Age and Sex composition

Ayman Zohry

CDC 103 – Lecture 4 – February 26, 2012

Importance of Age and Sex Composition

Why classifying population by age and sex is important in the study of demography?

- Separate data for males and females and for ages are important in themselves, for the analysis of other types of data, and for the evaluation of the completeness and accuracy of the census counts of population.
- Age and sex composition of population are important in planning for economic and social service provision.
- Breaking down population by age and sex enables researchers to analyze variations of social phenomena by specific groups of population such as infants and children, youth, adults, and the elderly.

Importance of Age and Sex Composition

Why classifying population by age and sex is important in the study of demography?

- Age is the most important variable in the study of mortality, fertility, nuptiality, and certain other areas of demographic analysis.
- The tabulation of population by sex is useful in the evaluation of census and survey data.

Definitions and Classification

- The definition of sex is easy, while the definition of age is more complicated.
- In some countries, individuals provide their age in terms of a lunar-based calendar. China is an example
- Under the lunar-based Chinese calendar system, an individual is assigned an age of 1 at birth. While age at birth is commonly zero under the Western definition.
- Also in Islamic country that use the lunar (Hijri) calendar, the year is 11 days less than the solar year.
- The United Nations recommendations favor the Western approach, defining age as "the interval of time between the date of birth and the date of the census, expressed in completed solar years."

Definitions and Classification

- The date-of birth gives more accurate data rather than asking about age in the last birthday (in complete years).
- When date-of-birth information is collected in a census or sample survey, the recommended method for converting it to age at last birthday is to subtract the exact date of birth from the date of the census or survey.
- The resulting ages, in complete years, could then be tabulated by single years or classified into age groups, as desired.

Sources of Data on Age and Sex

- Data on Age and sex are collected in the censuses and surveys all over the world.
- Comparable data by country can be found in the UN Demographic Yearbook – available in the CDC library.

Quality of Data on Age and Sex

- The principal problem relating to the quality of the data on sex collected in censuses concerns *the difference in the completeness of coverage of the two sexes*.
- Errors in the tabulated data on age may arise from the following types of errors of enumeration:
  - Coverage errors,
  - Failure to record age, and
  - Misreporting of age.

Analysis of Sex Composition

Numerical Measures

- The numerical measures of sex composition are few and simple to compute.
- They are:
  - (1) the percentage of males in the population, or *the masculinity proportion;*
  - (2) the *sex ratio*, or the masculinity ratio; and
  - (3) the *ratio of the excess* or deficit of males to the total population.

The Masculinity Proportion

- The *masculinity proportion* (or percentage male) is the measure of sex composition most often used.
- The formula for the masculinity proportion is:

$$\frac{P_m}{P_t} \times 100$$

- Where P_m represents the number of males and P_t the total population.

Sex Ratio

- The sex ratio is the main measure of sex composition used in demography.
- The sex ratio is usually defined as the number of males per 100 females.
- The formula for the sex ratio is:

$$\frac{P_m}{P_f} \times 100$$

- Where P_m represents the number of males and P_f the number of females.

Sex Ratio

- One hundred is the point of balance of the sexes according to this measure.
- A sex ratio above 100 denotes an *excess* of males, and a sex ratio below 100 denotes an *excess* of females.
- Accordingly, the greater the excess of males, the higher the sex ratio; the greater the excess of females, the lower the sex ratio.

The Ratio of the Excess or Deficit of Males to the Total Population

- The the excess (or deficit) of males as a percentage of the total population, is calculated using the following formula:

$$\frac{P_m - P_f}{P_t} \times 100$$

- Where P_m represents the number of males, P_f the number of females, and P_t the total population.

Analysis of Sex Ratios in Terms of Population Subgroups

- Because the sex ratio may vary widely from one population subgroup to another, it is frequently desirable to consider separately the sex ratios of the important component subgroups in any detailed analysis of the sex composition of a population group.
- Subgroups may include:
  - Geographical areas
  - Analysis by race
  - Analysis by origin
  - Sex ratio among migrants and citizens

Analysis of Change

- It is frequently desired to explain in demographic terms the change in the sex composition of the population from one census to another.
- What is called for is a quantitative indication of how the components of population change - births, deaths, migration - contributed to the change in sex composition.
- Gulf countries as an example

A Spreadsheet to Calculate Sex Ratio

Analysis of Deficiencies in Age Data

- We will consider the types of deficiencies in census tabulations of age under four general headings:
  - errors in single years of age,
  - errors in grouped data,
  - reporting of extreme old age, and
  - failure to report age.

