

Demographic Analysis

Reporting Problems
by Age and Sex

Evaluations based on one census
(5-year and single-year age groups),
and two or more censuses

Population Age-Sex Structure: Reporting Issues

The prior lesson introduced broad concepts and issues related to age-sex structure (the “population pyramid”). In this lesson, we explore:

- Common problems in age and sex reporting in censuses and surveys
- Spreadsheets to assist in identifying possible problems, **evaluating the quality** of such reporting, and possible methods of adjusting for flaws
 - Evaluations from a single census
 - Heaping by five-year age groups
 - Digit preference
 - Intercensal evaluations

Population Age-Sex Structure: Introduction

Motivation:

Given the importance of the age structure with respect to social and economic characteristics, it is imperative that the information on the population age and sex structure be as accurate as possible.

Yet questions are often raised about the accuracy of counts by age and sex.

How can we assess that accuracy?

Checking Census Data for Consistency

The comparative method:

The most basic tool for evaluation in the demographer's toolkit is to **compare** population data by age and sex related data, models, or assumptions. Here are three such methods (PAS):

- Comparisons age-sex distributions from one census to assumptions about the smoothness of populations and sex ratios (AGESEX.xls)
- Evaluation of digit preference across the life span in one census (SINGAGE.xls)
- Comparisons of counts of the same birth cohort in two or more censuses (GRPOP-YB.xls)

Categories of Errors in Reporting by Age

Reporting errors by age fall into two general categories:

- **Age misreporting** – when those at a particular age have been incorrectly reported as belonging to a different age group
- **Differences in reporting completeness by age** – when those reported at an age group are higher or lower than the actual number (and the difference does not simply reflect misreporting to/from other ages).

It is often difficult to distinguish between the above;
PAS AGESEX.xls gets us started by calculating age ratios
and sex ratios ...

Possible Indices of Misreporting: Age Ratios

One possible method to detect age misreporting is to use age ratios for 5-year age groups. An **age ratio** is defined as the ratio of those at a particular age group and the average of those at adjacent age groups. For 5-year age groups:

$${}_5AR_x = 100 \frac{{}_5P_x}{1/2 ({}_5P_{x-5} + {}_5P_{x+5})}$$

Where:

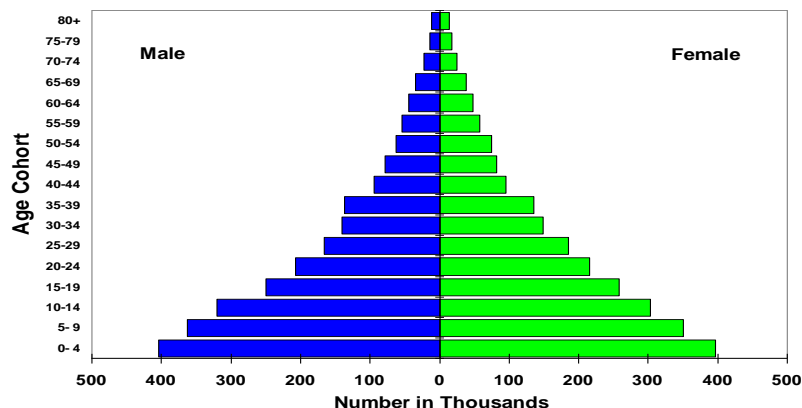
${}_5AR_x$ represents the age ratio for ages x to $x+4$;
 ${}_5P_x$ represents the population at ages x to $x+4$.

Possible Indices of Misreporting: Age Ratios

If fertility has not fluctuated much in the past and international migration has not been significant, the “walls” of the population pyramid should be fairly smooth. In such case, age ratios should be fairly close to 100.

Smooth Walls

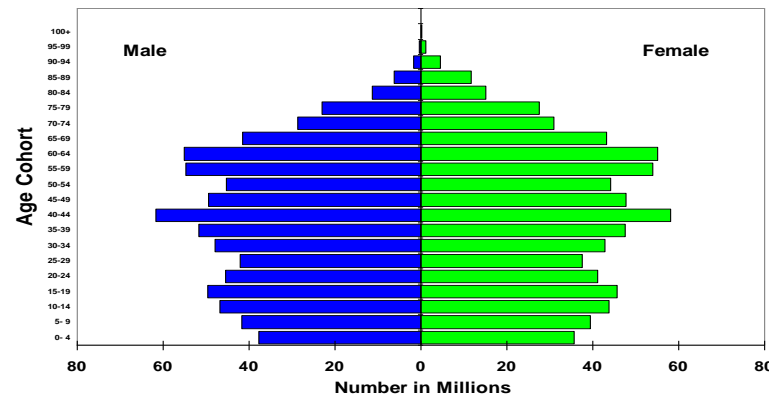
Population of Laos, 1995



Source: International Database, U.S. Census Bureau (<http://www.census.gov/ipc/www/idbnew.html>).

