Chapter 109
The STDRATE Procedure

Overview: STDRATE Procedure

Epidemiology is the study of the occurrence and distribution of health-related states or events in specified populations. Epidemiology also includes the study of the determinants that influence these states, and the application of this knowledge to control health problems (Porta 2008). It is a discipline that describes, quantifies, and postulates causal mechanisms for health phenomena in populations (Friss and Sellers 2009).
A common goal is to establish relationships between various factors (such as exposure to a specific chemical) and the event outcomes (such as incidence of disease). But the measure of an association between an exposure and an event outcome can be biased due to confounding. That is, the association of the exposure to some other variables, such as age, influences the occurrence of the event outcome. With confounding, the usual effect between an exposure and an event outcome can be biased because some of the effect might be accounted for by other variables. For example, with an event rate discrepancy among different age groups of a population, the overall crude rate might not provide a useful summary statistic to compare populations.

One strategy to control confounding is stratification. In stratification, a population is divided into several subpopulations according to specific criteria for the confounding variables, such as age and sex groups. The effect of the exposure on the event outcome is estimated within each stratum, and then these stratum-specific effect estimates are combined into an overall estimate.

Two commonly used event frequency measures are rate and risk:

- A rate is a measure of the frequency with which an event occurs in a defined population in a specified period of time. It measures the change in one quantity per unit of another quantity. For example, an event rate measures how fast the events are occurring. That is, an event rate of a population over a specified time period can be defined as the number of new events divided by population-time (Kleinbaum, Kupper, and Morgenstern 1982, p. 100) over the same time period.

- A risk is the probability that an event occurs in a specified time period. It is assumed that only one event can occur in the time period for each subject or item. The overall crude risk of a population over a specified time period is the number of new events in the time period divided by the population size at the beginning of the time period.

Standardized overall rate and risk estimates based on stratum-specific estimates can be derived with the effects of confounding variables removed. These estimates provide useful summary statistics and allow valid comparison of the populations. There are two types of standardization:

- Direct standardization computes the weighted average of stratum-specific estimates in the study population, using the weights from a standard or reference population. This standardization is applicable when the study population is large enough to provide stable stratum-specific estimates. The directly standardized estimate is the overall crude rate in the study population if it has the same strata distribution as the reference population. When standardized estimates for different populations are derived by using the same reference population, the resulting estimates can also be compared by using the estimated difference and estimated ratio statistics.

- Indirect standardization computes the weighted average of stratum-specific estimates in the reference population, using the weights from the study population. The ratio of the overall crude rate or risk in the study population and the corresponding weighted estimate in the reference population is the standardized morbidity ratio (SMR). This ratio is also the standardized mortality ratio if the event is death. SMR is used to compare rates or risks in the study and reference populations. With SMR, the indirectly standardized estimate is then computed as the product of the SMR and the overall crude estimate for the reference population. SMR and indirect standardization are applicable even when the study population is so small that the resulting stratum-specific rates are not stable.
Assuming that an effect, such as the rate difference between two populations, is homogeneous across strata, each stratum provides an estimate of the same effect. A pooled estimate of the effect can then be derived from these stratum-specific effect estimates. One way to estimate a homogeneous effect is the Mantel-Haenszel method (Greenland and Rothman 2008, p. 271). For a homogeneous rate difference effect between two populations, the Mantel-Haenszel estimate is identical to the difference between two directly standardized rates, but with weights derived from the two populations instead of from an explicitly specified reference population. The Mantel-Haenszel method can also be applied to other homogeneous effects between populations, such as the rate ratio, risk difference, and risk ratio.

The STDRATE procedure computes directly standardized rates and risks for study populations. For two study populations with the same reference population, PROC STDRATE compares directly standardized rates or risks from these two populations. For homogeneous effects across strata, PROC STDRATE computes Mantel-Haenszel estimates. The STDRATE procedure also computes indirectly standardized rates and risks, including SMR.

The attributable fraction measures the excess event rate or risk fraction in the exposed population that can be attributed to the exposure. The rate or risk ratio statistic is required in the attributable fraction computation, and the STDRATE procedure estimates the ratio by using either SMR or the rate ratio statistic in the Mantel-Haenszel estimates.

Although the STDRATE procedure provides useful summary standardized statistics, standardization is not a substitute for individual comparisons of stratum-specific estimates. PROC STDRATE provides summary statistics, such as rate and risk estimates and their confidence limits, in each stratum. In addition, PROC STDRATE also displays these stratum-specific statistics by using ODS Graphics.

Note that the term standardization has different meanings in other statistical applications. For example, the STDIZE procedure standardizes numeric variables in a SAS data set by subtracting a location measure and dividing by a scale measure.

---

### Getting Started: STDRATE Procedure

This example illustrates indirect standardization and uses the standardized mortality ratio to compare the death rate from skin cancer between people who live in the state of Florida and people who live in the United States as a whole.

The Florida_C43 data set contains the stratum-specific mortality information for skin cancer in year 2000 for the state of Florida (Florida Department of Health 2000, 2013). The variable Age is a grouping variable that forms the strata in the standardization, and the variables Event and PYear identify the number of events and total person-years, respectively. The COMMA9. format is specified in the DATA step to input numerical values that contain commas in PYear.
data Florida_C43;
  input Age $1-5 Event PYear:comma9.;
datalines;
  00-04   0   953,785
  05-14   0   1,997,935
  15-24   4   1,885,014
  25-34  14   1,957,573
  35-44  43   2,356,649
  45-54  72   2,088,000
  55-64  70   1,548,371
  65-74 126   1,447,432
  75-84 136   1,087,524
  85+    73   335,944
;

The US_C43 data set contains the corresponding stratum-specific mortality information for the United States in year 2000 (Miniño et al. 2002; US Bureau of the Census 2011). The variable Age is the grouping variable, and the variables Event and PYear identify the number of events and the total person-years, respectively.

data US_C43;
  input Age $1-5 Event:comma5. PYear:comma10.;
datalines;
  00-04   0   19,175,798
  05-14   1   41,077,577
  15-24  41   39,183,891
  25-34 186   39,892,024
  35-44 626   45,148,527
  45-54 1,199 37,677,952
  55-64 1,303 24,274,684
  65-74 1,637 18,390,986
  75-84 1,624 12,361,180
  85+   803   4,239,587
;

The following statements invoke the STDRATE procedure and request indirect standardization to compare death rates between the state of Florida and the United States:

ods graphics on;
proc stdrate data=Florida_C43 refdata=US_C43
  method=indirect
  stat=rate(mult=100000)
  plots=all;
  population event=Event total=PYear;
  reference event=Event total=PYear;
  strata Age / stats smr;
run;
The DATA= and REFDATA= options name the study data set and reference data set, respectively. The METHOD=INDIRECT option requests indirect standardization. The STAT=RATE option specifies the rate as the frequency measure for standardization, and the MULT=100000 suboption (which is the default) displays the rates per 100,000 person-years in the table output and graphics output. The PLOTS=ALL option requests all appropriate plots with indirect standardization.

The POPULATION statement specifies the options that are related to the study population, and the EVENT= and TOTAL= options specify variables for the number of events and person-years in the study population, respectively.

The REFERENCE statement specifies the options related to the reference population, and the EVENT= and TOTAL= options specify variables for the number of events and person-years in the reference population, respectively.

The STRATA statement lists the variable Age that forms the strata. The STATS option requests a strata information table that contains stratum-specific statistics such as rates, and the SMR option requests a table of stratum-specific SMR estimates.

