JOHNS HOPKINS
UNIVERSITY

Population Composition

Stan Becker, PhD
Bloomberg School of Public Health

JOHNS HOPKINS
UNIVERSITY

Section A

Sex and Age

Sex

- Importance
  - Variable is of prime importance
  - Separate data for males and females are important:
    - In themselves
    - For the analysis of other types of data
    - For the evaluation of the completeness and accuracy of population counts

Sex Ratio

- *Sex Ratio*—Proportion of males relative to females in a population
- General formula:

$$\frac{\text{Number of males}}{\text{Number of females}} * 100$$

Continued

Sex Ratio

- Pattern by age
  - Higher at very young ages
  - Gap narrows with increasing age
  - Goes below 100 by middle age
- Other characteristics
  - Varies from one population to another
  - Desirable to consider separately the sex ratio of important component subgroups (e.g. by race, by ethnicity)

Sex Ratio of Births

- *Sex Ratio of Births*—Number of male births per 100 female births

$$= \frac{\text{Number of male births}}{\text{Number of female births}} * 100$$

- Ratio < 100 $\rightarrow$ More female than male births
  = 100 $\rightarrow$ Same number of male and female births
  > 100 $\rightarrow$ More male than female births

Continued

Sex Ratio of Births

- In human populations, there are more male births than female births, with ratio generally between 104 and 107
- Important characteristics which distinguish births with respect to their sex ratio are age of mother, order of birth, or race

Sex Ratio of Deaths

$$\frac{\text{Number of male deaths}}{\text{Number of female deaths}} * 100$$

- Much more variable from area to area than sex ratio at birth
- Often well above 100, i.e. males have a higher mortality
- Important characteristics to include in further analyses are age, race, ethnic group, residence, marital status, and occupation

Sex Ratio of Migrants

$$\frac{\text{Number of male migrants}}{\text{Number of female migrants}} * 100$$

- Shows more extreme values than sex ratio of either birth or death
- Less uniform from area to area
- Patterns of sex-selectivity of migrants vary depending largely on types of occupational opportunities and on cultural factors

Age

- *Demographers' definition of age*— Completed age, i.e. age of an individual at last birthday
- Most important variable in demographic analyses

Data Collection on Age

- Data on age may be secured by
  - Asking a direct question on age
  - Asking a question on date of birth, or month and year of birth
  - Or a combination of these

Age-Reporting Errors

- Content
  - Centenarians
    - Those close to 100 years tend to overestimate their age
  - Understatement
    - Women tend to understate their age

Continued

Age-Reporting Errors

- Overstatement
  - Mothers tend to round up the age of their children
- Heaping/Digit preference
  - People tend to report certain ages at the expense of others
  - Can occur at any digit but happens most often with 0 and 5

Continued

Age-Reporting Errors

- Distributing unknown and/or unreported age
  - Needs to be imputed with care

Continued

Age-Reporting Errors

- *Coverage*—Missed or counted twice
  - There is a tendency to miss the people in certain age groups (e.g. young men)
  - Some people are counted twice