Fluctuating Walls

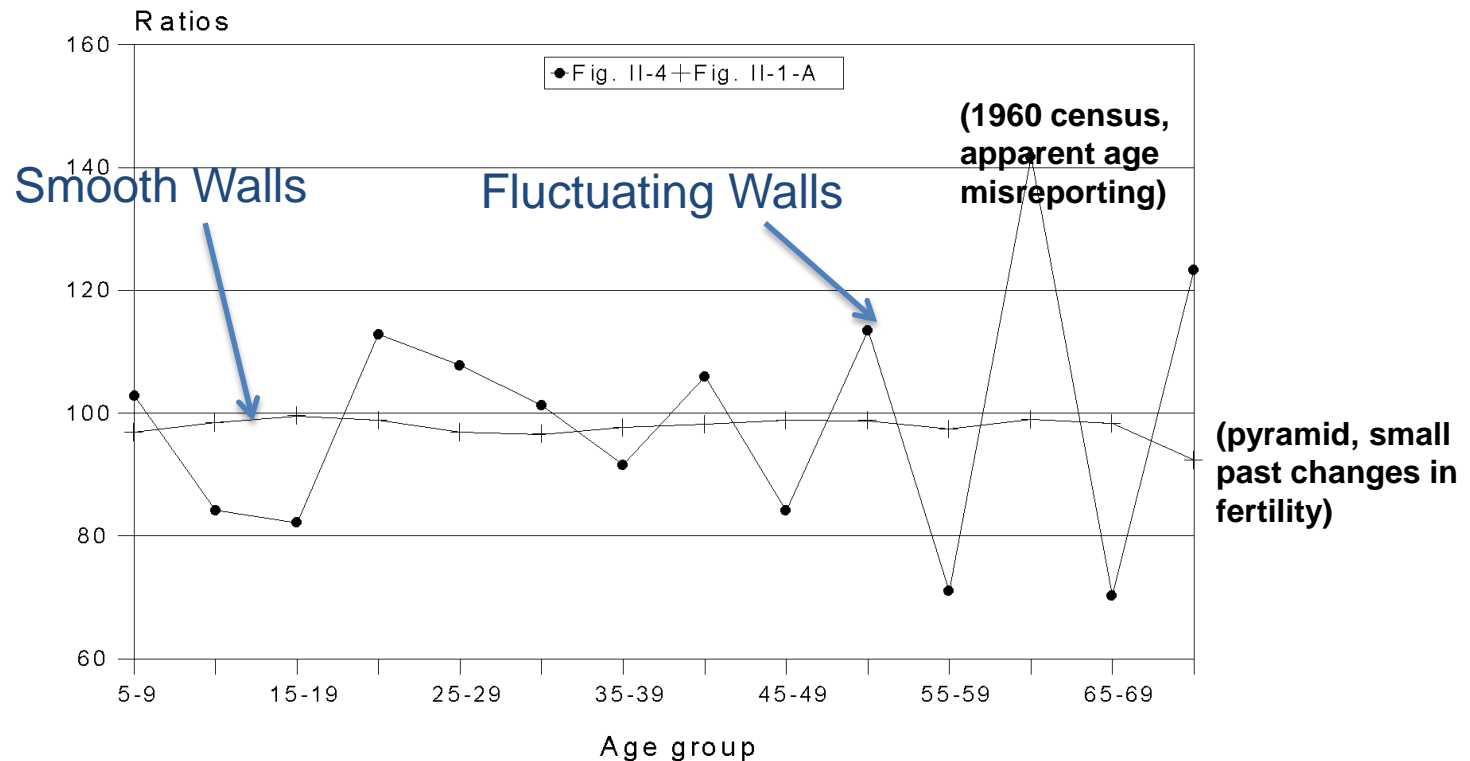
Population of China, 2030



Source: International Database, U.S. Census Bureau (<http://www.census.gov/ipc/www/idbnew.html>).

Possible Indices of Misreporting: Age Ratios

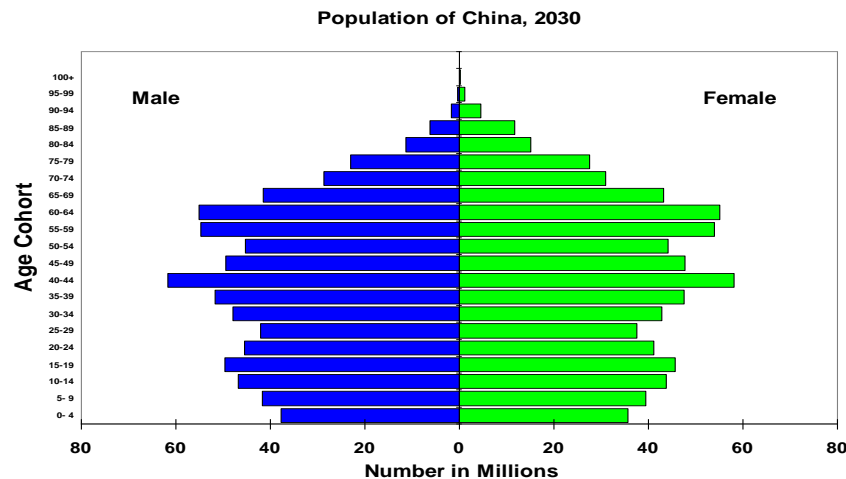
Example of how age ratios might vary for the two pyramids on the preceding slide



Note: Each ratio represents the population in a given age group divided by the average population of the two adjacent age groups, times 100.

Possible Indices of Misreporting: Age Ratios

Caution – Fluctuating age ratios across the age spectrum **DO NOT** necessarily indicate reporting problems. For instance, if fertility actually has fluctuated over time, the shifting age ratios may reflect actual historical realities (as we saw earlier for China and other countries).



Source: International Database, U.S. Census Bureau (<http://www.census.gov/ipc/www/idbnew.html>).

Possible Indices of Misreporting: Sex Ratios

Another method to detect possible age misreporting looks at sex ratios across age groups. At any 5-year age group:

$${}_5SR_x = 100 \frac{{}_5MP_x}{{}_5FP_x}$$

Where:

${}_5SR_x$ represents the sex ratio at ages x to $x+4$; and

${}_5MP_x$ and ${}_5FP_x$ represent the male and female populations, respectively, at ages x to $x+4$.