The “Standardization Information” table in Figure 109.1 displays the standardization information.

```
Figure 109.1 Standardization Information

The STDRATE Procedure

<table>
<thead>
<tr>
<th>Standardization Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Reference Data Set</td>
</tr>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Statistic</td>
</tr>
<tr>
<td>Number of Strata</td>
</tr>
<tr>
<td>Rate Multiplier</td>
</tr>
</tbody>
</table>
```

The STATS option in the STRATA statement requests that the “Indirectly Standardized Strata Statistics” table in Figure 109.2 display the strata information and expected number of events at each stratum. The MULT=100000 suboption in the STAT=RATE option requests that crude rates per 100,000 person-years be displayed. The Expected Events column displays the expected number of events when the stratum-specific rates in the reference data set are applied to the corresponding person-years in the study data set.
Chapter 109: The STDRATE Procedure

Figure 109.2 Strata Information (Indirect Standardization)

### The STDRATE Procedure

#### Indirectly Standardized Strata Statistics

Rate Multiplier = 100000

<table>
<thead>
<tr>
<th>Stratum Index</th>
<th>Age</th>
<th>Observed Events</th>
<th>Value</th>
<th>Proportion</th>
<th>Crude Rate</th>
<th>Standard Error</th>
<th>95% Normal Confidence Limits</th>
<th>Value</th>
<th>Proportion</th>
<th>Crude Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>00-04</td>
<td>0</td>
<td>953785</td>
<td>0.0609</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>19175798</td>
<td>0.0681</td>
</tr>
<tr>
<td>2</td>
<td>05-14</td>
<td>0</td>
<td>1997935</td>
<td>0.1276</td>
<td>0.2122</td>
<td>0.10610</td>
<td>0.0000</td>
<td>0.0000</td>
<td>41077577</td>
<td>0.1460</td>
</tr>
<tr>
<td>3</td>
<td>15-24</td>
<td>4</td>
<td>1885014</td>
<td>0.1204</td>
<td>0.1752</td>
<td>0.19114</td>
<td>0.3405</td>
<td>0.4202</td>
<td>39183891</td>
<td>0.1392</td>
</tr>
<tr>
<td>4</td>
<td>25-34</td>
<td>14</td>
<td>1957573</td>
<td>0.1250</td>
<td>0.7151</td>
<td>0.3405</td>
<td>1.0898</td>
<td>0.4202</td>
<td>39892024</td>
<td>0.1418</td>
</tr>
<tr>
<td>5</td>
<td>35-44</td>
<td>43</td>
<td>2356649</td>
<td>0.1505</td>
<td>1.8246</td>
<td>0.27825</td>
<td>1.2793</td>
<td>2.3700</td>
<td>45148527</td>
<td>0.1604</td>
</tr>
<tr>
<td>6</td>
<td>45-54</td>
<td>72</td>
<td>2088000</td>
<td>0.1333</td>
<td>3.4483</td>
<td>0.40638</td>
<td>2.6518</td>
<td>4.2448</td>
<td>37677952</td>
<td>0.1339</td>
</tr>
<tr>
<td>7</td>
<td>55-64</td>
<td>70</td>
<td>1548371</td>
<td>0.0989</td>
<td>4.5209</td>
<td>0.54035</td>
<td>3.4618</td>
<td>5.5799</td>
<td>24274684</td>
<td>0.0863</td>
</tr>
<tr>
<td>8</td>
<td>65-74</td>
<td>126</td>
<td>1447432</td>
<td>0.0924</td>
<td>8.7051</td>
<td>0.77551</td>
<td>7.1851</td>
<td>10.2250</td>
<td>18390986</td>
<td>0.0654</td>
</tr>
<tr>
<td>9</td>
<td>75-84</td>
<td>136</td>
<td>1087524</td>
<td>0.0695</td>
<td>12.5055</td>
<td>1.07234</td>
<td>10.4037</td>
<td>14.6072</td>
<td>12361180</td>
<td>0.0439</td>
</tr>
<tr>
<td>10</td>
<td>85+</td>
<td>73</td>
<td>335944</td>
<td>0.0215</td>
<td>21.7298</td>
<td>2.54328</td>
<td>16.7451</td>
<td>26.7146</td>
<td>4239587</td>
<td>0.0151</td>
</tr>
</tbody>
</table>

Indirectly Standardized Strata Statistics

Rate Multiplier = 100000

<table>
<thead>
<tr>
<th>Stratum Index</th>
<th>Expected Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>0.049</td>
</tr>
<tr>
<td>3</td>
<td>1.972</td>
</tr>
<tr>
<td>4</td>
<td>9.127</td>
</tr>
<tr>
<td>5</td>
<td>32.676</td>
</tr>
<tr>
<td>6</td>
<td>66.445</td>
</tr>
<tr>
<td>7</td>
<td>83.112</td>
</tr>
<tr>
<td>8</td>
<td>128.837</td>
</tr>
<tr>
<td>9</td>
<td>142.878</td>
</tr>
<tr>
<td>10</td>
<td>63.630</td>
</tr>
</tbody>
</table>

With ODS Graphics enabled, the PLOTS=ALL option displays all appropriate plots. With indirect standardization and a rate statistic, these plots include the strata distribution plot, the strata rate plot, and the strata SMR plot. By default, strata levels are displayed on the vertical axis for these plots.
The strata distribution plot displays proportions for stratum-specific person-years in the study and reference populations, as shown in Figure 109.3.

**Figure 109.3** Strata Distribution Plot

The strata distribution plot displays the proportions in the “Indirectly Standardized Strata Statistics” table in Figure 109.2. In the plot, the proportions of the study population are identified by the blue lines, and the proportions of the reference population are identified by the red lines. The plot shows that the study population has higher proportions in older age groups and lower proportions in younger age groups than the reference population.
The strata rate plot displays stratum-specific rate estimates in the study and reference populations, as shown in Figure 109.4. This plot displays the rate estimates in the “Indirectly Standardized Strata Statistics” table in Figure 109.2. In addition, the plot displays the confidence limits for the rate estimates in the study population and the overall crude rates for the two populations.

**Figure 109.4** Strata Rate Plot
The SMR option in the STRATA statement requests that the “Strata SMR Estimates” table in Figure 109.5 display the strata SMR at each stratum. The MULT=100000 suboption in the STAT=RATE option requests that the reference rates per 100,000 person-years be displayed.

**Figure 109.5** Strata SMR Information

<table>
<thead>
<tr>
<th>Stratum Index</th>
<th>Age</th>
<th>Observed Events</th>
<th>Population-Time</th>
<th>Reference Crude Rate</th>
<th>Expected Events</th>
<th>SMR</th>
<th>Standard Error</th>
<th>95% Normal Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>00-04</td>
<td>0</td>
<td>953785</td>
<td>0.0000</td>
<td>0.000</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td>05-14</td>
<td>0</td>
<td>1997935</td>
<td>0.0024</td>
<td>0.049</td>
<td>0.0000</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>15-24</td>
<td>4</td>
<td>1885014</td>
<td>0.1046</td>
<td>1.972</td>
<td>2.0280</td>
<td>1.0140</td>
<td>0.0406</td>
</tr>
<tr>
<td>4</td>
<td>25-34</td>
<td>14</td>
<td>1957573</td>
<td>0.4663</td>
<td>9.127</td>
<td>1.5339</td>
<td>0.4099</td>
<td>0.7304</td>
</tr>
<tr>
<td>5</td>
<td>35-44</td>
<td>43</td>
<td>2356649</td>
<td>1.3865</td>
<td>32.676</td>
<td>1.3160</td>
<td>0.2007</td>
<td>0.9226</td>
</tr>
<tr>
<td>6</td>
<td>45-54</td>
<td>72</td>
<td>2088000</td>
<td>3.1822</td>
<td>66.445</td>
<td>1.0836</td>
<td>0.1277</td>
<td>0.8333</td>
</tr>
<tr>
<td>7</td>
<td>55-64</td>
<td>70</td>
<td>1548371</td>
<td>5.3677</td>
<td>83.112</td>
<td>0.8422</td>
<td>0.1007</td>
<td>0.6449</td>
</tr>
<tr>
<td>8</td>
<td>65-74</td>
<td>126</td>
<td>1447432</td>
<td>8.9011</td>
<td>128.837</td>
<td>0.9780</td>
<td>0.0871</td>
<td>0.8072</td>
</tr>
<tr>
<td>9</td>
<td>75-84</td>
<td>136</td>
<td>1087524</td>
<td>13.1379</td>
<td>142.878</td>
<td>0.9519</td>
<td>0.0816</td>
<td>0.7919</td>
</tr>
<tr>
<td>10</td>
<td>85+</td>
<td>73</td>
<td>335944</td>
<td>18.9405</td>
<td>63.630</td>
<td>1.1473</td>
<td>0.1343</td>
<td>0.8841</td>
</tr>
</tbody>
</table>

The “Strata SMR Estimates” table shows that although SMR is less than 1 only at three age strata (55–64, 65–74, and 75–84), these three strata contain about 60% of the total events.
The strata SMR plot displays stratum-specific SMR estimates with confidence limits, as shown in Figure 109.6. The plot displays the SMR estimates in the “Strata SMR Estimates” table in Figure 109.5.

**Figure 109.6** Strata SMR Plot

![Strata SMR Estimates with 95% Normal Confidence Limits](image)

The METHOD=INDIRECT option requests that the “Standardized Morbidity/Mortality Ratio” table in Figure 109.7 be displayed. The table displays the SMR, its confidence limits, and the test for the null hypothesis \( H_0 : \text{SMR} = 1 \). The default ALPHA=0.05 option requests that 95% confidence limits be constructed.