Errors in Single Years of Age

- **Measurement of Age and Digit Preference**
- *Indexes of Age Preference*
- *Whipple's Index*
- *Myers's Blended Method (Myers's Index)*

Measurement of Age and Digit Preference

- See the distributed table and comment on the age distribution.
- It's noticed that ages that end with "0" and "5" have higher numbers than their neighboring ages.
- This phenomenon is called *digit preference*.
- In order to measure the degree of digit (age) preference we apply specific techniques.
- Consider the population of age 30.
- The population in this age are expected to be a number that's less than population in age 29 and greater than population in age 31.

Measurement of Age and Digit Preference

- So that, in order to measure the bias in age reporting for this age (30) we can apply the following equation (index):

$$\frac{P_{30}}{1/3(P_{29}+P_{30}+P_{31})}\times 100 = \frac{1,031,406}{1/3(928,327+1,031,406+831,571)}\times 100 = 110.9$$

- Or we may consider this age to be the average of five year ages; that is 28, 29, 30, 31, and 32.

$$\frac{P_{30}}{1/5(P_{28}+P_{29}+P_{30}+P_{31}+P_{32})}\times 100 = \frac{1,031,406}{1/5(907,680+928,327+1,031,406+831,571+810,274)}\times 100 = 114.4$$

Measurement of Age and Digit Preference

- In this case, the two indexes are similar whether a 3-year group or a 5-year group is used; both indicate substantial heaping on age 30.
- The higher the index, the greater the concentration on the age examined;
- An index of 100 indicates no concentration on this age.

Whipple's Index

- Whipple index is on of the widely used indexes to measure age misreporting.
- The index was invented by the American demographer George C. Whipple (1866–1924).
- It has been develop to reflect preference of terminal digits; 0 and 5.
- Calculation of the index:

$$\frac{\sum(P_{25}+P_{30}+.....+P_{55}+P_{60})}{(1/5)\sum(P_{23}+P_{24}+.....+P_{60}+P_{61}+P_{62})}\times 100$$

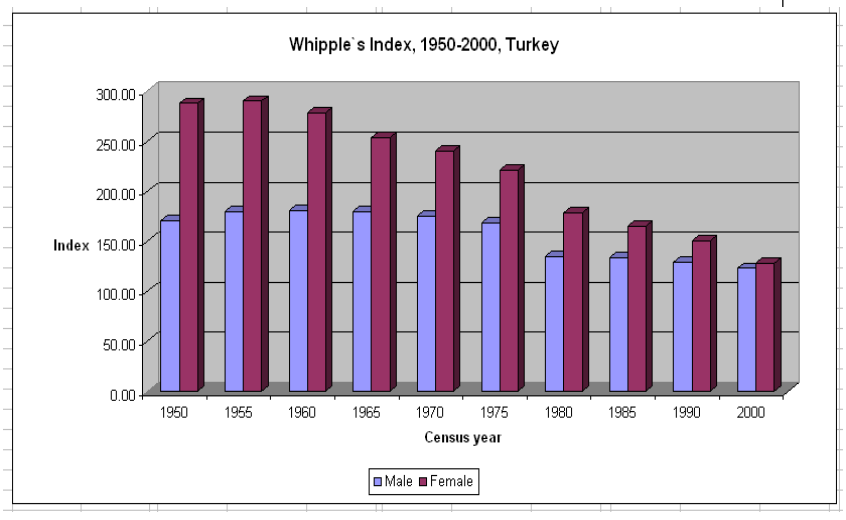
Whipple's Index

- The choice of the range 23 to 62 is largely arbitrary.
- In computing indexes of heaping, the ages of childhood and old age are often excluded because they are more strongly affected by other types of errors of reporting than by preference for specific terminal digits.

Whipple's Index

Whipple's Index	Quality of Data	Deviation from Perfect
< 105	very accurate	< 5%
105–110	relatively accurate	5–9.99%
110–125	OK	10–24.99%
125–175	bad	25–74.99%
> 175	very bad	≥ 75%

Whipple's Index



Myers's Blended Method (Myers's Index)

- Myers' method also measures preferences for each of the ten possible digits and proposes a blended index.
- It is based on the principle that in the absence of age heaping, the aggregate population of each age ending in one of the digits 0 to 9 should represent 10% of the total population.
- The index is calculated by summing the number of people whose age ends with a particular digit for the population aged 10 and over, and then for the population aged 20 and over.
- Each series is then weighted and the results are added to obtain a blended population.
- Myers' blended index is obtained by summing the absolute deviations between the aggregate and theoretical distributions (10%).

