# Data Evaluation with Consecutive Censuses: Adult Mortality and Census Coverage

## United Nations Statistics Division

United Nations Workshop on Census Data Evaluation for English Speaking African Countries
Kampala, Uganda
12 – 16 November 2012

# Outline: Data Evaluation with Consecutive Censuses

1. The Population Balancing Equation
2. Adult Mortality (Death Distribution Methods)
    a) General Growth Balance (GGB) method
3. Intercensal Cohort Survival Rates
4. Cohort Component Method

# Population Balancing Equation

## Census coverage

United Nations Workshop on Census Data Evaluation for English Speaking African Countries
Kampala, Uganda
12 – 16 November 2012

# Population balancing equation

If a country

- Has a relatively complete system of vital registration
- Has a fairly reliable estimate of the degree of under-registration

- Information on the number of intercensal births, deaths and net international migrants can be used in conjunction with the results of a previous census to evaluate the coverage of a subsequent or current census.

# Population balancing equation

$$P_1 = P_0 + B - D + M$$

Where:

$P_1$=the population enumerated in the census being evaluated

$P_0$= The population enumerated in a previous census

B = the number of births in the period between the two censuses

D= the number of deaths in the period between two censuses

M= the number of net international migrants in the period

M = I (Immigrants) – E (Emigrants)

# Population balancing equation

- The population balancing equation is the most fundamental equation in demographic analysis and is also used to estimate population growth. It is based on the logic that:
    - The population of a country can increase or decrease between any two points in time only as a result of births, deaths and movement of population across national boundaries
- Births and immigration add to the population
- Deaths and emigration reduce the population

# Population balancing equation

- For census evaluation purposes, there is a residual (*e*) needed to make the equation balance exactly
- "*e*" in the equation is referred to as the "error of closure" and represents the balance of errors in the data on births, deaths, net migration, and the coverage of the two censuses:

$$P_1 = P_0 + B - D + M + e$$

- *If a negative residual quantity e,* $P_1$ *is under-enumerated relative to* $P_0$
- *If a positive residual is required to balance the equation,* $P_1$ *is over-enumerated relative to* $P_0$

# Population balancing equation – Data required

- ☐ The population enumerated in two consecutive censuses
  - ☐ $P_1$: the census under evaluation
  - ☐ $P_0$: previous census

- ☐ The number of births, deaths and net international migration (immigrants-emigrants) during the intercensal period, adjusted for under-registration (to the extent possible)

# Population balancing equation – Computational procedure

1. Compile registered numbers of intercensal births, deaths and migrants
   - ❑ Vital registration system
   - ❑ Immigration record system (residence permit, border records, etc.)
     - ➢ Adjustment based on under-coverage of these systems including indirect estimates

2. Calculation of the “expected” census population (E(P1))
   - $E(P1) = P0 + B - D + M$
3. Calculation of the residual error or error of closure
   - $e = P_1 - E(P_1)$

# Population balancing equation – Interpretation of "e"

- If $P_0$ has been adjusted for net coverage error, the estimated residual error (*e*) will represent an estimate of net coverage error in $P_1$
  - If "*e*" is positive, $P_1$ is overenumerated
  - If "*e*" is negative, $P_1$ is underenumerated

- If $P_0$ is not adjusted, "e" will represent an estimate of the relative level of net coverage error in $P_1$ in comparison with $P_0$

# Population balancing equation – Example Sri Lanka, 1971 and 1981 census (1)

**For an unadjusted census**:

$E(P_1)$ = $P_0$ (unadjusted) + $B_{adj} - D_{adj} + M_{adj}$

= 12,689,897 + 3,716,878 – 1,002,108 + (-446,911)

= 14,957,756

$e = P_1 - E(P_1)$

=14,848,364 - 14,957,756

=-109,392 *0.7% of $E(P_1)$*

*$P_1$ is underenumerated relative to $P_0$*

Source: U.S. Census Bureau, 1985. *Evaluating Censuses of Population and Housing*

# Population balancing equation – Example Sri Lanka, 1971 and 1981 census (2)

**For an adjusted census count**:

$E(P_1)$ = P0 (adjusted) +$B_{adj}$ – $D_{adj}$ + $M_{adj}$

= 12,849,796 + 3,716,878 – 1,002,108 + (-446,911)

= 15,117,655

*e* = $P_1 - E(P_1)$ = 14,848,364 – 15,117,655

= -269,291 *1.8% of $E(P_1)$*

*$P_1$ is underenumerated*

# Population balancing equation - limitations

- Incomplete and defective data on the components of population change are very common
  - Applicability of the method is limited to countries with good vital registration coverage and migration data
- It is generally not useful for obtaining estimates of net census coverage error for sub-national populations (for example regions, provinces).
  - In addition to the components of population change considered, internal migration has to be considered.
  - For most practical purposes, the use of the population balancing equation is limited to analysis of net coverage error at the national level.

# Death Distribution Methods

## Completeness of reporting of adult mortality

United Nations Workshop on Census Data Evaluation for English Speaking African Countries
Kampala, Uganda
12 – 16 November 2012

# Death distribution methods

- "Death distribution methods" apply the logic of the population balancing equation to different age groups in the population
  - E.g. for the age group 40 – 50, the only way to enter the age group in a country is through aging or immigration, the only way to exit is through death or emigration
  - By comparing our expectation for the size of an age group at the time of the census to its actual enumerated size, we can get a sense of whether we have "missing" or "extra" people in the enumeration

# Death Distribution Methods - Advantages

- Can provide timely estimates of age-specific period mortality rates – here we will use the method to check estimates of completeness of death reporting
- Data requirements:
  - Population by sex and 5-year age groups
  - Deaths by sex and 5-year age groups
  - Can be computed with data from two consecutive censuses with an estimate of the number of deaths between the two censuses

Source: *IUSSP Tools for Demographic Estimation* http://demographicestimation.iussp.org/

# Death Distribution Methods – Assumptions and Violations (1)

- Completeness of deaths reporting is the same across ages
  - Generally does not hold for the oldest and youngest age groups
  - ➢ To avoid, usually truncate analysis to middle age ranges
- *(Two-census variant)* Coverage of both censuses is the same for all age groups
  - Census coverage evaluation will be discussed later in this session
- Age reporting (by 5-year age groups) is accurate
  - Can be checked through age-sex distribution techniques discussed in previous sessions

Source: *IUSSP Tools for Demographic Estimation* http://demographicestimation.iussp.org/

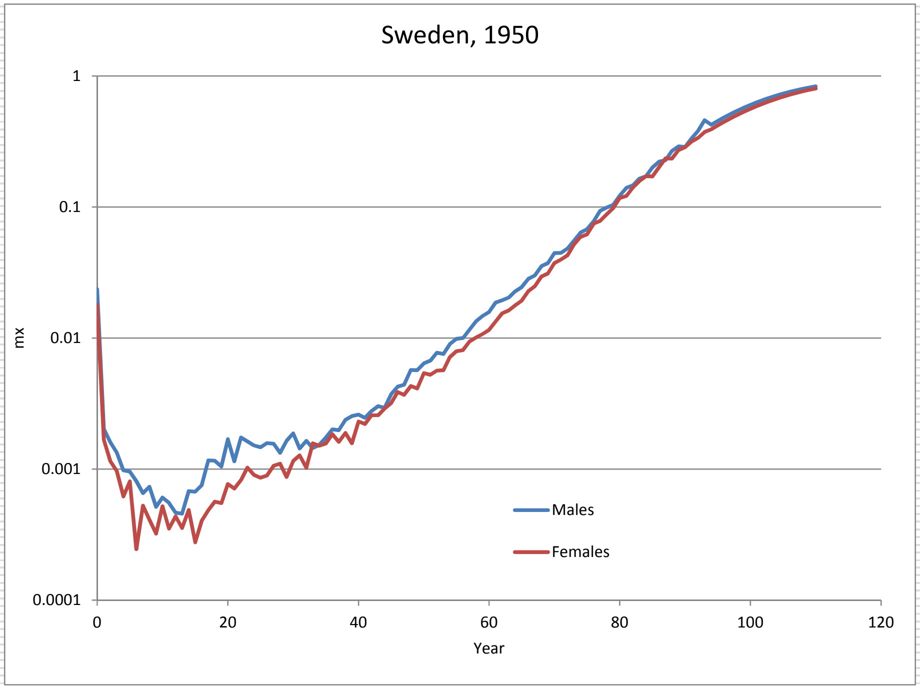
# Death Distribution Methods – Assumptions and Violations (2)

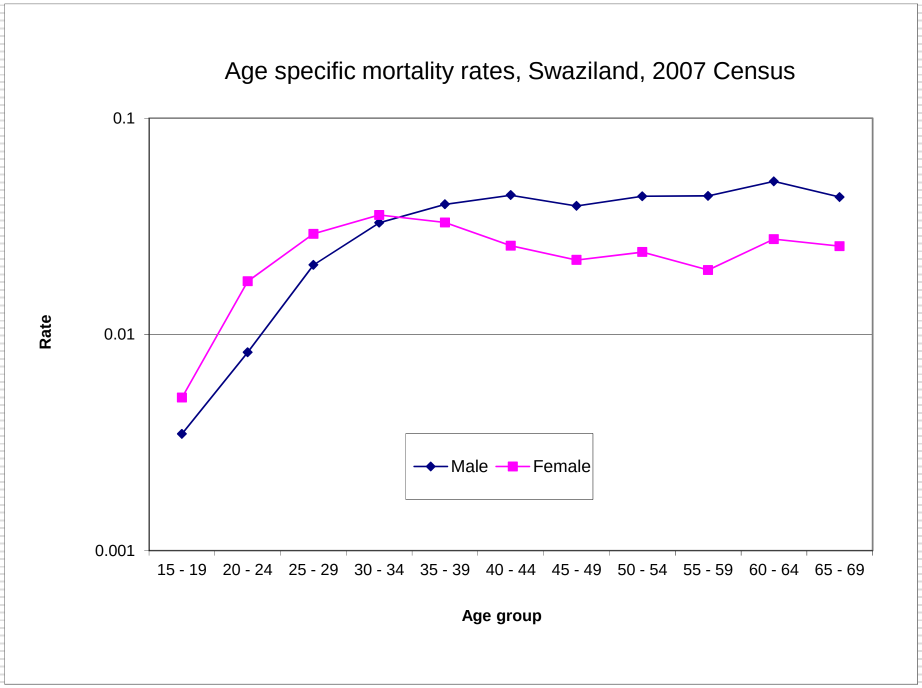
- Net in-migration is limited
  - Will depend on country context
- *(One-census variant)* population is stable (constant growth rate over past several decades)
  - Will depend on country context – in contexts with recent fertility decline, will not hold

# Data quality issues

- Common errors in data on recent deaths by age
    - Under-reporting, especially for child deaths and older age deaths
    - Reference period errors in reporting of deaths (i.e. reporting deaths that occurred prior to the usual 12-month reference period)
    - Death question omitted by interviewers
    - Household breaking up due to the death of a senior household member
        - In this case, any deaths in household will not be captured
    - Age-heaping and age exaggeration
- In addition to age-sex distribution checks discussed in previous sessions, the age and sex structure of reported deaths should be examined prior to conducting any analysis

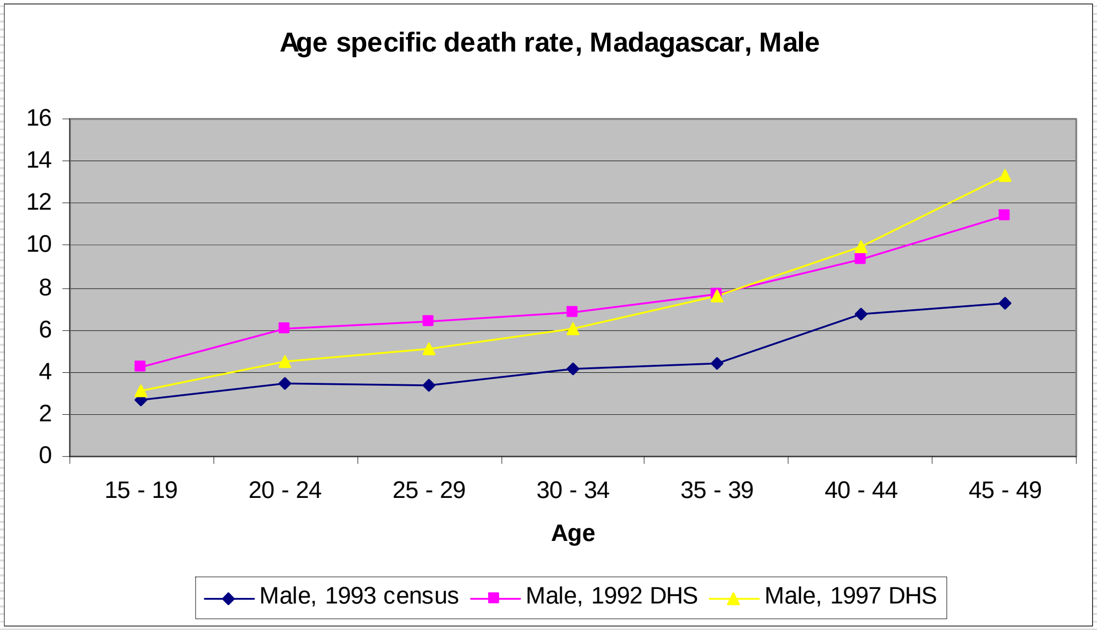
# Data quality checks: schedules of death rates by age and sex





Source: Graph produced based on data from the United Nations *Demographic Yearbook*

# Data quality checks: Comparison with other surveys



Source: Graph produced based on data collected by the United Nations *Demographic Yearbook and Measure DHS country report*

# General Growth Balance Method

Basis: The Balancing Equation of Population Change

$$P_2 = P_1 + B - D + G$$

Assumptions:

a) population is closed to migration, $G=0$;

b) completeness of first census, $k_1$, is independent of age;

c) completeness of second census, $k_2$, is independent of age;

d) completeness of intercensal deaths, $c$, is independent both of age and year;

# GGB regression: $b(x+) - r(x+) = \beta_0 + \beta_1 d(x+)$

Relative completeness of censuses

$$\frac{k_1}{k_2} = \exp(t\hat{\beta}_0)$$

Completeness of death registration

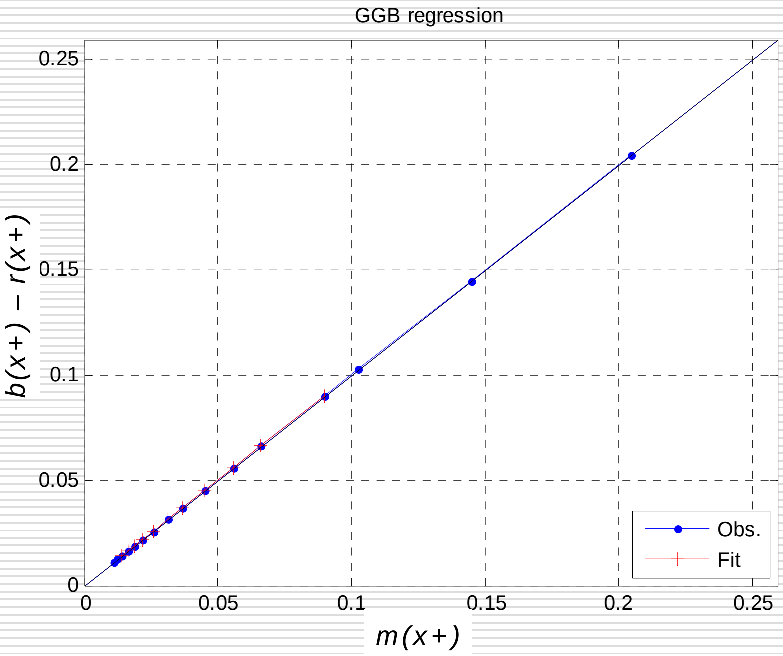
$$c = \frac{(k_1 k_2)^{\frac{1}{2}}}{\hat{\beta}_1}$$

Adjusting observed death rates

$$m = \hat{\beta}_1 m^*$$

*r(x+)* – population growth rate above age x
*b(x+)* – entry rate at age x+, “birth rate”
*d(x+)* – open age death rates
*m** – observed death rates
*m* – adjusted death rates

Ideal case:



Intercept = 0
Slope = 1

# GGB regression with migration:

$$b(x+) - r(x+) + \underline{g(x+)} = \beta_0 + \beta_1 d(x+)$$

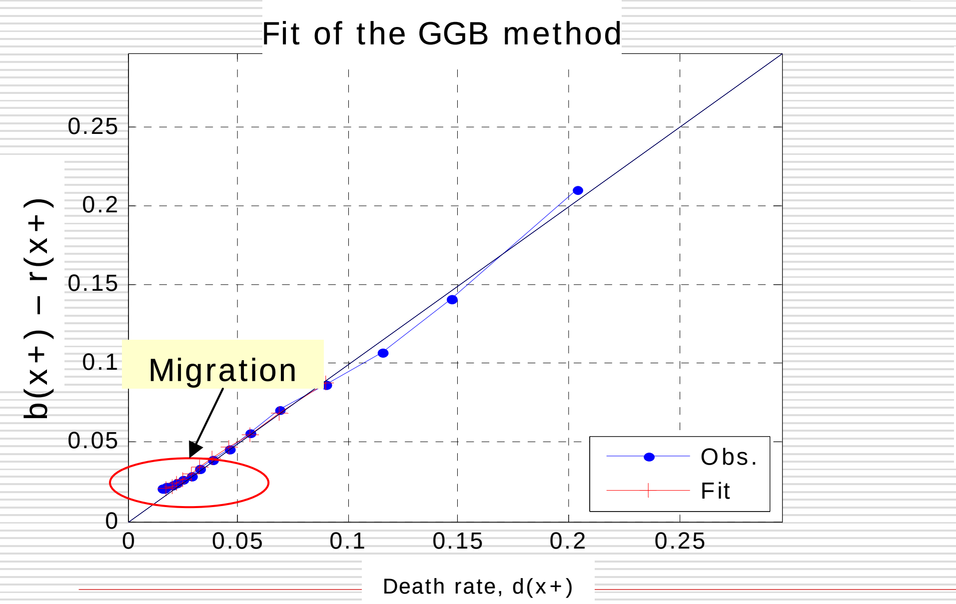
*g(x+)* – net migration rate, age x+

# Effect of unaccounted migration on the GGB estimates (simulation study)

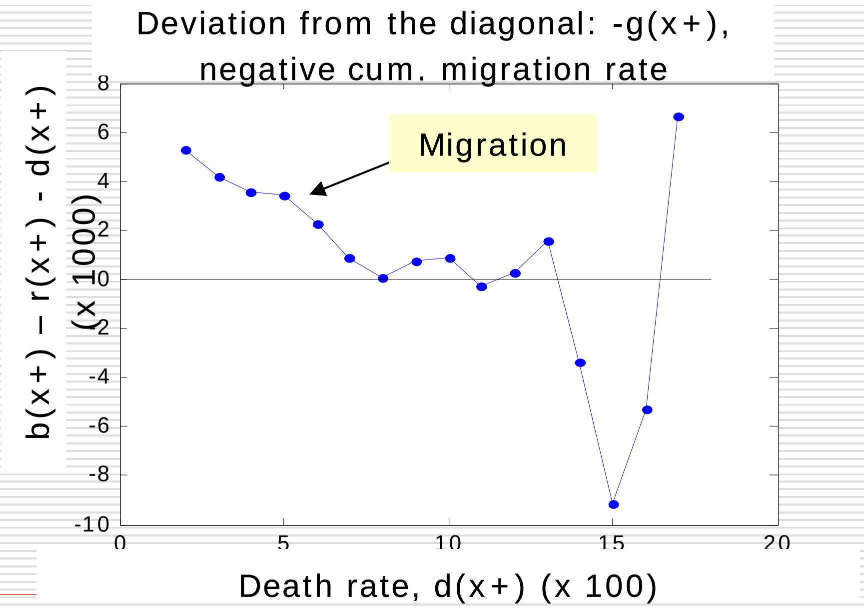
Simulation parameters:
1) Pop. proj., 20 years, by single year and age;
2) Age structure: Denmark, 1880-1900, Males;
3) Mortality: constant over time, e0 = 44;
4) Fertility: constant over time; TFR = 4;
5) Migration: labor out-migration based on residual adjusted migration pattern;

| Estimates\Dataset | DnkA | DnkA (g(x) = 0) |
|---|---|---|
| Slope | 0.9375 | |
| Intercept | 0.0035 | |
| k1 / k2 | 1.0677 | |
| c | 1.0323 | |
| Adult mortality, 45q1 | | |
| Observed | 0.391 | 0.390 |
| Adjusted | 0.371 | 0.385 |
| **Simulated** | **0.394** | **0.394** |

