**Appendix 2**

**SAS and STATA code**

This document contains SAS and Stata code for computing life expectancy, and decomposing the gap between two populations into age and cause components using Arriaga’s method.

Please cite this code as:

Auger N, Feuillet P, Martel S, Lo E, Barry AD, Harper S. Mortality inequality in populations with equal life expectancy: A practical decomposition method in SAS, Stata and Excel. Annals of Epidemiology 2014; doi: 10.1016/j.annepidem.2014.05.006.

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

SAS Macro

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

This macro requires that you prepare two separate datasets, one for each population, in the following format:

*in1 (e.g., Quebec)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| age | pop1 | death1 | c11 | c12 | … | c1n |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

*in2 (e.g., rest of Canada)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| age | pop2 | death2 | c21 | c22 | … | c2n |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Where:

age = age at the start of the age category (e.g., 0, 1, 5,… , 85, 90). Note that this macro is set up for 20 age groups, with a last age group of 90 years. Should you wish to change the number of age groups, please modify the syntax highlighted in yellow.

pop1 = number of individuals in population 1, by age category

pop2 = number of individuals in population 2, by age category

death1 = number of deaths from all causes in population 1, by age category

death2 = number of deaths from all causes in population 2, by age category

c11 to c1n: number of deaths from each specific cause in population 1, by age category (n causes of death in total)

c21 to c2n: number of deaths from each specific cause in population 2, by age category(n causes of death in total). For example, c25 is the observed number of deaths in population 2 for the fifth cause of death.

Macro %arriaga generates one output file containing life expectancy of the two populations and the decomposition of the life expectancy gap by age and cause.

The macro has five parameters:

1. in1 = name of input dataset for population 1
2. in2 = name of input dataset for population 2
3. c = total number of causes of death
4. output = name of output file
5. path = input and output file path \*/

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* Macro %arriaga \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

**%macro** arriaga (in1, in2, c, output, path);

libname x "&path";

/\* Preparation of the two datasets \*/

data data1; set x.&in1;

data data2; set x.&in2;

%do i=**1** %to **2**;

proc sort data=data&i;

by age;

%end;

data data12;

merge data1 data2;

by age;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

COMPUTE LIFE EXPECTANCY OF TWO POPULATIONS

(Adapted from: Yang Z, Sun X. SAS Macros for generating abridged and cause-eliminated life tables. Cary: SAS Institute Inc.; 2006)

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Compute age-specific death rate (m), probability of dying in the interval

(q), and probability of surviving in the interval (p) \*/

data table;

set data12;

n+**1**;

%do i=**1** %to **2**;

m&i= death&i/pop&i;

if age=**1** then q&i=(**2**\***4**\*m&i)/(**2**+**4**\*m&i);

else q&i=(**2**\***5**\*m&i)/(**2**+**5**\*m&i);

if age=**0** then q&i = m&i;

if age=**90** then q&i=**1**;

p&i=**1**-q&i;

%end;

/\* Compute individuals left alive at age x (sur) \*/

data temp1;

set table;

%do i=**1** %to **2**;

retain x&i **1**;

x&i=x&i\*p&i;

xx&i=lag(x&i);

if age=**0** then sur&i=**100000**;

if n=**1** then xx&i=**1**;

if n>**1** then sur&i=**100000**\*xx&i;

%end;

proc sort data=temp1;

by descending n;

/\* Compute individuals dying in the age group (death1) and person-years lived in the interval (l) \*/

data temp2;

set temp1;

%do i=**1** %to **2**;

lagsur&i=lag(sur&i);

if n=**20** then lagsur&i=**0**;

death1&i=sur&i-lagsur&i;

select(n);

when (**1**) l&i=**0.9**\*lagsur&i+**0.10**\*sur&i;

when (**2**) l&i= **2**\*(sur&i+lagsur&i);

when (**20**) l&i=sur&i/m&i;

otherwise l&i= **2.5**\*(sur&i+lagsur&i);

end;

%end;

/\* Compute years-lived above age x (t) and life expectancy (e) \*/

data temp3;

set temp2;

%do i=**1** %to **2**;

t&i+l&i;

e&i=t&i/sur&i;

%end;

proc sort data=temp3;

by n; run;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

CONTRIBUTION OF AGE to the life expectancy gap

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Compute direct effect (DE) of age \*/

data temp4;

set temp3;

DE=(sur1/**100000**)\*((l2/sur2)-(l1/sur1));

proc sort data=temp4;

by descending n;

