Session IX: Special topics

2. Subnational population projections

10 March 2016

Cheryl Sawyer, Lina Bassarsky
Population Estimates and Projections Section

www.unpopulation.org

Materials adapted from United Nations National Workshop on Subnational Population Projections using Census Data , Beijing, China, 2013 and US Census Bureau Subnational Projections Toolkit User's Guide version 2.0.

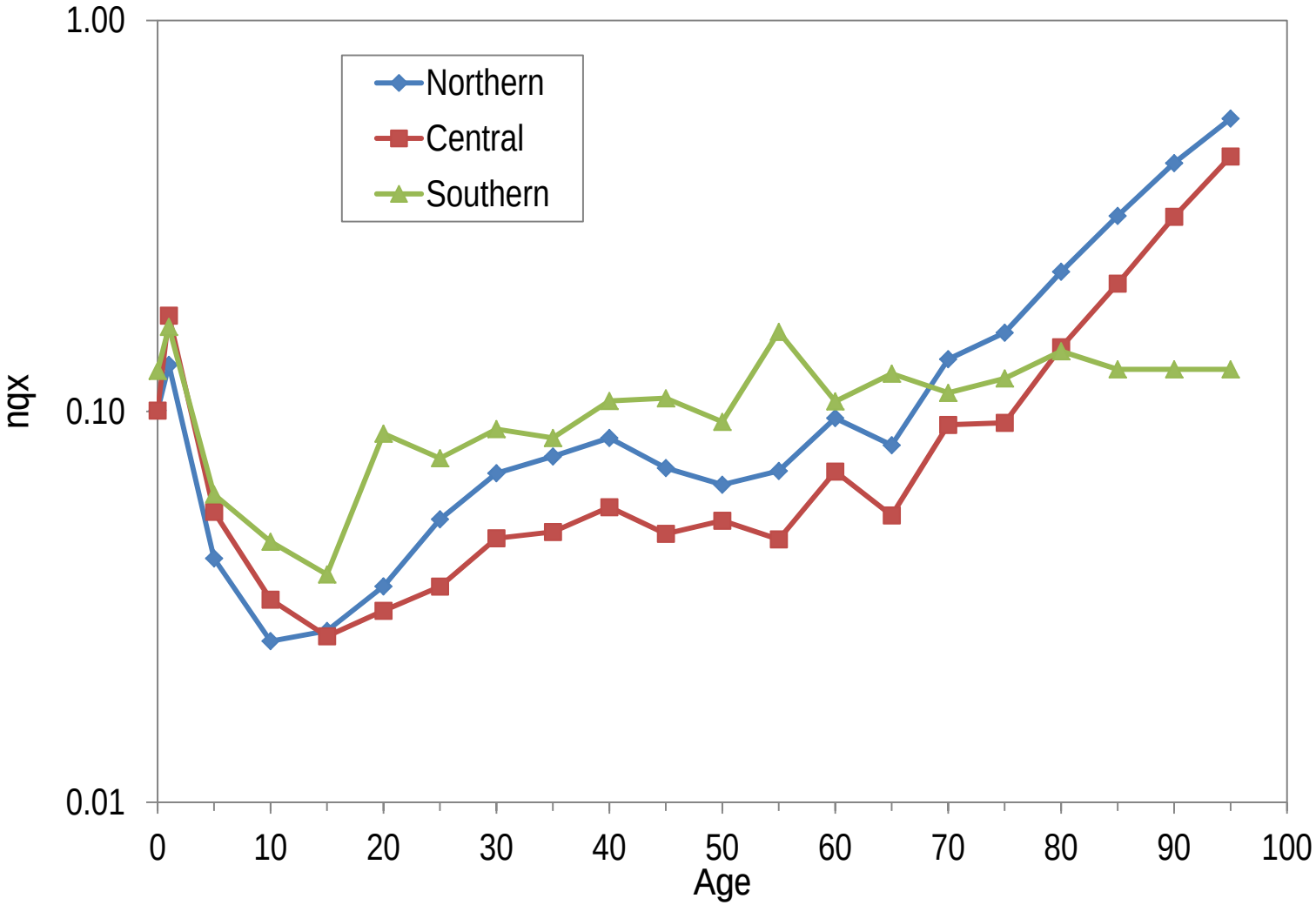
Demand for subnational projections

- Subnational population projections are increasingly in demand. They are an important information source for regional and local policy makers.
- Subnational population projections are also important for countries that are large and have a diverse population. In such circumstances, national population projections may not effectively and accurately reflect the demographic settings in its administrative or regional entities.