The GGB adjustment factor suggests revising the raw death rates downwards

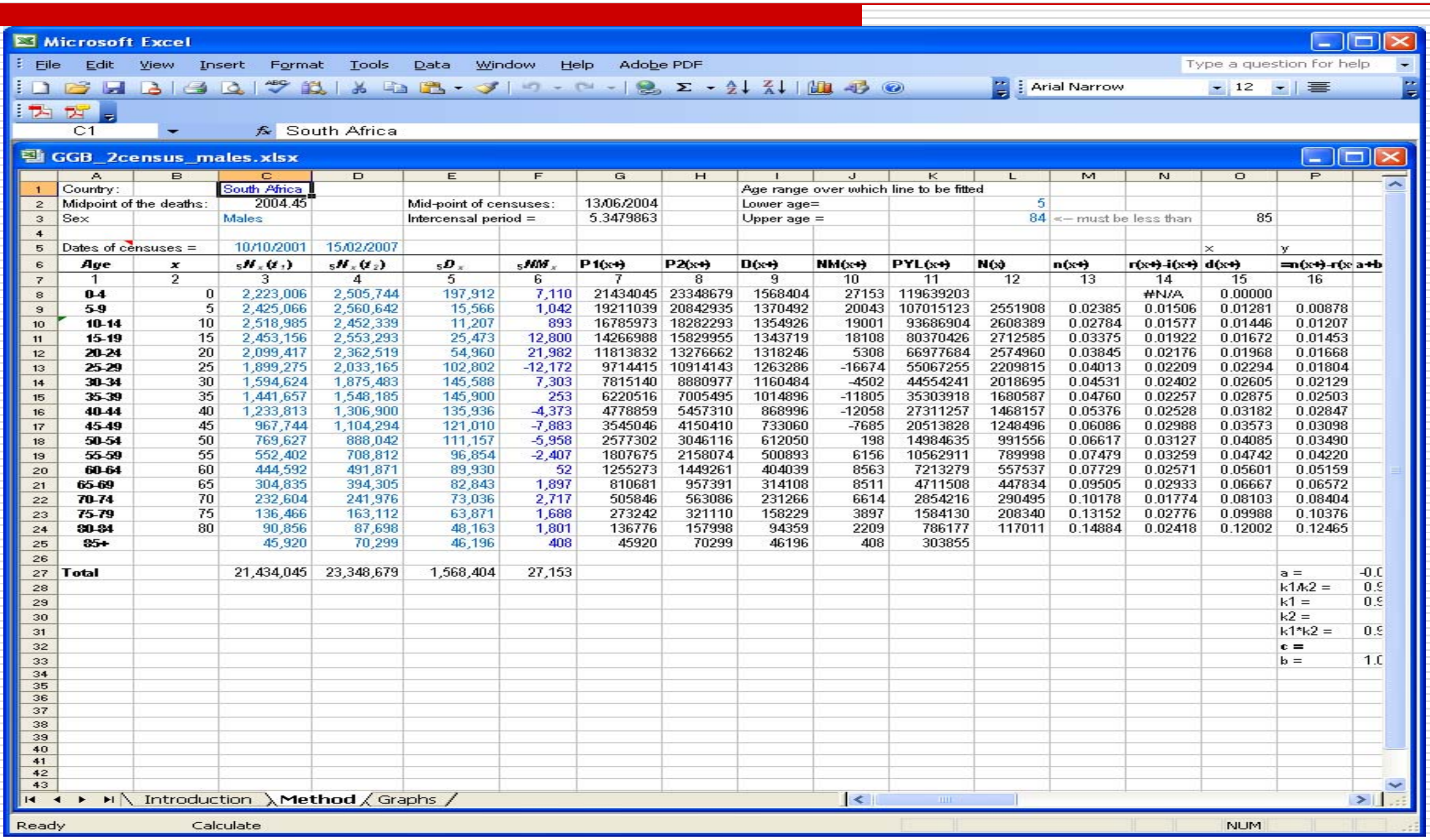

Fit of the GGB method

Migration


Deviation from the diagonal: -g(x+), negative cum. migration rate

Migration

# GGB Method – Application



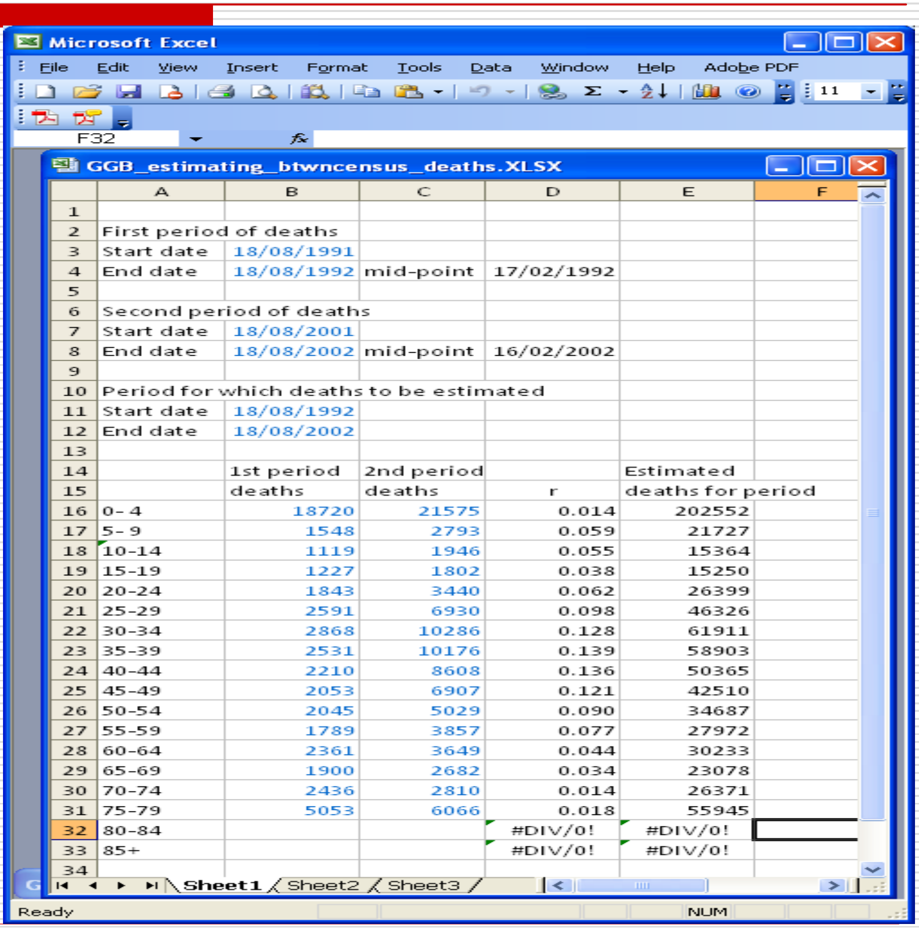
C1 South Africa

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Country: | | South Africa | | | | | | Age range over which line to be fitted | | | | | | | | |
| 2 | Midpoint of the deaths: | | 2004.45 | | Mid-point of censuses: | | 13/06/2004 | | Lower age= | | | 5 | | | | | |
| 3 | Sex | | Males | | Intercensal period = | | 5.3479863 | | Upper age = | | | 84 | <— must be less than | | 85 | | |
| 4 | | | | | | | | | | | | | | | | | |
| 5 | Dates of censuses = | | 10/10/2001 | 15/02/2007 | | | | | | | | | | | x | y | |
| 6 | **Age** | ***x*** | $_5N_x(t_1)$ | $_5N_x(t_2)$ | $_5D_x$ | $_5NM_x$ | **P1(x+)** | **P2(x+)** | **D(x+)** | **NM(x+)** | **PYL(x+)** | **N(x)** | **n(x+)** | **r(x+)-i(x+)** | **d(x+)** | **=n(x+)-r(x** | **a+b** |
| 7 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | |
| 8 | **0-4** | 0 | 2,223,006 | 2,505,744 | 197,912 | 7,110 | 21434045 | 23348679 | 1568404 | 27153 | 119639203 | | | #N/A | 0.00000 | | |
| 9 | **5-9** | 5 | 2,425,066 | 2,560,642 | 15,566 | 1,042 | 19211039 | 20842935 | 1370492 | 20043 | 107015123 | 2551908 | 0.02385 | 0.01506 | 0.01281 | 0.00878 | |
| 10 | **10-14** | 10 | 2,518,985 | 2,452,339 | 11,207 | 893 | 16785973 | 18282293 | 1354926 | 19001 | 93686904 | 2608389 | 0.02784 | 0.01577 | 0.01446 | 0.01207 | |
| 11 | **15-19** | 15 | 2,453,156 | 2,553,293 | 25,473 | 12,800 | 14266988 | 15829955 | 1343719 | 18108 | 80370426 | 2712585 | 0.03375 | 0.01922 | 0.01672 | 0.01453 | |
| 12 | **20-24** | 20 | 2,099,417 | 2,362,519 | 54,960 | 21,982 | 11813832 | 13276662 | 1318246 | 5308 | 66977684 | 2574960 | 0.03845 | 0.02176 | 0.01968 | 0.01668 | |
| 13 | **25-29** | 25 | 1,899,275 | 2,033,165 | 102,802 | -12,172 | 9714415 | 10914143 | 1263286 | -16674 | 55067255 | 2209815 | 0.04013 | 0.02209 | 0.02294 | 0.01804 | |
| 14 | **30-34** | 30 | 1,594,624 | 1,875,483 | 145,588 | 7,303 | 7815140 | 8880977 | 1160484 | -4502 | 44554241 | 2018695 | 0.04531 | 0.02402 | 0.02605 | 0.02129 | |
| 15 | **35-39** | 35 | 1,441,657 | 1,548,185 | 145,900 | 253 | 6220516 | 7005495 | 1014896 | -11805 | 35303918 | 1680587 | 0.04760 | 0.02257 | 0.02875 | 0.02503 | |
| 16 | **40-44** | 40 | 1,233,813 | 1,306,900 | 135,936 | -4,373 | 4778859 | 5457310 | 868996 | -12058 | 27311257 | 1468157 | 0.05376 | 0.02528 | 0.03182 | 0.02847 | |
| 17 | **45-49** | 45 | 967,744 | 1,104,294 | 121,010 | -7,883 | 3545046 | 4150410 | 733060 | -7685 | 20513828 | 1248496 | 0.06086 | 0.02988 | 0.03573 | 0.03098 | |
| 18 | **50-54** | 50 | 769,627 | 888,042 | 111,157 | -5,958 | 2577302 | 3046116 | 612050 | 198 | 14984635 | 991556 | 0.06617 | 0.03127 | 0.04085 | 0.03490 | |
| 19 | **55-59** | 55 | 552,402 | 708,812 | 96,854 | -2,407 | 1807675 | 2158074 | 500893 | 6156 | 10562911 | 789998 | 0.07479 | 0.03259 | 0.04742 | 0.04220 | |
| 20 | **60-64** | 60 | 444,592 | 491,871 | 89,930 | 52 | 1255273 | 1449261 | 404039 | 8563 | 7213279 | 557537 | 0.07729 | 0.02571 | 0.05601 | 0.05159 | |
| 21 | **65-69** | 65 | 304,835 | 394,305 | 82,843 | 1,897 | 810681 | 957391 | 314108 | 8511 | 4711508 | 447834 | 0.09505 | 0.02933 | 0.06667 | 0.06572 | |
| 22 | **70-74** | 70 | 232,604 | 241,976 | 73,036 | 2,717 | 505846 | 563086 | 231266 | 6614 | 2854216 | 290495 | 0.10178 | 0.01774 | 0.08103 | 0.08404 | |
| 23 | **75-79** | 75 | 136,466 | 163,112 | 63,871 | 1,688 | 273242 | 321110 | 158229 | 3897 | 1584130 | 208340 | 0.13152 | 0.02776 | 0.09988 | 0.10376 | |
| 24 | **80-84** | 80 | 90,856 | 87,698 | 48,163 | 1,801 | 136776 | 157998 | 94359 | 2209 | 786177 | 117011 | 0.14884 | 0.02418 | 0.12002 | 0.12465 | |
| 25 | **85+** | | 45,920 | 70,299 | 46,196 | 408 | 45920 | 70299 | 46196 | 408 | 303855 | | | | | | |
| 26 | | | | | | | | | | | | | | | | | |
| 27 | **Total** | | 21,434,045 | 23,348,679 | 1,568,404 | 27,153 | | | | | | | | | | a = | -0.0 |
| 28 | | | | | | | | | | | | | | | | k1/k2 = | 0.9 |
| 29 | | | | | | | | | | | | | | | | k1 = | 0.9 |
| 30 | | | | | | | | | | | | | | | | k2 = | |
| 31 | | | | | | | | | | | | | | | | k1*k2 = | 0.9 |
| 32 | | | | | | | | | | | | | | | | **c =** | |
| 33 | | | | | | | | | | | | | | | | b = | 1.0 |

Source: *IUSSP Tools for Demographic Estimation* http://demographicestimation.iussp.org/

# GGB Method – Estimating intercensal deaths

- ☐ If accurate data for intercensal deaths is not available, they can be estimated if deaths for two other well-defined periods are available – e.g., deaths in the year prior to two different censuses
- ☐ Worksheet will compute growth rates and deaths for the intercensal period



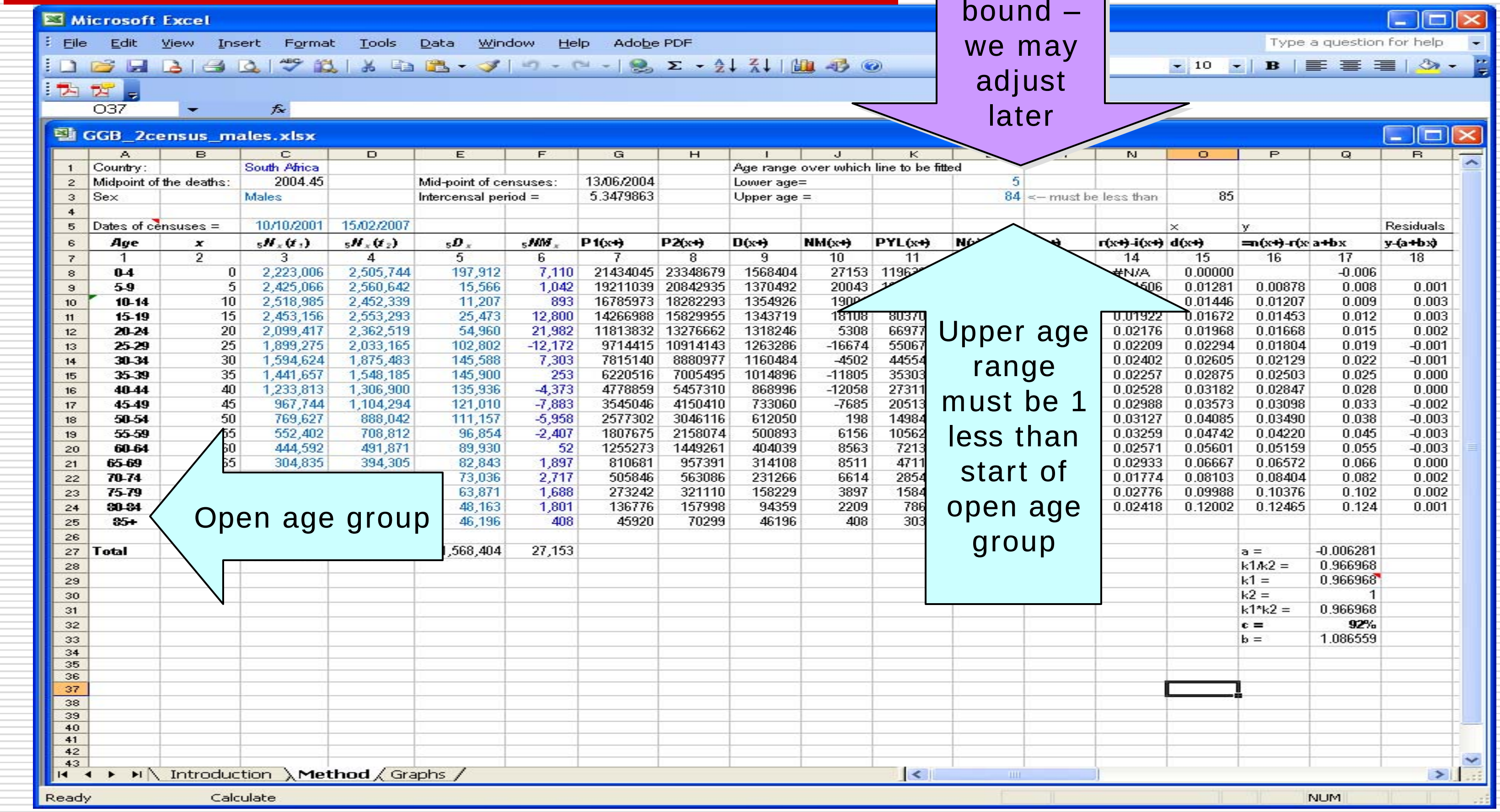
| | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| 1 | | | | | | |
| 2 | First period of deaths | | | | | |
| 3 | Start date | 18/08/1991 | | | | |
| 4 | End date | 18/08/1992 | mid-point | 17/02/1992 | | |
| 5 | | | | | | |
| 6 | Second period of deaths | | | | | |
| 7 | Start date | 18/08/2001 | | | | |
| 8 | End date | 18/08/2002 | mid-point | 16/02/2002 | | |
| 9 | | | | | | |
| 10 | Period for which deaths to be estimated | | | | | |
| 11 | Start date | 18/08/1992 | | | | |
| 12 | End date | 18/08/2002 | | | | |
| 13 | | | | | | |
| 14 | | 1st period | 2nd period | | Estimated | |
| 15 | | deaths | deaths | r | deaths for period | |
| 16 | 0- 4 | 18720 | 21575 | 0.014 | 202552 | |
| 17 | 5- 9 | 1548 | 2793 | 0.059 | 21727 | |
| 18 | 10-14 | 1119 | 1946 | 0.055 | 15364 | |
| 19 | 15-19 | 1227 | 1802 | 0.038 | 15250 | |
| 20 | 20-24 | 1843 | 3440 | 0.062 | 26399 | |
| 21 | 25-29 | 2591 | 6930 | 0.098 | 46326 | |
| 22 | 30-34 | 2868 | 10286 | 0.128 | 61911 | |
| 23 | 35-39 | 2531 | 10176 | 0.139 | 58903 | |
| 24 | 40-44 | 2210 | 8608 | 0.136 | 50365 | |
| 25 | 45-49 | 2053 | 6907 | 0.121 | 42510 | |
| 26 | 50-54 | 2045 | 5029 | 0.090 | 34687 | |
| 27 | 55-59 | 1789 | 3857 | 0.077 | 27972 | |
| 28 | 60-64 | 2361 | 3649 | 0.044 | 30233 | |
| 29 | 65-69 | 1900 | 2682 | 0.034 | 23078 | |
| 30 | 70-74 | 2436 | 2810 | 0.014 | 26371 | |
| 31 | 75-79 | 5053 | 6066 | 0.018 | 55945 | |
| 32 | 80-84 | | | #DIV/0! | #DIV/0! | |
| 33 | 85+ | | | #DIV/0! | #DIV/0! | |
| 34 | | | | | | |