Method for Detecting the Extent of Age Errors

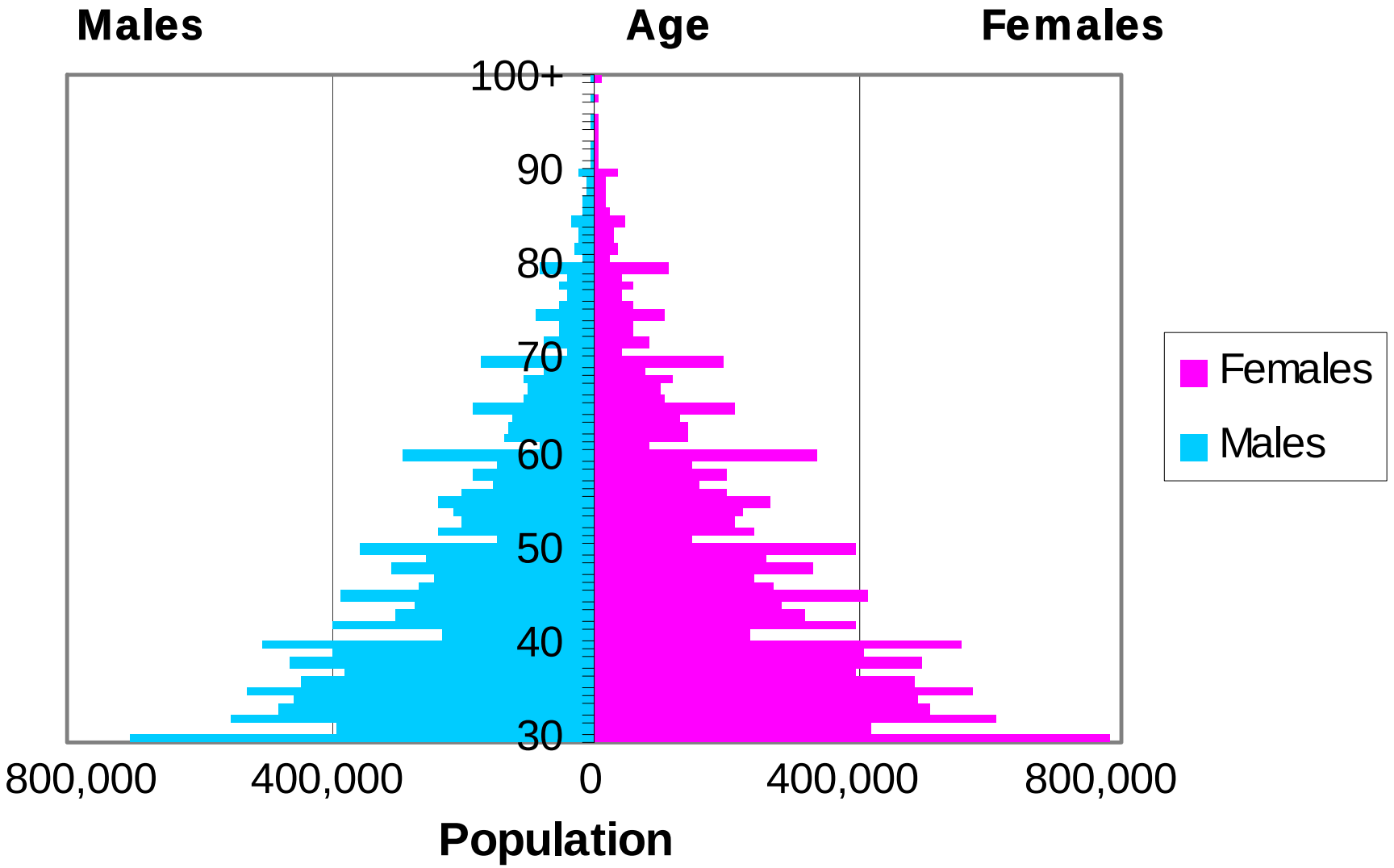
- Age ratios used to estimate data quality
- Two formulas available:
- Let a_x = number of persons of age x

$$\frac{a_x}{1/2\left(a_{x-1}+a_{x+1}\right)}*100$$

or

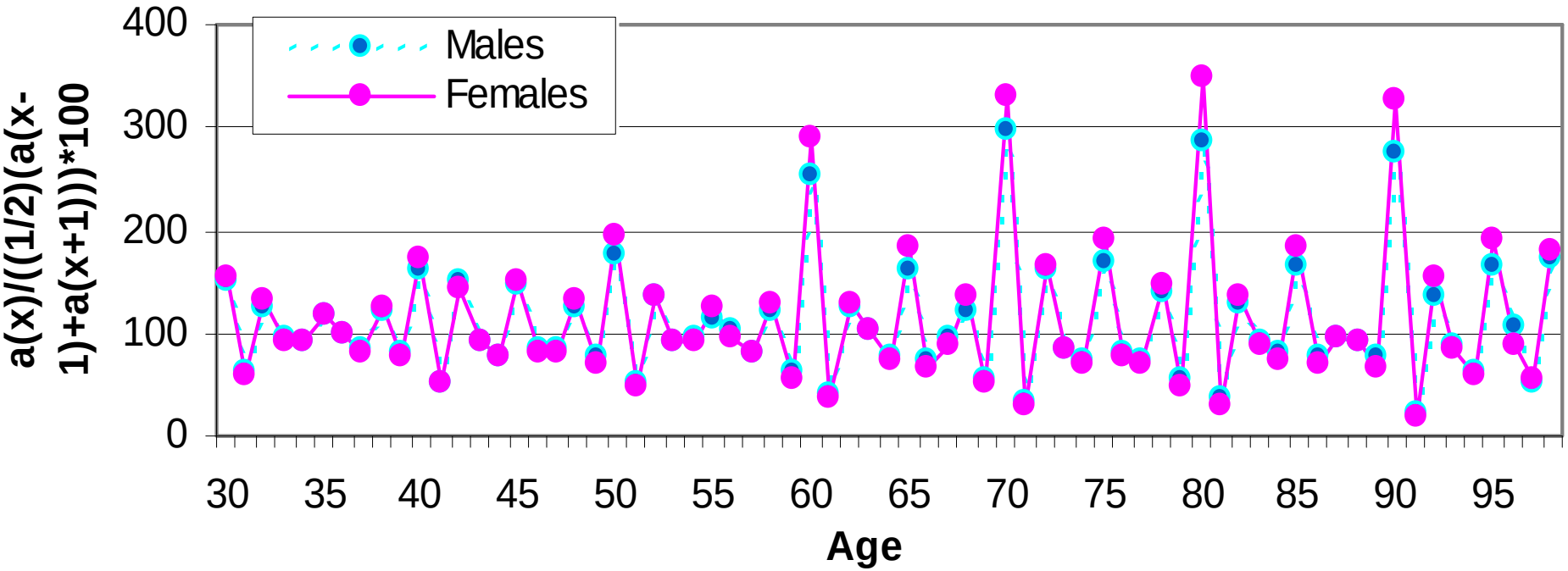
$$\frac{a_x}{1/3\left(a_{x-1}+a_x+a_{x+1}\right)}*100$$