**Figure 109.7** Standardized Morbidity/Mortality Ratio

<table>
<thead>
<tr>
<th>Observed Events</th>
<th>Expected Events</th>
<th>Standard Error</th>
<th>Standard Error</th>
<th>Normal Confidence Limits</th>
<th>Z Pr &gt; [Z]</th>
</tr>
</thead>
<tbody>
<tr>
<td>538</td>
<td>528.726</td>
<td>1.0175</td>
<td>0.0439</td>
<td>0.9316</td>
<td>1.1035</td>
</tr>
</tbody>
</table>

The 95% normal confidence limits contain the null hypothesis value SMR = 1, and the hypothesis of SMR = 1 is not rejected at the \( \alpha = 0.05 \) level from the normal test.
The “Indirectly Standardized Rate Estimates” table in Figure 109.8 displays the indirectly standardized rate and related statistics.

**Figure 109.8  Standardized Rate Estimates (Indirect Standardization)**

<table>
<thead>
<tr>
<th>Study Population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Indirectly Standardized Rate Estimates</strong></td>
</tr>
<tr>
<td><strong>Rate Multiplier = 100000</strong></td>
</tr>
<tr>
<td><strong>Observed Events</strong></td>
</tr>
<tr>
<td>---------------------------</td>
</tr>
<tr>
<td>538</td>
</tr>
</tbody>
</table>

The indirectly standardized rate estimate is the product of the SMR and the crude rate estimate for the reference population. The table shows that although the crude rate in the state of Florida (3.4359) is 30% higher than the crude rate in the U.S. (2.6366), the resulting standardized rate (2.6829) is close to the crude rate in the U.S.

**Syntax: STDRA TE Procedure**

The following statements are available in PROC STDRA TE:

```
PROC STDRA TE < options > ;
   BY variables ;
   POPULATION options ;
   REFERENCE options ;
   STRATA variables  < / option > ;
```

The PROC STDRA TE statement invokes the procedure, names the data sets, specifies the standardization method, and identifies the statistic for standardization. The BY statement requests separate analyses of groups defined by the BY variables. The required POPULATION statement specifies the rate or risk information in study populations, and the REFERENCE statement specifies the rate or risk information in the reference population. The STRATA statement lists the variables that form the strata.

The following sections describe the PROC STDRA TE statement and then describe the other statements in alphabetical order.
PROC STDRATE Statement

PROC STDRATE < options > ;

Table 109.1 summarizes the options in the PROC STDRATE statement.

Table 109.1  Summary of PROC STDRATE Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input Data Sets</td>
<td></td>
</tr>
<tr>
<td>DATA=</td>
<td>Names the SAS data set that contains the study populations</td>
</tr>
<tr>
<td>REFDATA=</td>
<td>Names the SAS data set that contains the reference population</td>
</tr>
<tr>
<td>Standardization Methods</td>
<td></td>
</tr>
<tr>
<td>METHOD=</td>
<td>Specifies the method for standardization</td>
</tr>
<tr>
<td>STAT=</td>
<td>Specifies the statistic for standardization</td>
</tr>
<tr>
<td>EFFECT</td>
<td>Specifies the test to compare study populations for direct standardization and Mantel-Haenszel estimation</td>
</tr>
<tr>
<td>Displayed Output</td>
<td></td>
</tr>
<tr>
<td>ALPHA=</td>
<td>Specifies the significance level for confidence intervals</td>
</tr>
<tr>
<td>CL=</td>
<td>Requests the confidence limits for the standardized estimates</td>
</tr>
<tr>
<td>PLOTS</td>
<td>Requests stratum-specific plots</td>
</tr>
</tbody>
</table>

You can specify the following options in the PROC STDRATE statement to compute standardized rates and risks in the procedure. They are listed in alphabetical order.

ALPHA=\(\alpha\)

requests that confidence limits be constructed with confidence level \(100(1 - \alpha)\%\), where \(0 < \alpha < 1\). The default is ALPHA=0.05. These confidence limits include confidence limits for the stratum-specific rates or risks, standardized rate and risk, standardized morbidity/mortality ratio, and population attributable rate and risk.

CL=GAMMA < (TYPE=AVERAGE | CONSERVATIVE) > | LOGNORMAL | NONE | NORMAL | POISSON

specifies the method to construct confidence limits for SMR and standardized rate and risk. You can specify the following values for this option:

GAMMA

requests confidence limits based on a gamma distribution for METHOD=DIRECT and METHOD=MH. This value applies only when STAT=RATE. You can specify the TYPE=CONSERVATIVE suboption to request conservative confidence limits that are based on a gamma distribution and were developed by Fay and Feuer (1997), or you can use the default TYPE=AVERAGE suboption to request modified confidence limits proposed by Tiwari, Clegg, and Zou (2006).
LOGNORMAL
requests confidence limits based on a lognormal distribution.

NONE
suppresses construction of confidence limits.

NORMAL
requests confidence limits based on a normal distribution.

POISSON
requests confidence limits based on a Poisson distribution. This value applies only when METHOD=INDIRECT.

The default is CL=NORMAL.

DATA=SAS-data-set
names the required SAS data set that contains the event information in the study populations.

EFFECT <=DIFF | RATIO >
displays a table of the effect estimate and associated confidence limits. This option applies only when METHOD=DIRECT with two study populations and when METHOD=MH, where two study populations are required.

EFFECT and EFFECT=RATIO display a test on the ratio effect of estimates between the study populations, and the EFFECT=DIFF option displays a test on the difference effect.

METHOD= DIRECT | INDIRECT <(AF)> | MH <(AF)>
M= DIRECT | INDIRECT <(AF)> | MH <(AF)>
specifies the required method for standardization. The AF suboption (available only for METHOD=INDIRECT or METHOD=MH) requests the attributable fraction, which measures how much of the excess event rate or risk fraction in the exposed population is attributable to the exposure. This suboption also requests the population attributable fraction, which measures how much of the excess event rate or risk fraction in the total population is attributable to the exposure.

You can specify the following values:

DIRECT
requests direct standardization.

INDIRECT
requests indirect standardization. If you specify the AF suboption, the study population is treated as the exposed population and the reference population is treated as the unexposed population.

MH
requests Mandel-Haenszel estimation. The order of the two study populations is indicated by the ORDER= suboption in the GROUP option in the POPULATION statement. If you specify the AF suboption, the exposed population is identified by the EXPOSED= suboption in the GROUP option in the POPULATION statement. If the EXPOSED= suboption is not specified, then the first study population is treated as the exposed population and the second study population is treated as the unexposed population.
Chapter 109: The STDRATE Procedure

PLOTS < ( global-options ) > <= plot-request >

specifies options that control the details of the plots. The default is PLOTS=RATE for STAT=RATE and PLOTS=RISK for STAT=RISK.

You can specify the following global-options:

DISPLAY=INDEX | LEVEL
specifies tick mark values for the strata axis. DISPLAY=LEVEL displays strata levels on the strata axis, and DISPLAY=INDEX displays strata indices of sequential strata identification numbers on the strata axis. The default is DISPLAY=LEVEL.

ONLY
suppresses the default plots and displays only plots that are specifically requested.

STRATUM=HORIZONTAL | VERTICAL
controls the orientation of the plots. STRATUM=VERTICAL places the strata information on the vertical axis, and STRATUM=HORIZONTAL places the strata information on the horizontal axis. The default is STRATUM=VERTICAL.

You can specify the following plot-requests:

ALL
produces all appropriate plots.

DIST | DISTRIBUTION
displays a plot of the proportions for stratum-specific exposed time or sample size.

EFFECT
displays a plot of the stratum-specific effect estimates and associated confidence limits. This option applies only when METHOD=DIRECT with two study populations and when METHOD=MH, where two study populations are required. If the EFFECT=DIFF option is specified, the stratum-specific rate or risk difference effects are displayed. Otherwise, the stratum-specific rate or risk ratio effects are displayed.

NONE
suppresses all plots.

RATE
displays a plot of the stratum-specific rates and associated confidence limits. This option applies only when STAT=RATE. If a confidence limits method is specified in the STATS(CL=) option in the STRATA statement, that method is used to compute the confidence limits. Otherwise, the normal approximation is used.