# GGB Method –Setting the age range

Note the lower age bound – we may adjust later

Upper age range must be 1 less than start of open age group

Open age group

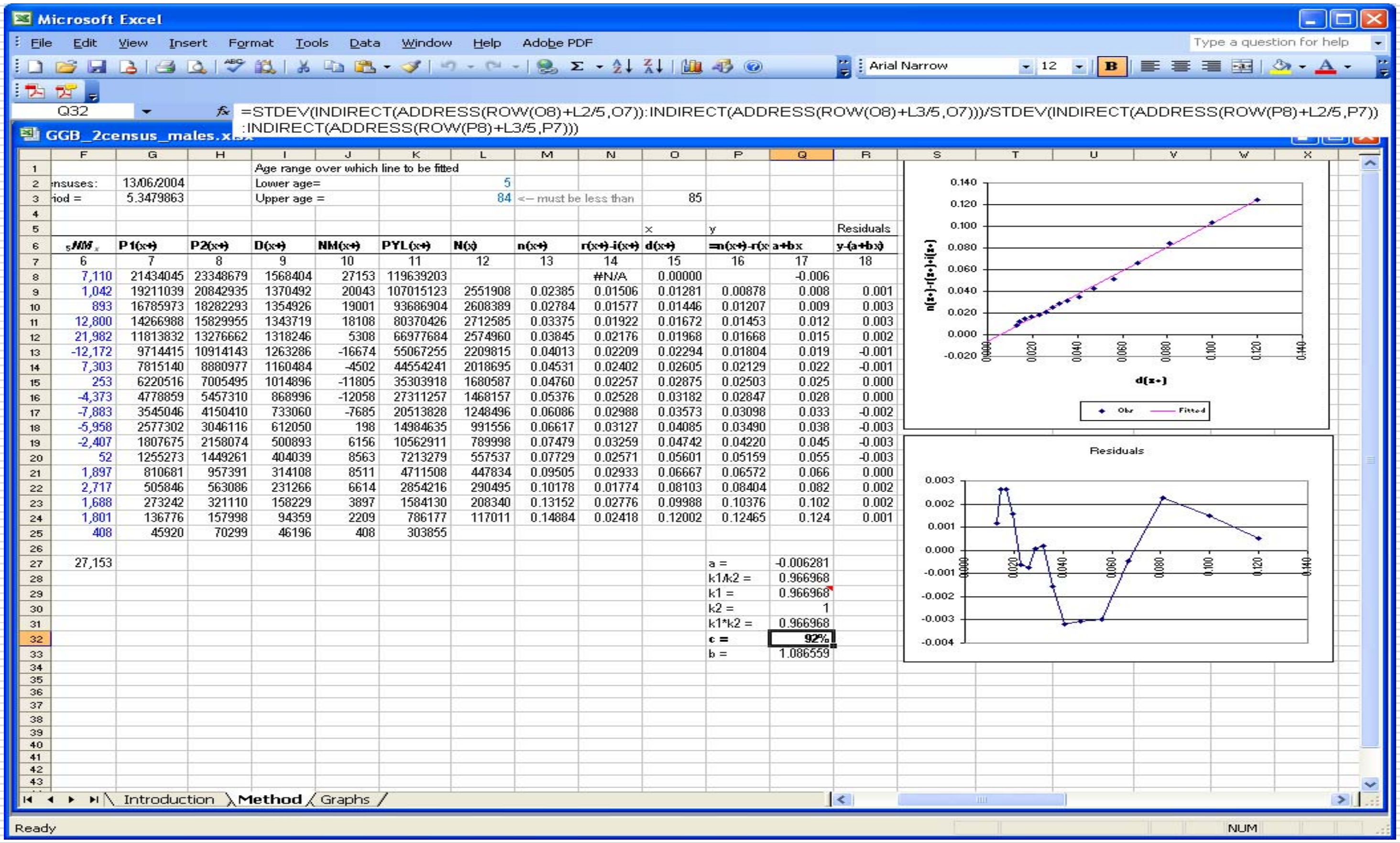


Country: South Africa
Midpoint of the deaths: 2004.45
Mid-point of censuses: 13/06/2004
Intercensal period = 5.3479863
Sex Males
Dates of censuses = 10/10/2001 15/02/2007

Age range over which line to be fitted
Lower age= 5
Upper age = 84 <-- must be less than 85

| | A | B | C | D | E | F | G | H | I | J | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | | | | | | | | | | | | x | y | | Residuals |
| 6 | **Age** | **x** | $_5N_x(t_1)$ | $_5N_x(t_2)$ | $_5D_x$ | $_5NM_x$ | **P1(x+)** | **P2(x+)** | **D(x+)** | **NM(x+)** | **r(x+)-i(x+)** | **d(x+)** | **=n(x+)-r(x** | **a+bx** | **y-(a+bx)** |
| 7 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 14 | 15 | 16 | 17 | 18 |
| 8 | **0-4** | 0 | 2,223,006 | 2,505,744 | 197,912 | 7,110 | 21434045 | 23348679 | 1568404 | 27153 | #N/A | 0.00000 | | -0.006 | |
| 9 | **5-9** | 5 | 2,425,066 | 2,560,642 | 15,566 | 1,042 | 19211039 | 20842935 | 1370492 | 20043 | [illegible] | 0.01281 | 0.00878 | 0.008 | 0.001 |
| 10 | **10-14** | 10 | 2,518,985 | 2,452,339 | 11,207 | 893 | 16785973 | 18282293 | 1354926 | [illegible] | [illegible] | 0.01446 | 0.01207 | 0.009 | 0.003 |
| 11 | **15-19** | 15 | 2,453,156 | 2,553,293 | 25,473 | 12,800 | 14266988 | 15829955 | 1343719 | 18108 | 0.01922 | 0.01672 | 0.01453 | 0.012 | 0.003 |
| 12 | **20-24** | 20 | 2,099,417 | 2,362,519 | 54,960 | 21,982 | 11813832 | 13276662 | 1318246 | 5308 | 0.02176 | 0.01968 | 0.01668 | 0.015 | 0.002 |
| 13 | **25-29** | 25 | 1,899,275 | 2,033,165 | 102,802 | -12,172 | 9714415 | 10914143 | 1263286 | -16674 | 0.02209 | 0.02294 | 0.01804 | 0.019 | -0.001 |
| 14 | **30-34** | 30 | 1,594,624 | 1,875,483 | 145,588 | 7,303 | 7815140 | 8880977 | 1160484 | -4502 | 0.02402 | 0.02605 | 0.02129 | 0.022 | -0.001 |
| 15 | **35-39** | 35 | 1,441,657 | 1,548,185 | 145,900 | 253 | 6220516 | 7005495 | 1014896 | -11805 | 0.02257 | 0.02875 | 0.02503 | 0.025 | 0.000 |
| 16 | **40-44** | 40 | 1,233,813 | 1,306,900 | 135,936 | -4,373 | 4778859 | 5457310 | 868996 | -12058 | 0.02528 | 0.03182 | 0.02847 | 0.028 | 0.000 |
| 17 | **45-49** | 45 | 967,744 | 1,104,294 | 121,010 | -7,883 | 3545046 | 4150410 | 733060 | -7685 | 0.02988 | 0.03573 | 0.03098 | 0.033 | -0.002 |
| 18 | **50-54** | 50 | 769,627 | 888,042 | 111,157 | -5,958 | 2577302 | 3046116 | 612050 | 198 | 0.03127 | 0.04085 | 0.03490 | 0.038 | -0.003 |
| 19 | **55-59** | [illegible] | 552,402 | 708,812 | 96,854 | -2,407 | 1807675 | 2158074 | 500893 | 6156 | 0.03259 | 0.04742 | 0.04220 | 0.045 | -0.003 |
| 20 | **60-64** | [illegible] | 444,592 | 491,871 | 89,930 | 52 | 1255273 | 1449261 | 404039 | 8563 | 0.02571 | 0.05601 | 0.05159 | 0.055 | -0.003 |
| 21 | **65-69** | [illegible] | 304,835 | 394,305 | 82,843 | 1,897 | 810681 | 957391 | 314108 | 8511 | 0.02933 | 0.06667 | 0.06572 | 0.066 | 0.000 |
| 22 | **70-74** | [illegible] | [illegible] | [illegible] | 73,036 | 2,717 | 505846 | 563086 | 231266 | 6614 | 0.01774 | 0.08103 | 0.08404 | 0.082 | 0.002 |
| 23 | **75-79** | [illegible] | [illegible] | [illegible] | 63,871 | 1,688 | 273242 | 321110 | 158229 | 3897 | 0.02776 | 0.09988 | 0.10376 | 0.102 | 0.002 |
| 24 | **80-84** | [illegible] | [illegible] | [illegible] | 48,163 | 1,801 | 136776 | 157998 | 94359 | 2209 | 0.02418 | 0.12002 | 0.12465 | 0.124 | 0.001 |
| 25 | **85+** | [illegible] | [illegible] | [illegible] | 46,196 | 408 | 45920 | 70299 | 46196 | 408 | | | | | |
| 27 | **Total** | | | | 1,568,404 | 27,153 | | | | | | | | | |
| 28 | | | | | | | | | | | | | a = | -0.006281 | |
| 29 | | | | | | | | | | | | | k1/k2 = | 0.966968 | |
| 30 | | | | | | | | | | | | | k1 = | 0.966968 | |
| 31 | | | | | | | | | | | | | k2 = | 1 | |
| 32 | | | | | | | | | | | | | k1*k2 = | 0.966968 | |
| 33 | | | | | | | | | | | | | **c =** | **92%** | |
| 34 | | | | | | | | | | | | | b = | 1.086559 | |

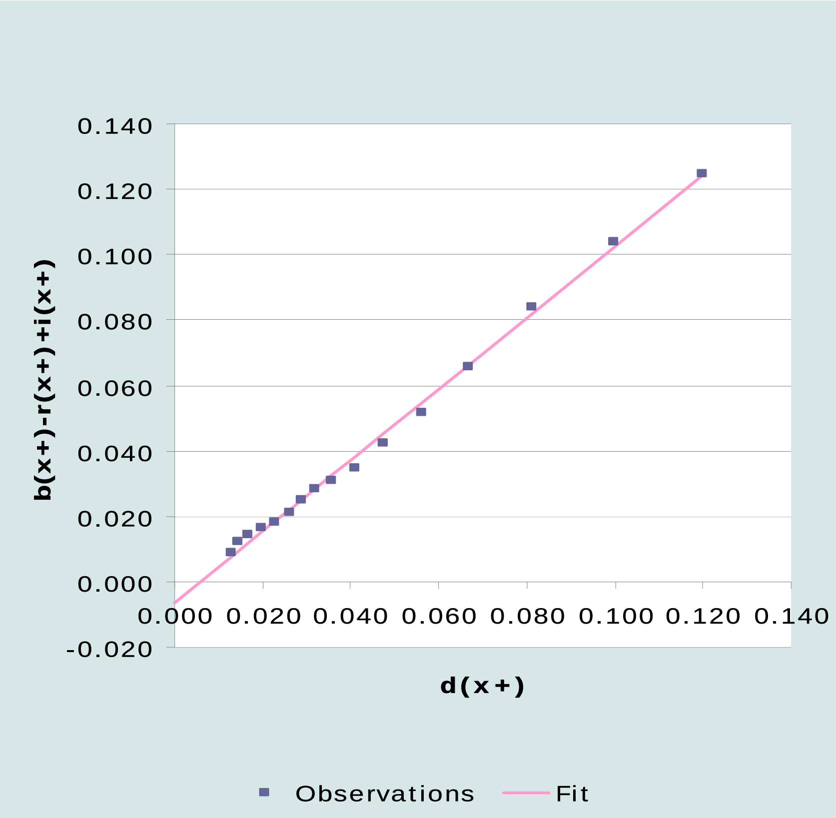
# GGB Method - Diagnostics

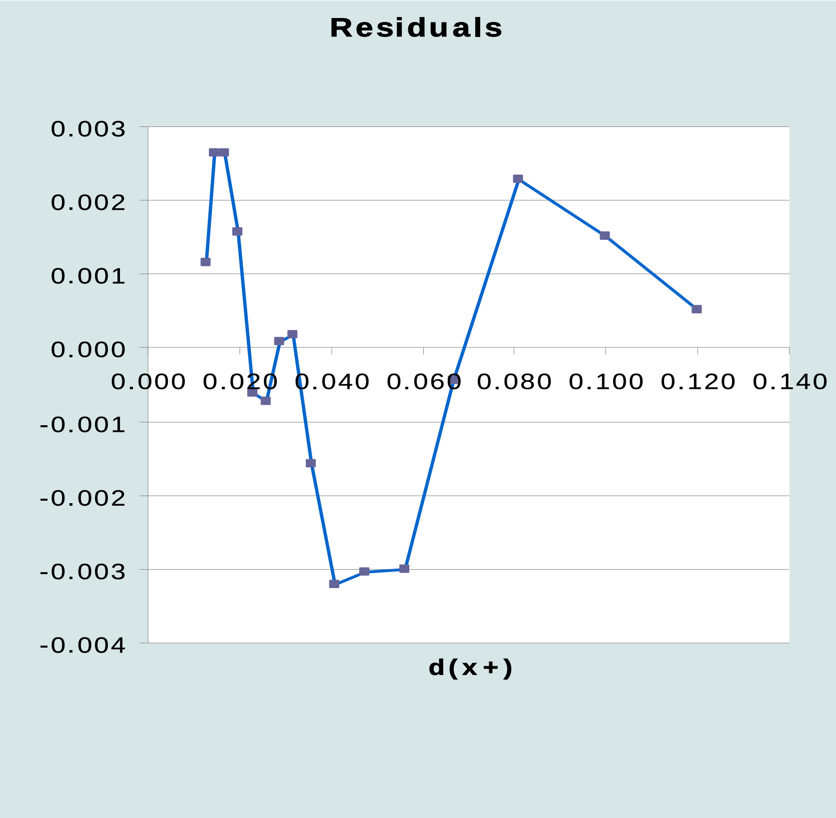


=STDEV(INDIRECT(ADDRESS(ROW(O8)+L2/5,O7)):INDIRECT(ADDRESS(ROW(O8)+L3/5,O7)))/STDEV(INDIRECT(ADDRESS(ROW(P8)+L2/5,P7)):INDIRECT(ADDRESS(ROW(P8)+L3/5,P7)))

| | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | | | Age range over which line to be fitted | | | | | | | | | |
| 2 | nsuses: | 13/06/2004 | | Lower age= | | | 5 | | | | | | |
| 3 | iod = | 5.3479863 | | Upper age = | | | 84 | <-- must be less than | | 85 | | | |
| 4 | | | | | | | | | | | | | |
| 5 | | | | | | | | | | x | y | | Residuals |
| 6 | **$_5NM_x$** | **P1(x+)** | **P2(x+)** | **D(x+)** | **NM(x+)** | **PYL(x+)** | **N(x)** | **n(x+)** | **r(x+)-i(x+)** | **d(x+)** | **=n(x+)-r(x** | **a+bx** | **y-(a+bx)** |
| 7 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| 8 | 7,110 | 21434045 | 23348679 | 1568404 | 27153 | 119639203 | | | #N/A | 0.00000 | | -0.006 | |
| 9 | 1,042 | 19211039 | 20842935 | 1370492 | 20043 | 107015123 | 2551908 | 0.02385 | 0.01506 | 0.01281 | 0.00878 | 0.008 | 0.001 |
| 10 | 893 | 16785973 | 18282293 | 1354926 | 19001 | 93686904 | 2608389 | 0.02784 | 0.01577 | 0.01446 | 0.01207 | 0.009 | 0.003 |
| 11 | 12,800 | 14266988 | 15829955 | 1343719 | 18108 | 80370426 | 2712585 | 0.03375 | 0.01922 | 0.01672 | 0.01453 | 0.012 | 0.003 |
| 12 | 21,982 | 11813832 | 13276662 | 1318246 | 5308 | 66977684 | 2574960 | 0.03845 | 0.02176 | 0.01968 | 0.01668 | 0.015 | 0.002 |
| 13 | -12,172 | 9714415 | 10914143 | 1263286 | -16674 | 55067255 | 2209815 | 0.04013 | 0.02209 | 0.02294 | 0.01804 | 0.019 | -0.001 |
| 14 | 7,303 | 7815140 | 8880977 | 1160484 | -4502 | 44554241 | 2018695 | 0.04531 | 0.02402 | 0.02605 | 0.02129 | 0.022 | -0.001 |
| 15 | 253 | 6220516 | 7005495 | 1014896 | -11805 | 35303918 | 1680587 | 0.04760 | 0.02257 | 0.02875 | 0.02503 | 0.025 | 0.000 |
| 16 | -4,373 | 4778859 | 5457310 | 868996 | -12058 | 27311257 | 1468157 | 0.05376 | 0.02528 | 0.03182 | 0.02847 | 0.028 | 0.000 |
| 17 | -7,883 | 3545046 | 4150410 | 733060 | -7685 | 20513828 | 1248496 | 0.06086 | 0.02988 | 0.03573 | 0.03098 | 0.033 | -0.002 |
| 18 | -5,958 | 2577302 | 3046116 | 612050 | 198 | 14984635 | 991556 | 0.06617 | 0.03127 | 0.04085 | 0.03490 | 0.038 | -0.003 |
| 19 | -2,407 | 1807675 | 2158074 | 500893 | 6156 | 10562911 | 789998 | 0.07479 | 0.03259 | 0.04742 | 0.04220 | 0.045 | -0.003 |
| 20 | 52 | 1255273 | 1449261 | 404039 | 8563 | 7213279 | 557537 | 0.07729 | 0.02571 | 0.05601 | 0.05159 | 0.055 | -0.003 |
| 21 | 1,897 | 810681 | 957391 | 314108 | 8511 | 4711508 | 447834 | 0.09505 | 0.02933 | 0.06667 | 0.06572 | 0.066 | 0.000 |
| 22 | 2,717 | 505846 | 563086 | 231266 | 6614 | 2854216 | 290495 | 0.10178 | 0.01774 | 0.08103 | 0.08404 | 0.082 | 0.002 |
| 23 | 1,688 | 273242 | 321110 | 158229 | 3897 | 1584130 | 208340 | 0.13152 | 0.02776 | 0.09988 | 0.10376 | 0.102 | 0.002 |
| 24 | 1,801 | 136776 | 157998 | 94359 | 2209 | 786177 | 117011 | 0.14884 | 0.02418 | 0.12002 | 0.12465 | 0.124 | 0.001 |
| 25 | 408 | 45920 | 70299 | 46196 | 408 | 303855 | | | | | | | |
| 26 | | | | | | | | | | | | | |
| 27 | 27,153 | | | | | | | | | | a = | -0.006281 | |
| 28 | | | | | | | | | | | k1/k2 = | 0.966968 | |
| 29 | | | | | | | | | | | k1 = | 0.966968 | |
| 30 | | | | | | | | | | | k2 = | 1 | |
| 31 | | | | | | | | | | | k1*k2 = | 0.966968 | |
| 32 | | | | | | | | | | | **c =** | **92%** | |
| 33 | | | | | | | | | | | b = | 1.086559 | |