Age Distribution of Males and Females Over Age 30 in Mexico 1990



Source: UN, Demographic yearbook 1990

Extent of Age Heaping in Mexico, 1990



Source: UN, Demographic yearbook 1990

Age

- The age ratio is then:
  - Compared to 100
  - Plotted by single year of age
  - Indices and methods have been developed to summarize the preference of or avoidance of particular terminal digits, e.g. Whipple's index, Myer's blended method

Correction of Age Errors

- Grouping
  - Generally by 5 or 10 years
  - Avoid problem of fluctuations in single year data
- Interpolation
  - To be developed later on in this course

Analysis

- When working with several populations, comparison of age distributions is a classic demographic analysis

Index of Dissimilarity (ID)

$$ID = \frac{1}{2} \sum \left| r_{2a} - r_{1a} \right|$$

- Where r_{ia} = Percent of age group a in population i

  $r_{1a} < r_{2a} \rightarrow$ Lower proportion in group a of pop. 1 compared to pop. 2

  $r_{1a} > r_{2a} \rightarrow$ Higher proportion in group a of pop. 1 compared to pop. 2

Continued

Index of Dissimilarity (ID)

- *Index of Dissimilarity (ID)*—Indicates the percent of one population that needs to be redistributed to have the age distribution of the other
- Can be calculated for any two comparable percent distributions
- Does not tell which population has the older population

Density and Distribution Functions

- f(x) = Discrete density function of x
  = probability that a case has value x
- F(x) = Cumulative distribution function of x
  = Probability that a case has value at or below x

$$= \sum_{j=0}^{x} f(j)$$

Mean Age

- There are two ways of calculating the mean of any distribution:

$$\sum xf(x) \quad \text{or} \quad \sum[1-F(x)]$$

Example of Calculation

x	f(x)	F(x)	xf(x)	1-F(x)
0	0.00	0.00	0.00	1.00
1	0.50	0.50	0.50	0.50
2	0.25	0.75	0.50	0.25
3	0.25	1.00	0.75	0.00
Mean			1.75	1.75

Note that F(0) = 0

Median Age

- Let ℓ_{Md} = Lower limit of class containing the middle (N/2th) item

  N = Sum of all frequencies, i.e. population size

  Σf_x = Sum of frequencies in all classes preceding the one containing the N/2th item

Continued

Median Age

f_{Md} = Frequency of class containing the $N/2^{th}$ item

i = Width of class interval containing the $N/2^{th}$ item

Median

- *Median*—Age which divides the population into two equal-size groups, one younger and the other older

$$= \ell_{Md} + \left(\frac{\frac{N}{2} - \sum f_x}{f_{Md}} \right) * i$$

Continued

Median

- The median is preferred to the mean because of:
  - The marked skewness of the age distribution
  - The calculation of mean is often complicated by open-ended age groups (e.g. 80+)

Continued

Median

Table: Median Age for Various Countries: Around 1960

Country and Year	*Median Age*	*Country and Year*	*Median Age*
Chile (1960)	23.3	Sweden (1960)	36.2
France (1961)	33.3	Syria (1960)	17.2
Ghana (1960)	18.4	Taiwan (1956)	17.9
Honduras (1961)	16.1	U.S.S.R. (1959)	26.6
India (1961)	20.5	United Arab Rep. (1960)	19.4
Iran (1966)	17.3	United States (1960)	29.5
Italy (1961)	31.6	Venezuela (1961)	17.8
Japan (1960)	25.6	Yugoslavia (1961)	26.4

Source: Shryock, S. and Siegel, J.S., The methods and materials of demograhy, (1976). Academic press. Basic data from UN, Demographic year book 1962, 1964.

Age

- $P_{0\text{-}14}$ = Proportion of children under 15
  P_{65+} = Proportion of aged persons
  $P_{15\text{-}64}$ = Proportion of persons of "working age"

Age-Dependency Ratios

- *Age-Dependency Ratios*—Proportion of children less than 15 and elderly 65+ relative to the population of "working ages"

$$= \frac{P_{0-14} + P_{65+}}{P_{15-64}} * 100$$

Child-Dependency Ratio

- *Child-Dependency Ratio*—Proportion of children less than 15 relative to the population of "working ages"

$$= \frac{P_{0-14}}{P_{15-64}} * 100$$

Old-Age Dependency Ratio

- *Old-Age Dependency Ratio*—Proportion of adults over age 65 relative to the population of "working ages"

$$= \frac{P_{65+}}{P_{15-64}} * 100$$

Age Pyramid

- Gives a detailed picture of the age-sex structure of a population
- Consists of bars representing age groups in ascending order from lowest to highest pyramided on one another
- Bars are generally by single or 5-year age groups