Challenges

- Generating sub-national projections that are both internally consistent and consistent with a national projection is more challenging than preparing a national projection.
- Each subnational entity presents the same data problems as the national projection but, in addition, preserving consistency across regions and dealing with data problems that are often more severe than those at the national level adds to the challenge.

Regional probabilities of dying ($_nq_x$)
Malawi, 1998 Census

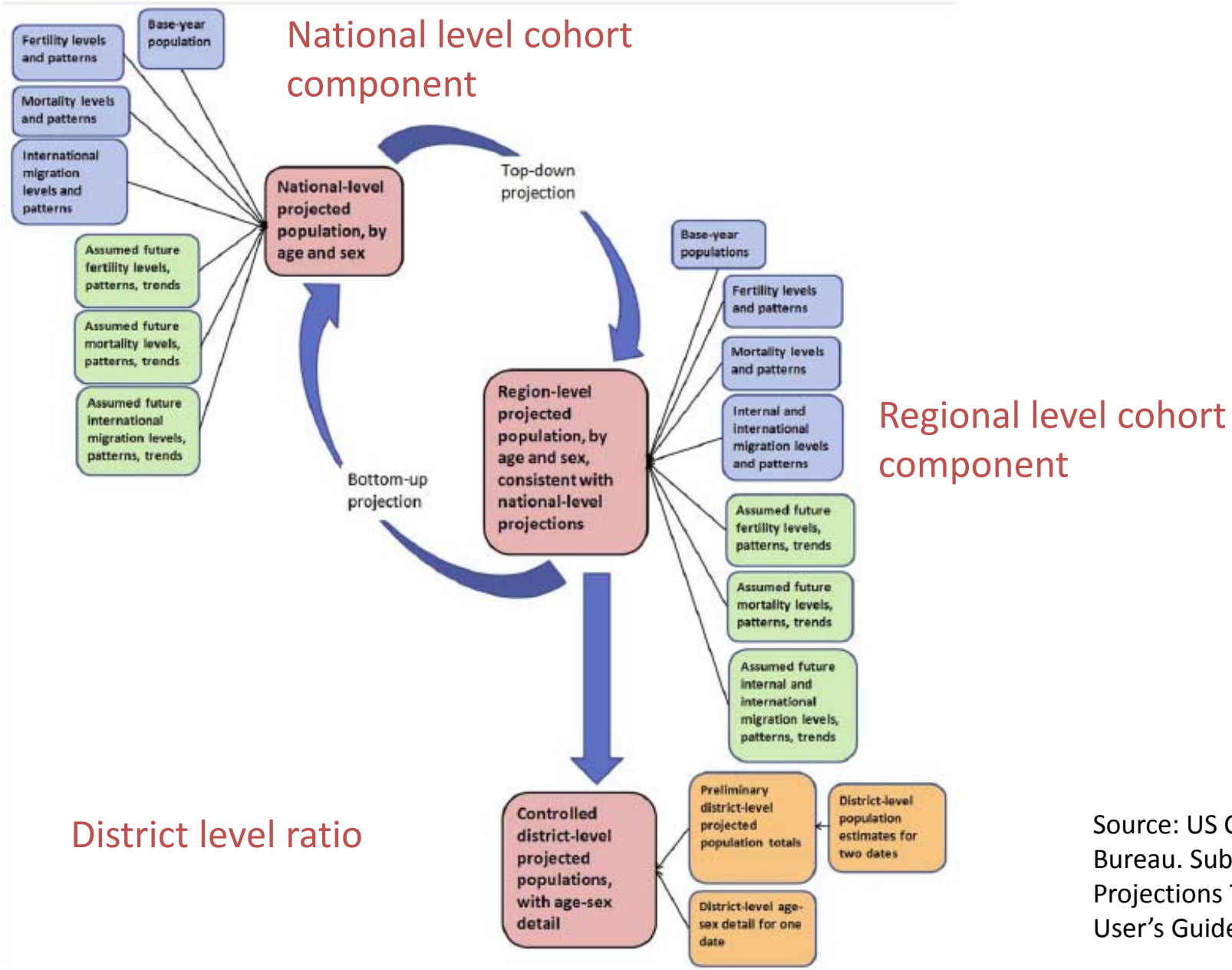


Bottom-up vs. top-down

- Bottom-up
  - Cohort-component projections made for subnational areas
  - National projection created as sum of subnational projections
- Top-down
  - Controls subnational area projections to the national total
  - Two possible approaches for projecting subnational areas:
    - Mathematical extrapolation (ratio) methods
    - Cohort component method