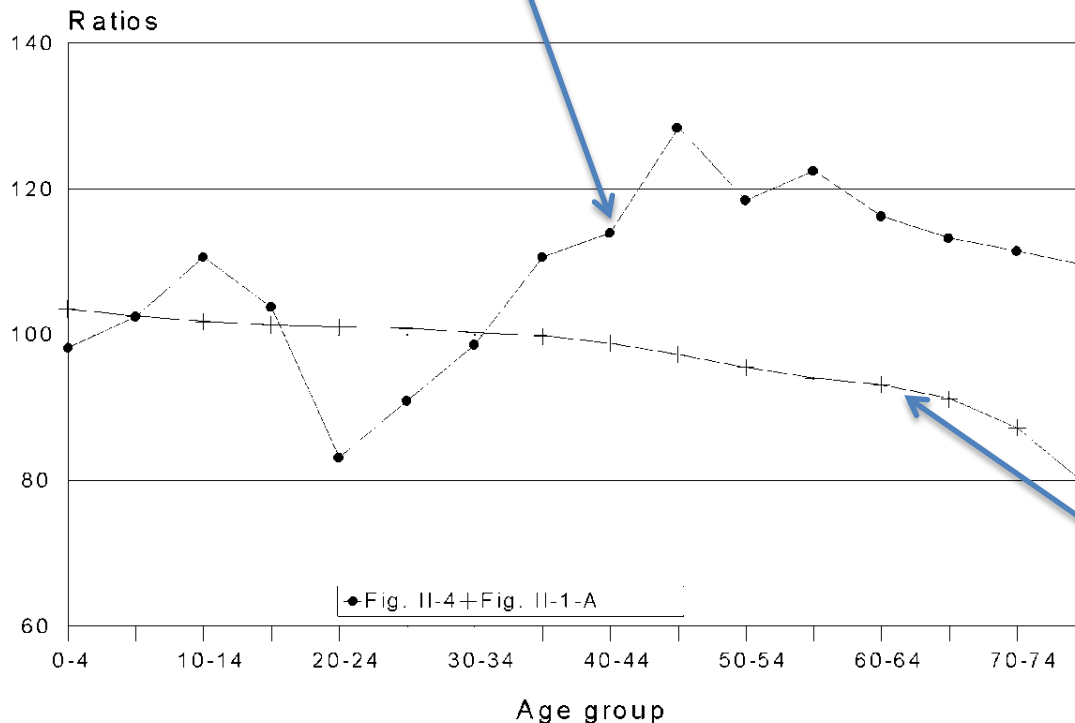
Possible Indices of Misreporting: Sex Ratios

What to look for in sex ratios across age groups -

- As in the case of age ratios, the larger the abrupt departure of this ratio from values close to 100, the larger the possibility of errors in the data.
- The standard comparative pattern is fairly clear:
 - In all societies, the expected sex ratio at birth is 103-106 males per 100 females.
 - As the population ages, sex ratios gradually tend to become more feminine because male mortality tends to exceed that of females.

Possible Indices of Misreporting: Sex Ratios

Example of reported sex ratios by 5-year age groups



“Model” pattern of sex ratios by age

Note: Each ratio refers to the number of males per 100 females in each age group.

Age-Ratio and Sex-Ratio Summary Scores

Summary scores for age and sex ratios can be calculated by adding up the absolute values of differences across the age spectrum as follows:

- The index of **sex-ratio score** (SRS) is defined as:
The mean difference in the absolute value of sex ratios for successive age groups.
- The index of **age-ratio score** (ARS) is defined as:
The mean absolute deviation of the age ratios from 100 percent.

The Composite Age-Sex Accuracy Index

In the early 1950's, the United Nations suggested a joint accuracy index to summarize the values of the age and sex ratios.

Based on empirical relationships between the sex-ratio scores and the age-ratio scores, the following index was defined as the **joint score** (JS) or age-sex accuracy index (three times the sex ratio score plus the age-ratio score for males and the age-ratio score for females):

$$\text{JS} = 3 \times \text{SRS} + \text{ARSM} + \text{ARSF}$$

The Composite Age-Sex Accuracy Index

Based on empirical analysis of the age and sex declaration in censuses from different developed and developing countries, the United Nations suggested that the age and sex structure of a population will be:

- (a) *accurate* if the joint score index is under 20,
- (b) *inaccurate* if the joint score index is between 20 and 40,
and
- (c) *highly inaccurate* if the index value is over 40.

These indices are useful mainly in international or historical comparative analyses. These measures are computed in the PAS workbook **AGESEX.xls**.

Indices of Misreporting: Digit Preference

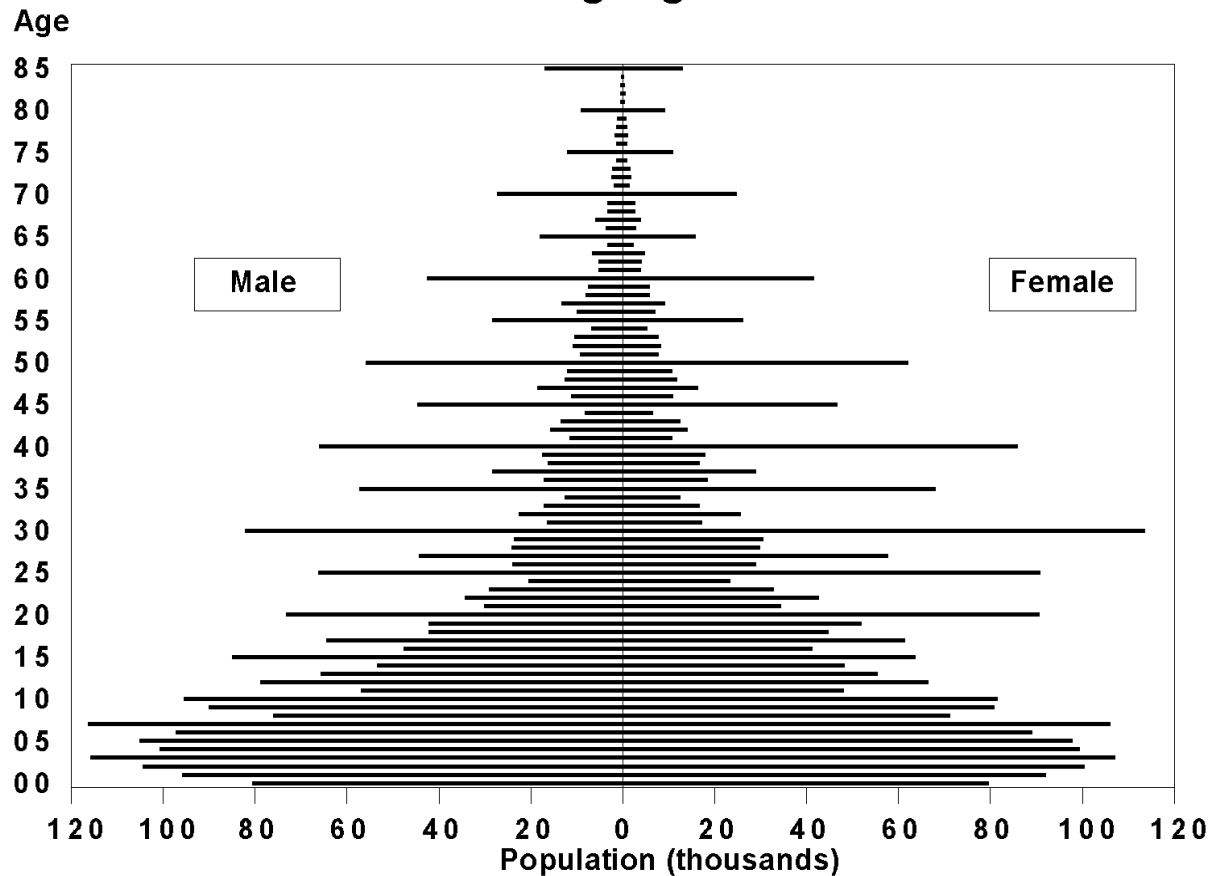
In addition to possible reporting anomalies across 5-year age groups, examination of reported data by single years often reveals digit preference.