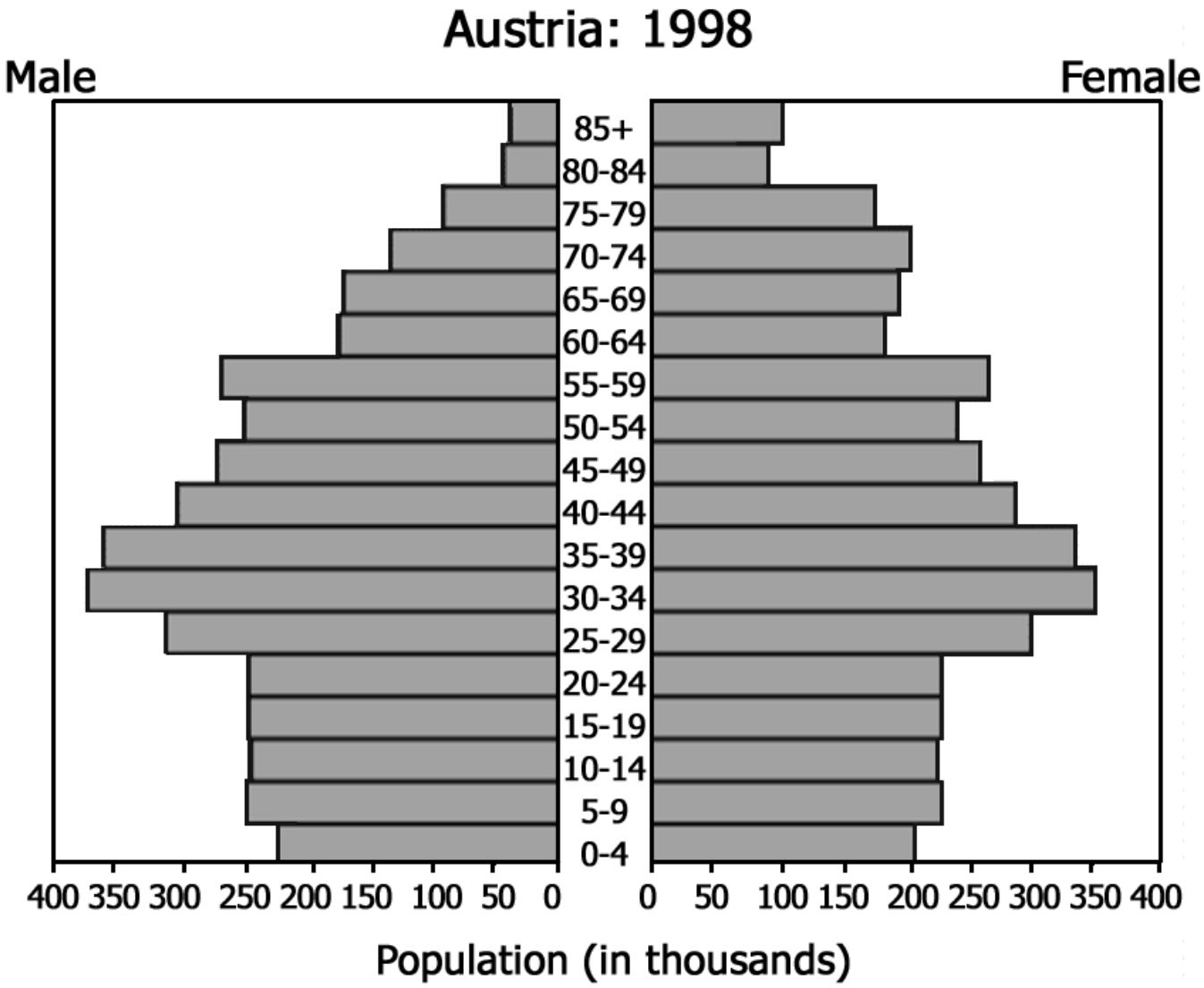
Continued

Age Pyramid

- Number (or percentage) of persons in an age group is indicated by length of its bar from central axis
- Males on left side; females on right side
- Pyramids with absolute numbers show differences in overall sizes of total populations and in number at each age
- Percent pyramids show relative differences in population size at each age-sex group