# GGB Method - Diagnostics





# GGB Method – Two census (6) – Interpretation

- Check the estimate of completeness of death reporting and reasonableness of the analysis
  - Compare with the results for the opposite sex – unless we have reason to believe that completeness will vary significantly by sex, should be fairly close
  - Compare with results of Synthetic Extinct Generations approach (worksheets also available through IUSSP)

# Synthetic Extinct Generations (Bennett & Horiuchi) method

- Used for estimating completeness of death registration, with different of inputs
- Population at exact age, *N*, can be computed from registered deaths, *D*, and intercensal rates of increase *r*:

$$N(a) = \int_a^{\infty} D(x) e^{\int_a^x r(u)du} dx$$

- Software: Ken Hill's spreadsheet, `Death_dist_method all template.xls`

# Cohort Survival Ratios

# Mortality and census coverage

United Nations Workshop on Census Data Evaluation for English Speaking African Countries
Kampala, Uganda
12 – 16 November 2012

# Cohort survival ratios

- This technique is based on a comparison of the size of birth cohorts enumerated in successive censuses
- In the absence of census errors, the ratio of the number of persons in a cohort enumerated in the second census to the number enumerated in the first census should approximate the survival rate that would be expected on the basis of mortality conditions
  - E.g. we have a cohort of males aged 40 – 44 at the time of the first census, say in 2000
  - If the next census is held exactly 10 years later, in 2010, this cohort will be aged 50 – 54
  - In the absence of other factors, we expect their numbers to have been reduced only by the life table quantity ${}_{10}d_x = l_x - l_{x+n}$, the number of deaths to those aged x over the subsequent 10 years

Source: U.S. Census Bureau, 1985. *Evaluating Censuses of Population and Housing* and IUSSP Tools for Demographic Estimation http://demographicestimation.iussp.org/

# Cohort Survival Ratios: Caveats

- The method is less useful when other factors make it difficult to determine whether deviations from the expected CSR are due to census error or something else
    - Substantial net migration (unless there are accurate estimates of net migration by age)
    - Changes in country borders between censuses
    - Changes in the population groups included in the two censuses (e.g. active military, nomadic groups) if the size of these groups is substantial

# Calculating CSRs (1)

☐ Intercensal cohort survival rates are defined as:

$${}_{n}CSR_{x}\ (a) = \frac{{}_{n}P_{x+a}\ (t+a)}{{}_{n}P_{x}\ (t)}$$

Where:

- $t$ = time of first census
- $a$ = number of years between censuses
- ${}_{n}P_{x}\ (t)$ = size of the cohort at the time of the first census
- ${}_{n}P_{x+a}\ (t+a)$ = size of the cohort at the time of the second census

# Calculating CSRs (2)

☐ The ratio of the observed intercensal cohort survival rate to the corresponding life-table survival rate

$$ {}_nR_x = \frac{{}_nP_{x+a}\,(t+a) \,/\, {}_nP_x\,(t)}{{}_nL_{x+a} \,/\, {}_nL_x} $$

Where:

- ${}_nP_{x+a}$ (t+a) = size of the cohort at time of the second census
- ${}_nP_x$ (t) = size of the cohort at the time of the first census
- ${}_nL_{x+i}$= the life table number of person-years lived in the age interval x+a to x+a+n years
- ${}_nL_x$ = the life table number of person-years lived in the age interval x to x+n years

# Cohort Survival Ratio - Interpretation

- ❑ In the absence of census error, the expected value of the ratio ($_{n}R_{x}$) would be 1.0

- ❑ Ratio values for any particular cohort which exceed 1.0 would indicate over-enumeration of the cohort in the second census relative to the first census

- ❑ Ratio values of less than 1.0 would indicate under-enumeration of the cohort in the second census relative the first census

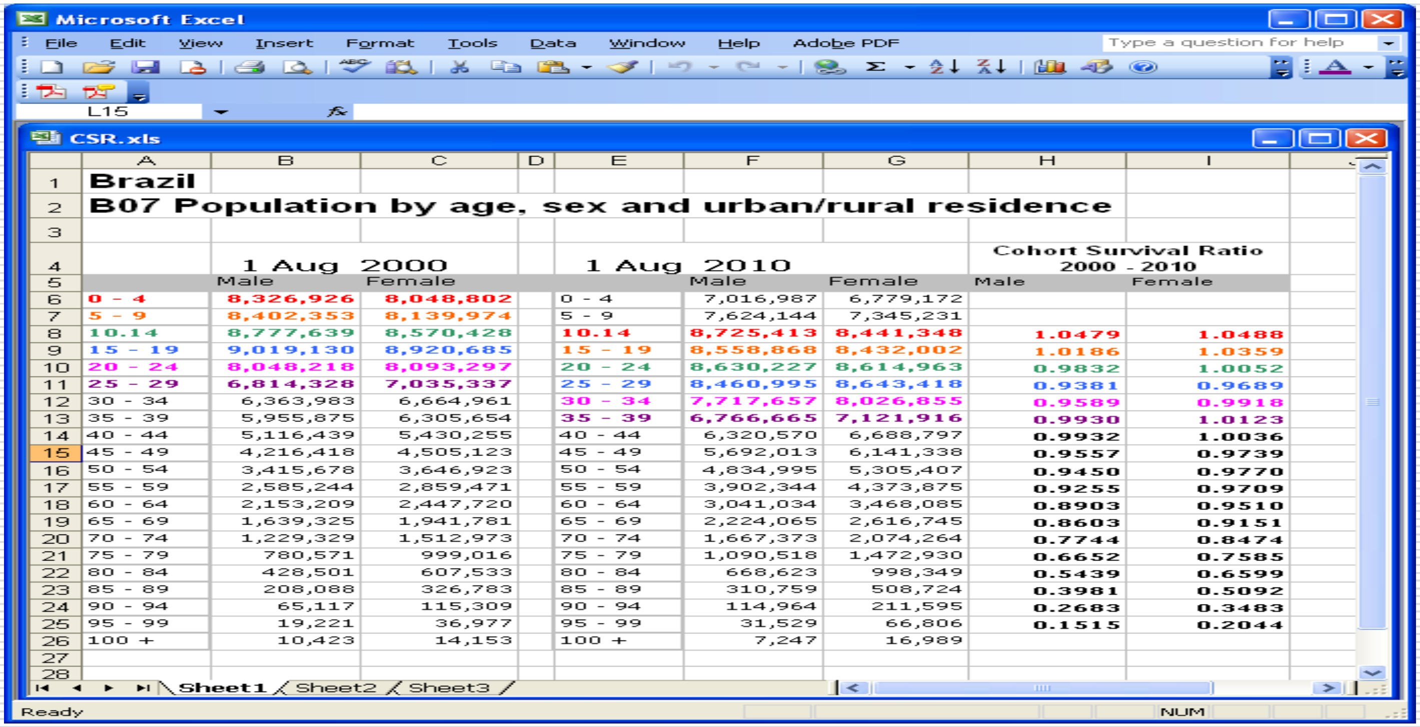
# Cohort survival ratios – Example (1)

**Step 1: Adjustment for migration (if appropriate)**

- In countries experiencing net immigration significant levels of net intercensal immigration, the number of net immigrants in each cohort may either added to the cohort enumerated in the first census or subtracted from the cohort enumerated in the second census
- In cohorts experiencing net intercensal emigration, the number of net intercensal emigrants can either added to the second census or subtracted from the first census
- Should be confident that migration data is reasonably accurate before making any adjustments

# Cohort survival ratios – Example (2)

- ☐ **Step 2:** Calculation of census survival rates using two consecutive censuses ${}_nP_{x+a}(t+a)$ / ${}_nP_x$ (t)



**Brazil**

**B07 Population by age, sex and urban/rural residence**

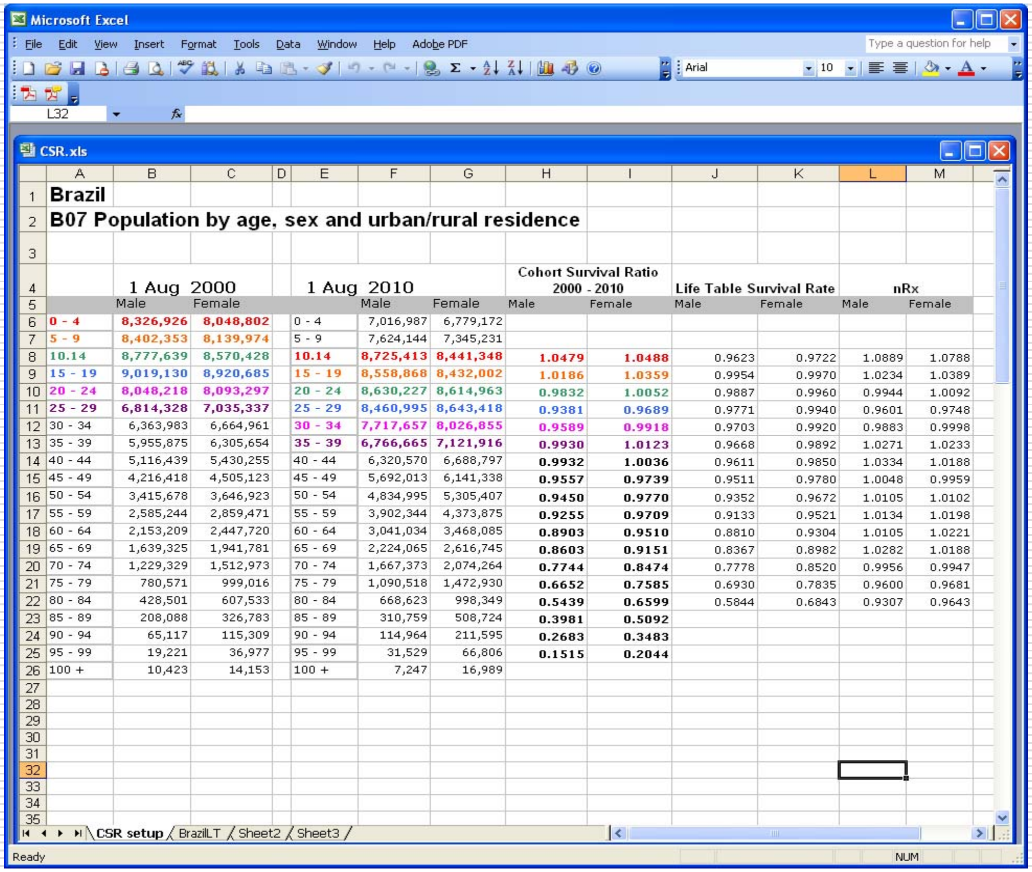
| | 1 Aug 2000 | | | 1 Aug 2010 | | Cohort Survival Ratio 2000 - 2010 | |
|---|---|---|---|---|---|---|---|
| | Male | Female | | Male | Female | Male | Female |
| 0 - 4 | 8,326,926 | 8,048,802 | 0 - 4 | 7,016,987 | 6,779,172 | | |
| 5 - 9 | 8,402,353 | 8,139,974 | 5 - 9 | 7,624,144 | 7,345,231 | | |
| 10.14 | 8,777,639 | 8,570,428 | 10.14 | 8,725,413 | 8,441,348 | 1.0479 | 1.0488 |
| 15 - 19 | 9,019,130 | 8,920,685 | 15 - 19 | 8,558,868 | 8,432,002 | 1.0186 | 1.0359 |
| 20 - 24 | 8,048,218 | 8,093,297 | 20 - 24 | 8,630,227 | 8,614,963 | 0.9832 | 1.0052 |
| 25 - 29 | 6,814,328 | 7,035,337 | 25 - 29 | 8,460,995 | 8,643,418 | 0.9381 | 0.9689 |
| 30 - 34 | 6,363,983 | 6,664,961 | 30 - 34 | 7,717,657 | 8,026,855 | 0.9589 | 0.9918 |
| 35 - 39 | 5,955,875 | 6,305,654 | 35 - 39 | 6,766,665 | 7,121,916 | 0.9930 | 1.0123 |
| 40 - 44 | 5,116,439 | 5,430,255 | 40 - 44 | 6,320,570 | 6,688,797 | 0.9932 | 1.0036 |
| 45 - 49 | 4,216,418 | 4,505,123 | 45 - 49 | 5,692,013 | 6,141,338 | 0.9557 | 0.9739 |
| 50 - 54 | 3,415,678 | 3,646,923 | 50 - 54 | 4,834,995 | 5,305,407 | 0.9450 | 0.9770 |
| 55 - 59 | 2,585,244 | 2,859,471 | 55 - 59 | 3,902,344 | 4,373,875 | 0.9255 | 0.9709 |
| 60 - 64 | 2,153,209 | 2,447,720 | 60 - 64 | 3,041,034 | 3,468,085 | 0.8903 | 0.9510 |
| 65 - 69 | 1,639,325 | 1,941,781 | 65 - 69 | 2,224,065 | 2,616,745 | 0.8603 | 0.9151 |
| 70 - 74 | 1,229,329 | 1,512,973 | 70 - 74 | 1,667,373 | 2,074,264 | 0.7744 | 0.8474 |
| 75 - 79 | 780,571 | 999,016 | 75 - 79 | 1,090,518 | 1,472,930 | 0.6652 | 0.7585 |
| 80 - 84 | 428,501 | 607,533 | 80 - 84 | 668,623 | 998,349 | 0.5439 | 0.6599 |
| 85 - 89 | 208,088 | 326,783 | 85 - 89 | 310,759 | 508,724 | 0.3981 | 0.5092 |
| 90 - 94 | 65,117 | 115,309 | 90 - 94 | 114,964 | 211,595 | 0.2683 | 0.3483 |
| 95 - 99 | 19,221 | 36,977 | 95 - 99 | 31,529 | 66,806 | 0.1515 | 0.2044 |
| 100 + | 10,423 | 14,153 | 100 + | 7,247 | 16,989 | | |

# Cohort survival ratios – Example (3)

**Step 3:** Calculation of life table survival rates based on the expected level of mortality

nSx= ( nLx+a / nLx)

**Step 4:** Calculation of cohort survival ratios (nRx)

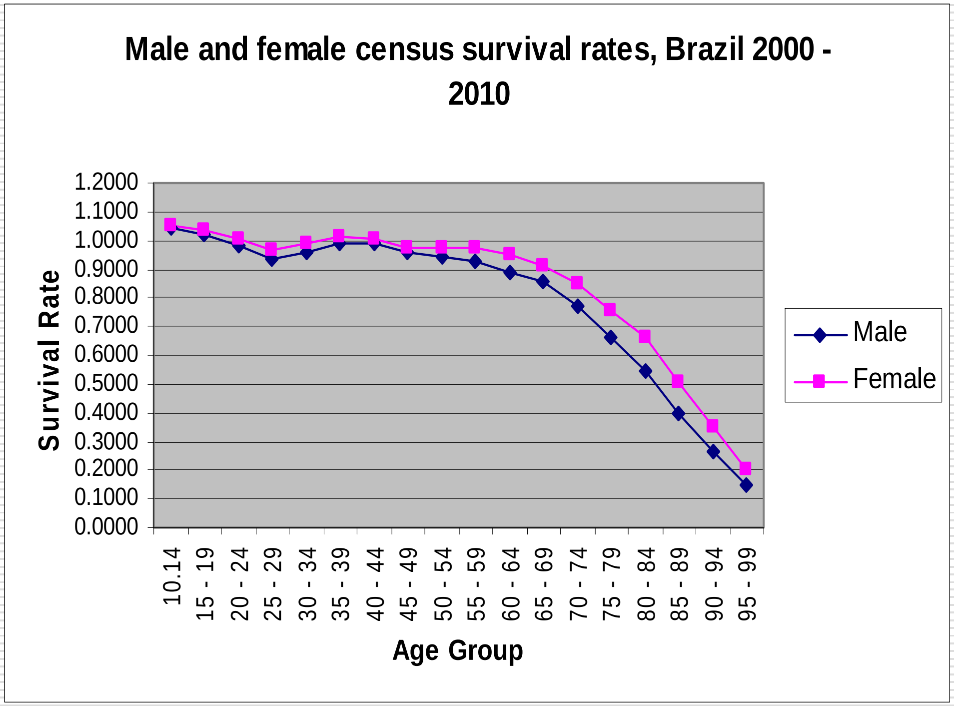


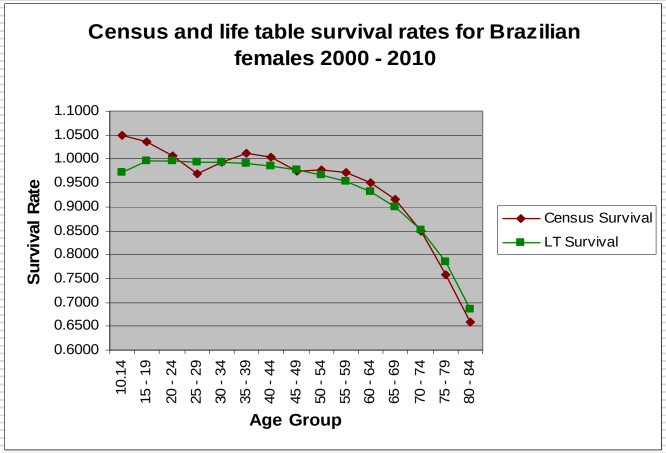
Brazil

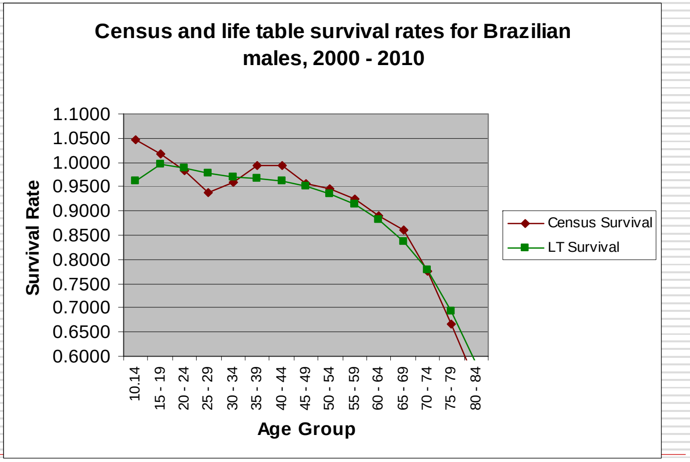
B07 Population by age, sex and urban/rural residence

