enrolment in grade g of individuals age g+6 in year t is given by:

$${}_g\hat{E}_{g+6}^t = P_5^{t-g-1} \prod_{i=0}^{g-1} \left[(1-d_i-r_i)\,(1+p_{5+i}^{t-g-1+i}) \right] r_g\,(1+p_{g+5}^{t-1}) \;+$$
$$\sum_{i=0}^{g-1} P_5^{t-g-1} \left[\prod_{j=0}^{i-1} \left[(1-d_j-r_j)\,(1+p_{5+j}^{t-g-1+j}) \right] r_i\,(1+p_{5+i}^{t-g-1+i}) \prod_{k=i}^{g-1} \left[(1-d_k)\,(1+p_{6+k}^{t-g+k}) \right] \right]$$

As estimated elementary enrolment is zero for age a unless g=a-5 or g=a-6, the estimate of elementary enrolment age a in year t is given by:

$$\hat{E}_a^t = {}_{a-5}\hat{E}_a^t + {}_{a-6}\hat{E}_a^t$$

The estimates $\hat{E}_a^t$ of Sections 2.1, 2.1.0 and 2.1.1 can be reconciled with total elementary enrolment $ELEM^t$ as described above to produce adjusted estimates $\widehat{ELEM}_a^t$.

2.2 Historical enrolments by grade available

Inputs:

- P_a^t: population estimate of individuals age a in year t
- ${}_gE^t$: elementary enrolment in grade g in year t
- r_0: repeater rate for grade 0 (kindergarten), applicable in all years
- d_g: dropout rate for grade g, applicable in all years

Outputs:

- r_g^*: estimate of repeater rate for grade g, for g=1,2,...
- ${}_g\hat{E}_a^t$: estimate of ${}_gE_a^t$ for g=a-5 and g=a-6 (i.e. a=g+5 and a=g+6)
- $\hat{E}_a^t$: estimate of elementary enrolment age a in year t

As described in the text, the elementary enrolment in grade (g+1) in year (t+1) can be decomposed into those who successfully completed grade g in year t and those who are repeating grade (g+1), each multiplied by a population growth factor. This relation is given by:

$$ {}_{g+1}E^{t+1} = (1+p_{g+5}^t)(1-r_g-d_g)\,{}_gE^t + r_{g+1}(1+p_{g+6}^t)\,{}_{g+1}E^t \quad (*) $$

for g=0, 1, ... Rewriting the above we obtain:

$$ r_{g+1} = \frac{{}_{g+1}E^{t+1} - (1+p_{g+5}^t)(1-r_g-d_g)\,{}_gE^t}{(1+p_{g+6}^t)\,{}_{g+1}E^t} $$

Note that in the above the unknowns are d_g, r_g and r_{g+1} if population estimates and enrolments by grade are known. We observe that to calculate the repeater rate for grade (g+1) we require the repeater and dropout rates for grade g be specified. Assuming dropout rates for all grades are specified, if the repeater rate for grade 0 is specified the repeater rate for grade 1 can be calculated, and using the calculated repeater rate for grade 1 the repeater rate for grade 2 can be calculated, etc. Thus specifying the dropout rates for all grades and the repeater rate for kindergarten (grade 0) is sufficient to calculate repeater rates for all grades. Let r_0^*, r_1^*, ... be estimates of the repeater rates by grade, obtained as above or as averages (over various years) of values obtained as above. These will be used to break down the known elementary enrolment ${}_{g+1}E^{t+1}$ into those repeating grade g+1 and those who successfully completed grade g last year. Assuming the former are aged g+7 and the latter aged g+6, elementary enrolments by age can be estimated as:

$${}_{g+1}\hat{E}^{t+1}_{g+6} = \frac{(1+p^t_{g+5})(1-r^*_g-d_g)\,{}_gE^t}{(1+p^t_{g+5})(1-r^*_g-d_g)\,{}_gE^t + r^*_{g+1}(1+p^t_{g+6})\,{}_{g+1}E^t} * {}_{g+1}E^{t+1}$$

and

$${}_{g+1}\hat{E}^{t+1}_{g+7} = \frac{r^*_{g+1}(1+p^t_{g+6})\,{}_{g+1}E^t}{(1+p^t_{g+5})(1-r^*_g-d_g)\,{}_gE^t + r^*_{g+1}(1+p^t_{g+6})\,{}_{g+1}E^t} * {}_{g+1}E^{t+1}$$