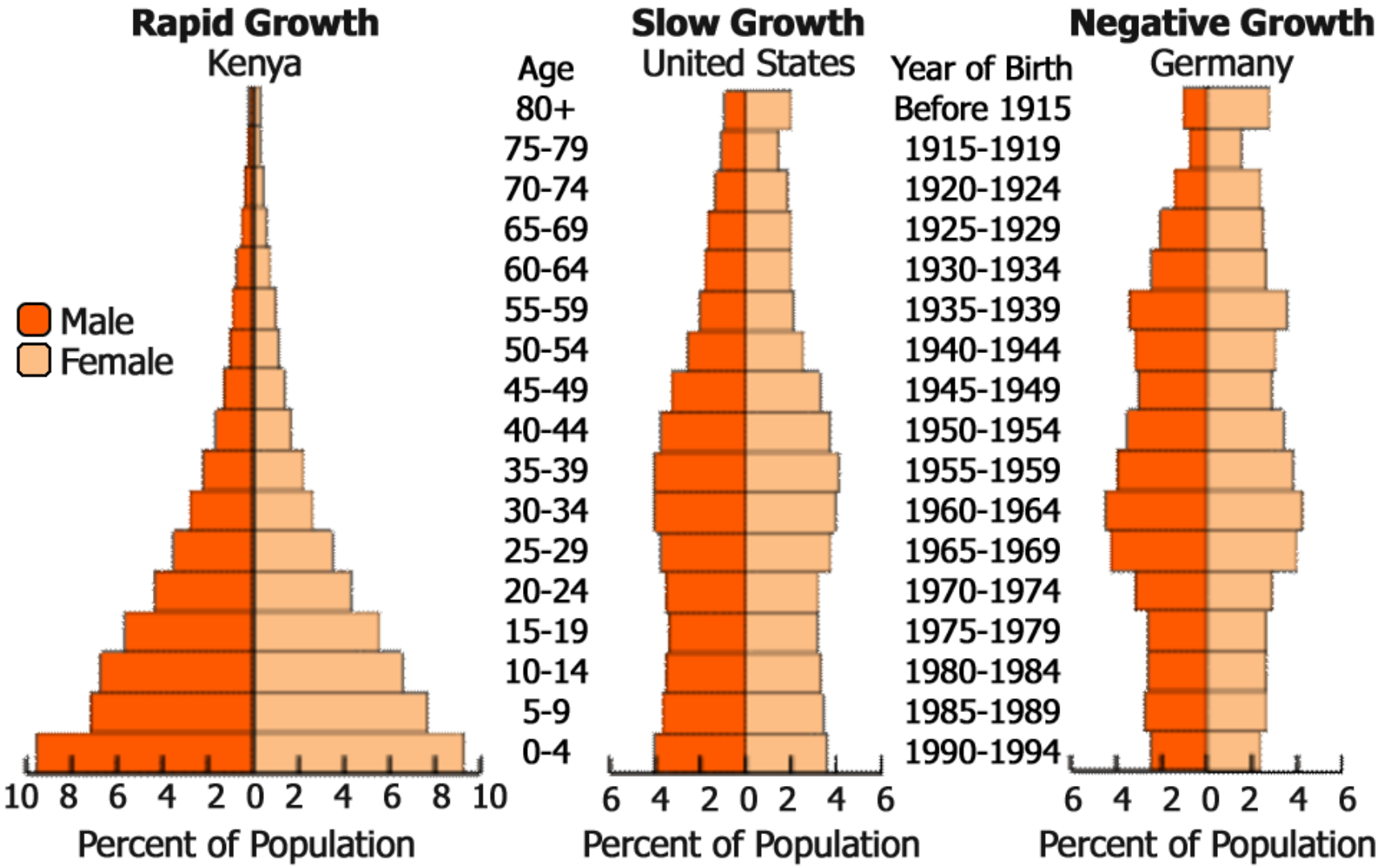
Continued

Age Pyramid



Source: U.S. Census Bureau, International Data Base

Three Patterns of Population Change



Source: United Nations. The Sex and Age Determinants of the World Population

Summary

- Several characteristics are commonly used by demographers to describe or compare populations; sex and age are two of the most often used besides age and sex
- These characteristics are often cross-tabulated by the economic characteristics of the individuals to measure labor force, economic status, occupation, etc. by sub-groups of the population

Continued

Summary

- Indices and techniques to obtain better information on the composition of the population have been developed and are widely used

JOHNS HOPKINS
UNIVERSITY

Section B

Race, Ethnicity, Nationality, and Marital Status

Definitions

- Race
- Ethnic group
- Nationality

Race

- There are no standard definitions of race and ethnicity

Ethnic Group

- *Ethnic Group*—Sub-Race
- Covers racial, national, cultural, or linguistic groups
- To varying extents, each such group will have a common descent, history, and habitat

Nationality

- *Nationality*—Country of present citizenship, or
- Country or other area of origin, sometimes an extinct country or a country that once had very different boundaries

Country of Birth

- Inclusion in censuses recommended by United Nations
- Ask for country of birth for foreign born
- Used to identify and describe immigrant minorities and throw light on ethnic composition

United States Practices

- *Native*—Persons born in the U.S., the Commonwealth of Puerto Rico, or a possession of the U.S.
- Also includes persons born in a foreign country or at sea who have at least one native American parent