What is Digit Preference? - when respondents are more likely to report an age ending in a particular number (e.g., a number ending in a zero or five - 10, 15, 20, 25).

Digit preference is generally easy to diagnose by plotting a reported age-sex distribution.

Indices of Misreporting: Digit Preference

Figure II-5. Single-Year Age Distribution Indicating Digit Preference



Note: Figure shows an excessive number of persons declaring ages ending in digits 0 and 5.

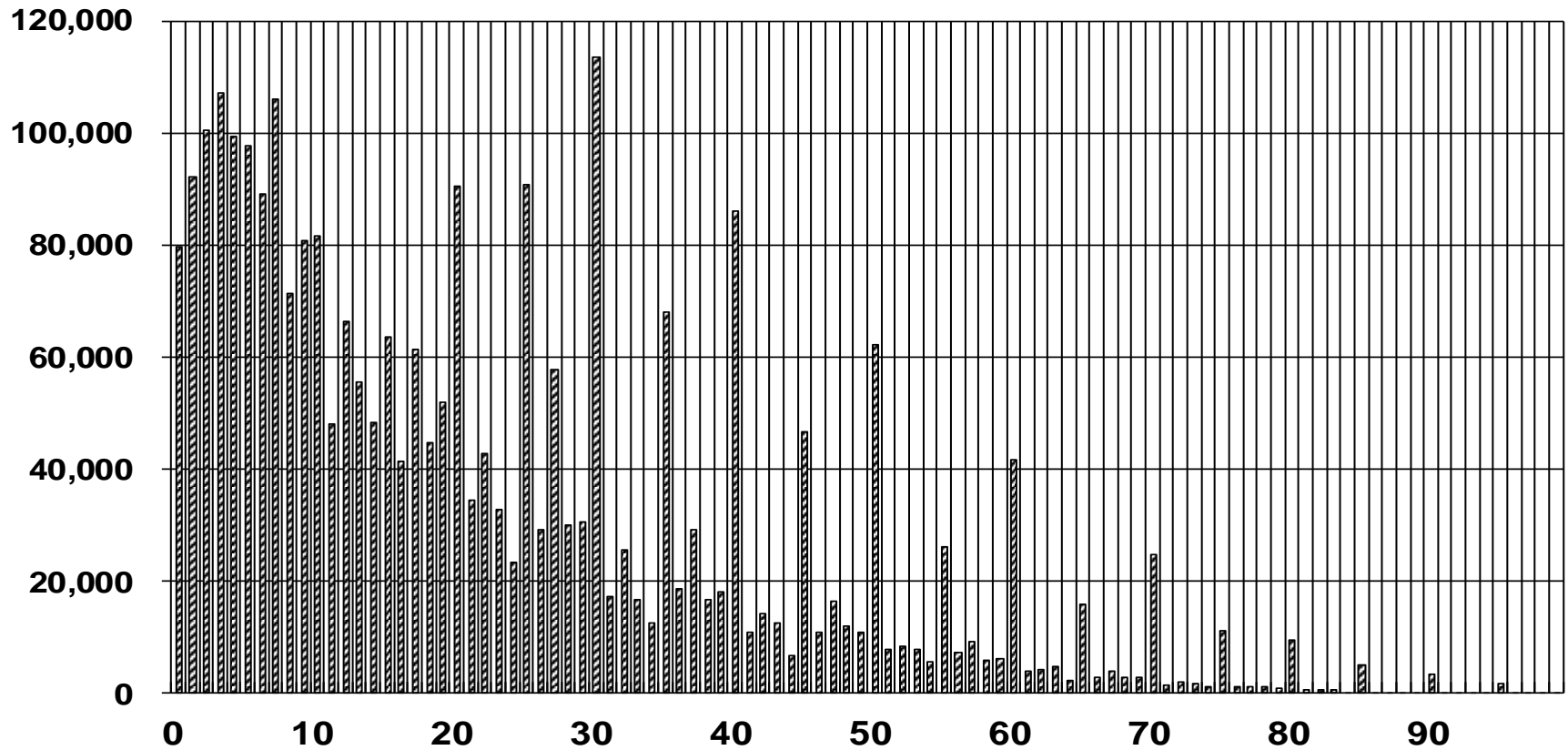
Indices of Misreporting: Digit Preference

There are several indices for detecting digit preference: Myers (1940), Whipple (U.S. Bureau of the Census, 1971), Bachi (1951, 1953), Carrier (1959), and Ramachandran (U.S. Bureau of the Census, 1971).