/\* Compute indirect and interactions effects (IE) of age \*/

data temp5;

set temp4;

lagt=lag(t2);

if age^=**90** then IE=(lagT/**100000**)\*((sur1/sur2)-(lagsur1/lagsur2));

else IE=**0**;

proc sort data=temp5;

by n;

/\* Compute total effect (TE) of age \*/

data temp6;

set temp5;

TE=DE+IE;

DATA temp7;

set temp6;run;

proc sort data=temp7;

by age;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

CONTRIBUTION OF CAUSES OF DEATH to the life expectancy gap

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

data temp8;

set temp7;

%do i=**1** %to &c;

if death1=**0** then r1&i=**0**;

else r1&i=c1&i/death1;

if death2=**0** then r2&i=**0**;

else r2&i=c2&i/death2;

tau1&i=r1&i\*m1;

tau2&i=r2&i\*m2;

taux&i=tau2&i-tau1&i;

rate=m2-m1;

cause&i=TE\*(taux&i/rate);

%end;

run;

data temp9;

set temp8;

%do i=**1** %to &c;

keep age e1 e2 TE cause&i ;

%end;

run;

PROC TRANSPOSE DATA = temp9 OUT = temp10;

RUN;

data temp11;

set temp10;

Total=sum(of col1-col20);

run;

PROC TRANSPOSE DATA=temp11 OUT = CAUSE;

RUN;

data x.&output;

set cause;

drop \_NAME\_;

if age=**856** then do;

age=**.**;

e1=**.**;

e2=**.**;

end;

run;

**%mend** arriaga;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

SYNTAX TO LAUNCH MACRO %arriaga

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/

/\* Insert the names of your own parameters after the equal sign\*/

%***arriaga***

(in1 = ,

in2 = ,

c = ,

output = ,

path = );

**run**;

/\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Stata code

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

capture log close

log using arriaga-decomp.txt, replace text

// program: arriaga-decomp.do

// task: create life tables and decompose difference in e0

// input: pop1.txt pop2.txt

// output: decomp.dta

// project: decomposition of life expectancy gap across populations

// author: sam harper \ 3apr2014

This code assumes that you have prepared two separate tab-delimited datasets,

one for each population, in the following format:

pop1 (e.g., Canada) is the first set of records with higher life expectancy

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

age | pop | death | c1 | c2 | … | cn |

pop2 (e.g., Quebec) is the second set of records with lower life expectancy

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

age | pop | death | c1 | c2 | … | cn |

Where:

age = age at the start of the age category (e.g., 0, 1, 5,… , 85, 90).

Note that this code is set up for 20 age groups, with a last age group

of 90 years.

pop = number of individuals in population, by age category

death = number of deaths from all causes in the population, by age category

c1 to cn: number of deaths from each specific cause in the population,

by age category (n causes of death in total)

\*/

// #0

// program setup

version 12

set linesize 80

clear all

macro drop \_all

// #1

// bring in data on deaths and population

\* user should supply the file names

local name1 "pop1.txt"

local name2 "pop2.txt"

\* temporary dataset names

tempfile data causes

forvalues i=1/2 {

tempfile in`i'

insheet using `name`i'', clear

qui reshape long c, i(age) j(cause)

rename c cause\_

qui sum cause

local ncauses=r(max)

qui sum age

local maxage=r(max)

reshape wide cause\_, i(age) j(cause)

forvalues j=1/`ncauses' {

disp `j'

label var cause\_`j' "deaths from cause `j'"

gen pc\_`j'\_ = cause\_`j'/death

label var pc\_`j'\_ "% of deaths from cause `j'"

}

unab pclist : pc\_1\_ - pc\_`ncauses'\_

gen group=`i'

order group age death

save `in`i'', replace

}

use `in1'

append using `in2'

save `data', replace

\* create and save dataset of fractions of death by cause for each age group

drop death pop cause\_\*

reshape wide `pclist', i(age) j(group)

save `causes', replace

// #2

// set up for life table calculation

use `data', clear

drop cause\* pc\* // drop cause-specific deaths

\* define number of years in age interval, assuming that n=1 for the

\* first interval for infant mortality

sort group age

bysort group: gen n=age[\_n+1] - age[\_n]

replace n=1 if age==0 | age==`maxage'

label var n "no. years in age interval"

\* average person-years contributed by those dying within interval

gen ax=0.1 if age==0

replace ax=0.5 if age>0 & age<=`maxage'

\* life table variables

foreach var in m q p l d L T e {

qui gen `var'x=.