RISK
displays a plot of the stratum-specific risks and associated confidence limits. This option applies only when STAT=RISK. If a confidence limits method is specified in the STATS(CL=) option in the STRATA statement, that method is used to compute the confidence limits. Otherwise, the normal approximation is used.
SMR
displays a plot of the stratum-specific SMR estimates and associated confidence limits. This option applies only when METHOD=INDIRECT. If a method is specified in the SMR(CL=) option in the STRATA statement, that method is used to compute the confidence limits. Otherwise, the normal approximation is used.

REFDATA=SAS-data-set
names the required SAS data set that contains the event information in the reference population.

STAT=RATE <( MULT =c )>
STAT=RISK
specifies the statistic for standardization. STAT=RATE computes standardized rates, and STAT=RISK computes standardized risks. The default is STAT=RATE.

The MULT= suboption in the STAT=RATE option specifies a power of 10 constant c, and requests that rates per c population-time units be displayed in the output tables and graphics. The default is MULT=100000, which specifies rates per 100,000 population-time units.

BY Statement

BY variables ;
You can specify a BY statement with PROC STDRATE to obtain separate analyses of observations in groups that are defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If you specify more than one BY statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.
- Specify the NOTSORTED or DESCENDING option in the BY statement for the STDRATE procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables by using the DATASETS procedure (in Base SAS software).

For more information about BY-group processing, see the discussion in SAS Language Reference: Concepts. For more information about the DATASETS procedure, see the discussion in the Base SAS Procedures Guide.
POPULATION Statement

POPULATION < options > ;

The required POPULATION statement specifies the information in the study data set. You can specify the following options in the POPULATION statement:

EVENT=variable
specifies the variable for the number of events in the study data set.

GROUP < ( group-options ) > = variable
specifies the variable whose values identify the various populations. The GROUP= option is required when METHOD=MH and also applies when METHOD=DIRECT in the PROC STDRATE statement. You can specify the following group-options:

EXPOSED='group'
identifies the exposed group in the derivation of the attributable fraction. This option applies only when you specify METHOD=MH(AF). If you do not specify the EXPOSED= option, the first study population, as indicated by the ORDER= option, is treated as the exposed population.

ORDER=DATA | FORMATTED | INTERNAL
specifies the order in which the values of the variable are to be displayed. You can specify the following values for the ORDER= suboption:

DATA sorts by the order in which the values appear in the input data set.
FORMATTED sorts by their external formatted values.
INTERNAL sorts by the unformatted values, which yields the same order that the SORT procedure does.

By default, ORDER=INTERNAL. For ORDER=FORMATTED and ORDER=INTERNAL, the sort order is machine-dependent.

POPEVENT=number
specifies the total number of events in the study data set. This option applies only when METHOD=INDIRECT is specified in the PROC STDRATE statement and the total number of events is not available in the study data set.

RATE < ( MULT=c ) > = variable
specifies the variable for the observed rate in the study data set. This option applies only when STAT=RATE is specified in the PROC STDRATE statement. The MULT=c suboption specifies a power of 10 constant c and requests that the rates per c population-time units be read from the data set. The default is the value of the MULT= suboption used in the STAT=RATE option in the PROC STDRATE statement.

RISK=variable
specifies the variable for the observed risk in the study data set. This option applies only when STAT=RISK is specified in the PROC STDRATE statement.
TOTAL=variable
    specifies the variable for either the population-time (STAT=RATE) or the number of observations (STAT=RISK) in the study data set.

REFERENCE Statement

REFERENCE <options> ;

The REFERENCE statement specifies the information in the reference data set. This statement is required when METHOD=DIRECT or METHOD=INDIRECT is specified in the PROC STDRATE statement.

You can specify the following options in the REFERENCE statement:

EVENT=variable
    specifies the variable for the number of events in the reference data set.

RATE <(MULT=c)> = variable
    specifies the variable for the observed rate in the reference data set. This option applies only when STAT=RATE is specified in the PROC STDRATE statement. The MULT=c suboption specifies a power of 10 constant c and requests that the rates per c population-time units be read from the data set. The default is the value of the MULT= suboption used in the STAT=RATE option in the PROC STDRATE statement.

RISK=variable
    specifies the variable for the observed risk in the reference data set. This option applies only when STAT=RISK is specified in the PROC STDRATE statement.

TOTAL=variable
    specifies the variable for either the population-time (STAT=RATE) or the number of observations (STAT=RISK) in the reference data set.

When METHOD=INDIRECT is specified in the PROC STDRATE statement, the overall reference population rate and risk are needed to compute indirect standardized rate and risk, respectively. If the information is not available in the reference data set, you can specify the following options for overall reference population rate and risk.

REFEVENT=number
    specifies the total number of events in the reference data set.

REFRATE <(MULT=c)> = number
    specifies the crude rate in the reference data set. This option applies only when STAT=RATE is specified in the PROC STDRATE statement. The MULT=c suboption specifies a power of 10 constant c, and the number is the crude rate per c population-time units in the data set. The default is the value of the MULT= suboption in the STAT=RATE option in the PROC STDRATE statement.

REFRISK=number
    specifies the crude risk in the reference data set. This option applies only when STAT=RISK is specified in the PROC STDRATE statement.
REFTOTAL=number

specifies either the total population-time (STAT=RATE) or the total number of observations (STAT=RISK) in the reference data set.

When STAT=RATE, the REFRATE= option specifies the crude reference rate for the indirect standardized rate. If the REFRATE= option is not specified, the REFEVENT= and REFTOTAL options can be used to compute the crude reference rate. Similarly, when STAT=RISK, the REFRISK= option specifies the crude reference risk for the indirect standardized rate. If the REFRISK= option is not specified, the REFEVENT= and REFTOTAL options can be used to compute the crude reference risk.

### STRATA Statement

**STRATA variables */ options > ;**

The STRATA statement names variables that form the strata in the standardization. The combinations of categories of STRATA variables define the strata in the population.

The STRATA variables are one or more variables in all input data sets. These variables can be either character or numeric. The formatted values of the STRATA variables determine the levels. Thus, you can use formats to group values into levels. See the FORMAT procedure in the *Base SAS Procedures Guide* and the FORMAT statement and SAS formats in *SAS Language Reference: Dictionary* for more information.

When the STRATA statement is not specified or the statement is specified without variables, all observations in a data set are treated as though they are from a single stratum.

You can specify the following options in the STRATA statement after a slash (/):

- **EFFECT**
  
  displays a table of the stratum-specific effect estimates and associated confidence limits. This option applies only when METHOD=DIRECT with two study populations and when METHOD=MH, where two study populations are required. If the EFFECT=DIFF option in the PROC STDRATE statement is specified, the stratum-specific rate or risk difference effects are displayed. Otherwise, the stratum-specific rate or risk ratio effects are displayed.

- **MISSING**
  
  treats missing values as a valid (nonmissing) category for all STRATA variables. When PROC STDRATE determines levels of a STRATA variable, an observation with missing values for that STRATA variable is excluded, unless the MISSING option is specified.

- **ORDER=DATA | FORMATTED | INTERNAL**
  
  specifies the order in which the values of the categorical variables are to be displayed. You can specify the following values for the ORDER= option:

  - **DATA**
    
    sorts by the order in which the values appear in the input data set.

  - **FORMATTED**
    
    sorts by their external formatted values.

  - **INTERNAL**
    
    sorts by the unformatted values, which yields the same order that the SORT procedure does.

  By default, ORDER=INTERNAL. For ORDER=FORMATTED and ORDER=INTERNAL, the sort order is machine-dependent.
STATS <( CL=LOGNORMAL | NONE | NORMAL | POISSON )>

displays tables for stratum-specific statistics such as stratum-specific rates and risks. You can specify the following values of the CL= suboption to request confidence limits for the rate or risk estimate in each stratum:

LOGNORMAL
requests confidence limits based on a lognormal approximation.

NONE
suppresses confidence limits.

NORMAL
requests confidence limits based on a normal approximation and also displays the standard error for the rate estimate in each stratum.

POISSON
requests confidence limits based on a Poisson distribution for stratum-specific rates. This value applies only when STAT=RATE in the PROC STDRATE statement.

The default is CL=NORMAL.

SMR <( CL=LOGNORMAL | NONE | NORMAL | POISSON )>

displays tables for stratum-specific SMR estimates. This option applies only when METHOD=INDIRECT is specified in the PROC STDRATE statement. You can specify the following values of the CL= suboption to request confidence limits for the SMR estimate in each stratum:

LOGNORMAL
requests confidence limits based on a lognormal approximation.

NONE
suppresses confidence limits.