These indices show preference for digits ending in a certain number as well as summary measures of the overall extent of misreporting. The PAS workbook ***SINGAGE.xls*** includes the Whipple, Myers, and Bachi indices.

Formulas for Myers and Bachi indices differ slightly – for Bachi, the summary measure shows an estimate of the percent of people whose reported ages are incorrect.

Indices of Misreporting: Digit Preference – SINGAGE.xls

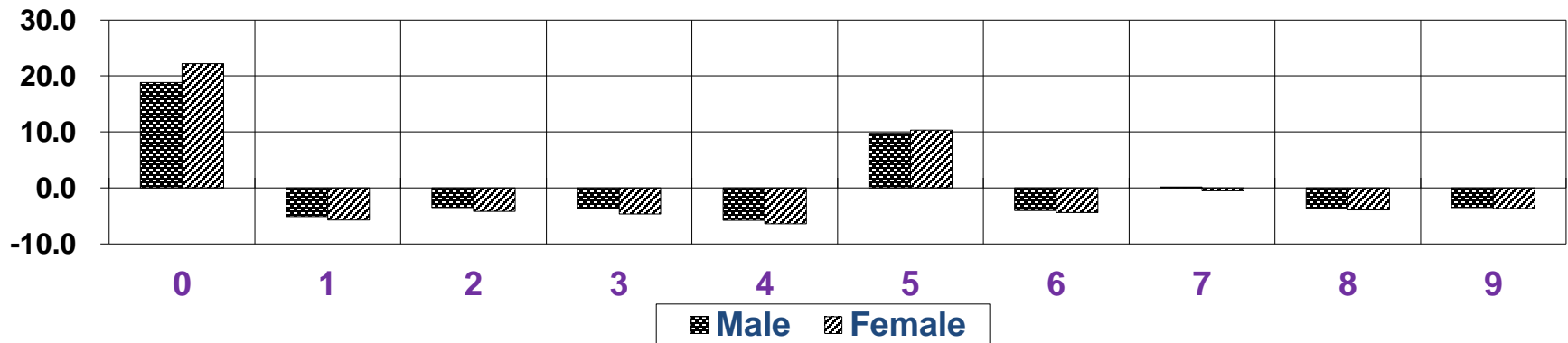


Indices of Misreporting: Digit Preference – SINGAGE.xls

Bachi overall Index (male= 29, female= 33) – a minimum of about 30% of ages reported are incorrect

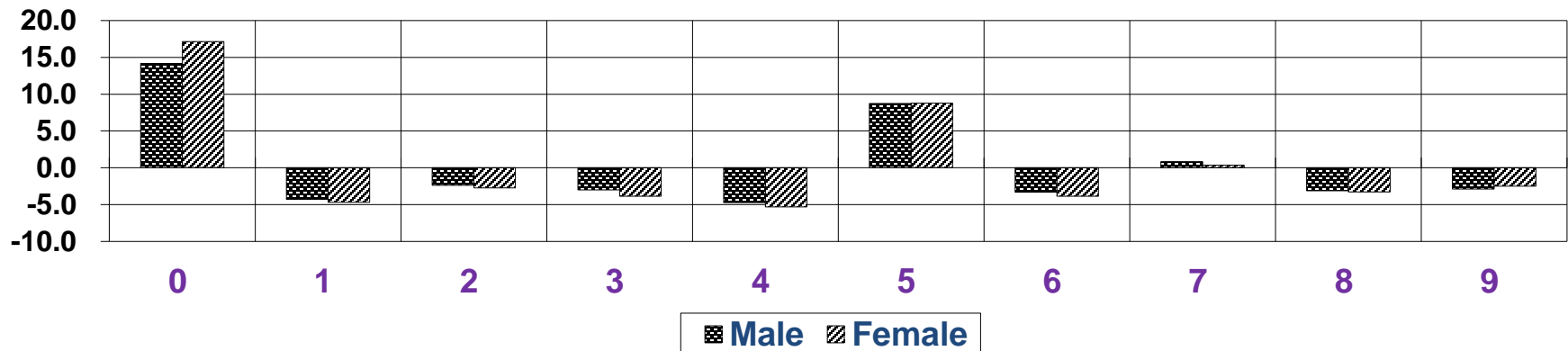
COUNTRY: YEAR

5. *Bachi* Preference by Digit



COUNTRY: YEAR

4. *Myers* Preference by Digit



Indices of Misreporting: Digit Preference – SINGAGE.xls

B. Measures of Digit Preference

Method and terminal digit	Male	Female	Both sexes
WHIPPLE METHOD (23-62)			
Index	2.34	2.55	2.45
MYERS METHOD			
Index *	47.5	52.5	50.1
0	14.2	17.1	15.7
1	-4.3	-4.7	-4.5
2	-2.4	-2.7	-2.6
3	-3.0	-3.9	-3.5
4	-4.7	-5.3	-5.0
5	8.7	8.8	8.8
6	-3.3	-3.9	-3.6
7	0.8	0.4	0.6
8	-3.1	-3.3	-3.2
9	-2.9	-2.5	-2.7
BACHI METHOD			
Index **	29.0	33.0	31.0
0	18.8	22.2	20.6
1	-5.1	-5.8	-5.4
2	-3.5	-4.2	-3.9
3	-3.7	-4.6	-4.2

Evaluation of Reporting in Two or More Censuses

Intercensal Cohort Comparisons

A key method for evaluating census quality is to compare the count of each birth cohort to the count of that cohort from a prior census when they were younger. Example: 55-59 in 2010 vs. 45-49 in 2000.

If counts were perfectly complete in both censuses, cohorts counted at older ages would normally be smaller. Why?