}

\* labels

label var ax "avg time contributed by deaths"

label var mx "death rate at age x"

label var qx "probability of death at age x"

label var px "probability of survival at age x"

label var lx "number alive at age x"

label var dx "expected deaths at age x"

label var Lx "person-years lived in interval"

label var Tx "time lived beyond age x"

label var ex "life expectancy at age x"

label var group "population group"

label var age "age group"

label var death "total deaths"

label var pop "total population"

\* sort by age

sort group age

// #3

// calculate life table values by group

qui levelsof group, local(levels)

foreach l of local levels {

\* mortality rate

qui replace mx=death/pop if group==`l'

\* probability of death

qui replace qx=n\*mx/(1+n\*(1-ax)\*mx) if group==`l'

qui replace qx = 1 if age==`maxage' & group==`l'

\* conditional prob of survival

qui replace px=1-qx if group==`l'

\* no. alive at beginning of interval

qui replace lx = 100000 if age==0 & group==`l'

qui replace lx = lx[\_n-1] \* px[\_n-1] if age>0 & group==`l'

\* Generate deaths by differencing the number of survivors and

\* noting that everyone dies in the end

qui replace dx = lx - lx[\_n+1] if group==`l'

qui replace dx = lx if age==`maxage' & group==`l'

\* Compute person-years lived in each age group

\* n for those who survive the age group and nax for those who die

qui replace Lx = n \* (lx[\_n+1] + (ax\*dx)) if group==`l'

qui replace Lx = lx/mx if age==`maxage' & group==`l'

/\* Accumulating from the bottom up is a bit tricky because Stata likes to sum from the top down. You could sort the data from oldest to youngest, sum, and then sort again. I will subtract the cumulative sum from the total.\*/

qui sum Lx if group==`l'

qui replace Tx = r(sum) - sum(Lx) + Lx if group==`l'

\* Compute life expectancy (time lived after each age / survivors to that age)

qui replace ex = Tx/lx if group==`l'

}

\* specify a few formats to make the presentation nicer

format %6.3f ax ex

format %8.6f mx qx px

format %9.0fc pop death lx dx Lx Tx

\* table of life expectancy at each age by group

table age group, c(mean ex)

// #3

// decompose life expectancy difference by age using formulas from

// Arriaga EE. Measuring and explaining the change in life expectancies.

// Demography 1984;21:83-96.

\*drop unnecessary variables and reshape the data to wide format with

\*rows for each sex year age and colums for each race group

drop n dx qx

reshape wide pop death ax mx px lx Tx Lx ex, i(age) j(group)

\* generate direct effect in each age group

gen de=(lx2/100000) \* ((Lx1/lx1) - (Lx2/lx2))

label var de "direct effect"

\* generate indirect effect and interaction term \*/

gen ie=(Tx1[\_n+1]/100000) \* ((lx2/lx1) - (lx2[\_n+1]/lx1[\_n+1])) if age!=`maxage'

replace ie=0 if age==`maxage'

label var ie "indirect effect+interact"

\* total effect (direct + indirect + interaction)

gen te=de+ie

label var te "diff in life exp"

\* calculate total LE gap by group

egen gap=total(te)

label var gap "total life expectancy gap"

gen pctgap=te/gap\*100

label var pctgap "percentage contribution to the gap by age"

\* table of contribution of each age group to the total gap

table age, c(sum te sum pctgap) row format(%5.3f)

// #4

// decomposition by cause of death using formulas from Arriaga (1988)

// Changing trends in mortality decline during the last decades. In:

// Differential mortality: methodological issues and biosocial factors, edited

// by Ruzicka et al., 105-29. Oxford, UK: Clarendon Press, 1989 \*/

\* merge age-specific contributions with cause-specific mortality data

merge 1:1 age using `causes'

drop \_merge

local i = 1

while `i' <= `ncauses' {

gen cause`i' = te \* (((mx1\*(pc\_`i'\_1) - (mx2\*pc\_`i'\_2)))) / (mx1 - mx2)

local ++i

}

// #5

// make some tables to describe results

\*reshape the data from wide to long, with each observation now

\*consisting of the age-cause-specific contribution to the gap

drop pc\*

reshape long cause, i(age) j(cod)

rename cause cont

label var cont "contribution to LE gap"

label var cod "causes of death"

\* net effect of age groups and causes of death, by period

table age, contents(sum cont) row format(%4.3f)

table cod, contents(sum cont) row format(%4.3f)

\* save as Stata dataset

save decomp, replace

log close

exit