As estimated elementary enrolment is zero for age a unless g=a-5 or g=a-6, the estimate of elementary enrolment age a in year t is given by:

$$\hat{E}^t_a = {}_{a-5}\hat{E}^t_a + {}_{a-6}\hat{E}^t_a$$

We note that individuals enrolled in grade g+1 are not necessarily of ages g+6 and g+7 as we no longer assume one repeated grade per lifetime. However the number of individuals older than age g+7 in grade g should be not significant for expected repeater rates.

The estimates $\hat{E}^t_a$ can be reconciled with total elementary enrolment $ELEM^t$ as described above

to produce adjusted estimates $\widehat{ELEM}^t_a$.

It should also be noted that the calculation of the repeater rates could be repeated under the assumption of only one repeated grade per lifetime, as in Section 2.1.1, although the calculations would be more unwieldy than those just presented. To proceed, replace (*) by:

$${}_{g+1}E^{t+1} = (1+p^t_{g+5})(1-r_g-d_g)\,{}_gE^t_{g+5} + r_{g+1}(1+p^t_{g+6})\,{}_{g+1}E^t_{g+6} + (1-d_g)(1+p^t_{g+6})\,{}_gE^t_{g+6}$$

and rewrite r_{g+1} in terms of r_g, d_g and known elementary enrolments and population growth factors. After calculating the r^*_g, break down ${}_{g+1}E^{t+1}$ into the number who are repeating grade g+1 or have repeated a grade (the second and third terms in the sum above) and the number who have never repeated a grade (the first term in the sum). Now necessarily the former are of age g+7 and the latter of age g+6 due to the assumption of only one repeated grade per individual.

2.3 Historical enrolments by age available

Inputs:

P_a^t: population estimate of individuals age a in year t
E_t^y: elementary enrolment age a in year y, for certain years y

Outputs:

$\hat{e}_a^t$: estimate of elementary enrolment rate for age a in year t

$\hat{E}_a^t$: estimate of elementary enrolment age a in year t

Recall the elementary enrolment rate e_a^t is given by $e_a^t = E_a^t / P_a^t$.

Suppose that elementary enrolments by age are available for a single year t_0. If the elementary enrolment rates for all ages in t and t_0 are considered similar, an estimate of elementary enrolment age a in year t is given by:

$$\hat{E}_a^t = e_a^{t_0} * P_a^t = \frac{E_a^{t_0}}{P_a^{t_0}} * P_a^t$$

Reconciling this estimate with known total elementary enrolment $ELEM^t$ gives the adjusted estimate:

$$\widehat{ELEM}_a^t = \hat{E}_a^t * \frac{ELEM^t}{\sum_{a=5}^{18} \hat{E}_a^t} = \frac{E_a^{t_0}}{P_a^{t_0}} * P_a^t * \frac{ELEM^t}{\sum_{a=5}^{18} \frac{E_a^{t_0}}{P_a^{t_0}} * P_a^t}$$

Suppose that between years t_0 and t the population distribution by age for the cohort age 5-18 remains stable, in addition to the assumption of similar elementary enrolment rates for each age. That is, for each age between 5 and 18, the percentage of the cohort of that age is similar in years t and t_0. Then for each age a the ratio of the population in year t to that in year t_0 is a constant depending on t and t_0 but not on a, i.e.:

$$\frac{P_a^t}{P_a^{t_0}} = c(t,t_0)$$

From substitution of this in the above it follows that :

$$\widehat{ELEM}_a^t = \frac{E_a^{t_0}}{\sum_{a=5}^{18} E_a^{t_0}} * ELEM^t$$

If only the <u>elementary enrolment distribution by age</u> (i.e. for each age, the proportion of elementary enrolment) is known, rather than elementary enrolment counts by age, then the above formula can be used for estimation of elementary enrolment by age for year t, if it is believed the elementary enrolment distribution by age is similar between year t_0 and year t. Note that the proportion of elementary enrolment age a for year t_0 is given by the first term on the right side of the above.