Checking Census Counts by Birth Cohort

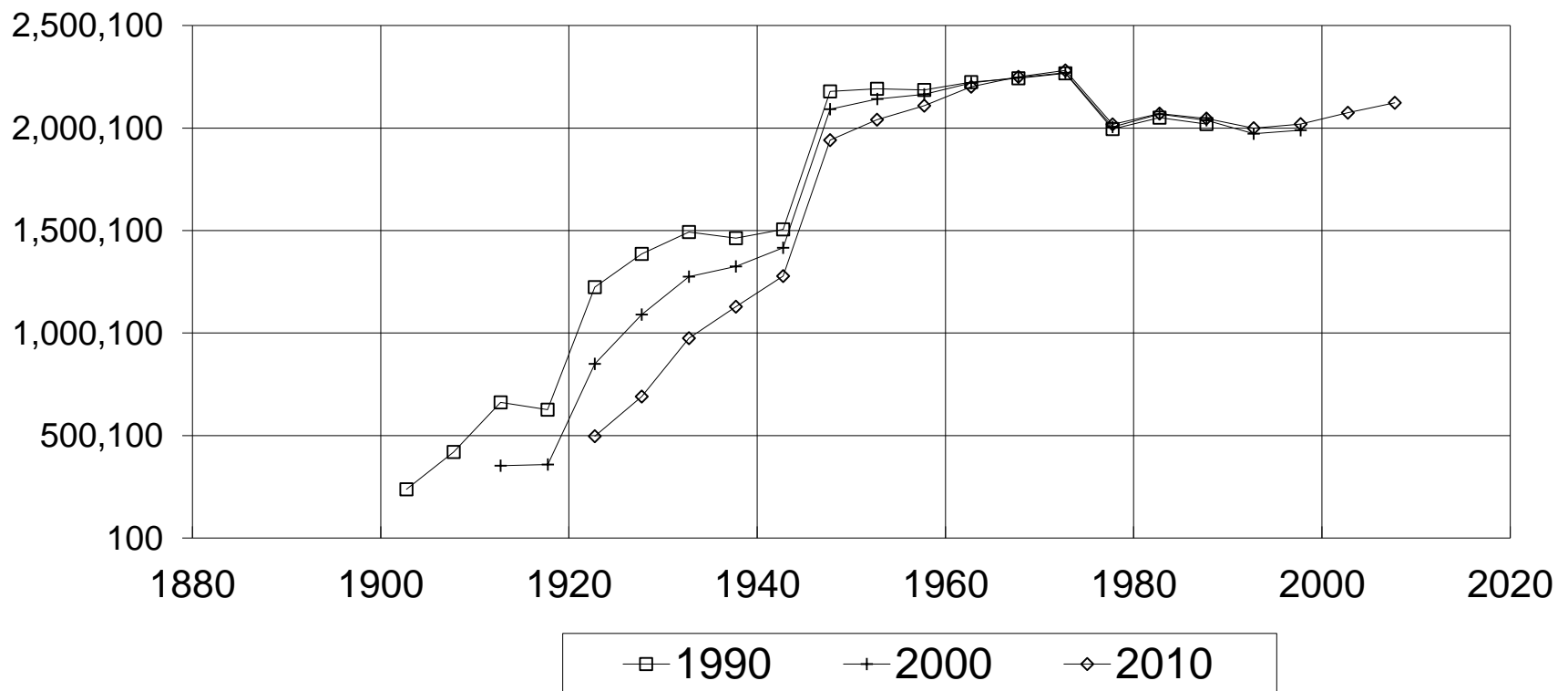
PAS GRPOP-YB.xls performs this simple comparison by plotting populations counted in successive censuses **by year of birth in the X axis**. Benefits of doing so:

- Counts of the same birth cohort across censuses can be readily compared by a vertical line
- A “best case” scenario (that would indicate consistency - and *possible* accuracy - of counts):
 - No crossovers
 - Sharp drop off in size of cohorts at oldest ages
 - Decline in count of the same birth cohort as it ages across successive censuses

A Better Case Scenario

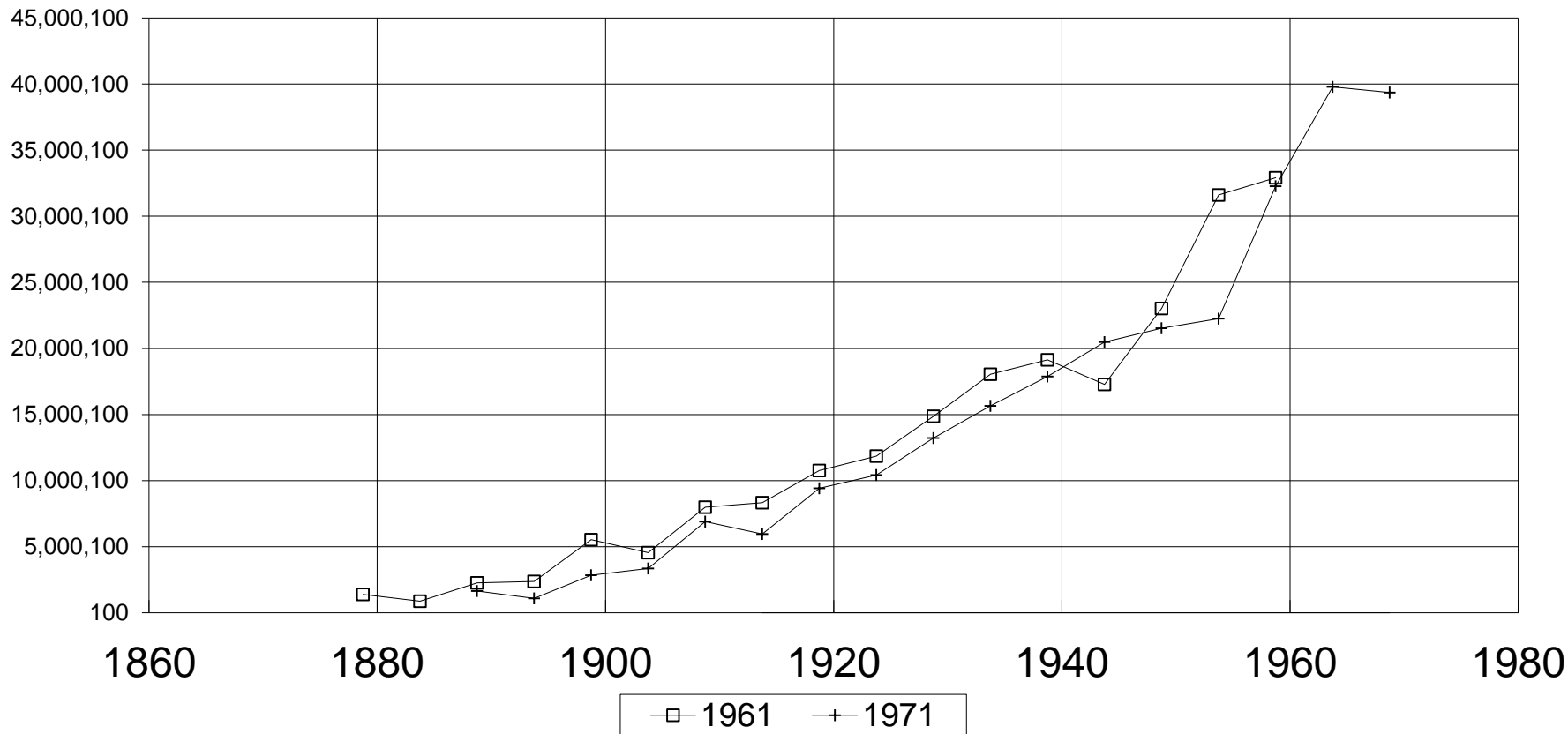
Pretty Good Intercensal Fit in the **Land of Consistencia** –
Males Counted in 1990, 2000, and 2010 Censuses (April 1)

