Adding the Feature of Age Adjustment for Survey-related Procedures in SAS--Age Adjusting Prevalence Estimates from Population Based Surveys

Zhiwei Zeng, County of Los Angeles, Los Angeles, CA

ABSTRACT

Estimating the prevalence of health conditions from population based surveys is very common and useful in health surveillance. However, as many medical conditions are strongly associated with age, comparisons can not be made about the relative health of population groups or change over time without age adjustments, which involves age grouping, population standardizing and some calculations. While age adjustment for survey data analysis is so useful and highly utilized in practice, there is no such feature available at all current survey-related procedures in SAS (version 9). This paper describes how to use SAS Macro and SQL to generate the age-adjusted prevalence estimation and 95% confidence intervals directly from a population based survey dataset, providing a useful and practical tool for age adjustment. The SAS codes developed in this paper is easy for implementing, and can be modified and applied for other similar adjustments such as race and gender. In a way, it could be served as an added feature or routine for the survey-related procedures in the current version’s SAS.

Keywords: Age adjustment, SAS Macro, SQL

INTRODUCTION

Many medical conditions, such as hypertension, diabetes and heart disease, are typically more prevalent in older populations than younger ones. When making comparisons among populations with different age structures, age factor is particularly relevant. For example, Florida has a relatively old population and Alaska has a relatively young population. From each state’s crude prevalence rate, it may appear as if Florida has a higher rate for these conditions. However, this is because the population of Florida is older, and the risk of developing many health conditions increases with age. This different age structure makes the overall medical prevalence in Florida appear higher than in Alaska. Therefore, a strategy called age-adjustment is implemented so that the prevalence of specific disease among different states or regions can be compared among people of similar age.

While age adjustment is highly demanded and utilized for survey data analysis, it is not included as a regular feature or option in the survey-related procedures in the current SAS
software (version 9). The SAS program developed in this paper provides a practical tool for age adjustment that can be used on a regular basis.

Age Adjustment Methods

To compare the relative health of population groups or to assess change over time, there are two criteria that should be considered. (1) First, the effect of the population’s age distribution must be taken into account, in which the population needs to be standardized. Second, rates should relate the number of events to the population. The simplest prevalence rate is the crude rate, defined as the total number of cases divided by the population. Crude rates for individual age cohorts, called age-specific rates, are the ratio of the cases in a given age group to the population of that age group.

There are two basic methods of age-adjustment or standardization, known as the direct and indirect methods. For direct standardization, the observed age-specific rates and a standard population are used. For indirect standardization, the observed population and a standard set of age-specific rates are used. In this paper, the direct method is used for age adjustment.

Calculation Formula

Age adjustment, using the direct method, is the application of observed age-specific rates to an age distribution from a standard reference population to eliminate differences in crude rates in populations of interest that may result from differences in the populations’ age distributions. Its formula is:

\[ R = \sum_{i=1}^{n} ri \times \left( \frac{pi}{P} \right) \]  

where R = age-adjusted rate  
\( ri \) = age-specific rates for the population of interest  
\( pi \) = standard population in age group \( i \)

\[ P = \sum_{i=1}^{n} pi \]  

for the age groups that comprise the age range of the rate being age adjusted

\( n \) = total number of age groups

Selection of Standard Population

In 1998, the Secretary of the Department of Health and Human Service (DHHS) issued a policy statement directing all DHHS agencies to use the year 2000 projected U.S. population as the 2000 Standard Population for age adjusting mortality rate (Table 1).(3)
Table 1. Master list: 2000 U.S. projected population and adjustment weights

<table>
<thead>
<tr>
<th>Agegroup</th>
<th>Population in thousands</th>
<th>Adjustment weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>All ages</td>
<td>274,634</td>
<td>1</td>
</tr>
<tr>
<td>Under 1 year</td>
<td>3,795</td>
<td>0.013818</td>
</tr>
<tr>
<td>1 year</td>
<td>3,759</td>
<td>0.013687</td>
</tr>
<tr>
<td>2-4 years</td>
<td>11,433</td>
<td>0.04163</td>
</tr>
<tr>
<td>5 years</td>
<td>3,896</td>
<td>0.014186</td>
</tr>
<tr>
<td>6-8 years</td>
<td>11,800</td>
<td>0.042966</td>
</tr>
<tr>
<td>9 years</td>
<td>4,224</td>
<td>0.01538</td>
</tr>
<tr>
<td>10-11 years</td>
<td>8,258</td>
<td>0.030069</td>
</tr>
<tr>
<td>12-14 years</td>
<td>11,799</td>
<td>0.042963</td>
</tr>
<tr>
<td>15-17 years</td>
<td>11,819</td>
<td>0.043035</td>
</tr>
<tr>
<td>18-19 years</td>
<td>8,001</td>
<td>0.029133</td>
</tr>
<tr>
<td>20-24 years</td>
<td>18,257</td>
<td>0.066478</td>
</tr>
<tr>
<td>25-29 years</td>
<td>17,722</td>
<td>0.06453</td>
</tr>
<tr>
<td>30-34 years</td>
<td>19,511</td>
<td>0.071044</td>
</tr>
<tr>
<td>35-39 years</td>
<td>22,180</td>
<td>0.080762</td>
</tr>
<tr>
<td>40-44 years</td>
<td>22,479</td>
<td>0.081851</td>
</tr>
<tr>
<td>45-49 years</td>
<td>19,806</td>
<td>0.072118</td>
</tr>
<tr>
<td>50-54 years</td>
<td>17,224</td>
<td>0.062716</td>
</tr>
<tr>
<td>55-59 years</td>
<td>13,307</td>
<td>0.048454</td>
</tr>
<tr>
<td>60-64 years</td>
<td>10,654</td>
<td>0.038793</td>
</tr>
<tr>
<td>65-69 years</td>
<td>9,410</td>
<td>0.034264</td>
</tr>
<tr>
<td>70-74 years</td>
<td>8,726</td>
<td>0.031773</td>
</tr>
<tr>
<td>75-79 years</td>
<td>7,415</td>
<td>0.027</td>
</tr>
<tr>
<td>80-84 years</td>
<td>4,900</td>
<td>0.017842</td>
</tr>
<tr>
<td>85 years and over</td>
<td>4,259</td>
<td>0.015508</td>
</tr>
</tbody>
</table>