NORMAL
requests confidence limits based on a normal approximation and also displays the standard error for the SMR estimate in each stratum.

POISSON
requests confidence limits based on a Poisson distribution for stratum-specific SMR estimates. This value applies only when STAT=RATE in the PROC STDRATE statement.

The default is CL=NORMAL.
Rate

A major task in epidemiology is to compare event frequencies for groups of people. Both rate and risk are commonly used to measure event frequency in the comparison. Rate is a measure of change in one quantity per unit of another quantity. An event rate measures how fast the events are occurring. In contrast, an event risk is the probability that an event occurs over a specified follow-up time period.

An event rate of a population over a specified time period can be defined as the number of new events divided by the population-time of the population over the same time period,

$$\hat{\lambda} = \frac{d}{T}$$

where $d$ is the number of events and $T$ is the population-time that is computed by adding up the time contributed by each subject in the population over the specified time period.

For a general population, the subsets (strata) might not be homogeneous enough to have a similar rate. Thus, the rate for each stratum should be computed separately to reflect this discrepancy. For a population that consists of $K$ homogeneous strata (such as different age groups), the stratum-specific rate for the $j$th stratum in a population is computed as

$$\hat{\lambda}_j = \frac{d_j}{T_j}$$

where $d_j$ is the number of events and $T_j$ is the population-time for subjects in the $j$th stratum of the population.

Assuming the number of events in the $j$th stratum, $d_j$, has a Poisson distribution, the variance of $\hat{\lambda}_j$ is

$$V(\hat{\lambda}_j) = V\left(\frac{d_j}{T_j}\right) = \frac{1}{T_j^2} V(d_j) = \frac{d_j}{T_j^2} = \frac{\hat{\lambda}_j}{T_j}$$

By using the method of statistical differentials (Elandt-Johnson and Johnson 1980, pp. 70–71), the variance of the logarithm of rate can be estimated by

$$V(\log(\hat{\lambda}_j)) = \frac{1}{\hat{\lambda}_j^2} V(\hat{\lambda}_j) = \frac{1}{\hat{\lambda}_j^2} \frac{\hat{\lambda}_j}{T_j} = \frac{1}{\hat{\lambda}_j T_j} = \frac{1}{d_j}$$

Because the rate value can be very small, especially for rare events, it is sometimes expressed in terms of the product of a multiplier and the rate itself. For example, a rate can be expressed as the number of events per 100,000 person-years.
Normal Distribution Confidence Interval for Rate

A $(1 - \alpha)$ confidence interval for $\hat{\lambda}_j$ based on a normal distribution is given by

$$\left( \hat{\lambda}_j - z \sqrt{V(\hat{\lambda}_j)}, \hat{\lambda}_j + z \sqrt{V(\hat{\lambda}_j)} \right)$$

where $z = \Phi^{-1}(1 - \alpha/2)$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution.

Lognormal Distribution Confidence Interval for Rate

A $(1 - \alpha)$ confidence interval for $\log(\hat{\lambda}_j)$ based on a normal distribution is given by

$$\left( \log(\hat{\lambda}_j) - z \sqrt{V(\log(\hat{\lambda}_j))}, \log(\hat{\lambda}_j) + z \sqrt{V(\log(\hat{\lambda}_j))} \right)$$

where $z = \Phi^{-1}(1 - \alpha/2)$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution and the variance $V(\log(\hat{\lambda}_j)) = 1/d_j$.

Thus, a $(1 - \alpha)$ confidence interval for $\hat{\lambda}_j$ based on a lognormal distribution is given by

$$\left( \hat{\lambda}_j e^{-\frac{z^2}{2d_j}}, \hat{\lambda}_j e^{\frac{z^2}{2d_j}} \right)$$

Poisson Distribution Confidence Interval for Rate

Denote the $(\alpha/2)$ quantile for the $\chi^2$ distribution with $2d_j$ degrees of freedom by

$$q_{lj} = (\chi^2_{2d_j})^{-1}(\alpha/2)$$

Denote the $(1 - \alpha/2)$ quantiles for the $\chi^2$ distribution with $2(d_j + 1)$ degrees of freedom by

$$q_{uj} = (\chi^2_{2(d_j + 1)})^{-1}(1 - \alpha/2)$$

Then a $(1 - \alpha)$ confidence interval for $\hat{\lambda}_j$ based on the $\chi^2$ distribution is given by

$$\left( \frac{q_{lj}}{2T_j}, \frac{q_{uj}}{2T_j} \right)$$

Confidence Interval for Rate Difference Statistic

For rate estimates from two independent samples, $\hat{\lambda}_{1j}$ and $\hat{\lambda}_{2j}$, a $(1 - \alpha)$ confidence interval for the rate difference $\hat{\lambda}_{dj} = \hat{\lambda}_{1j} - \hat{\lambda}_{2j}$ is

$$\left( \hat{\lambda}_{dj} - z \sqrt{V(\hat{\lambda}_{dj})}, \hat{\lambda}_{dj} + z \sqrt{V(\hat{\lambda}_{dj})} \right)$$

where $z = \Phi^{-1}(1 - \alpha/2)$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution and the variance $V(\hat{\lambda}_{dj}) = V(\hat{\lambda}_{1j}) + V(\hat{\lambda}_{2j})$
Confidence Interval for Rate Ratio Statistic

For rate estimates from two independent samples, \( \hat{\lambda}_{1j} \) and \( \hat{\lambda}_{2j} \), a \((1 - \alpha)\) confidence interval for the log rate ratio statistic \( \log(\hat{\lambda}_{1j}/\hat{\lambda}_{2j}) \) is

\[
\left( \log(\hat{\lambda}_{rj}) - z \sqrt{V(\log(\hat{\lambda}_{rj}))}, \log(\hat{\lambda}_{rj}) + z \sqrt{V(\log(\hat{\lambda}_{rj}))} \right)
\]

where \( z = \Phi^{-1}(1 - \alpha/2) \) is the \((1 - \alpha/2)\) quantile of the standard normal distribution and the variance \( V(\log(\hat{\lambda}_{rj})) = V(\log(\hat{\lambda}_{1j})) + V(\log(\hat{\lambda}_{2j})) \).

Thus, a \((1 - \alpha)\) confidence interval for the rate ratio statistic \( \hat{\lambda}_{rj} \) is given by

\[
\left( \frac{\hat{\lambda}_{1j}}{\hat{\lambda}_{2j}} e^{-z\sqrt{V(\log(\hat{\lambda}_{rj}))}}, \frac{\hat{\lambda}_{1j}}{\hat{\lambda}_{2j}} e^{z\sqrt{V(\log(\hat{\lambda}_{rj}))}} \right)
\]

Confidence Interval for Rate SMR

At stratum \( j \), a stratum-specific standardized morbidity/mortality ratio is

\( R_j = \frac{d_j}{E_j} \)

where \( E_j \) is the expected number of events.

With the rate \( \hat{\lambda}_j = \frac{d_j}{T_j} \), SMR can be expressed as

\( R_j = \frac{T_j}{E_j} \hat{\lambda}_j \)

Thus, a \((1 - \alpha)\) confidence interval for \( R_j \) is given by

\[
\left( \frac{T_j}{E_j} \hat{\lambda}_{jl}, \frac{T_j}{E_j} \hat{\lambda}_{ju} \right)
\]

where \(( \hat{\lambda}_{jl}, \hat{\lambda}_{ju} \) is a \((1 - \alpha)\) confidence interval for the rate \( \hat{\lambda}_j \).

Risk

An event risk of a population over a specified time period can be defined as the number of new events in the follow-up time period divided by the event-free population size at the beginning of the time period,

\( \hat{y} = \frac{d}{N} \)

where \( N \) is the population size.
For a general population, the subsets (strata) might not be homogeneous enough to have a similar risk. Thus, the risk for each stratum should be computed separately to reflect this discrepancy. For a population that consists of $K$ homogeneous strata (such as different age groups), the stratum-specific risk for the $j$th stratum in a population is computed as

$$\hat{\gamma}_j = \frac{d_j}{N_j}$$

where $N_j$ is the population size in the $j$th stratum of the population.