1. Male Population, 5-Year Cohorts



A Case Indicating Reporting Issues

Intercensal Discrepancies and Crossovers Among Females Counted in Censuses of 1961 and 1971



Checking Census Data for Consistency – GRPOP-YB.xls

Why might deviations from the best case scenario occur? Possible explanations

- Genuine reporting problems in one or more censuses
- Actual historical demographic changes in migration, mortality, or fertility ...

How do we distinguish between the above explanations? ...

Census Errors vs. Actual Demographic Change

To distinguish reporting problems from actual demographic change, here are some common patterns to look for:

Observation

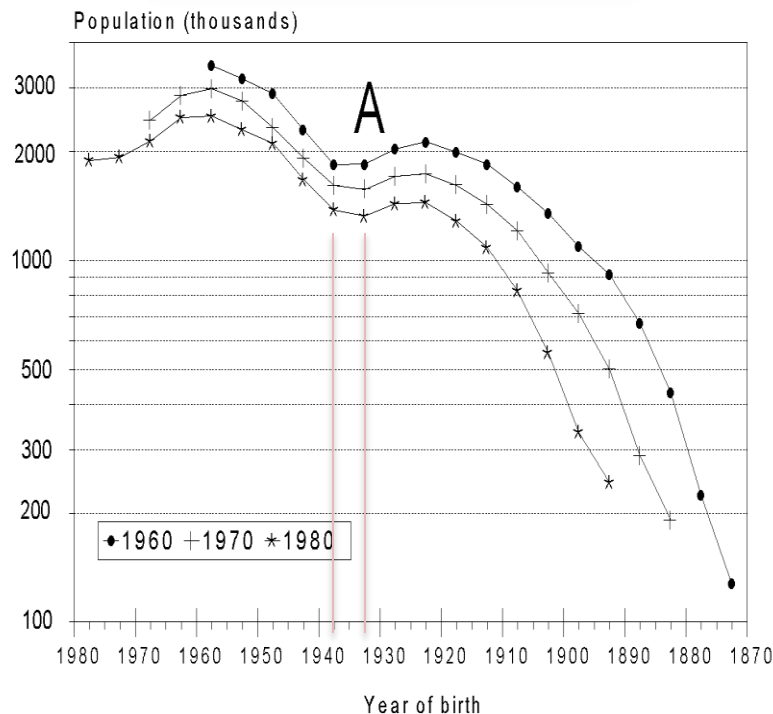
1. Consistent dips/bumps in cohort changes
2. Parallel shifts by age
3. Crossover – e.g., new counts of those who did not exist in the past

Possible Explanation

- May indicate baby boom or true shift
- May indicate age-specific reporting problems
- May indicate underreporting in the prior census (or overreporting in the latter census)