Continued

United States Practices

- *Foreign Born*—Persons not classified as native i.e. persons who report a foreign country as place of birth

Citizenship

- *Citizenship*—Legal nationality
- *Alien*—Person included in census, register, etc… of a country but non-citizen thereof
  - An alien, however, is usually citizen of another country; that country is his/her legal nationality

Continued

Citizenship

- May be acquired by birth or naturalization
- Almost all people born in some countries are automatically citizens, or at least have that option; this is not so in some other countries

Continued

Citizenship

- Useful in connection with problems of legal status and civil rights of immigrants, and for studies of naturalization and assimilation of alien populations
- Limitations
  - Confusion in definition with country of birth
  - Also sensitive where minorities are insecure in their status

Language

- One of the most sensitive indices of ethnic origin because linguistic differences tend to persist until complete assimilation

Types of Language Data Recommended by UN

- Language spoken at home in early childhood or language of parents (mother tongue)
  - Is an indicator of ethnic origin

Continued

Types of Language Data Recommended by UN

- Language currently spoken, or most often spoken in present home (usual language)
  - Is an indicator of ethnic origin
  - Can also be used to see degree of assimilation and integration of foreign stock or of specific ethnic minorities

Continued

Types of Language Data Recommended by UN

- Ability to speak various languages (designated languages)
  - Indicates linguistic skills of population, both native and foreign born

Religion

- Item mentioned by the United Nations in its consideration of ethnic origin
- If race or country of origin is also known, religion can be used to make further distinctions among ethnic groups
- Useful personal characteristic because it is associated with a variety of differences in attitudes, statuses, and behavior

Continued

Religion

- Of great interest for its hypothesized relationship with fertility, mortality, nuptiality, and migration
- For the benefit of users of the data who may not be familiar with all the religions or sects within a country, as well as for purposes of international comparability, the classifications of the data should show each sect as a sub-category of the religion of which it forms a part

Marital Status

- Categories
  - Minimum list of groups recommended by the United Nations
    - Single (never married)
    - Married and not legally separated
    - Married but legally separated
    - Widowed and not remarried
    - Divorced and not remarried

Continued

Marital Status

- Note: Ever married = married + widowed + divorced + separated
- Additional options to take into account when appropriate:
  - *Consensual unions*—Common-law, extra-legal or de facto unions
  - Practices such as polygamy, concubinage and inherited widows
  - Same sex

Singulate Mean Age at Marriage (SMAM)

- *Singulate Mean Age at Marriage (SMAM)*— Estimate of the mean age at 1st marriage approximated by indirect method from data on marital status by age — no marriage records required
- Application of the formula $\sum_{x}(1-F(x))$ for the mean

Continued

Singulate Mean Age at Marriage (SMAM)

- *SMAM*—Mean age at marriage of women marrying before they reach 50
- More stable estimate than from retrospective data

Continued

Singulate Mean Age at Marriage (SMAM)

- Assumptions
  - The risk of marriage has remained constant (otherwise we estimate the mean for some average cohort)
  - The change in the proportion single from age x to age x+1 is a measure of the proportion of a birth cohort who married at that age, i.e. no woman dies between her 15th and 55th birthday

Continued

Singulate Mean Age at Marriage (SMAM)

- Steps for calculation—data in 5-year age groups

1. Sum the percentages single from age group 15–19 through age group 45–49 and multiply the sum by 5 (because of 5–year age group)
2. Add 1500 (i.e. years lived by the cohort before the 15th birthday)
3. Average the percentage single for age group 45–49 and 50–54

Continued

Singulate Mean Age at Marriage (SMAM)

4. Multiply the result of (3) by 50 (Equals number of years lived by those who did not marry before age 50)
5. Subtract result of (4) from (2) (Equals number of years lived by those who marry by age 50)
6. Subtract result of (3) from 100 (Equals average percentage married by age 50)
7. Divide result of (5) by result of (6) (Equals singulate mean age at marriage)

Continued

Singulate Mean Age at Marriage (SMAM)

$$SMAM = \frac{\left[\left(\sum_{i=1}^{i=7} S_i * 5\right) + 1{,}500\right] - \left[\left(\frac{S_7 + S_8}{2}\right) * 50\right]}{100 - \left(\frac{S_7 + S_8}{2}\right)}$$

Where S_i equal the proportion of women single in age group i

i=1 if age group = 15-19
i=2 if age group = 20-24
i=3 if age group = 25-29
i=4 if age group = 30-34

i=5 if age group = 35-39
i=6 if age group = 40-44
i=7 if age group = 45-49
i=8 if age group = 50-54

Order of Marriage

- The variables most often used in demographic analysis of marriage and of the marital factor in fertility are:
- *Order of Marriage*—Number of times married, or whether married more than once