Estimation of elementary enrolment rate

Let us assume elementary enrolments by age are available for all years y, $y<=t_0$. We wish to estimate e_a^t, the elementary enrolment rate for age a for some year t, $t>t_0$.

Some estimates for e_a^t are:

i) the elementary enrolment rate from the last year of available data (as presented above), i.e.:

$$\hat{e}_a^t = e_a^{t_0}$$

ii) the average of the elementary enrolment rates of the last n years, where n is the number of years of historical data to be included:

$$\hat{e}_a^t = \frac{1}{n} * \sum_{i=1}^{n} e_a^{(t_0+1)-i}$$

ii) a weighted average of the elementary enrolment rates of the last n years, where n is the number of years of historical data to be included:

$$\hat{e}_a^t = \sum_{i=1}^{n} w_i * e_a^{(t_0+1)-i}$$

where $w_1, w_2, \ldots, w_n$ satisfy:

$$\sum_{i=1}^{n} w_i = 1$$

One possible choice of the weights w_i is motivated by a process known as single exponential smoothing. In single exponential smoothing,

$$w_i = \frac{\alpha}{1-(1-\alpha)^n}(1-\alpha)^{i-1}$$

where α is known as the smoothing constant ($0<\alpha<1$). The National Centre for Education Statistics (NCES) uses α =0.4 in "Projections of Education Statistics to 2007" (p.121). In single exponential smoothing, more weight is placed on recent observations than on earlier ones, and consequently older observations have less influence. Single exponential smoothing is appropriate when the historical data have basically a horizontal pattern (p. 121).

Estimate elementary enrolment for age a in year t as:

$$\hat{E}_a^t = \hat{e}_a^t * P_a^t$$

The estimates $\hat{E}_a^t$ can be reconciled with total elementary enrolment $ELEM^t$ as described above

to produce adjusted estimates $\widehat{ELEM}_a^t$.

2.4 Secondary enrolments

Reconciliation of estimated secondary enrolment by age with known total secondary enrolment

Analogously to the above definitions for elementary enrolment, let the following be defined for secondary enrolment:

$SECY^t$: total secondary enrolment in year t
S_a^t: secondary enrolment in year t of individuals age a
$\hat{S}_a^t$: estimate of secondary enrolment age a in year t

Estimates of secondary enrolment by age are calculated using the methods of elementary enrolment estimation described above. In order that the calculated estimates $\hat{S}_a^t$ of secondary enrolment by age S_a^t will be consistent with known total secondary enrolment $SECY^t$, adjusted estimates $\widehat{SECY}_a^t$ given by:

$$\widehat{SECY}_a^t = \frac{SECY^t}{\sum_{a=12}^{34} \hat{S}_a^t} * \hat{S}_a^t$$

are produced for ages a=12, 13, ...,34. The choice of notation is motivated by the fact that the adjusted estimate can be viewed as an estimate of the portion of total secondary enrolment which is age a, the logical notation for which would be $SECY_a^t$. Note that:

$$\sum_{a=12}^{34} \widehat{SECY}_a^t = SECY^t$$

Appendix B
Estimation of university enrolment by age

Definitions and derived quantities

In addition to quantities specified in Appendix A for elementary enrolment, there are some additional quantities to be defined for university enrolment.