Myers's Blended Method (Myers's Index)

Terminal digit, *a*	Population with terminal digit, *a*		Weights for—		Blended population		Deviation of percentage from 10.00[1]
	Starting at age 10 + a (1)	Starting at age 20 + a (2)	Column 1 (3)	Column 2 (4)	Number (1) × (3) + (2) × (4) = (5)	Percent distribution (6)	(6) – 10.00 = (7)
0	5,794,442	4,144,526	1	9			
1	4,735,734	3,243,767	2	8			
2	4,706,488	3,200,533	3	7			
3	4,371,722	2,962,601	4	6			
4	4,382,456	2,973,683	5	5			
5	4,534,455	3,158,357	6	4			
6	3,926,253	2,623,463	7	3			
7	4,028,469	2,672,365	8	2			
8	3,767,867	2,438,758	9	1			
9	3,780,227	2,503,677	10	0			
Total	(X)	(X)	(X)	(X)			
Summary index of age preference = Total ÷ 2	(X)	(X)	(X)	(X)			

Myers's Blended Method (Myers's Index)

Terminal digit, *a*	Population with terminal digit, *a*: Starting at age 10 + a (1)	Population with terminal digit, *a*: Starting at age 20 + a (2)	Weights for—: Column 1 (3)	Weights for—: Column 2 (4)	Blended population: Number (1) × (3) + (2) × (4) = (5)	Blended population: Percent distribution (6)	Deviation of percentage from 10.00[1] (6) – 10.00 = (7)
0	5,794,442	4,144,526	1	9	43,095,176		
1	4,735,734	3,243,767	2	8	35,421,604		
2	4,706,488	3,200,533	3	7	36,523,195		
3	4,371,722	2,962,601	4	6	35,262,494		
4	4,382,456	2,973,683	5	5	36,780,695		
5	4,534,455	3,158,357	6	4	39,840,158		
6	3,926,253	2,623,463	7	3	35,354,160		
7	4,028,469	2,672,365	8	2	37,572,482		
8	3,767,867	2,438,758	9	1	36,349,561		
9	3,780,227	2,503,677	10	0	37,802,270		
Total	(X)	(X)	(X)	(X)	374,001,795		
Summary index of age preference = Total ÷ 2	(X)	(X)	(X)	(X)	(X)		

Myers's Blended Method (Myers's Index)

Terminal digit, *a*	Population with terminal digit, *a*: Starting at age 10 + a (1)	Population with terminal digit, *a*: Starting at age 20 + a (2)	Weights for—: Column 1 (3)	Weights for—: Column 2 (4)	Blended population: Number (1) × (3) + (2) × (4) = (5)	Blended population: Percent distribution (6)	Deviation of percentage from 10.00[1] (6) – 10.00 = (7)
0	5,794,442	4,144,526	1	9	43,095,176	11.52	
1	4,735,734	3,243,767	2	8	35,421,604	9.47	
2	4,706,488	3,200,533	3	7	36,523,195	9.77	
3	4,371,722	2,962,601	4	6	35,262,494	9.43	
4	4,382,456	2,973,683	5	5	36,780,695	9.83	
5	4,534,455	3,158,357	6	4	39,840,158	10.65	
6	3,926,253	2,623,463	7	3	35,354,160	9.45	
7	4,028,469	2,672,365	8	2	37,572,482	10.05	
8	3,767,867	2,438,758	9	1	36,349,561	9.72	
9	3,780,227	2,503,677	10	0	37,802,270	10.11	
Total	(X)	(X)	(X)	(X)	374,001,795	100.00	
Summary index of age preference = Total ÷ 2	(X)	(X)	(X)	(X)	(X)	(X)	

Myers's Blended Method (Myers's Index)