| | 1 Aug 2000 | | | | 1 Aug 2010 | | Cohort Survival Ratio 2000 - 2010 | | Life Table Survival Rate | | nRx | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | | | Male | Female | Male | Female | Male | Female | Male | Female |
| 0 - 4 | 8,326,926 | 8,048,802 | | 0 - 4 | 7,016,987 | 6,779,172 | | | | | | |
| 5 - 9 | 8,402,353 | 8,139,974 | | 5 - 9 | 7,624,144 | 7,345,231 | | | | | | |
| 10.14 | 8,777,639 | 8,570,428 | | 10.14 | 8,725,413 | 8,441,348 | 1.0479 | 1.0488 | 0.9623 | 0.9722 | 1.0889 | 1.0788 |
| 15 - 19 | 9,019,130 | 8,920,685 | | 15 - 19 | 8,558,868 | 8,432,002 | 1.0186 | 1.0359 | 0.9954 | 0.9970 | 1.0234 | 1.0389 |
| 20 - 24 | 8,048,218 | 8,093,297 | | 20 - 24 | 8,630,227 | 8,614,963 | 0.9832 | 1.0052 | 0.9887 | 0.9960 | 0.9944 | 1.0092 |
| 25 - 29 | 6,814,328 | 7,035,337 | | 25 - 29 | 8,460,995 | 8,643,418 | 0.9381 | 0.9689 | 0.9771 | 0.9940 | 0.9601 | 0.9748 |
| 30 - 34 | 6,363,983 | 6,664,961 | | 30 - 34 | 7,717,657 | 8,026,855 | 0.9589 | 0.9918 | 0.9703 | 0.9920 | 0.9883 | 0.9998 |
| 35 - 39 | 5,955,875 | 6,305,654 | | 35 - 39 | 6,766,665 | 7,121,916 | 0.9930 | 1.0123 | 0.9668 | 0.9892 | 1.0271 | 1.0233 |
| 40 - 44 | 5,116,439 | 5,430,255 | | 40 - 44 | 6,320,570 | 6,688,797 | 0.9932 | 1.0036 | 0.9611 | 0.9850 | 1.0334 | 1.0188 |
| 45 - 49 | 4,216,418 | 4,505,123 | | 45 - 49 | 5,692,013 | 6,141,338 | 0.9557 | 0.9739 | 0.9511 | 0.9780 | 1.0048 | 0.9959 |
| 50 - 54 | 3,415,678 | 3,646,923 | | 50 - 54 | 4,834,995 | 5,305,407 | 0.9450 | 0.9770 | 0.9352 | 0.9672 | 1.0105 | 1.0102 |
| 55 - 59 | 2,585,244 | 2,859,471 | | 55 - 59 | 3,902,344 | 4,373,875 | 0.9255 | 0.9709 | 0.9133 | 0.9521 | 1.0134 | 1.0198 |
| 60 - 64 | 2,153,209 | 2,447,720 | | 60 - 64 | 3,041,034 | 3,468,085 | 0.8903 | 0.9510 | 0.8810 | 0.9304 | 1.0105 | 1.0221 |
| 65 - 69 | 1,639,325 | 1,941,781 | | 65 - 69 | 2,224,065 | 2,616,745 | 0.8603 | 0.9151 | 0.8367 | 0.8982 | 1.0282 | 1.0188 |
| 70 - 74 | 1,229,329 | 1,512,973 | | 70 - 74 | 1,667,373 | 2,074,264 | 0.7744 | 0.8474 | 0.7778 | 0.8520 | 0.9956 | 0.9947 |
| 75 - 79 | 780,571 | 999,016 | | 75 - 79 | 1,090,518 | 1,472,930 | 0.6652 | 0.7585 | 0.6930 | 0.7835 | 0.9600 | 0.9681 |
| 80 - 84 | 428,501 | 607,533 | | 80 - 84 | 668,623 | 998,349 | 0.5439 | 0.6599 | 0.5844 | 0.6843 | 0.9307 | 0.9643 |
| 85 - 89 | 208,088 | 326,783 | | 85 - 89 | 310,759 | 508,724 | 0.3981 | 0.5092 | | | | |
| 90 - 94 | 65,117 | 115,309 | | 90 - 94 | 114,964 | 211,595 | 0.2683 | 0.3483 | | | | |
| 95 - 99 | 19,221 | 36,977 | | 95 - 99 | 31,529 | 66,806 | 0.1515 | 0.2044 | | | | |
| 100 + | 10,423 | 14,153 | | 100 + | 7,247 | 16,989 | | | | | | |

# Cohort survival ratios – Example (4)


Male and female census survival rates, Brazil 2000 - 2010


Census and life table survival rates for Brazilian females 2000 - 2010


Census and life table survival rates for Brazilian males, 2000 - 2010

# Cohort Survival: Uses and limitations

- It is a widely applicable approach for examining error in consecutive censuses
- Method requires relatively little information
  - Information on the level of fertility is not required since the method does not assess the coverage of the population born between two censuses
- Method is complicated by migration etc. as discussed
- When only two censuses are available, the method suffers from the limitations shared by many demographic methods, namely difficulties in separating census errors from real irregularities caused by extraordinary events
  - The utility of census survival approaches increases significantly when three or more censuses are available

# Cohort Component Method

# Census coverage

United Nations Workshop on Census Data Evaluation for English Speaking African Countries
Kampala, Uganda
12 – 16 November 2012

# Overview of cohort-component method

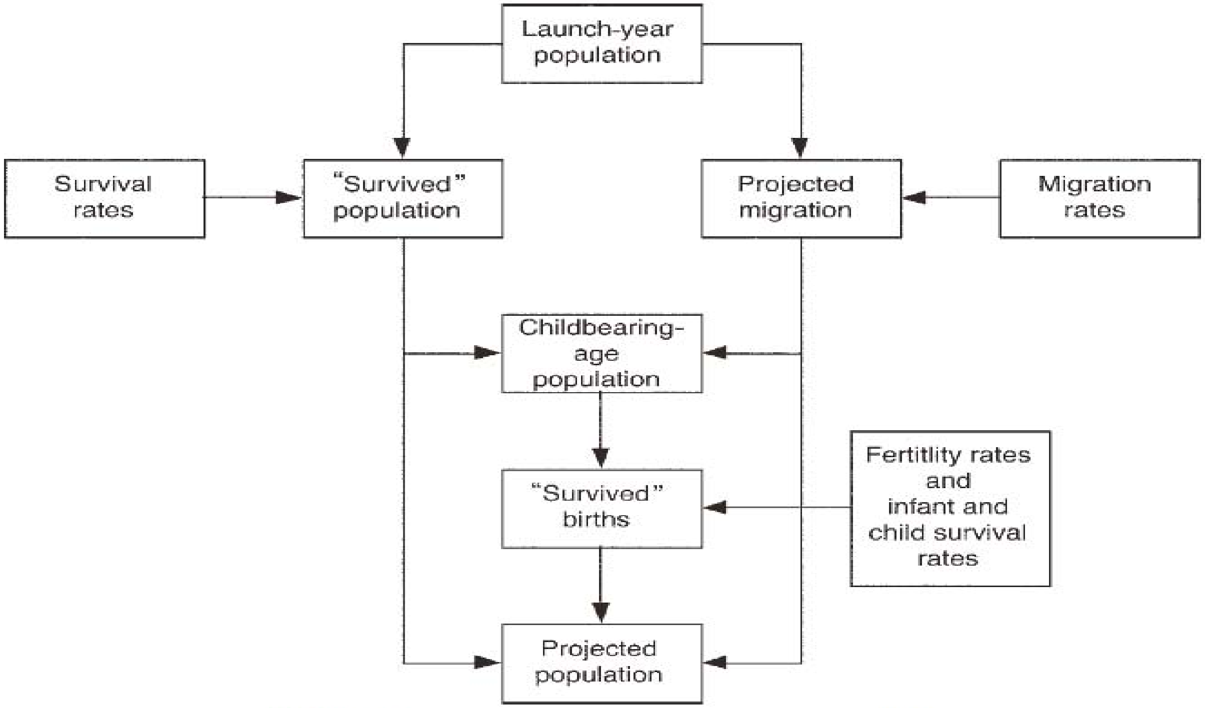


FIGURE 21.2 Overview of the cohort-component method for a projection interval.
Source: Smith *et al*. (2001, p. 47).

Smith, S., J. Tayman, and D. A. Swanson. 2001. *State and Local Population Projections: Methodology and Analysis*. New York: Kluwer Academic/Plenum Publishers.

# Cohort component method

- ☐ In this approach, the population enumerated in the first census is projected to the reference date of the second census based on estimated levels and age schedules of fertility, mortality and migration during the intercensal period
- ☐ The expected population from the projection is then compared with the actual population enumerated in the second census
- ☐ Data for intercensal births, deaths and migration are taken from estimates and/or assumptions regarding the level and age schedules of these parameters rather than directly available data based on registration systems

Source: U.S. Census Bureau, 1985. *Evaluating Censuses of Population and Housing*

# Cohort component method – data required

1. The population enumerated in two censuses by age and sex
2. Age specific fertility rates for women aged 15 to 49 (in 5-year age groups), assumed to represent the level and age structure of fertility during the intercensal period
3. Life table survival rates for males and females, assumed to be representative of mortality conditions during the intercensal period
4. An estimate of sex ratio at birth
5. Estimates of the level and age pattern of net international migration during the intercensal period if the level of net migration is substantial

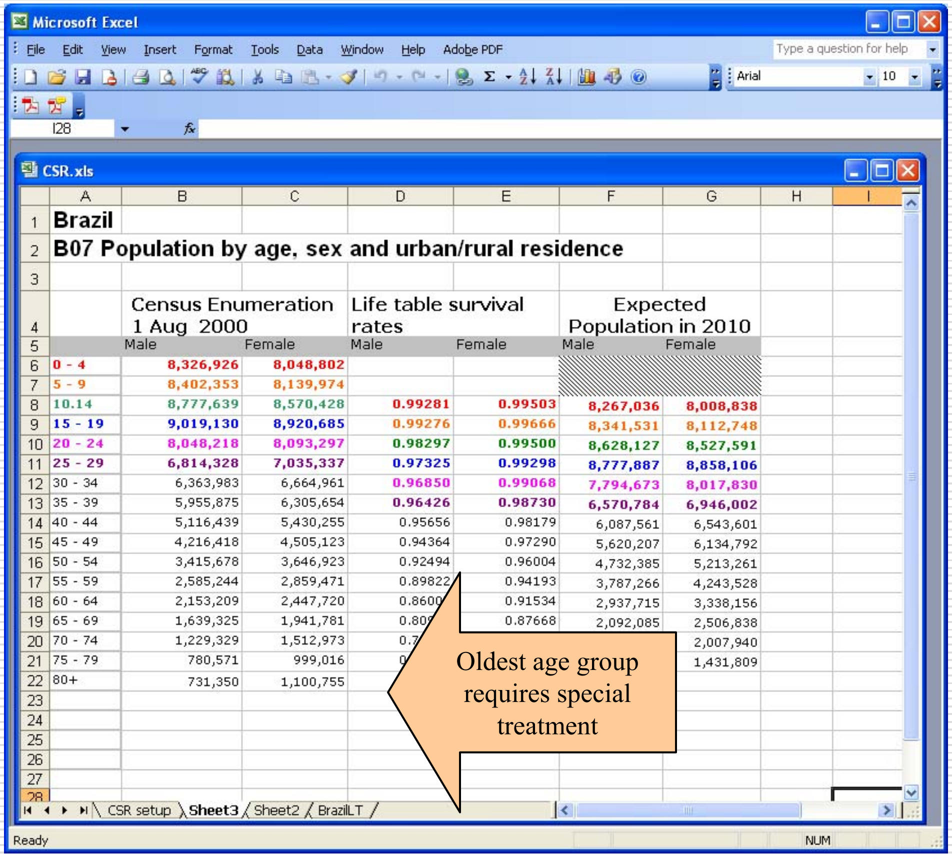
# Cohort component method – overview of computational steps

1. "Survive" the age distribution at the initial census to the time of the second census
    1. Multiply each age group population by life table survival rates
    2. Open-ended interval requires special handling
2. Make any necessary adjustments for migration
3. Calculate the number of births during the period
    1. Average initial and projected population for each age group between 15 – 49 to estimate mid-period female population
    2. Apply age-specific birth rates to these populations to generate total numbers of births during time period
    3. Apply sex ratio factor to get female and male births from total births
4. Apply life table survivorship to these births to determine number that survive to time of the second census
5. Compare the estimated female population by age group with the enumerated female population

# Cohort component method – Step 1 (survive initial age distribution)

1. “Survive” the age distribution at the initial census to the time of the second census

$${}_nS_x = \frac{{}_nL_{x+a}}{{}_nL_x}$$



**Brazil**

**B07 Population by age, sex and urban/rural residence**

| | Census Enumeration 1 Aug 2000 | | Life table survival rates | | Expected Population in 2010 | |
|---|---|---|---|---|---|---|
| | Male | Female | Male | Female | Male | Female |
| **0 - 4** | **8,326,926** | **8,048,802** | | | ▨ | ▨ |
| **5 - 9** | **8,402,353** | **8,139,974** | | | ▨ | ▨ |
| **10.14** | **8,777,639** | **8,570,428** | **0.99281** | **0.99503** | **8,267,036** | **8,008,838** |
| **15 - 19** | **9,019,130** | **8,920,685** | **0.99276** | **0.99666** | **8,341,531** | **8,112,748** |
| **20 - 24** | **8,048,218** | **8,093,297** | **0.98297** | **0.99500** | **8,628,127** | **8,527,591** |
| **25 - 29** | **6,814,328** | **7,035,337** | **0.97325** | **0.99298** | **8,777,887** | **8,858,106** |
| 30 - 34 | 6,363,983 | 6,664,961 | **0.96850** | **0.99068** | **7,794,673** | **8,017,830** |
| 35 - 39 | 5,955,875 | 6,305,654 | **0.96426** | **0.98730** | **6,570,784** | **6,946,002** |
| 40 - 44 | 5,116,439 | 5,430,255 | 0.95656 | 0.98179 | 6,087,561 | 6,543,601 |
| 45 - 49 | 4,216,418 | 4,505,123 | 0.94364 | 0.97290 | 5,620,207 | 6,134,792 |
| 50 - 54 | 3,415,678 | 3,646,923 | 0.92494 | 0.96004 | 4,732,385 | 5,213,261 |
| 55 - 59 | 2,585,244 | 2,859,471 | 0.89822 | 0.94193 | 3,787,266 | 4,243,528 |
| 60 - 64 | 2,153,209 | 2,447,720 | 0.8600 | 0.91534 | 2,937,715 | 3,338,156 |
| 65 - 69 | 1,639,325 | 1,941,781 | 0.80 | 0.87668 | 2,092,085 | 2,506,838 |
| 70 - 74 | 1,229,329 | 1,512,973 | 0.7 | | | 2,007,940 |
| 75 - 79 | 780,571 | 999,016 | 0 | | | 1,431,809 |
| 80+ | 731,350 | 1,100,755 | | | | |

Oldest age group requires special treatment

# Cohort component method – Step 1 (survive initial age distribution)

2. For the oldest age category (open-ended)

$$ {}_wS_x = \frac{{}_wT_{x+a}}{{}_wT_x} $$

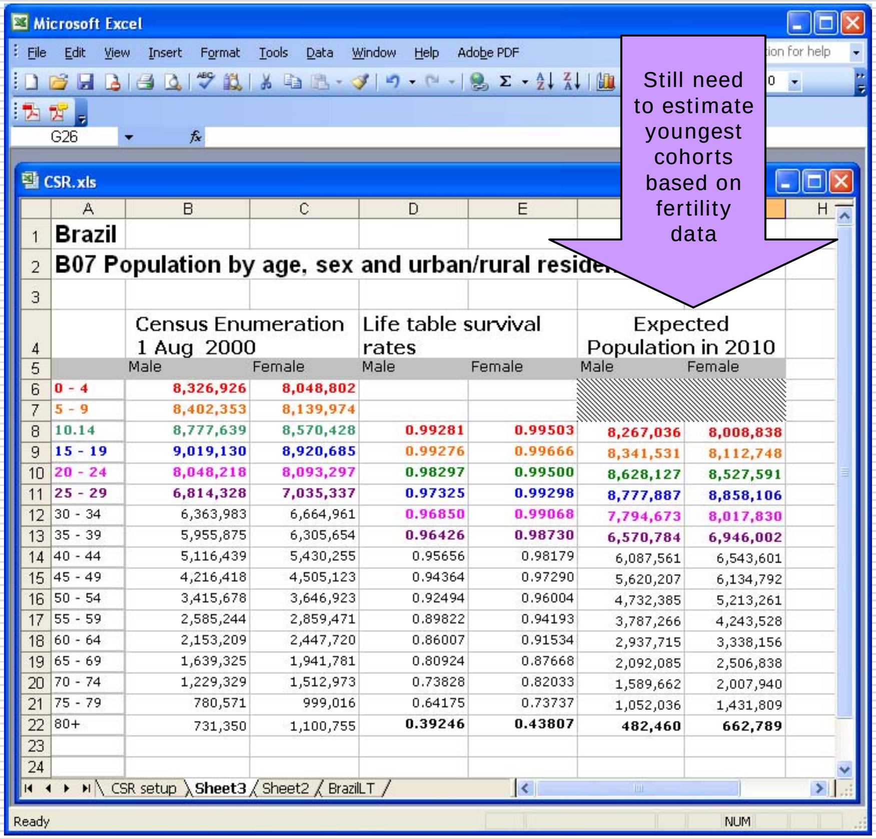
w= the oldest age attainable in the population

a= the length of the projection interval

${}_wS_x$= the life table survival ratio for the population aged x and above

${}_wT_x$= the number of life table persons lived at ages x and above

${}_wT_{x+a}$= the number of life table person-years lived at ages x+a and above



Still need to estimate youngest cohorts based on fertility data

**Brazil**

**B07 Population by age, sex and urban/rural reside**

| | Census Enumeration 1 Aug 2000 | | Life table survival rates | | Expected Population in 2010 | |
|---|---|---|---|---|---|---|
| | Male | Female | Male | Female | Male | Female |
| **0 - 4** | **8,326,926** | **8,048,802** | | | ▨ | ▨ |
| **5 - 9** | **8,402,353** | **8,139,974** | | | ▨ | ▨ |
| **10.14** | **8,777,639** | **8,570,428** | **0.99281** | **0.99503** | **8,267,036** | **8,008,838** |
| **15 - 19** | **9,019,130** | **8,920,685** | **0.99276** | **0.99666** | **8,341,531** | **8,112,748** |
| **20 - 24** | **8,048,218** | **8,093,297** | **0.98297** | **0.99500** | **8,628,127** | **8,527,591** |
| **25 - 29** | **6,814,328** | **7,035,337** | **0.97325** | **0.99298** | **8,777,887** | **8,858,106** |
| 30 - 34 | 6,363,983 | 6,664,961 | **0.96850** | **0.99068** | **7,794,673** | **8,017,830** |
| 35 - 39 | 5,955,875 | 6,305,654 | **0.96426** | **0.98730** | **6,570,784** | **6,946,002** |
| 40 - 44 | 5,116,439 | 5,430,255 | 0.95656 | 0.98179 | 6,087,561 | 6,543,601 |
| 45 - 49 | 4,216,418 | 4,505,123 | 0.94364 | 0.97290 | 5,620,207 | 6,134,792 |
| 50 - 54 | 3,415,678 | 3,646,923 | 0.92494 | 0.96004 | 4,732,385 | 5,213,261 |
| 55 - 59 | 2,585,244 | 2,859,471 | 0.89822 | 0.94193 | 3,787,266 | 4,243,528 |
| 60 - 64 | 2,153,209 | 2,447,720 | 0.86007 | 0.91534 | 2,937,715 | 3,338,156 |
| 65 - 69 | 1,639,325 | 1,941,781 | 0.80924 | 0.87668 | 2,092,085 | 2,506,838 |
| 70 - 74 | 1,229,329 | 1,512,973 | 0.73828 | 0.82033 | 1,589,662 | 2,007,940 |
| 75 - 79 | 780,571 | 999,016 | 0.64175 | 0.73737 | 1,052,036 | 1,431,809 |
| 80+ | 731,350 | 1,100,755 | **0.39246** | **0.43807** | **482,460** | **662,789** |