1. Consistent Dips/Bumps in Cohort Change

Maidupland



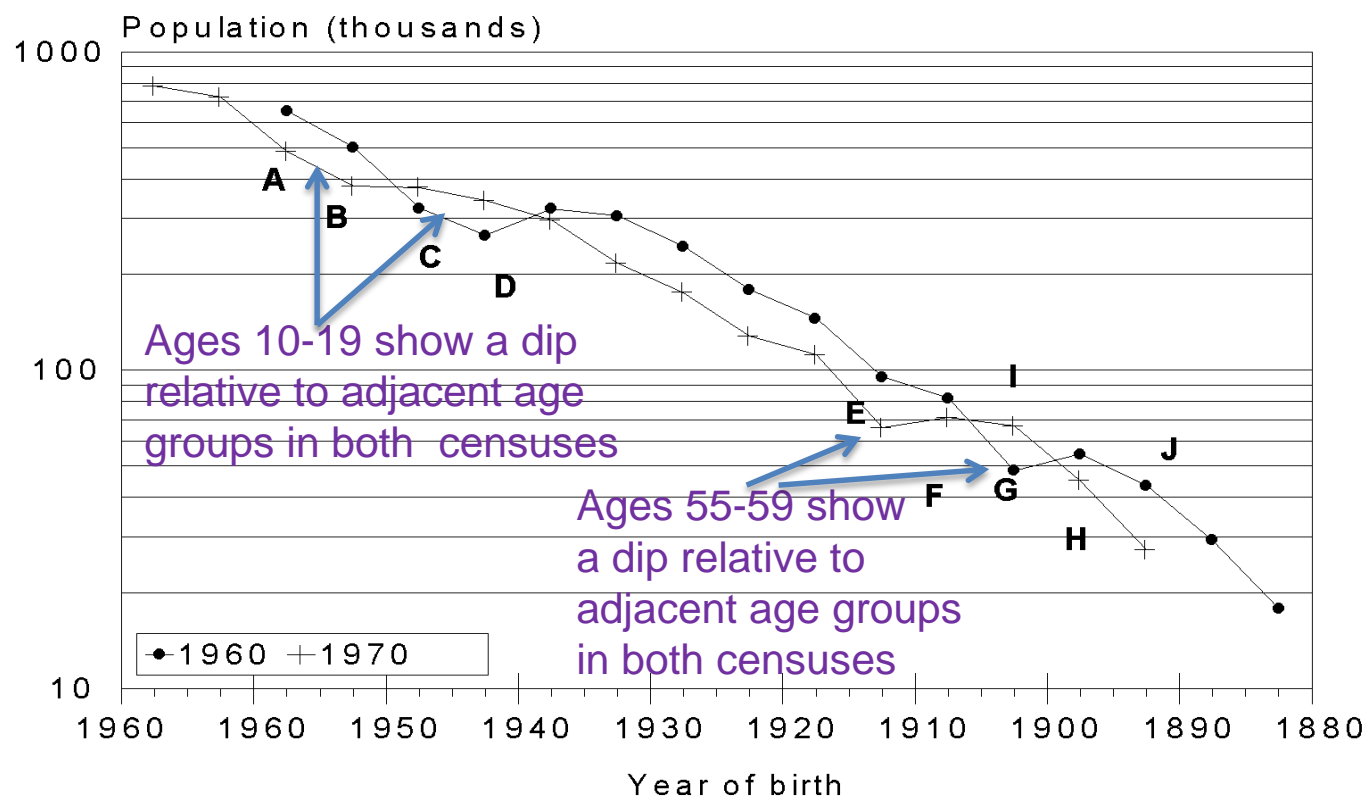
Successive census counts all suggest a dip in the cohort born in the 1930s. This cohort (A) was aged:

20-29 in 1960 census,
30-39 in 1970 census,
40-49 in 1980 census.

The consistency of the dip (counted at different age groups) likely suggests that this cohort was the product of a *temporary baby bust*.

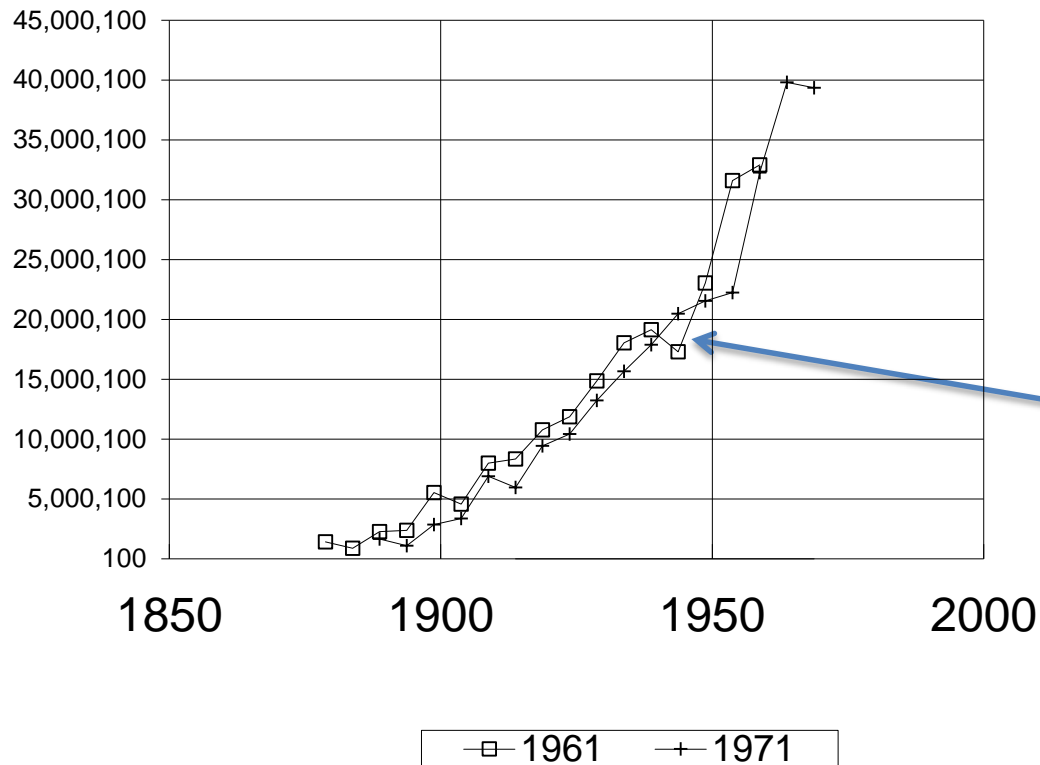
2. Parallel Shifts – Age-Specific Reporting Problems?

Figure II-4. Census Population by Year of Birth Indicating Age Misreporting



3. Crossovers – and What Causes Them?

Female Counts in 1961 and 1971 Censuses



In this example, more females were counted at ages 25-29 in the 1971 census than at 15-19 in the 1961 census. Why?

*Spontaneous generation?
In-migration?
Undercount in 1961?
Double counted in 1971?
Age misreporting?*

Exercises

- Applications of PAS workbooks:
 - AGESEX.xls – age ratios and sex ratios across age groups
 - SINGAGE.xls – digit preference
 - GRPOP-YB.xls – intercensal counts by birth cohort