Assuming the number of events, $d_j$, has a binomial distribution, then a variance estimate of $\hat{\gamma}_j$ is

$$V(\hat{\gamma}_j) = \frac{\hat{\gamma}_j(1 - \hat{\gamma}_j)}{N_j}$$

By using the method of statistical differentials (Elandt-Johnson and Johnson 1980, pp. 70–71), the variance of the logarithm of risk can be estimated by

$$V(\log(\hat{\gamma}_j)) = \frac{1}{\hat{\gamma}_j^2} V(\hat{\gamma}_j) = \frac{1}{\hat{\gamma}_j^2} \frac{\hat{\gamma}_j(1 - \hat{\gamma}_j)}{N_j} = \frac{1 - \hat{\gamma}_j}{\hat{\gamma}_j N_j} = \frac{1}{d_j} - \frac{1}{N_j}$$

**Normal Distribution Confidence Interval for Risk**

A $(1 - \alpha)$ confidence interval for $\hat{\gamma}_j$ based on a normal distribution is given by

$$\left( \hat{\gamma}_j - z \sqrt{V(\hat{\gamma}_j)}, \hat{\gamma}_j + z \sqrt{V(\hat{\gamma}_j)} \right)$$

where $z = \Phi^{-1}(1 - \alpha/2)$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution.

**Lognormal Distribution Confidence Interval for Risk**

A $(1 - \alpha)$ confidence interval for $\log(\hat{\gamma}_j)$ based on a normal distribution is given by

$$\left( \log(\hat{\gamma}_j) - z \sqrt{V(\log(\hat{\gamma}_j))}, \log(\hat{\gamma}_j) + z \sqrt{V(\log(\hat{\gamma}_j))} \right)$$

where $z = \Phi^{-1}(1 - \alpha/2)$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution and the variance $V(\log(\hat{\gamma}_j)) = 1/d_j - 1/N_j$.

Thus, a $(1 - \alpha)$ confidence interval for $\hat{\gamma}_j$ based on a lognormal distribution is given by

$$\left( \hat{\gamma}_j e^{-z \sqrt{\frac{1}{d_j} - \frac{1}{N_j}}}, \hat{\gamma}_j e^{z \sqrt{\frac{1}{d_j} - \frac{1}{N_j}}} \right)$$

**Confidence Interval for Risk Difference Statistic**

For rate estimates from two independent samples, $\hat{\gamma}_{1j}$ and $\hat{\gamma}_{2j}$, a $(1 - \alpha)$ confidence interval for the risk difference $\hat{\gamma}_{dj} = \hat{\gamma}_{1j} - \hat{\gamma}_{2j}$ is

$$\left( \hat{\gamma}_{dj} - z \sqrt{V(\hat{\gamma}_{dj})}, \hat{\gamma}_{dj} + z \sqrt{V(\hat{\gamma}_{dj})} \right)$$

where $z = \Phi^{-1}(1 - \alpha/2)$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution and the variance $V(\hat{\gamma}_{dj}) = V(\hat{\gamma}_{1j}) + V(\hat{\gamma}_{2j})$
Confidence Interval for Risk Ratio Statistic

For rate estimates from two independent samples, \( \hat{\gamma}_{1j} \) and \( \hat{\gamma}_{2j} \), a \((1 - \alpha)\) confidence interval for the log risk ratio statistic \( \log(\hat{\gamma}_{rj}) = \log(\hat{\gamma}_{1j} / \hat{\gamma}_{2j}) \) is

\[
\left( \log(\hat{\gamma}_{rj}) - z \sqrt{V(\log(\hat{\gamma}_{rj}))}, \log(\hat{\gamma}_{rj}) + z \sqrt{V(\log(\hat{\gamma}_{rj}))} \right)
\]

where \( z = \Phi^{-1}(1 - \alpha/2) \) is the \((1 - \alpha/2)\) quantile of the standard normal distribution and the variance

\[
V(\log(\hat{\gamma}_{rj})) = V(\log(\hat{\gamma}_{1j})) + V(\log(\hat{\gamma}_{2j}))
\]

Thus, a \((1 - \alpha)\) confidence interval for the risk ratio statistic \( \hat{\gamma}_{rj} \) is given by

\[
\left( \frac{\hat{\gamma}_{1j}}{\hat{\gamma}_{2j}} e^{-z \sqrt{V(\log(\hat{\gamma}_{rj}))}} , \frac{\hat{\gamma}_{1j}}{\hat{\gamma}_{2j}} e^{z \sqrt{V(\log(\hat{\gamma}_{rj}))}} \right)
\]

Confidence Interval for Risk SMR

At stratum \( j \), a stratum-specific standardized morbidity/mortality ratio is

\[
\mathcal{R}_j = \frac{d_j}{\mathcal{E}_j}
\]

where \( \mathcal{E}_j \) is the expected number of events.

With the risk

\[
\hat{\gamma}_j = \frac{d_j}{N_j}
\]

SMR can be expressed as

\[
\mathcal{R}_j = \frac{N_j}{\mathcal{E}_j} \hat{\gamma}_j
\]

Thus, a \((1 - \alpha)\) confidence interval for \( \mathcal{R}_j \) is given by

\[
\left( \frac{N_j}{\mathcal{E}_j} \hat{\gamma}_{jl} , \frac{N_j}{\mathcal{E}_j} \hat{\gamma}_{ju} \right)
\]

where \(( \hat{\gamma}_{jl} , \hat{\gamma}_{ju} \) is a \((1 - \alpha)\) confidence interval for the risk \( \hat{\gamma}_j \).

Direct Standardization

Direct standardization uses the weights from a reference population to compute the standardized rate of a study group as the weighted average of stratum-specific rates in the study population. The standardized rate is computed as

\[
\hat{\lambda}_{ds} = \frac{\sum_j T_{rj} \hat{\lambda}_{sj}}{T_r}
\]
where \( \hat{\lambda}_{sj} \) is the rate in the \( j \)th stratum of the study population, \( T_{rj} \) is the population-time in the \( j \)th stratum of the reference population, and \( T_r = \sum_k T_{rk} \) is the population-time in the reference population.

Similarly, direct standardization uses the weights from a reference population to compute the standardized risk of a study group as the weighted average of stratum-specific risks in the study population. The standardized risk is computed as

\[
\hat{\gamma}_{ds} = \frac{\sum_j N_{rj} \hat{\gamma}_{sj}}{N_r}
\]

where \( \hat{\gamma}_{sj} \) is the risk in the \( j \)th stratum of the study population, \( N_{rj} \) is the number of observations in the \( j \)th stratum of the reference population, and \( N_r = \sum_k N_{rk} \) is the total number of observations in the reference population.

That is, the directly standardized rate and risk of a study population are weighted averages of the stratum-specific rates and risks, respectively, where the weights are the corresponding strata population sizes in the reference population. The direct standardization can be used when the study population is large enough to provide stable stratum-specific rates or risks. When the same reference population is used for multiple study populations, directly standardized rates and risks provide valid comparisons between study populations.

The variances of the directly standardized rate and risk are

\[
V(\hat{\lambda}_{ds}) = V\left( \frac{\sum_j T_{rj} \hat{\lambda}_{sj}}{T_r} \right) = \frac{\sum_j T_{rj}^2 V(\hat{\lambda}_{sj})}{T_r^2}
\]

\[
V(\hat{\gamma}_{ds}) = V\left( \frac{\sum_j N_{rj} \hat{\gamma}_{sj}}{N_r} \right) = \frac{\sum_j N_{rj}^2 V(\hat{\gamma}_{sj})}{N_r^2}
\]

By using the method of statistical differentials (Elandt-Johnson and Johnson 1980, pp. 70–71), the variance of the logarithm of directly standardized rate and risk can be estimated by

\[
V(\log(\hat{\lambda}_{ds})) = \frac{1}{\hat{\lambda}_{ds}^2} V(\hat{\lambda}_{ds})
\]

\[
V(\log(\hat{\gamma}_{ds})) = \frac{1}{\hat{\gamma}_{ds}^2} V(\hat{\gamma}_{ds})
\]

The confidence intervals for \( \hat{\lambda}_{ds} \) and \( \hat{\gamma}_{ds} \) can be constructed based on normal and lognormal distributions. A gamma distribution confidence interval can also be constructed for \( \hat{\lambda}_{ds} \).

In the next four subsections, \( \beta = \lambda \) denotes the rate statistic and \( \beta = \gamma \) denotes the risk statistic.