Combination of these two approaches common

Illustration of integrated approach combining cohort component and trend methods



Source: US Census Bureau. Subnational Projections Toolkit User's Guide v. 2.0

Mathematical extrapolation (ratio method) vs cohort-component

- Decision of which method to apply for lower administrative units has to be made on a country-by-country basis
- Consideration to population size, data availability and data quality
- Data appropriate for the cohort-component method may be limited at the regional level

The ratio method

- The ratio method is applied mainly for projecting the population of small areas within a country for which all inputs required by the component method are not always readily available
- Three variants presented here
  - Constant share
  - Shift share
  - Logistic implementation

The ratio method: constant share

- The constant share method holds the smaller area's share of the parent population constant at a certain point in time, usually at the base year.
- Data requirements are light: Population figures for the smaller areas are required for one point in time only, and for the parent area a population projection.
- Like all variations of the ratio method, the constant-share method is usually applied to total populations, sometimes disaggregated by sex

The ratio method: constant share

$\overline{P}_t$ = Total Population at time t (Parent Population)

P_t^i = Sub Population in region i at time t

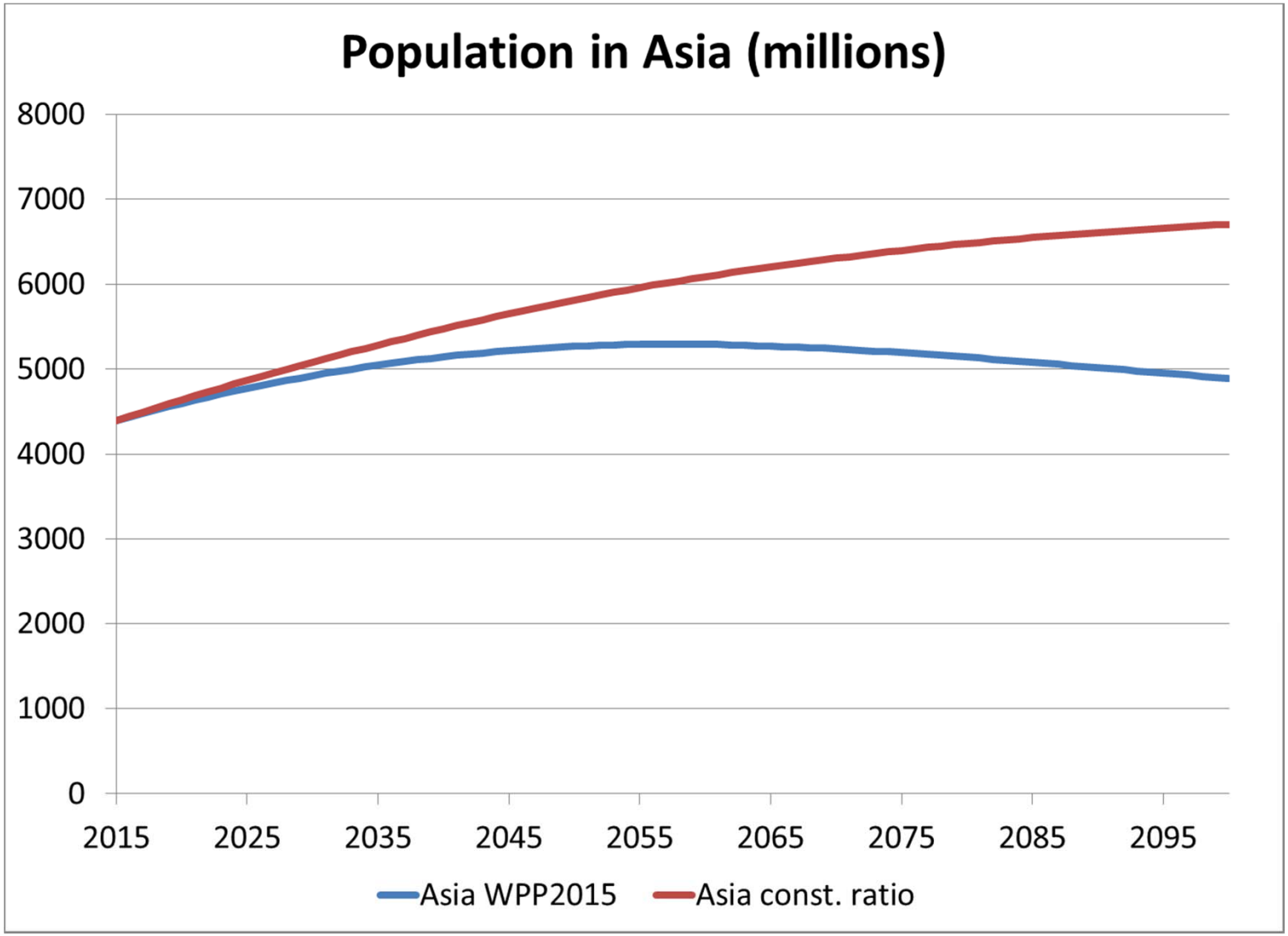
$X_0^i = \dfrac{P_{t=0}^i}{\overline{P}_{t=0}}$ Ratio (child i to parent at time t=0)