# Cohort component method – Step 2 (adjust for migration)

- If net international migration is substantial, the “survived” cohort population must be adjusted to reflect the effects of migration
  - The introduction of net migrants by age group at the mid-point of the projection period and the survival of net migrants to the end of the period:

$$
{}_n\hat{M}_{x+i} = \frac{1}{4}\,{}_nM_x\left(1+{}_nS_x\right)+\frac{1}{4}\,{}_nM_{x+i}\left(1+{}_nS_{x+i}\right)
$$

- ***Assumptions:*** i) An equal distribution of net migrants across years of the intercensal period, ii) Migrants have the same fertility and mortality level as the enumerated population

# Cohort component method – Step 3 (calculate births)

1. Calculate the average number of women in each childbearing age group (15 – 49) during the intercensal period in order to estimate the number of births during the projection period

$$ {}_{n}\overline{P}_{x} = \frac{{}_{n}P^{0}{}_{x} + {}_{n}P^{1}{}_{x}}{2} $$

${}_{n}\overline{P}_{x}$ = average number of females aged x to x+n in the projection period

${}_{n}P^{0}{}_{x}$ = number of females aged x to x+n at the beginning of the projection period

${}_{n}P^{1}{}_{x}$ = projected number of females aged x to x+n at the end of the projection period

# Cohort component method – Step 3 (calculate births)

## 2. Calculate total births during the period

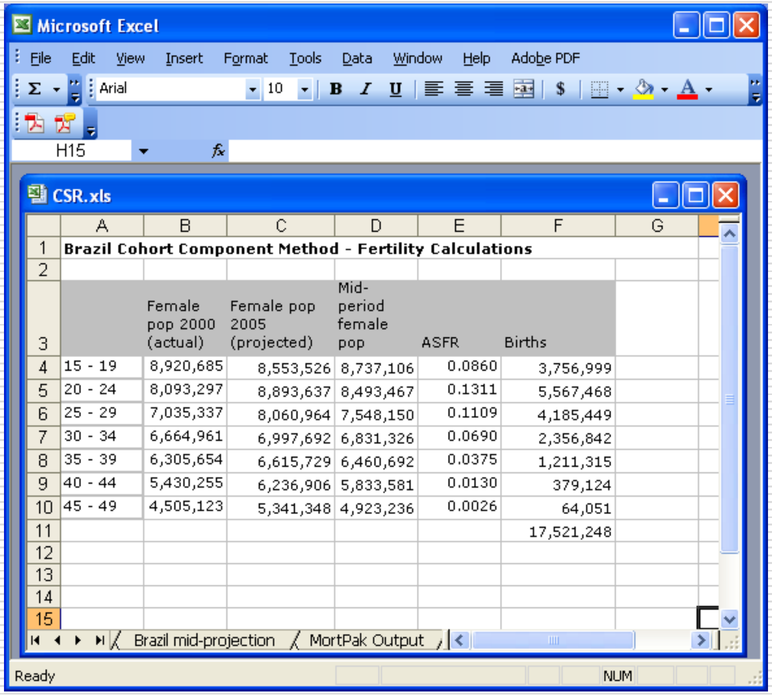
$$B = \sum_{X=15}^{49} ({}_n\bar{P}_x * {}_nf_x) \text{ for 1-year projection}$$

$$B = 5 * \sum_{X=15}^{49} ({}_n\bar{P}_x * {}_nf_x) \text{ for 5-year projection period}$$

$B$ = the estimated number of births during the projection period

${}_n\bar{P}_x$ = the average number of women in the age group x to x+n years during the projection period

${}_nf_x$ = the age specific fertility rate (per woman) for women age x to x+n years during the projection period



Brazil Cohort Component Method - Fertility Calculations

| | Female pop 2000 (actual) | Female pop 2005 (projected) | Mid-period female pop | ASFR | Births |
|---|---|---|---|---|---|
| 15 - 19 | 8,920,685 | 8,553,526 | 8,737,106 | 0.0860 | 3,756,999 |
| 20 - 24 | 8,093,297 | 8,893,637 | 8,493,467 | 0.1311 | 5,567,468 |
| 25 - 29 | 7,035,337 | 8,060,964 | 7,548,150 | 0.1109 | 4,185,449 |
| 30 - 34 | 6,664,961 | 6,997,692 | 6,831,326 | 0.0690 | 2,356,842 |
| 35 - 39 | 6,305,654 | 6,615,729 | 6,460,692 | 0.0375 | 1,211,315 |
| 40 - 44 | 5,430,255 | 6,236,906 | 5,833,581 | 0.0130 | 379,124 |
| 45 - 49 | 4,505,123 | 5,341,348 | 4,923,236 | 0.0026 | 64,051 |
| | | | | | 17,521,248 |

# Cohort component method – Step 3 (calculate births)

☐ 3. Calculate proportion of male and female births

$$B^f = 1 - \frac{SRB}{1 + SRB}$$

$B^f = 1 - (1.05)/(1+1.05) = .488$

$= .488*17,521,248 = 8,550,369$

$$B^m = \frac{SRB}{1 + SRB}$$

$B^m = (1.05)/(1+1.05) = .512$

$= .512*17,521,248 = 8,970,879$

# Cohort component method – Step 4 (survive intercensal births)

- 3. Apply life table survivorship to these births to determine number that survive to time of the second census

- ${}_5P^f{}_0 = B^f * {}_5S_0$

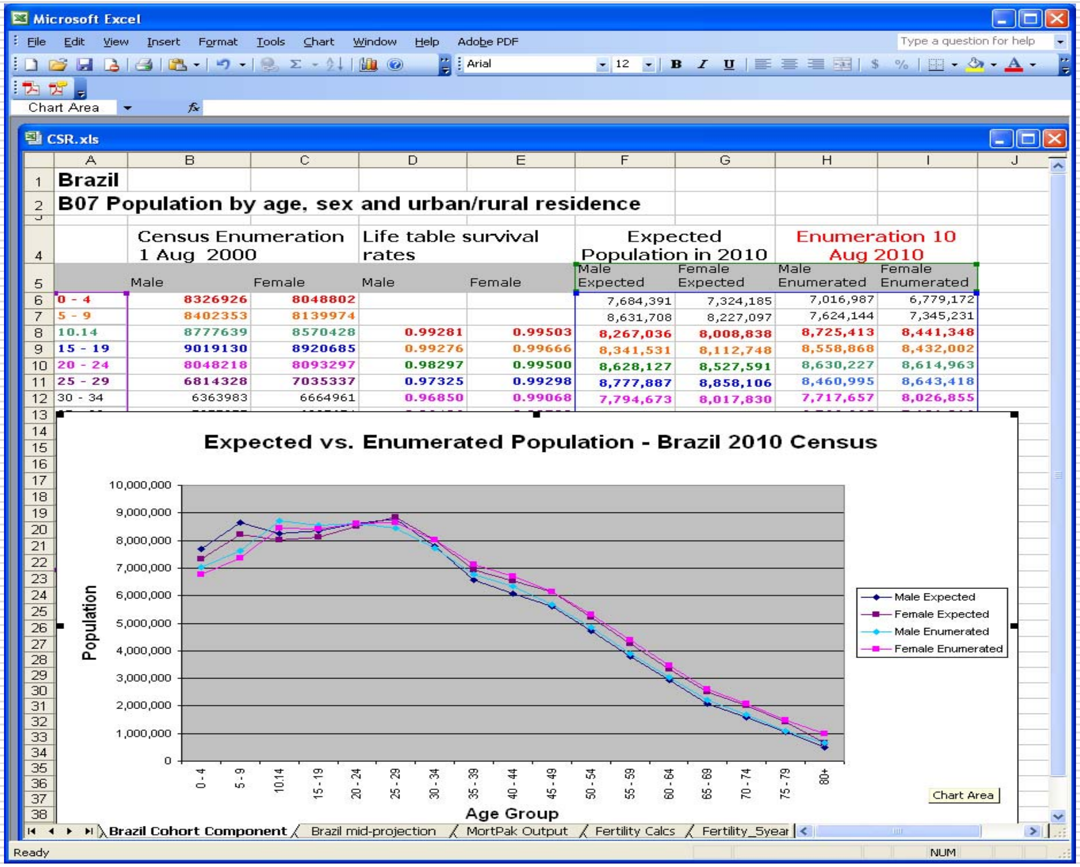
${}_5S_0 = 484{,}129/500{,}000 = .968$

${}_5P^f{}_0 = .968 * 8{,}550{,}369 = 8{,}276{,}757$

- $${}_5S_0 = \frac{{}_5L_0}{5 \text{ X } l_0}$$

# Cohort component method – Step 5 (compare with enumerated population)

- Final step in procedure is to compare the compare the enumerated population by age and sex in the second population with the expected population

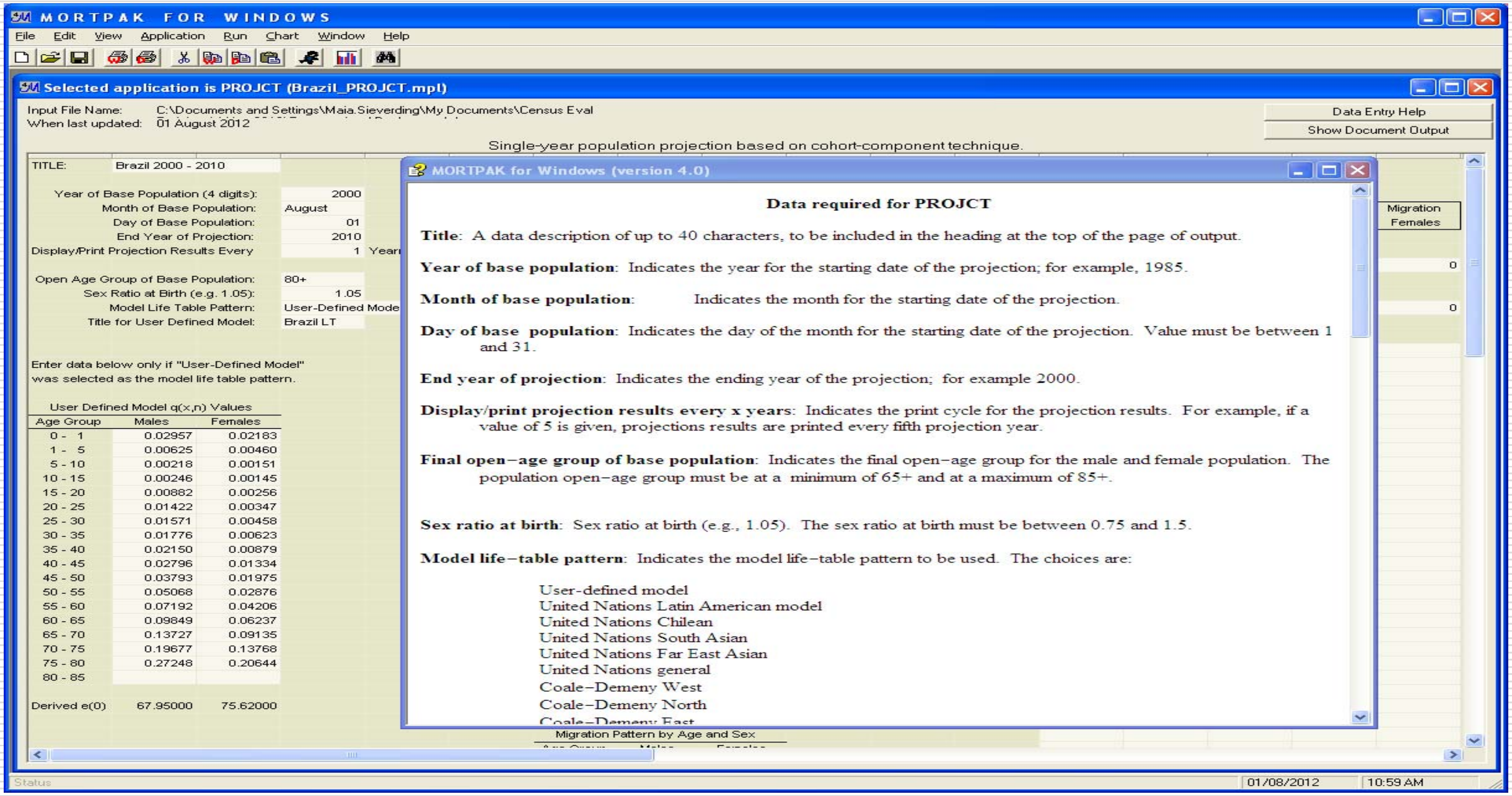


**Brazil**

**B07 Population by age, sex and urban/rural residence**

| | Census Enumeration 1 Aug 2000 | | Life table survival rates | | Expected Population in 2010 | | Enumeration 10 Aug 2010 | |
|---|---|---|---|---|---|---|---|---|
| | Male | Female | Male | Female | Male Expected | Female Expected | Male Enumerated | Female Enumerated |
| 0 - 4 | 8326926 | 8048802 | | | 7,684,391 | 7,324,185 | 7,016,987 | 6,779,172 |
| 5 - 9 | 8402353 | 8139974 | | | 8,631,708 | 8,227,097 | 7,624,144 | 7,345,231 |
| 10.14 | 8777639 | 8570428 | 0.99281 | 0.99503 | 8,267,036 | 8,008,838 | 8,725,413 | 8,441,348 |
| 15 - 19 | 9019130 | 8920685 | 0.99276 | 0.99666 | 8,341,531 | 8,112,748 | 8,558,868 | 8,432,002 |
| 20 - 24 | 8048218 | 8093297 | 0.98297 | 0.99500 | 8,628,127 | 8,527,591 | 8,630,227 | 8,614,963 |
| 25 - 29 | 6814328 | 7035337 | 0.97325 | 0.99298 | 8,777,887 | 8,858,106 | 8,460,995 | 8,643,418 |
| 30 - 34 | 6363983 | 6664961 | 0.96850 | 0.99068 | 7,794,673 | 8,017,830 | 7,717,657 | 8,026,855 |

# Cohort component method in MortPak (1)



Input File Name: C:\Documents and Settings\Maia.Sieverding\My Documents\Census Eval
When last updated: 01 August 2012

Single-year population projection based on cohort-component technique.

TITLE: Brazil 2000 - 2010

Year of Base Population (4 digits): 2000
Month of Base Population: August
Day of Base Population: 01
End Year of Projection: 2010
Display/Print Projection Results Every 1 Year

Open Age Group of Base Population: 80+
Sex Ratio at Birth (e.g. 1.05): 1.05
Model Life Table Pattern: User-Defined Mode
Title for User Defined Model: Brazil LT

Enter data below only if "User-Defined Model" was selected as the model life table pattern.

User Defined Model q(x,n) Values

| Age Group | Males | Females |
|---|---|---|
| 0 - 1 | 0.02957 | 0.02183 |
| 1 - 5 | 0.00625 | 0.00460 |
| 5 - 10 | 0.00218 | 0.00151 |
| 10 - 15 | 0.00246 | 0.00145 |
| 15 - 20 | 0.00882 | 0.00256 |
| 20 - 25 | 0.01422 | 0.00347 |
| 25 - 30 | 0.01571 | 0.00458 |
| 30 - 35 | 0.01776 | 0.00623 |
| 35 - 40 | 0.02150 | 0.00879 |
| 40 - 45 | 0.02796 | 0.01334 |
| 45 - 50 | 0.03793 | 0.01975 |
| 50 - 55 | 0.05068 | 0.02876 |
| 55 - 60 | 0.07192 | 0.04206 |
| 60 - 65 | 0.09849 | 0.06237 |
| 65 - 70 | 0.13727 | 0.09135 |
| 70 - 75 | 0.19677 | 0.13768 |
| 75 - 80 | 0.27248 | 0.20644 |
| 80 - 85 | | |
| Derived e(0) | 67.95000 | 75.62000 |

**Data required for PROJCT**

**Title**: A data description of up to 40 characters, to be included in the heading at the top of the page of output.

**Year of base population**: Indicates the year for the starting date of the projection; for example, 1985.

**Month of base population**: Indicates the month for the starting date of the projection.

**Day of base population**: Indicates the day of the month for the starting date of the projection. Value must be between 1 and 31.

**End year of projection**: Indicates the ending year of the projection; for example 2000.

**Display/print projection results every x years**: Indicates the print cycle for the projection results. For example, if a value of 5 is given, projections results are printed every fifth projection year.