**Normal Distribution Confidence Intervals for Standardized Rate and Risk**

A \((1 - \alpha)\) confidence interval for \( \hat{\beta}_{ds} \) based on a normal distribution is then given by

\[
\left( \hat{\beta}_{ds} - z \sqrt{V(\hat{\beta}_{ds})}, \ \hat{\beta}_{ds} + z \sqrt{V(\hat{\beta}_{ds})} \right)
\]

where \( z = \Phi^{-1}(1 - \alpha/2) \) is the \((1 - \alpha/2)\) quantile of the standard normal distribution.
Lognormal Distribution Confidence Intervals for Standardized Rate and Risk

A $(1 - \alpha)$ confidence interval for $\log(\hat{\beta}_{ds})$ based on a normal distribution is given by

$$
\left( \log(\hat{\beta}_{ds}) - z \sqrt{V(\log(\hat{\beta}_{ds}))}, \log(\hat{\beta}_{ds}) + z \sqrt{V(\log(\hat{\beta}_{ds}))} \right)
$$

where $z = \Phi^{-1}(1 - \alpha/2)$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution.

Thus, a $(1 - \alpha)$ confidence interval for $\hat{\beta}_{ds}$ based on a lognormal distribution is given by

$$
\left( \hat{\beta}_{ds} e^{-z \sqrt{V(\log(\hat{\beta}_{ds}))}}, \hat{\beta}_{ds} e^{z \sqrt{V(\log(\hat{\beta}_{ds}))}} \right)
$$

Gamma Distribution Confidence Interval for Standardized Rate

Fay and Feuer (1997) use the relationship between the Poisson and gamma distributions to derive approximate confidence intervals for the standardized rate $\hat{\lambda}_{ds}$ based on the gamma distribution. As in the construction of the asymptotic normal confidence intervals, it is assumed that the number of events has a Poisson distribution, and the standardized rate is a weighted sum of independent Poisson random variables. A confidence interval for $\hat{\lambda}_{ds}$ is then given by

$$
\left( \frac{v}{2\hat{\lambda}_{ds}} \left( \chi^2 \right)^{-1}_{2\hat{\lambda}_{ds} + w_X} \left( \frac{\alpha}{2} \right), \frac{v + w_X^2}{2(\hat{\lambda}_{ds} + w_X)} \left( \chi^2 \right)^{-1}_{2(\hat{\lambda}_{ds} + w_X)} \left( \frac{1 - \alpha}{2} \right) \right)
$$

where

$$
v = \sum_j w_j \frac{\hat{\lambda}_{sj}}{T_{sj}}
$$

$$
w_j = \frac{T_{rij}}{T_r} \frac{1}{T_{sj}}
$$

and $w_X$ is the maximum $w_j$.

Tiwari, Clegg, and Zou (2006) propose a less conservative confidence interval for $\hat{\lambda}_{ds}$ with a different upper confidence limit,

$$
\left( \frac{v}{2\hat{\lambda}_{ds}} \left( \chi^2 \right)^{-1}_{2\hat{\lambda}_{ds} + w_m} \left( \frac{\alpha}{2} \right), \frac{v + w_m^2}{2(\hat{\lambda}_{ds} + w_m)} \left( \chi^2 \right)^{-1}_{2(\hat{\lambda}_{ds} + w_m)} \left( 1 - \frac{\alpha}{2} \right) \right)
$$

where $w_m$ is the average $w_j$ and $w_{2m}$ is the average $w_j^2$.

Comparing Standardized Rates and Comparing Standardized Risks

By using the same reference population, two directly standardized rates or risks from different populations can be compared. Both the difference and ratio statistics can be used in the comparison. Assume that $\hat{\beta}_1$ and $\hat{\beta}_2$ are directly standardized rates or risks for two populations with variances $V(\hat{\beta}_1)$ and $V(\hat{\beta}_2)$, respectively. The difference test assumes that the difference statistic

$$
\hat{\beta}_1 - \hat{\beta}_2
$$
Mantel-Haenszel Effect Estimation

In direct standardization, the derived standardized rates and risks in a study population are the weighted average of the stratum-specific rates and risks in the population, respectively, where the weights are given by the population-time for standardized rate and the number of observations for standardized risk in a reference population.

Assuming that an effect, such as rate difference, rate ratio, risk difference, and risk ratio between two populations, is homogeneous across strata, the Mantel-Haenszel estimates of this effect can be constructed from directly standardized rates or risks in the two populations, where the weights are constructed from the stratum-specific population-times for rate and number of observations for risk of the two populations.

That is, for population \( k, k=1 \) and \( 2 \), the standardized rate and risk are

\[
\hat{\lambda}_k = \frac{\sum_j w_j \hat{\lambda}_{kj}}{\sum_j w_j} \quad \text{and} \quad \hat{\gamma}_k = \frac{\sum_j w_j \hat{\gamma}_{kj}}{\sum_j w_j}
\]

where the weights are

\[
w_j = \frac{T_{1j} T_{2j}}{T_{1j} + T_{2j}}
\]

for standardized rate, and

\[
w_j = \frac{N_{1j} N_{2j}}{N_{1j} + N_{2j}}
\]

for standardized risk.

Rate and Risk Difference Statistics

Denote \( \beta = \lambda \) for rate and \( \beta = \gamma \) for risk. The variance is

\[
V(\hat{\beta}_k) = V \left( \frac{\sum_j w_j \hat{\beta}_{kj}}{\sum_j w_j} \right) = \frac{1}{(\sum_j w_j)^2} \sum_j w_j^2 V(\hat{\beta}_{kj})
\]
The Mantel-Haenszel difference statistic is
\[ \hat{\beta}_1 - \hat{\beta}_2 \]
with variance
\[ V(\hat{\beta}_1 - \hat{\beta}_2) = V(\hat{\beta}_1) + V(\hat{\beta}_2) \]

Under the null hypothesis \( H_0 : \beta_1 = \beta_2 \), the difference statistic \( \hat{\beta}_1 - \hat{\beta}_2 \) has a normal distribution with mean 0.

**Rate Ratio Statistic**

The Mantel-Haenszel rate ratio statistic is \( \hat{\lambda}_1 / \hat{\lambda}_2 \), and the log ratio statistic is
\[ \log \left( \frac{\hat{\lambda}_1}{\hat{\lambda}_2} \right) \]

Under the null hypothesis \( H_0 : \lambda_1 = \lambda_2 \) (or equivalently, \( \log(\lambda_1 / \lambda_2) = 0 \)), the log ratio statistic has a normal distribution with mean 0 and variance
\[ V \left( \log \left( \frac{\hat{\lambda}_1}{\hat{\lambda}_2} \right) \right) = \frac{\sum_j w_j \hat{\lambda}_{pj}}{(\sum_j w_j \hat{\lambda}_1j) (\sum_j w_j \hat{\lambda}_2j)} \]

where
\[ \hat{\lambda}_{pj} = \frac{d_{1j} + d_{2j}}{T_{1j} + T_{2j}} \]
is the combined rate estimate in stratum \( j \) under the null hypothesis of equal rates (Greenland and Robins 1985; Greenland and Rothman 2008, p. 273).

**Risk Ratio Statistic**

The Mantel-Haenszel risk ratio statistic is \( \hat{\gamma}_1 / \hat{\gamma}_2 \), and the log ratio statistic is
\[ \log \left( \frac{\hat{\gamma}_1}{\hat{\gamma}_2} \right) \]

Under the null hypothesis \( H_0 : \gamma_1 = \gamma_2 \) (or equivalently, \( \log(\gamma_1 / \gamma_2) = 0 \)), the log ratio statistic has a normal distribution with mean 0 and variance
\[ V \left( \log \left( \frac{\hat{\gamma}_1}{\hat{\gamma}_2} \right) \right) = \frac{\sum_j w_j (\hat{\gamma}_{pj} - \hat{\gamma}_{1j}\hat{\gamma}_{2j})}{(\sum_j w_j \hat{\gamma}_{1j}) (\sum_j w_j \hat{\gamma}_{2j})} \]

where
\[ \hat{\gamma}_{pj} = \frac{d_{1j} + d_{2j}}{N_{1j} + N_{2j}} \]
is the combined risk estimate in stratum \( j \) under the null hypothesis of equal risks (Greenland and Robins 1985; Greenland and Rothman 2008, p. 275).
Indirect Standardization and Standardized Morbidity/Mortality Ratio

Indirect standardization compares the rates of the study and reference populations by applying the stratum-specific rates in the reference population to the study population, where the stratum-specific rates might not be reliable.

The expected number of events in the study population is

\[ E = \sum_j T_{sj} \hat{\lambda}_{rj} \]

where \( T_{sj} \) is the population-time in the \( j \)th stratum of the study population and \( \hat{\lambda}_{rj} \) is the rate in the \( j \)th stratum of the reference population.