$P_t^i = X_0^i * \overline{P}_t$

The ratio method: constant share

Year	2015	2020	2025	2030
World population	7,349 (P_0)	7,758 (P_1)	8,141 (P_2)	8,501 (P_3)
Asian Population	4,393	-	-	-
Ratio of Asia to World, 2015	0.5977 (X_0^i)	-	-	-
Projected Asian population		4,638 ($P_1 \times X_0^i$)	4,867 ($P_2 \times X_0^i$)	5.082 ($P_3 \times X_0^i$)

The ratio method: constant share



- The shift-share method allows for changes in population shares of the smaller areas over time.
- The trends of changing shares may be determined by inspecting past trends or by formulating appropriate assumptions.
- The shift-share method does not guarantee plausible results, so care must be taken and constraints introduced to avoid extremely large subpopulations or even negative population figures.

The ratio method: shift share

$\overline{P}_t$ = Total Population at time t (Parent Population)

P_t^i = Sub Population at time t

t = Year

b = Base year

z = Number of years in the projection period

y = Number of years in the base period

$$P_t^i = \overline{P}_t \left[\frac{P_{t=0}^i}{\overline{P}_{t=0}} + \frac{z}{y} * \left(\frac{P_{t=0}^i}{\overline{P}_{t=0}} - \frac{P_{t=b}^i}{\overline{P}_{t=b}} \right) \right]$$

The ratio method: logistic implementation

- US Census Bureau tool SALGST.xls uses logistic fits to project subarea proportions from two censuses and control to national population projection total

Table 1
COUNTRY: YEARS

Project

B. Total Projected Population

Area	2010.5	2015.5	2020.5	2025.5	2030.5
Total	335,612	561,096	938,076	1,568,334	2,622,039

C. Population by District

	Population		Proportion		Asymptotes		Logistic
District	2000.42	2010.41	2000.42	2010.41	Lower	Upper	growth rate
Total	300,000	332,648	1.0000	1.0000	0	1	
District 1	100,000	100,000	0.3333	0.3006	0	1	-0.0151
District 2	100,000	122,134	0.3333	0.3672	0	1	0.0149
District 3	100,000	110,514	0.3333	0.3322	0	1	-0.0005

The ratio method: logistic implementation

Table 1
COUNTRY: YEARS
D. Projection of the proportion using the logistic formula

District	2010.5	2015.5	2020.5	2025.5	2030.5
Total	1.000	1.001	1.003	1.006	1.010
District 1	0.300	0.285	0.270	0.255	0.241
District 2	0.367	0.385	0.403	0.421	0.439
District 3	0.332	0.332	0.331	0.331	0.330

Result of logistic fit for individual areas

Table 1
COUNTRY: YEARS
E. Adjustment of the unrounded proportions to sum to 1

District	2010.5	2015.5	2020.5	2025.5	2030.5
Total	1.000	1.000	1.000	1.000	1.000
District 1	0.300	0.284	0.269	0.253	0.239
District 2	0.367	0.384	0.401	0.418	0.435
District 3	0.332	0.331	0.330	0.329	0.327

Adjust proportions to sum to 1

Table 1
COUNTRY: YEARS
F. Projected Population by District

District	2010.5	2015.5	2020.5	2025.5	2030.5
Total	335,612	561,096	938,076	1,568,334	2,622,039
District 1	100,797	159,540	252,022	397,359	625,372
District 2	123,321	215,697	376,475	655,718	1,139,714
District 3	111,494	185,859	309,579	515,257	856,953

Projected population by district

Ratio methods: advantages and shortcomings

- Minimum of data
- Ease of use

- Problematic for longer projection periods
- No or little information on the demographic characteristics of the population

- Challenges in the preparation of cohort component inputs for subnational areas
  - Because subnational populations are smaller than the national one, data may show more irregularities and fluctuations. This affects all elements of demographic components: fertility, mortality and migration.
  - If a population is small, its age composition is also subject to random fluctuations.
  - Projections of TFR, life expectancy at subnational level should be consistent with national projection
    - US Census Bureau subnational toolkit has tools for projecting consistent TFR (PROJTFR32) or e_0 (PROJE032) given subnational estimates for base year and projected national trend