$UNIV^t$: total university enrolment in year t

U_a^t: university enrolment in year t of individuals age a
${}_iU_a^t$: university enrolment in year t in year of study i of individuals age a

Define the <u>university participation rate</u> u_a^t for age a in year t as $u_a^t = U_a^t / P_a^t$.

The following may be defined for a particular program of study or for university in general:

G_a^t: number of graduates of age a in year t
NE_a^t: number of new entrants of age a in year t

Reconciliation of estimated university enrolment by age with known total university enrolment

Similarly to elementary enrolment estimates, in order to sum to the known total university enrolment $UNIV^t$, the estimates $\hat{U}_a^t$ are adjusted to produce reconciled estimates given by:

$$\widehat{UNIV}_a^t = \hat{U}_a^t * \frac{UNIV^t}{\sum_{a=16}^{34} \hat{U}_a^t}$$

for a=16,17,...,34. Note that:

$$\sum_{a=16}^{34} \widehat{UNIV}_a^t = UNIV^t$$

3.1 Numbers of new entrants and graduates available

Inputs:

G_a^t: the number of graduates of age a in year t from this program of study
NE_a^t: the number of new entrants of age a in year t to this program of study
D: the duration in years of the program of study

Outputs:

${}_i\hat{U}_a^t$: estimate of enrolment age a in year of study i in year t in the program of study

$\hat{U}_a^t$: estimate of enrolment age a in year t in the program of study

Assuming the dropouts (i.e. the difference between the number of new entrants and the number of graduates D years later) are equally distributed during the D years of study, an estimate of enrolment by age and year of study is given by:

$${}_{i+1}\hat{U}_{a+i}^{t+i} = NE_a^t - \frac{i}{D} * (NE_a^t - G_{a+D}^{t+D})$$

for i=0,1,2,...,D-1.

For the example given in the text (a=18, D=4):

$${}_{i+1}\hat{U}_{18+i}^{t+i} = NE_{18}^t - \frac{i}{4} * (NE_{18}^t - G_{22}^{t+4})$$

for i=0,1,2,3. The term in brackets is the number who drop out, and we assume one-quarter of this number drop out during each of the first, second, third and fourth years. Thus the enrolment in any year of the program is the number of new entrants less the number which have dropped out so far. Substituting $NE_{18}^t = 1000$, $G_{22}^{t+4} = 600$ and dropping the "hat" from the "U" for now gives: ${}_1U_{18}^t = 1000$, ${}_2U_{19}^{t+1} = 900$, ${}_3U_{20}^{t+2} = 800$, ${}_4U_{21}^{t+3} = 700$, as described in the text.

For any year t individuals age a are (conceivably) enrolled in any of the D years of study, depending on when they entered the program. Thus to estimate enrolment age a, sum over D years of study to give:

$$\hat{U}_a^t = \sum_{i=1}^{D} {}_i\hat{U}_a^t$$

The estimates $\hat{U}_a^t$ can be reconciled with total university enrolment $UNIV^t$ as described above

to produce adjusted estimates $\widehat{UNIV}_a^t$.

The assumption of a standard duration of completion is restrictive; however, with nothing other than new entrant and graduation counts by age available, there is little basis for calculation of university enrolments by age and certain assumptions are necessary. If the distribution of duration to completion is known (e.g. 80% complete in four years, 20% in five years) the above can be modified to reflect this, although this short note will not go into the mathematical details.

3.2 Enrolments by year of study available for two consecutive years and number of new entrants available

Inputs:

$S_{i,t}$: university enrolment in ith year of study in year t (notation of TGM-97-4)
NE_a^t: number of new entrants age a in year t

Outputs:

${}_i\hat{U}_a^t$: estimate of enrolment age a in year of study i in year t in the program of study

$\hat{U}_a^t$: estimate of enrolment age a in year t in the program of study

D: estimate of average duration of studies (notation of TGM-97-4)