Terminal digit, *a*	Population with terminal digit, *a* — Starting at age 10 + a (1)	Population with terminal digit, *a* — Starting at age 20 + a (2)	Weights for— Column 1 (3)	Weights for— Column 2 (4)	Blended population — Number (1) × (3) + (2) × (4) = (5)	Blended population — Percent distribution (6)	Deviation of percentage from 10.00[1] (6) − 10.00 = (7)
0	5,794,442	4,144,526	1	9	43,095,176	11.52	1.52
1	4,735,734	3,243,767	2	8	35,421,604	9.47	0.53
2	4,706,488	3,200,533	3	7	36,523,195	9.77	0.23
3	4,371,722	2,962,601	4	6	35,262,494	9.43	0.57
4	4,382,456	2,973,683	5	5	36,780,695	9.83	0.17
5	4,534,455	3,158,357	6	4	39,840,158	10.65	0.65
6	3,926,253	2,623,463	7	3	35,354,160	9.45	0.55
7	4,028,469	2,672,365	8	2	37,572,482	10.05	0.05
8	3,767,867	2,438,758	9	1	36,349,561	9.72	0.28
9	3,780,227	2,503,677	10	0	37,802,270	10.11	0.11
Total	(X)	(X)	(X)	(X)	374,001,795	100.00	4.66
Summary index of age preference = Total ÷ 2	(X)	(X)	(X)	(X)	(X)	(X)	2.33

Myers's Blended Method (Myers's Index)

- The theoretical range of Myers's index is 0, representing no heaping, to 90, which would result if all ages were reported at a single digit.

A Spreadsheet to calculate Whipple and Myers indexes

Errors in Grouped Data

Intercensal Cohort Analysis:

- Intercensal means between two successive censuses.
- If we are dealing with a closed population (no migration), the difference between the total number of population in two successive censuses will be solely attributed to births (increase) and deaths (decrease).
- The basics:
  - Population Increases by births and decreases by deaths
  - Population increases by immigrants and decreased by emigrants

- Then expected population in age group a in the second census (Pta+10) is the population of age group a+10 in the second census – deaths that occurred in the intercensal period + migration to the country in the intercensal period – migration from the country in the intercensal period.
- That is:
- $P_{ta+10} = P_{0a} - D_{a\,(0-t)} + I_{a\,(0-t)} - E_{a\,(0-t)}$
- Where:
- P_{ta+10} = Population of age group a+10 in the second census
- P_{0a} = Population of age group a+10 in the second census
- $D_{a\,(0-t)}$ = Deaths occurred in the intercensal period
- $I_{a\,(0-t)}$ = Migrants arrived in the country in the inetrcensal period
- $E_{a\,(0-t)}$ = Migrants who left the country in the inetrcensal period

Reporting of Extreme Old Age

- Census age distributions at advanced ages, say for those 85 years old and over, suffer from serious reporting problems, with age exaggeration in older ages generally considered to be common.
- To avoid this problem we tend to make the last age group as an open ended group.
- Example: 65+ or 70+

Failure to Report Age

- Age is not always reported in a census, even though the enumerator may be instructed to secure an estimate from the respondent or to estimate it as well as possible while enumerating.
- National censuses always include a category of "unknown" age.
- To avoid this category, we distribute the unknown ages assuming that those of unknown age have the same percentage distribution by age as those of known age.

Failure to Report Age

Procedure for Prorating Ages Not Reported, for Zimbabwe: 1992

Age (years)	Population as enumerated	Population with ages not reported distributed over all ages
Total	10,412,548	10,412,548
Under 5	1,584,691	
5 to 9	1,653,788	
10 to 14	1,456,751	
15 to 19	1,248,238	
20 to 24	989,897	
25 to 34	1,318,573	
35 to 44	852,690	
45 to 54	569,478	
55 to 64	361,165	
65 and over	343,291	
Age not reported	33,986	(X)

Total population of reported age = 10,378,562

$$= \frac{\text{Total population}}{\text{Total population of reported age}} = \frac{10,412,548}{10,378,562} = 1.003274635$$

Failure to Report Age

Procedure for Prorating Ages Not Reported, for Zimbabwe: 1992

Age (years)	Population as enumerated	Population with ages not reported distributed over all ages
Total	10,412,548	10,412,548
Under 5	1,584,691	1,589,880
5 to 9	1,653,788	1,659,203
10 to 14	1,456,751	1,461,521
15 to 19	1,248,238	1,252,326
20 to 24	989,897	993,139
25 to 34	1,318,573	1,322,891
35 to 44	852,690	855,482
45 to 54	569,478	571,343
55 to 64	361,165	362,348
65 and over	343,291	344,415
Age not reported	33,986	(X)

Total population of reported age = 10,378,562