- Migration in subnational cohort component projections
  - International migration for national populations is usually integrated into the projection as net migration, that is the balance of immigration and emigration. If migration is of relatively small magnitude, international migration has often even been ignored (e.g. assumed to be zero).
  - Ignoring migration is almost always not possible for subnational projections. The magnitude of internal migration is in most cases larger than international migration, and therefore has significant demographic impact that cannot be ignored.
    - US Census Bureau subnational toolkit has tool for projecting subnational net migration (MIGSUB) using input on population by duration of residence

Spectrum

- Create separate projection files for each subnational area.
- Files are aggregated using Tools→More tools→Aggregate (for details, see section 7.2 of the Spectrum manual)
- Spectrum can be used for explorative subnational projections that can be compared and aggregated.
- However, Spectrum does not (yet) provide consistency mechanisms that guarantee that individual projections are consistent with and add up to the national total.

US Census Bureau Subnational Projections Toolkit

- Well-developed and supported toolkit
- Series of Excel sheets to support preparation of inputs for subnational files
- Uses RUP projection software, RUPAGG to aggregate input files for subnational regions
- Provides workbook for adjusting subnational population by age and sex to national totals

Controlling subarea results to national total

- Projected population totals by age and sex for subunits need to be controlled to national totals for consistency
- Iteratively adjust columns and rows of a two-way table. Called by various terms: iterative proportional fitting, contingency table, two-way raking
- US Census Bureau workbook CTBL32.xls provides ability to adjust up to 32 subpopulations by age

Controlling subarea results to national total

Table
COUNTY: YEAR
Contingency Table Adjustment
A. Initial Population by Subpopulation and Age

Age	Desired total	Row sum	Adj. Factor	Subpop 1	Subpop 2	Subpop 3	Subpop 4	S
0-4	2,488,460	2,662,652	0.9346	1,318,884	696,769	647,000	0	
5-9	2,448,469	2,472,954	0.99					
10-14	2,266,525	2,130,534	1.06					
15-19	2,033,047	2,073,708	0.98					
20-24	1,688,670	1,587,350	1.06					
25-29	1,314,334	1,196,044	1.09					
30-34	1,005,099	1,005,099	1.00					
35-39	875,681	849,411	1.03					
40-44	802,584	818,636	0.98					
45-49	699,060	720,032	0.97					
50-54	561,363	555,749	1.01					
55-59	409,943	401,744	1.02					
60-64	310,059	285,254	1.08					
65-69	223,745	225,982	0.99					
70-74	154,793	147,053	1.05					
75-79	88,353	84,819	1.04					
80+	64,818	58,984	1.09					
Col. sum	17,435,003	17,276,005						
Alternate col. totals		17,276,005						
Adjusted								

Input

Output

Table
COUNTY: YEAR
B. Adjusted Population by Subpopulation and Age

Age	Desired total	Row sum	Adj. Factor	Subpop 1	Subpop 2	Subpop 3	S
All ages	17,435,003	17,435,003		8,761,379	4,505,795	4,167,829	
0-4	2,488,460	2,488,460	1.0000	1,231,072	652,627	604,761	
5-9	2,448,469	2,448,469	1.0000	1,162,115	680,245	606,109	
10-14	2,266,525	2,266,525	1.0000	1,180,182	483,373	602,970	
15-19	2,033,047	2,033,047	1.0000	1,015,242	559,315	458,490	
20-24	1,688,670	1,688,670	1.0000	861,221	468,108	359,341	
25-29	1,314,334	1,314,334	1.0000	706,928	304,017	303,390	
30-34	1,005,099	1,005,099	1.0000	491,886	261,903	251,311	
35-39	875,681	875,681	1.0000	432,783	235,235	207,663	
40-44	802,584	802,584	1.0000	408,658	212,922	181,004	
45-49	699,060	699,060	1.0000	345,700	190,452	162,908	
50-54	561,363	561,363	1.0000	300,202	119,359	141,801	
55-59	409,943	409,943	1.0000	192,163	121,565	96,215	
60-64	310,059	310,059	1.0000	171,693	70,942	67,424	
65-69	223,745	223,745	1.0000	108,410	62,163	53,172	
70-74	154,793	154,793	1.0000	76,485	42,457	35,851	
75-79	88,353	88,353	1.0000	43,201	23,981	21,170	
80+	64,818	64,818	1.0000	33,437	17,133	14,248	

Note: Computed by 5 cycles of row and column raking.

Thank you

Questions?

>> until 11 March:

>> After 11 March: sawyerc@un.org
bassarsky@un.org