Define the transition ratio from the (i-1)th to ith year of study for year t as:

$$a_{i,t}=\frac{S_{i,t}}{S_{i-1,t-1}}$$

and the conditional probability to the ith year of study in year t as:

$$q_{i,t}=\prod_{j=1}^{i} a_{j,t}$$

Define $a_{1,t}$=1 and $q_{1,t}$=1. It can be shown that the conditional probability to the ith year of study in year t gives the proportion of new entrants in year t-(i-1) who are enrolled in year of study i in year t. For the latter is given by:

$$\frac{S_{i,t}}{S_{1,t-(i-1)}}=\frac{S_{i,t}}{S_{i-1,t-1}}*\frac{S_{i-1,t-1}}{S_{i-2,t-2}}*\frac{S_{i-2,t-2}}{S_{i-3,t-3}}*\ldots*\frac{S_{3,t-(i-3)}}{S_{2,t-(i-2)}}*\frac{S_{2,t-(i-2)}}{S_{1,t-(i-1)}}$$

$$=a_{i,t}*a_{i-1,t-1}*a_{i-2,t-2}*\ldots*a_{3,t-(i-3)}*a_{2,t-(i-2)}$$

By the assumption of similar transition behaviour between cohorts, $a_{i-j,t-j}=a_{i-j,t}$ for all j, i.e. the proportion of students passing from the (i-j-1)th year of study to the (i-j)th year of study between years t-j-1 and t-j is the same as the proportion of students passing between the same years of study between years t-1 and t . Thus:

$$\frac{S_{i,t}}{S_{1,t-(i-1)}}=a_{i,t}*a_{i-1,t}*a_{i-2,t}*\ldots*a_{3,t}*a_{2,t}=\prod_{j=1}^{i}a_{j,t}$$

since $a_{1,t}=1$. Note also that the assumption of similar transition behaviour among cohorts implies that the conditional probability to the ith year of study is independent of year t, i.e. $q_{i,t}=q_{i,s}$ for all years of study i and all years t and s.

For an arbitrary entry cohort of size N, the number N_i reaching year of study i is given by:

$$N_i=N*q_{i,t}$$

If the number of new entrants by age NE_a^t is known, applying this methodology yields estimates of university enrolment by age and year of study given by:

$${}_i\hat{U}_a^t=q_{i,t}*NE_{a-(i-1)}^{t-(i-1)}$$

for a=16,...,34 and i=1,2,...,10 (assume a maximum of 10 years of study). To estimate university enrolment by age sum over years of study i=1,2,...,10, to get:

$$\hat{U}_a^t=\sum_{i=1}^{10}{}_i\hat{U}_a^t$$

The estimates $\hat{U}_a^t$ can be reconciled with total university enrolment $UNIV^t$ as described above

to produce adjusted estimates $\widehat{UNIV}_a^t$.

As illustrated in TGM-97-4, this methodology can be used to estimate the average duration of studies. Although not necessary for estimation of enrolment by age, the method will be included here. Returning to the notation of TGM-97-4, let us consider an arbitrary cohort of size N. The number completing exactly i years of study[4] is given by:

$$(N_i - N_{i+1}) = (q_{i,t} - q_{i+1,t}) * N$$

Assuming a maximum of 10 years of study, the average number of years of study is calculated as:

$$D = \frac{\sum_{i=1}^{10} i * \{\text{\# students who complete exactly i years of study}\}}{N}$$

$$= \sum_{i=1}^{10} i * (q_{i,t} - q_{i+1,t}) = \sum_{i=1}^{10} q_{i,t} * (i - (i-1))$$

provided $N_{11}=0$ and $q_{11,t}=0$. It follows the average duration of studies D is given by

$$D = \sum_{i=1}^{10} q_{i,t}$$

where 10 is the imposed maximum the number of years of study and the formula is independent of year t. The latter follows from the assumption of similar transition behaviour for all cohorts. See TGM-97-4 for further clarification. Note that this is the duration of study for both graduates and dropouts, and should not be confused with the duration of program, for which the symbol D was used in Section 3.1.