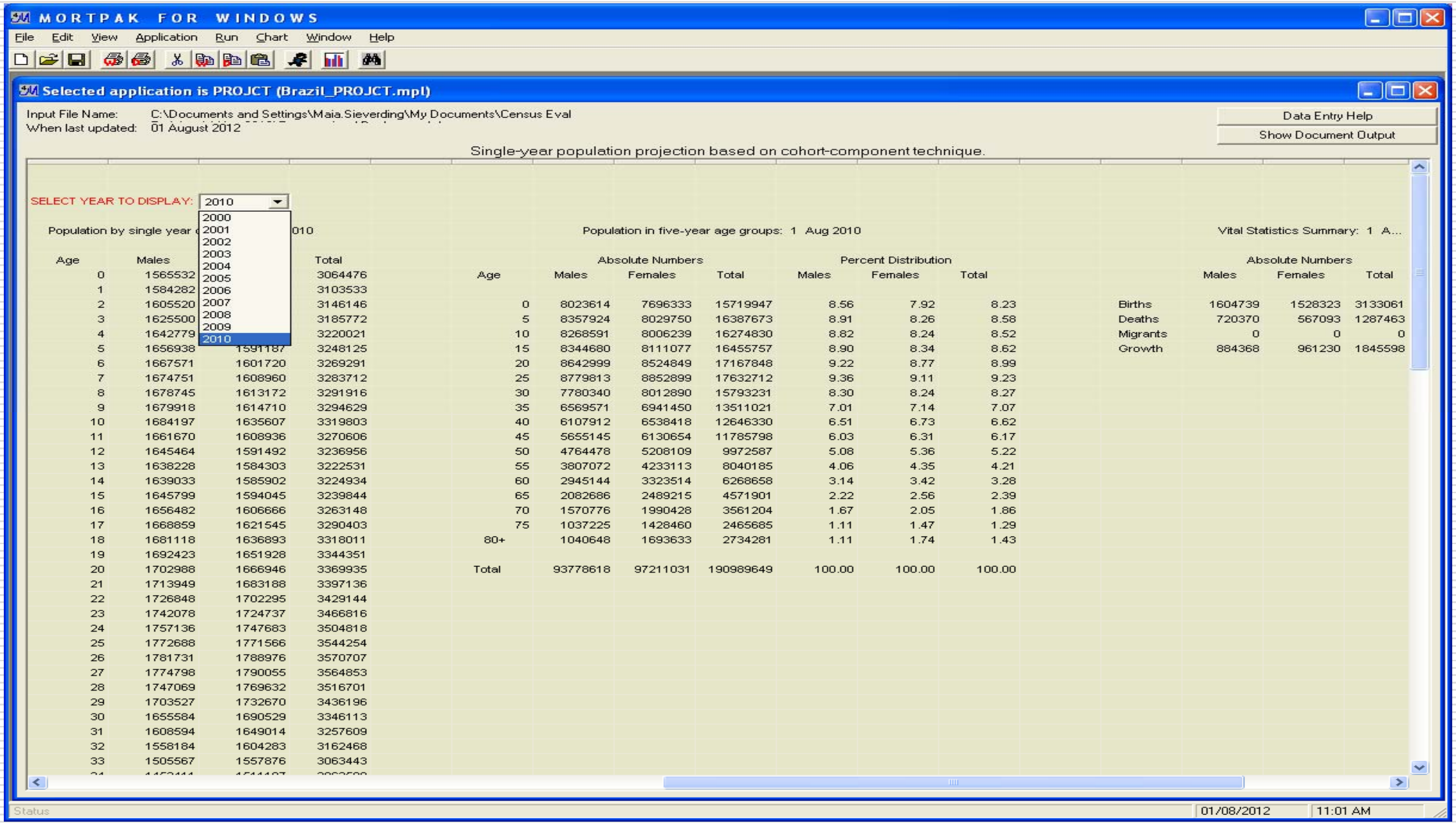
**Final open−age group of base population**: Indicates the final open−age group for the male and female population. The population open−age group must be at a minimum of 65+ and at a maximum of 85+.

**Sex ratio at birth**: Sex ratio at birth (e.g., 1.05). The sex ratio at birth must be between 0.75 and 1.5.

**Model life−table pattern**: Indicates the model life−table pattern to be used. The choices are:

User-defined model
United Nations Latin American model
United Nations Chilean
United Nations South Asian
United Nations Far East Asian
United Nations general
Coale−Demeny West
Coale−Demeny North
Coale−Demeny East

# Cohort component method in MortPak (2)



Selected application is PROJCT (Brazil_PROJCT.mpl)

Input File Name: C:\Documents and Settings\Maia.Sieverding\My Documents\Census Eval
When last updated: 01 August 2012

Single-year population projection based on cohort-component technique.

SELECT YEAR TO DISPLAY: 2010

Population by single year 010

| Age | Males | | Total |
|---|---|---|---|
| 0 | 1565532 | [illegible] | 3064476 |
| 1 | 1584282 | [illegible] | 3103533 |
| 2 | 1605520 | [illegible] | 3146146 |
| 3 | 1625500 | [illegible] | 3185772 |
| 4 | 1642779 | [illegible] | 3220021 |
| 5 | 1656938 | 1591187 | 3248125 |
| 6 | 1667571 | 1601720 | 3269291 |
| 7 | 1674751 | 1608960 | 3283712 |
| 8 | 1678745 | 1613172 | 3291916 |
| 9 | 1679918 | 1614710 | 3294629 |
| 10 | 1684197 | 1635607 | 3319803 |
| 11 | 1661670 | 1608936 | 3270606 |
| 12 | 1645464 | 1591492 | 3236956 |
| 13 | 1638228 | 1584303 | 3222531 |
| 14 | 1639033 | 1585902 | 3224934 |
| 15 | 1645799 | 1594045 | 3239844 |
| 16 | 1656482 | 1606666 | 3263148 |
| 17 | 1668859 | 1621545 | 3290403 |
| 18 | 1681118 | 1636893 | 3318011 |
| 19 | 1692423 | 1651928 | 3344351 |
| 20 | 1702988 | 1666946 | 3369935 |
| 21 | 1713949 | 1683188 | 3397136 |
| 22 | 1726848 | 1702295 | 3429144 |
| 23 | 1742078 | 1724737 | 3466816 |
| 24 | 1757136 | 1747683 | 3504818 |
| 25 | 1772688 | 1771566 | 3544254 |
| 26 | 1781731 | 1788976 | 3570707 |
| 27 | 1774798 | 1790055 | 3564853 |
| 28 | 1747069 | 1769632 | 3516701 |
| 29 | 1703527 | 1732670 | 3436196 |
| 30 | 1655584 | 1690529 | 3346113 |
| 31 | 1608594 | 1649014 | 3257609 |
| 32 | 1558184 | 1604283 | 3162468 |
| 33 | 1505567 | 1557876 | 3063443 |

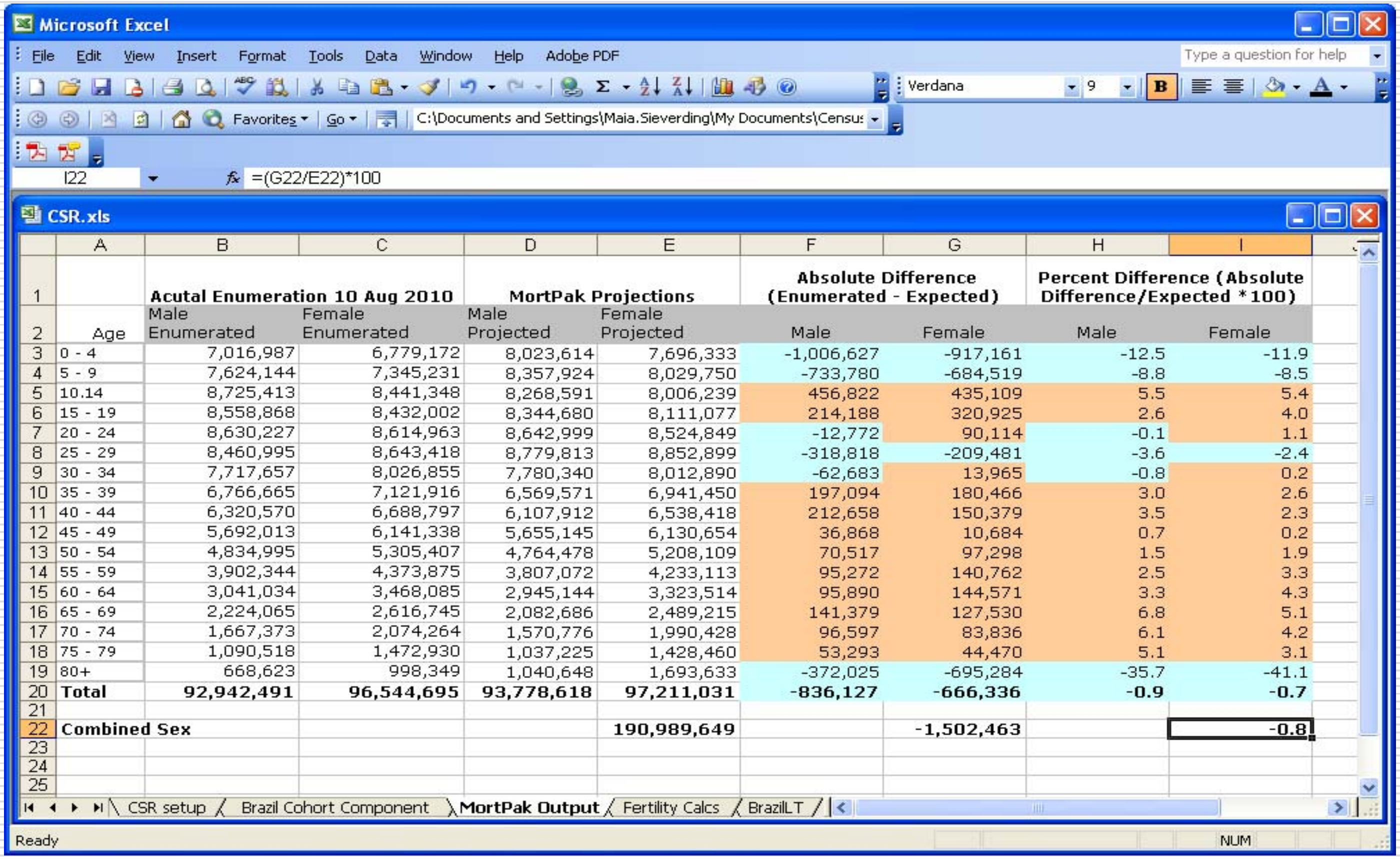
Population in five-year age groups: 1 Aug 2010

| | Absolute Numbers | | | Percent Distribution | | |
|---|---|---|---|---|---|---|
| Age | Males | Females | Total | Males | Females | Total |
| 0 | 8023614 | 7696333 | 15719947 | 8.56 | 7.92 | 8.23 |
| 5 | 8357924 | 8029750 | 16387673 | 8.91 | 8.26 | 8.58 |
| 10 | 8268591 | 8006239 | 16274830 | 8.82 | 8.24 | 8.52 |
| 15 | 8344680 | 8111077 | 16455757 | 8.90 | 8.34 | 8.62 |
| 20 | 8642999 | 8524849 | 17167848 | 9.22 | 8.77 | 8.99 |
| 25 | 8779813 | 8852899 | 17632712 | 9.36 | 9.11 | 9.23 |
| 30 | 7780340 | 8012890 | 15793231 | 8.30 | 8.24 | 8.27 |
| 35 | 6569571 | 6941450 | 13511021 | 7.01 | 7.14 | 7.07 |
| 40 | 6107912 | 6538418 | 12646330 | 6.51 | 6.73 | 6.62 |
| 45 | 5655145 | 6130654 | 11785798 | 6.03 | 6.31 | 6.17 |
| 50 | 4764478 | 5208109 | 9972587 | 5.08 | 5.36 | 5.22 |
| 55 | 3807072 | 4233113 | 8040185 | 4.06 | 4.35 | 4.21 |
| 60 | 2945144 | 3323514 | 6268658 | 3.14 | 3.42 | 3.28 |
| 65 | 2082686 | 2489215 | 4571901 | 2.22 | 2.56 | 2.39 |
| 70 | 1570776 | 1990428 | 3561204 | 1.67 | 2.05 | 1.86 |
| 75 | 1037225 | 1428460 | 2465685 | 1.11 | 1.47 | 1.29 |
| 80+ | 1040648 | 1693633 | 2734281 | 1.11 | 1.74 | 1.43 |
| Total | 93778618 | 97211031 | 190989649 | 100.00 | 100.00 | 100.00 |

Vital Statistics Summary: 1 A...

| | Absolute Numbers | | |
|---|---|---|---|
| | Males | Females | Total |
| Births | 1604739 | 1528323 | 3133061 |
| Deaths | 720370 | 567093 | 1287463 |
| Migrants | 0 | 0 | 0 |
| Growth | 884368 | 961230 | 1845598 |

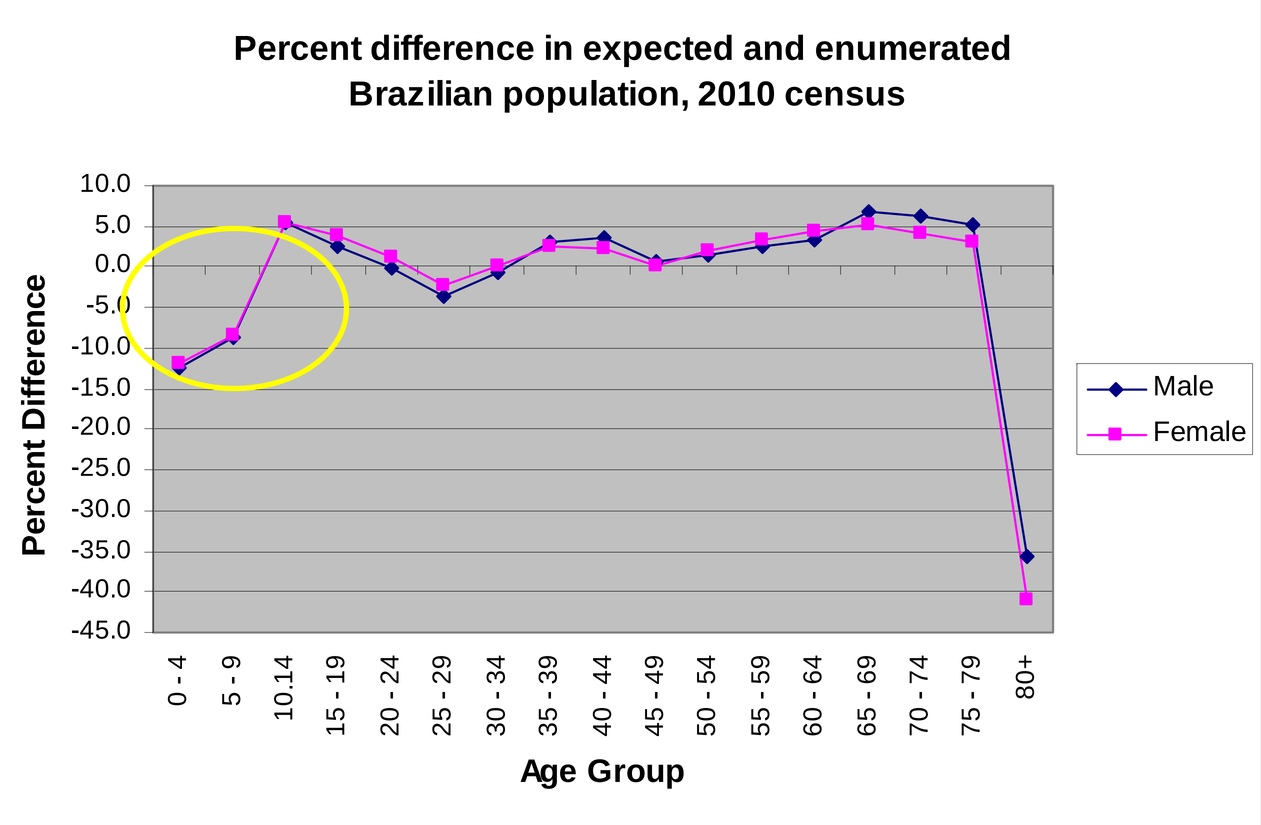
# Brazil Example with MortPak - Results



I22 =(G22/E22)*100

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | | **Acutal Enumeration 10 Aug 2010** | | **MortPak Projections** | | **Absolute Difference (Enumerated - Expected)** | | **Percent Difference (Absolute Difference/Expected *100)** | |
| 2 | Age | Male Enumerated | Female Enumerated | Male Projected | Female Projected | Male | Female | Male | Female |
| 3 | 0 - 4 | 7,016,987 | 6,779,172 | 8,023,614 | 7,696,333 | -1,006,627 | -917,161 | -12.5 | -11.9 |
| 4 | 5 - 9 | 7,624,144 | 7,345,231 | 8,357,924 | 8,029,750 | -733,780 | -684,519 | -8.8 | -8.5 |
| 5 | 10.14 | 8,725,413 | 8,441,348 | 8,268,591 | 8,006,239 | 456,822 | 435,109 | 5.5 | 5.4 |
| 6 | 15 - 19 | 8,558,868 | 8,432,002 | 8,344,680 | 8,111,077 | 214,188 | 320,925 | 2.6 | 4.0 |
| 7 | 20 - 24 | 8,630,227 | 8,614,963 | 8,642,999 | 8,524,849 | -12,772 | 90,114 | -0.1 | 1.1 |
| 8 | 25 - 29 | 8,460,995 | 8,643,418 | 8,779,813 | 8,852,899 | -318,818 | -209,481 | -3.6 | -2.4 |
| 9 | 30 - 34 | 7,717,657 | 8,026,855 | 7,780,340 | 8,012,890 | -62,683 | 13,965 | -0.8 | 0.2 |
| 10 | 35 - 39 | 6,766,665 | 7,121,916 | 6,569,571 | 6,941,450 | 197,094 | 180,466 | 3.0 | 2.6 |
| 11 | 40 - 44 | 6,320,570 | 6,688,797 | 6,107,912 | 6,538,418 | 212,658 | 150,379 | 3.5 | 2.3 |
| 12 | 45 - 49 | 5,692,013 | 6,141,338 | 5,655,145 | 6,130,654 | 36,868 | 10,684 | 0.7 | 0.2 |
| 13 | 50 - 54 | 4,834,995 | 5,305,407 | 4,764,478 | 5,208,109 | 70,517 | 97,298 | 1.5 | 1.9 |
| 14 | 55 - 59 | 3,902,344 | 4,373,875 | 3,807,072 | 4,233,113 | 95,272 | 140,762 | 2.5 | 3.3 |
| 15 | 60 - 64 | 3,041,034 | 3,468,085 | 2,945,144 | 3,323,514 | 95,890 | 144,571 | 3.3 | 4.3 |
| 16 | 65 - 69 | 2,224,065 | 2,616,745 | 2,082,686 | 2,489,215 | 141,379 | 127,530 | 6.8 | 5.1 |
| 17 | 70 - 74 | 1,667,373 | 2,074,264 | 1,570,776 | 1,990,428 | 96,597 | 83,836 | 6.1 | 4.2 |
| 18 | 75 - 79 | 1,090,518 | 1,472,930 | 1,037,225 | 1,428,460 | 53,293 | 44,470 | 5.1 | 3.1 |
| 19 | 80+ | 668,623 | 998,349 | 1,040,648 | 1,693,633 | -372,025 | -695,284 | -35.7 | -41.1 |
| 20 | **Total** | **92,942,491** | **96,544,695** | **93,778,618** | **97,211,031** | **-836,127** | **-666,336** | **-0.9** | **-0.7** |
| 21 | | | | | | | | | |
| 22 | **Combined Sex** | | | | **190,989,649** | | **-1,502,463** | | **-0.8** |
| 23 | | | | | | | | | |
| 24 | | | | | | | | | |
| 25 | | | | | | | | | |