Duration of Marriage

- Also of interest are age at and date of marriage
- Can be asked for first marriage or most recent marriage
- Date of first marriage is used to identify 1st marriage cohorts, i.e. persons who entered marriage for the 1st time in a specified period such as a calendar year

Importance for Fertility

- Age at first marriage is probably the most useful fact about women's marital history for the study of their fertility
- Data on spacing of children in relation to date of first marriage, or number of years since first marriage, are highly useful for analysis of fertility

Summary

- Several characteristics are commonly used by demographers to describe a population: race, nationality and ethnicity, and marital status are among the most often used
- These characteristics are often cross-tabulated by the economic characteristics of the individuals to measure labor force, economic status, occupation, etc by sub-groups of the population

Continued

Summary

- Indices and techniques to obtain better information on the composition of the population have been developed and are widely used

JOHNS HOPKINS
UNIVERSITY

Section C

Education and Economic Status

Education

- Important variable in accounting for demographic behavior

School Enrollment

- Importance
  - Measures of school enrollment usually relate to a point in time or a very short period of time

Continued

School Enrollment

- Data used to measure extent of the participation in the school systems of an area:
  - By persons of school age
  - The potential future participation of such persons
  - The relative participation of different segments of the population

Continued

School Enrollment

- *School Enrollment*—Enrollment at any regular educational institution, public or private, for systematic instruction at any level of education during a well-defined and recent time period (United Nations)

School Attendance

- *School Attendance*—While enrollment data are available at yearly intervals, attendance data may be collected daily in school classrooms

School Enrollment Rates

- Let E = Total number enrolled at all levels and ages covered

  P = Midyear population

  $\sum_{5}^{34} P$ = Midyear population in ages 5 to 34; these ages are arbitrary and can be adapted to country-specific situation

Crude Enrollment Rate

- *Crude Enrollment Rate*—Percentage of population enrolled in any regular educational institution

$$= \frac{E}{P} * 100$$

General Enrollment Rate

- *General Enrollment Rate*—Percentage of population 5 to 34 years old enrolled in any regular educational institution

$$= \frac{E}{\sum_{a=5}^{34} P_a} * 100$$

- (May use 5 to 29 years also)

Age-Specific Enrollment Rate

- *Age-Specific Enrollment Rate*—Percentage of persons age a enrolled in any regular educational institution

$$= \frac{E_a}{P_a} * 100$$

- Where E_a = Enrollment of persons of age a
  P_a = Midyear population at age a

Sex-Specific Enrollment Rate

- *Sex-Specific Enrollment Rate*—Percentage of persons of sex s enrolled in any regular educational institution

$$= \frac{E_s}{P_s} * 100$$

- Where E_s = Enrollment of persons of sex s
  P_s = Midyear population of persons of sex s

Age-Sex-Specific Enrollment Rate

- *Age-Sex-Specific Enrollment Rate*—Percentage of persons age a and sex s enrolled in any regular educational institution

$$= \frac{E_{a,s}}{P_{a,s}} * 100$$

- Where $E_{a,s}$ = Enrollment of persons of age a and sex s

  $P_{a,s}$ = Midyear population at age a and sex s

Level-Specific Enrollment Rate

- *Level-Specific Enrollment Rate*—Percentage of persons age a enrolled in a school at level ℓ

$$= \frac{E_{\ell}}{P_a} * 100$$

- Where E_{ℓ} = Enrollment at school level ℓ
  P_a = Midyear population in age group a corresponding to school level in numerator

Literacy

- *Literacy*—Ability of a person to both read and write, with understanding, a short statement on his everyday life (United Nations)

Continued

Literacy

- An illiterate person may not read and write at all, or may read and write only figures and his/her own name, or may only read and write a ritual phrase which has been memorized
- The language or languages in which a person can read and write are not a factor in determining literacy

Crude Literacy Rate

- *Crude Literacy Rate*—Percentage of the population who are literate

$$= \frac{\ell}{P} * 100$$

- Where ℓ = Number of literates in population covered

  P = Population covered

Age-Specific Literacy Rate

- *Age-Specific Literacy Rate*—Percentage of persons age a who are literate

$$= \frac{\ell_a}{P_a} * 100$$

- Where ℓ_a = Number of literates in age group a

  P_a = Population in age group a

Sex-Specific Literacy Rate

- *Sex-Specific Literacy Rate*—Percentage of persons of sex s who are literate

$$= \frac{\ell_s}{P_s} * 100$$

- Where ℓ_s = Number of literates of sex s

  P_s = Mid-year population of sex s

Age-Sex-Specific Literacy Rate

- *Age-Sex-Specific Literacy Rate*—Percentage of persons age a and sex s who are literate

$$= \frac{\ell_{a,s}}{P_{a,s}} * 100$$

- Where $\ell_{a,s}$ = Number of literates of sex s in age group a

  $P_{a,s}$ = Population of sex s in age group a

Economic Participation

- Importance:
  - Many demographic characteristics are cross-tabulated with the economic characteristics of the individuals
  - Although all persons consume goods and services, only a part of the entire population of a country is engaged in producing such goods and services

Labor Force

- *Labor Force*—Involves the carrying on of an activity from which the person derives, or attempts to derive, pay or profit
- In censuses and surveys, the entire population can be subdivided into the employed and the unemployed

Employed

- *Employed*—Persons who worked for pay or profit during the time-reference period or who had a job or business that period but were absent from it because of vacation, illness, etc.