3.3 Historical enrolments by age available

Inputs:

- P_a^t: population estimate for age a in year t
- U_t^y: number of individuals of age a enrolled in university in year y, for certain years y

Outputs:

- $\hat{u}_a^t$: estimate of university participation rate for age a in year t

---

[4]"completing exactly i years of study" is defined as reaching ith year of study but not (i+1)st and thus does not imply the individual successfully completed the ith year of study.

$\hat{U}_a^t$: estimate of university participation age a in year t

Recall the university participation rate is defined as $u_a^t = U_a^t / P_a^t$.

Suppose that university enrolments by age are available for a single year t_0. If the university participation rates for all ages in t and t_0 are considered similar, an estimate of university enrolment age a in year t is given by:

$$\hat{U}_a^t = u_a^{t_0} * P_a^t = \frac{U_a^{t_0}}{P_a^{t_0}} * P_a^t$$

Reconciling this estimate with known total university enrolment $UNIV^t$ gives the adjusted estimate:

$$\widehat{UNIV}_a^t = \hat{U}_a^t * \frac{UNIV^t}{\sum_{a=16}^{34} \hat{U}_a^t} = \frac{U_a^{t_0}}{P_a^{t_0}} * P_a^t * \frac{UNIV^t}{\sum_{a=16}^{34} \frac{U_a^{t_0}}{P_a^{t_0}} * P_a^t}$$

Suppose that between years t_0 and t the population distribution by age for the cohort age 16-34 remains stable, in addition to the assumption of similar university participation rates. That is, for each age between 16 and 34, the percentage of the cohort of that age is similar in years t and t_0. Then for each age a the ratio of the population in year t to that in year t_0 is a constant depending on t and t_0 but not on a, i.e.:

$$\frac{P_a^t}{P_a^{t_0}} = c\,(t,t_0)$$

From substitution of this in the above it follows that :

$$\widehat{UNIV}_a^t = \frac{U_a^{t_0}}{\sum_{a=16}^{34} U_a^{t_0}} * UNIV^t$$

If only the <u>university enrolment distribution by age</u> (i.e. for each age, the proportion of university enrolment) is known, rather than university enrolment counts by age, then the above formula can be used for estimation of university enrolment by age for year t, if it is believed the university enrolment distribution by age is similar between year t_0 and year t. Note that the proportion of university enrolment age a for year t_0 is given by the first term on the right side of the above.

Let us assume university enrolments by age are available for all years y, $y<=t_0$. We wish to estimate u_a^t, the university participation rate for age a for some year t, $t>t_0$.

The methods given in Appendix A for estimation of the elementary enrolment rate from historical enrolment and population counts can be applied to produce estimates of the university participation rate. However, unlike for the elementary enrolment rate, it may be desirable to include auxiliary variables reflecting the direct and "opportunity" cost of attending university, such as the unemployment rate, average disposable income, average tuition fees and availability of student funding.

If $\hat{u}_a^t$ denotes an estimate of the university participation rate for age a in year t, estimate university enrolment for age a in year t as:

$$\hat{U}_a^t = \hat{u}_a^t * P_a^t$$

The estimates $\hat{U}_a^t$ can be reconciled with total university enrolment $UNIV^t$ as described above to produce adjusted estimates $\widehat{UNIV}_a^t$.

Separate estimate of non-permanent resident university enrolment

Suppose a separate estimate of non-permanent resident (npr) university enrolment exists. University enrolment age a can be estimated as:

$$\hat{U}_a^t = \hat{u}_a^t * P_a^t + \hat{F}_a^t$$

where the participation rate estimate $\hat{u}_a^t$ and population estimate P_a^t both apply to the non-npr component of the population, and $\hat{F}_a^t$ is the estimate of the number of npr students in year t.

The estimates $\hat{U}_a^t$ can be reconciled with total university enrolment $UNIV^t$ as described above

to produce adjusted estimates $\widehat{UNIV}_a^t$.