Age-specific rates based on small numbers will exhibit a large amount of random variation and the data should be aggregated.(1) Hence, different age groupings and corresponding calculations of age-adjustment weights may be needed for different applications. For the Los Angeles County Health Survey (LACHS), the following age groups and age-adjustment weights for the adult population are used (Table 2):

Table 2. 2000 U.S. Standard Adult Population and Adjustment Weights for Los Angeles County Health Survey Data

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Population in thousands</th>
<th>Adjustment weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 years and over</td>
<td>203,851</td>
<td>1</td>
</tr>
<tr>
<td>18-24 years</td>
<td>26,258</td>
<td>0.128810</td>
</tr>
<tr>
<td>25-29 years</td>
<td>17,722</td>
<td>0.086936</td>
</tr>
<tr>
<td>30-39 years</td>
<td>41,691</td>
<td>0.204517</td>
</tr>
<tr>
<td>40-49 years</td>
<td>42,285</td>
<td>0.207431</td>
</tr>
</tbody>
</table>
SAS Programs for Age Adjustment

The following SAS code is used to generate the age-adjusted prevalence and the 95% confidence intervals directly from the LACHS survey dataset, and will be illustrated by following logical steps.

libname in 'Y:\HA_shared\Share\LASurvey05\Adult\Dataset';
options fmtsearch=(in);

Since age adjustment is used frequently and repeatedly for different applications, it is written in SAS macro, so as to be easily applied by others.

%macro adjusting (disease);

First, age specific rates are generated by the FREQ procedure.

proc freq data=in.adult05 noprint;
tables agegroup*&disease/out=a outpct;
weight pop_wgt;
where &disease not in (8,9); /*excluding 'do not know' and 'refused'*/
run;

Second, age-adjustment weights, based on 2000 U.S. standard population by age grouping (18-24 years, 25-29 years, 30-39 years, 40-49 years, 50-59 years, 60-64 years, 65 years and over) are included in the calculation based on the direct method.

data b;
set a;
where &disease in (1);
if agegroup in (1) then adjust=pct_row*0.12881;
if agegroup in (2) then adjust=pct_row*0.086936;
if agegroup in (3) then adjust=pct_row*0.204517;
if agegroup in (4) then adjust=pct_row*0.207431;
if agegroup in (5) then adjust=pct_row*0.149771;
if agegroup in (6) then adjust=pct_row*0.052264;
if agegroup in (7) then adjust=pct_row*0.170271;
footnote 'Based on 2000 U.S. Standard Population';
footnote2 'By Age Grouping 18-24, 25-29, 30-39, 40-49, 50-59, 60-64, 65+';
run;

Third, in preparation for calculating 95% confidence intervals, standard errors are derived from the SURVEYMEANS procedure, where numerical variables are normally used, but categorical variables could also be taken by the CLASS statement.

ods listing close;
proc surveymeans data=in.adult05;
var &disease;
class &disease;
where &disease not in (8,9); /*excluding ‘Do not know’ and ‘Refused’*/
weight pop_wgt;
ods output statistics = c;
run;
ods listing;

Last, the SQL procedure is used to calculate and output the resulting age adjusted rates and 95% confidence intervals.

proc sql;
title "&disease Age-adjusted Rate and 95% CI";
select distinct sum (b.adjust) as AdjRate,
    sum(b.adjust)-50*(c.upperclmean-c.lowerclmean) as Low95CI,
    sum(b.adjust)+50*(c.upperclmean-c.lowerclmean) as High95CI
from b, c where c.varlevel='Yes';
quit;
%mend adjusting;
%adjusting (diabetes)

Then, the macro will be resolved getting age-adjusted rate (AdjRate), low limit and high limit of 95% confidence intervals (Low95CI and High 95CI) for diabetes as an example of medical prevalence.

CONCLUSION

While age adjustment for survey data analysis is very common and highly utilized in practice, no such function is available within the current survey-related procedures in SAS version 9. The SAS program developed in this paper provides a useful tool to generate age-adjusted estimates and 95% confidence intervals directly from a population based survey dataset. The SAS program illustrated here is one approach to age adjustment which can easily be modified and applied to other factors such as race and gender. In a way, it can be incorporated as an additional routine for the survey-related procedures of the current SAS software (version 9).

REFERENCES


4. SAS Guide to the SQL procedure. SAS Institute, Inc. 1996

ACKNOWLEDGEMENTS

The author thanks Dr. Margaret Shih for her valuable comments and appreciates supports from Office of Health Assessment and Epidemiology for this project.

CONTACT INFORMATION

Comments and questions are valued and encouraged. Please contact the author at:

Zhiwei Zeng
Department of Public Health
County of Los Angeles
2615 S. Grand Ave. Room 500
Los Angeles, CA 90007
E-mail: zzeng@ladhs.org