Unemployed

- *Unemployed*—Persons who, during the reference period, were not working but who were seeking work for pay or profit, including those who never worked before

Not in Labor Force

- *Not in Labor Force*—All persons who are neither employed nor unemployed
- Note:
  - The difference between the amount of work performed by persons in employment and the amount of work they would normally be able or willing to perform is a measure of underemployment

Status in Labor Force

- *Status in Labor Force*—Refers to the status of a person in the labor force with respect to his/her employment, that is, whether s/he is (or was, if unemployed) an employer, own-account worker, employee, unpaid family worker, or a member of a producers' cooperative (United Nations)

Continued

Status in Labor Force

- *Employer*—A person who operates his/her own economic enterprise or engages independently in a profession or trade and hires one or more employees
- *Own-Account Worker*—A person who operates his/her own economic enterprise or engages independently in a profession or trade and hires no employees

Continued

Status in Labor Force

- *Employee*—A person who works for a public or private employer and receives remuneration in wages, salary, commission, tips, piece-rates, or pay in kind

Continued

Status in Labor Force

- *Unpaid Family Worker*—A person who works a specified minimum amount of time (at least 1/3 of normal working hours), without pay, in an economic enterprise operated by a related person living in the same household

Continued

Status in Labor Force

- *Member of Producers' Co-operative*—A person who is an active member of a producer's co-operative, regardless of the industry in which it is established
- *Persons Not Classified by Status*—Experienced workers with status unknown or inadequately described and unemployed persons not previously employed

Continued

Status in Labor Force

- Note:
  - Categories can be combined or further subdivided to better reflect a country's reality

Problems of Labor Force Concept

- Formal vs. informal sector
  - Barter
- Mismatch of skills and work
  - PhDs driving taxis
- Persons move in and out of labor force
- Persons working part-time who want full-time employment

Occupation

- Refers to the kind of work done during the time-reference period established for data on economic characteristics by the person employed (or performed previously by the unemployed) irrespective of the industry or the status in which the person should be classified (United Nations)
  - Examples: Teacher, sales clerk, farmer, car washer, etc.

Personal Income

- *Personal Income*—One of the best measures of economic well-being; vies with educational attainment and occupation as a measure of socioeconomic status
- No international standard definition and sometimes difficult to collect
- In principle, income from all sources should be counted, both cash income and income in kind
- Reference period = usually one year

Household Income

- *Household Income*—Income data collected for each person in the household may be added to produce total for the household

Wealth

- *Wealth*—All goods and resources having value in terms of exchange or use
  - Land holdings, stocks, and bonds
  - Jewelry and other possessions, livestock

Measures of Labor Force Participation

- *Crude Activity Rate*—Percentage of total population in the labor force
  - Equals crude labor force participation rate
  - Is greatly influenced by the age composition of the population

Crude Activity Rate

$$= \frac{L}{P} * 100$$

- Where L = Number of persons in the labor force
  P = Total population

General Activity Rate

- *General Activity Rate*—Activity rate of persons of working age

$$= \frac{L_{14+}}{P_{14+}} * 100$$

- Where L_{14+} = Number of persons 14 and older in the labor force

  P_{14+} = Population aged 14 and older

(Ages are set arbitrarily and can be adapted)

Sex-Specific Activity Rate

- Both of the previous rates can be calculated by sex
- Calculated for two different reasons
  1. Levels are usually higher for men than for women

Continued

Sex-Specific Activity Rate

2. Variations in definitions and operational procedures that most affect measures of economic activity have their greatest impact on figures for women because women's involvement in the labor force is often marginal and intermittent
   - Men are less subject to temporary or spurious variations in employment
   - Women are the most dynamic part of the labor force

Age-Sex Specific Activity Rate

- Labor force participation rates
  - Far more widely used measure of economic activity
  - Basic rates studied and projected in analyses of population in the labor force

Continued

Age-Sex Specific Activity Rate

$$= \frac{A_{a,s}}{P_{a,s}} * 100$$

- Where $A_{a,s}$ = Active population of age group a and sex s

  $P_{a,s}$ = Population of age group a and sex s

Summary

- Several characteristics are commonly used by demographers to describe a population. Education and economic participation are two characteristics often used
- These characteristics are often cross-tabulated by other demographic and economic characteristics to measure different aspects of sub-groups of the population

Continued

Summary

- Indices and techniques to obtain better information on the composition of the population as well as on the economic participation of the individuals have been developed and are widely used