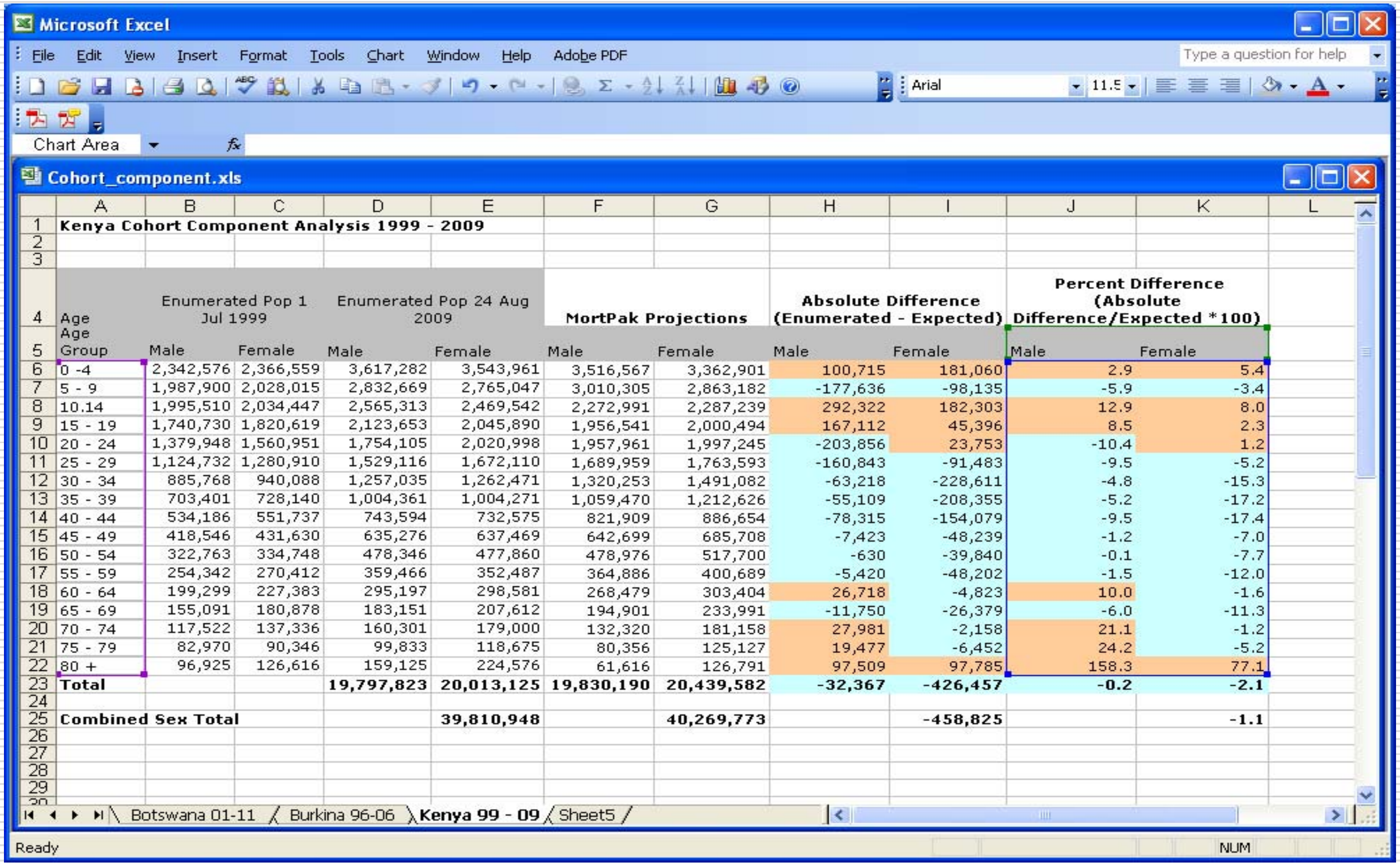
# Brazil Example with MortPak - Results



# Main findings from Brazil example

- Suggests underenumeration by about 1.5 million people, or 0.8% of the population
- Significant underenumeration of youngest two age groups, particularly children 0 - 4
- Some overenumeration of 10 – 14 year olds – could be a result of the underenumeration of this group (as 0 – 4 year olds) in the original census in 2000
- Seeing same under-enumeration of 25 – 29 year olds of both sexes as when we calculated by hand – needs to be explored
- Consistent but fairly low level of overenumeration of adults age 35 – 39 to 75 – 79
  - Could potentially indicate in-migration, ideally want to incorporate migration data
- Significant overenumeration of older people
  - Very likely that there is age exaggeration
  - Also might consider that our life table is not accurate for these ages

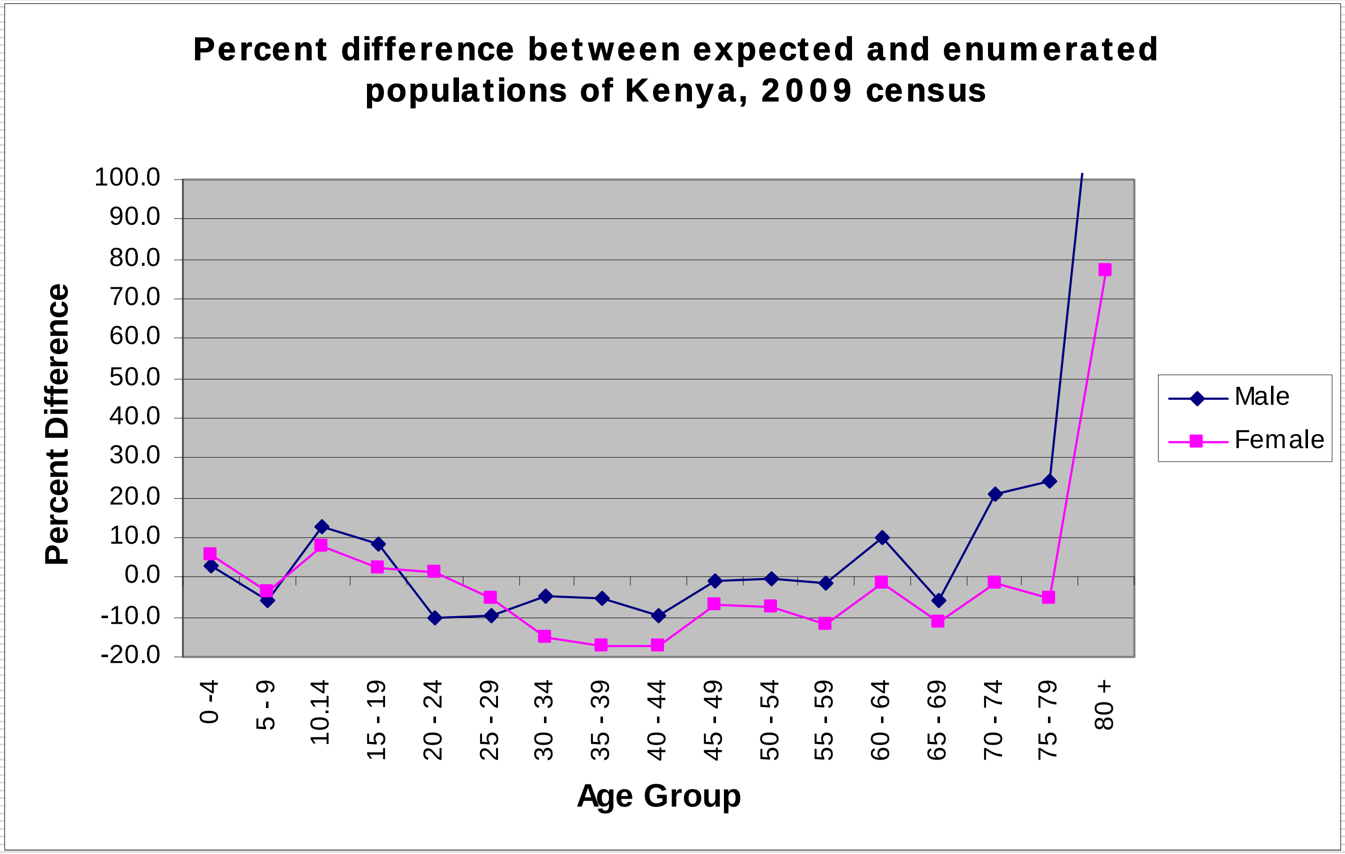
# Kenya (1) – 2009 census



**Kenya Cohort Component Analysis 1999 - 2009**

| Age<br>Age Group | Enumerated Pop 1 Jul 1999 |  | Enumerated Pop 24 Aug 2009 |  | MortPak Projections |  | Absolute Difference (Enumerated - Expected) |  | Percent Difference (Absolute Difference/Expected *100) |  |
|---|---|---|---|---|---|---|---|---|---|---|
|  | Male | Female | Male | Female | Male | Female | Male | Female | Male | Female |
| 0 -4 | 2,342,576 | 2,366,559 | 3,617,282 | 3,543,961 | 3,516,567 | 3,362,901 | 100,715 | 181,060 | 2.9 | 5.4 |
| 5 - 9 | 1,987,900 | 2,028,015 | 2,832,669 | 2,765,047 | 3,010,305 | 2,863,182 | -177,636 | -98,135 | -5.9 | -3.4 |
| 10.14 | 1,995,510 | 2,034,447 | 2,565,313 | 2,469,542 | 2,272,991 | 2,287,239 | 292,322 | 182,303 | 12.9 | 8.0 |
| 15 - 19 | 1,740,730 | 1,820,619 | 2,123,653 | 2,045,890 | 1,956,541 | 2,000,494 | 167,112 | 45,396 | 8.5 | 2.3 |
| 20 - 24 | 1,379,948 | 1,560,951 | 1,754,105 | 2,020,998 | 1,957,961 | 1,997,245 | -203,856 | 23,753 | -10.4 | 1.2 |
| 25 - 29 | 1,124,732 | 1,280,910 | 1,529,116 | 1,672,110 | 1,689,959 | 1,763,593 | -160,843 | -91,483 | -9.5 | -5.2 |
| 30 - 34 | 885,768 | 940,088 | 1,257,035 | 1,262,471 | 1,320,253 | 1,491,082 | -63,218 | -228,611 | -4.8 | -15.3 |
| 35 - 39 | 703,401 | 728,140 | 1,004,361 | 1,004,271 | 1,059,470 | 1,212,626 | -55,109 | -208,355 | -5.2 | -17.2 |
| 40 - 44 | 534,186 | 551,737 | 743,594 | 732,575 | 821,909 | 886,654 | -78,315 | -154,079 | -9.5 | -17.4 |
| 45 - 49 | 418,546 | 431,630 | 635,276 | 637,469 | 642,699 | 685,708 | -7,423 | -48,239 | -1.2 | -7.0 |
| 50 - 54 | 322,763 | 334,748 | 478,346 | 477,860 | 478,976 | 517,700 | -630 | -39,840 | -0.1 | -7.7 |
| 55 - 59 | 254,342 | 270,412 | 359,466 | 352,487 | 364,886 | 400,689 | -5,420 | -48,202 | -1.5 | -12.0 |
| 60 - 64 | 199,299 | 227,383 | 295,197 | 298,581 | 268,479 | 303,404 | 26,718 | -4,823 | 10.0 | -1.6 |
| 65 - 69 | 155,091 | 180,878 | 183,151 | 207,612 | 194,901 | 233,991 | -11,750 | -26,379 | -6.0 | -11.3 |
| 70 - 74 | 117,522 | 137,336 | 160,301 | 179,000 | 132,320 | 181,158 | 27,981 | -2,158 | 21.1 | -1.2 |
| 75 - 79 | 82,970 | 90,346 | 99,833 | 118,675 | 80,356 | 125,127 | 19,477 | -6,452 | 24.2 | -5.2 |
| 80 + | 96,925 | 126,616 | 159,125 | 224,576 | 61,616 | 126,791 | 97,509 | 97,785 | 158.3 | 77.1 |
| **Total** |  |  | **19,797,823** | **20,013,125** | **19,830,190** | **20,439,582** | **-32,367** | **-426,457** | **-0.2** | **-2.1** |
|  |  |  |  |  |  |  |  |  |  |  |
| **Combined Sex Total** |  |  |  | **39,810,948** |  | **40,269,773** |  | **-458,825** |  | **-1.1** |

# Kenya (2)


Percent difference between expected and enumerated populations of Kenya, 2009 census
Percent Difference
100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
-10.0
-20.0
0 -4
5 - 9
10.14
15 - 19
20 - 24
25 - 29
30 - 34
35 - 39
40 - 44
45 - 49
50 - 54
55 - 59
60 - 64
65 - 69
70 - 74
75 - 79
80 +
Age Group
Male
Female

# Kenya (3)

- Overall suggests net undercount of 1.1%, about 460,000 people
- Most of undercount is coming from males aged 20 – 45 and females aged 25 – 59
  - Migration may account for some of this difference
  - The lifetable used (based on Kenya 1999) census may not accurately represent changing mortality conditions over the 10 year period due to the HIV/AIDS epidemic
  - There may be a ‘true’ undercount of these age-sex groups

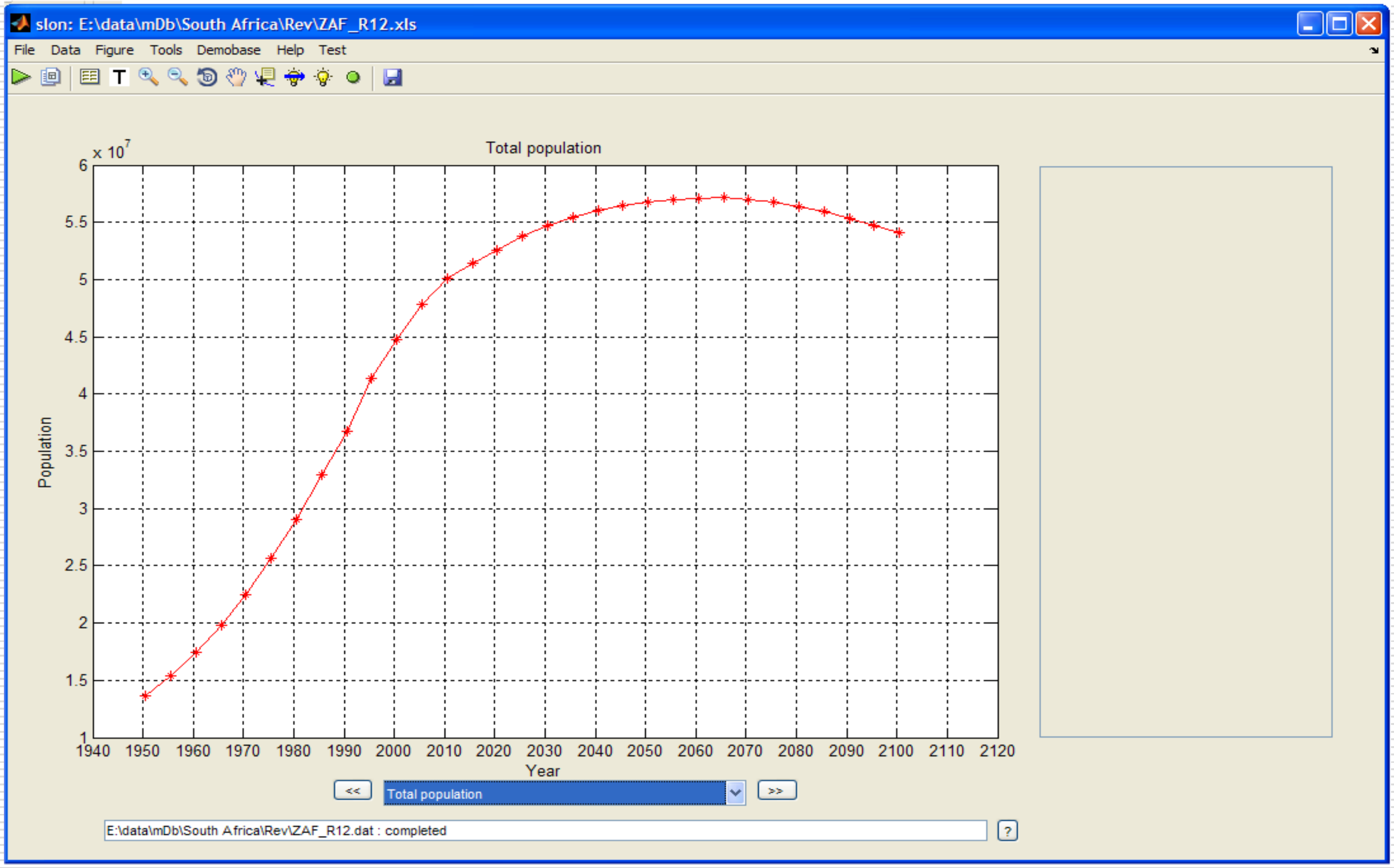
# Cohort component method – uses and limitations

- It is applicable when registration data are not-existent or deficient to such an extent that satisfactory adjustment is not possible
- Sufficient information to derive estimates of fertility and mortality levels should be available
  - Mortality estimates can be complicated by HIV/AIDS – with a generalized epidemic, one life table is generally not sufficient to model mortality patterns over a 10 year period
- Lack of information on international migration is often a problematic issue when applying this method
- In case where sufficient information exists to derive reliable estimates of demographic parameters, the method is perhaps the most powerful among the alternative demographic approaches for the evaluation of censuses, since it provides age and sex specific estimates of net census error

# Tools

- In addition to MortPak, the DemProj module of Spectrum can be used for population projections
- http://www.futuresinstitute.org/spectrum.aspx
  - DemProj is recommended for projection in contexts in which HIV/AIDS prevalence exceeds a few percent – better modeling of mortality conditions
  - Requires more data input, including prevalence and treatment estimates for HIV/AIDS
  - Data input options somewhat less flexible than MortPak

# SLON application, UN Population Division



E:\data\mDb\South Africa\Rev\ZAF_R12.dat : completed

# SLON application

Features

- Projections for 5-year periods and 5-year age groups
- Projection input parameters are stored in Excel files
- Used for preparation of World Population Prospects
- Includes a number of templates for interpolation, smoothing, mortality scatter plots, model life table validation etc.

Under development

- Projections for single year and single age groups
- Displays of empirical information together with projection input parameters

Contact information: andreev@un.org

# *THANK YOU …..*