With the expected number of events, \( E \), the standardized morbidity ratio or standardized mortality ratio can be expressed as

\[ R_{sm} = \frac{D}{E} \]

where \( D \) is the observed number of events (Breslow and Day 1987, p. 65).

The ratio \( R_{sm} > 1 \) indicates that the mortality rate or risk in the study population is larger than the estimate in the reference population, and \( R_{sm} < 1 \) indicates that the mortality rate or risk in the study population is smaller than the estimate in the reference population.

With the ratio \( R_{sm} \), an indirectly standardized rate for the study population is computed as

\[ \hat{\lambda}_{is} = R_{sm} \hat{\lambda}_r \]

where \( \hat{\lambda}_r \) is the overall crude rate in the reference population.

Similarly, to compare the risks of the study and reference populations, the stratum-specific risks in the reference population are used to compute the expected number of events in the study population

\[ E = \sum_j N_{sj} \hat{y}_{rj} \]

where \( N_{sj} \) is the number of observations in the \( j \)th stratum of the study population and \( \hat{y}_{rj} \) is the risk in the \( j \)th stratum of the reference population.

Also, with the standardized morbidity ratio \( R_{sm} = D/E \), an indirectly standardized risk for the study population is computed as

\[ \hat{y}_{is} = R_{sm} \hat{y}_r \]

where \( \hat{y}_r \) is the overall crude risk in the reference population.

The observed number of events in the study population is \( D = \sum_j d_{sj} \), where \( d_{sj} \) is the number of events in the \( j \)th stratum of the population. For the rate estimate, if \( d_{sj} \) has a Poisson distribution, then the variance of the standardized mortality ratio \( R_{sm} = D/E \) is

\[ V(R_{sm}) = \frac{1}{E^2} \sum_j V(d_{sj}) = \frac{1}{E^2} \sum_j d_{sj} = \frac{D}{E^2} = \frac{R_{sm}}{E} \]
For the risk estimate, if \( d_{sj} \) has a binomial distribution, then the variance of \( R_{sm} = \frac{D}{E} \) is

\[
V(R_{sm}) = V \left( \frac{1}{E} \sum_j d_{sj} \right) = \frac{1}{E^2} \sum_j V(d_{sj}) = \frac{1}{E^2} \sum_j N_{jx}^2 V(\hat{y}_{sj})
\]

where

\[
V(\hat{y}_{sj}) = \frac{\hat{y}_{sj}(1 - \hat{y}_{sj})}{N_{sj}}
\]

By using the method of statistical differentials (Elandt-Johnson and Johnson 1980, pp. 70–71), the variance of the logarithm of \( R_{sm} \) can be estimated by

\[
V(\log(R_{sm})) = \frac{1}{R_{sm}^2} V(R_{sm})
\]

For the rate estimate,

\[
V(\log(R_{sm})) = \frac{1}{R_{sm}^2} V(R_{sm}) = \frac{1}{R_{sm}^2} \frac{R_{sm}}{E} = \frac{1}{R_{sm}} \frac{1}{E} = \frac{1}{D}
\]

The confidence intervals for \( R_{sm} \) can be constructed based on normal, lognormal, and Poisson distributions.

**Normal Distribution Confidence Interval for SMR**

A \((1 - \alpha)\) confidence interval for \( R_{sm} \) based on a normal distribution is given by

\[
(R_R, R_U) = \left( R_{sm} - z \sqrt{V(R_{sm})}, \ R_{sm} + z \sqrt{V(R_{sm})} \right)
\]

where \( z = \Phi^{-1}(1 - \alpha/2) \) is the \((1 - \alpha/2)\) quantile of the standard normal distribution.

A test statistic for the null hypothesis \( H_0 : \text{SMR} = 1 \) is then given by

\[
\frac{R_{sm} - 1}{\sqrt{V(R_{sm})}}
\]

The test statistic has an approximate standard normal distribution under \( H_0 \).

**Lognormal Distribution Confidence Interval for SMR**

A \((1 - \alpha)\) confidence interval for \( \log(R_{sm}) \) based on a normal distribution is given by

\[
\left( \log(R_{sm}) - z \sqrt{V(\log(R_{sm}))}, \ \log(R_{sm}) + z \sqrt{V(\log(R_{sm}))} \right)
\]

where \( z = \Phi^{-1}(1 - \alpha/2) \) is the \((1 - \alpha/2)\) quantile of the standard normal distribution.

Thus, a \((1 - \alpha)\) confidence interval for \( R_{sm} \) based on a lognormal distribution is given by

\[
\left( R_{sm} e^{-z \sqrt{V(\log(R_{sm}))}}, \ R_{sm} e^{z \sqrt{V(\log(R_{sm}))}} \right)
\]

A test statistic for the null hypothesis \( H_0 : \text{SMR} = 1 \) is then given by

\[
\frac{\log(R_{sm})}{\sqrt{V(\log(R_{sm}))}}
\]

The test statistic has an approximate standard normal distribution under \( H_0 \).
**Poisson Distribution Confidence Interval for SMR**

Denote the \((\alpha/2)\) quantile for the \(\chi^2\) distribution with \(2D\) degrees of freedom by

\[
ql = (\chi^2_{2D})^{-1} (\alpha/2)
\]

Denote the \((1 - \alpha/2)\) quantiles for the \(\chi^2\) distribution with \(2(D + 1)\) degrees of freedom by

\[
q_u = (\chi^2_{2(D+1)})^{-1} (1 - \alpha/2)
\]

Then a \((1 - \alpha)\) confidence interval for \(R_{sm}\) based on the \(\chi^2\) distribution is given by

\[
(R_l, R_u) = \left( \frac{ql}{2E}, \frac{q_u}{2E} \right)
\]

A \(p\)-value for the test of the null hypothesis \(H_0 : SMR = 1\) is given by

\[
2 \min \left( \sum_{k=0}^{D} \frac{e^{-\bar{E}k} k!}{k!}, \sum_{k=D}^{\infty} \frac{e^{-\bar{E}k} k!}{k!} \right)
\]

**Indirectly Standardized Rate and Its Confidence Interval**

With a rate-standardized mortality ratio \(R_{sm}\), an indirectly standardized rate for the study population is computed as

\[
\hat{\lambda}_{is} = R_{sm} \hat{\lambda}_r
\]

where \(\hat{\lambda}_r\) is the overall crude rate in the reference population.

The \((1 - \alpha/2)\) confidence intervals for \(\hat{\lambda}_{is}\) can be constructed as

\[
(R_l \hat{\lambda}_r, R_u \hat{\lambda}_r)
\]

where \((R_l, R_u)\) is the confidence interval for \(R_{sm}\).

**Indirectly Standardized Risk and Its Confidence Interval**

With a risk-standardized mortality ratio \(R_{sm}\), an indirectly standardized risk for the study population is computed as

\[
\hat{y}_{is} = R_{sm} \hat{y}_r
\]

where \(\hat{y}_r\) is the overall crude risk in the reference population.

The \((1 - \alpha/2)\) confidence intervals for \(\hat{y}_{is}\) can be constructed as

\[
(R_l \hat{y}_r, R_u \hat{y}_r)
\]

where \((R_l, R_u)\) is the confidence interval for \(R_{sm}\).
Attributable Fraction and Population Attributable Fraction

The attributable fraction measures the excess event rate or risk fraction in the exposed population that is attributable to the exposure. That is, it is the proportion of event rate or risk in the exposed population that would be reduced if the exposure were not present. In contrast, the population attributable fraction measures the excess event rate or risk fraction in the total population that is attributable to the exposure.

In the STDRATE procedure, you can compute the attributable fraction by using either indirect standardization or Mantel-Haenszel estimation.

Indirect Standardization

With indirect standardization, you specify a study population that consists of subjects who are exposed to a factor, such as smoking, and a reference population that consists of subjects who are not exposed to the factor. Denote the numbers of events in the study and reference populations by $D_s$ and $D_r$, respectively.

For the rate estimate, denote the population-times in the study and reference populations by $T_s$ and $T_r$, respectively. Then the event rates in the two populations can be expressed as the following equations, respectively:

$$\hat{\lambda}_s = \frac{D_s}{T_s} \quad \text{and} \quad \hat{\lambda}_r = \frac{D_r}{T_r}$$

Similarly, for the risk estimate, denote the numbers of observations in the study and reference populations by $N_s$ and $N_r$, respectively. Then the event risks in the two populations can be expressed as the following equations, respectively:

$$\hat{\gamma}_s = \frac{D_s}{N_s} \quad \text{and} \quad \hat{\gamma}_r = \frac{D_r}{N_r}$$

In the next two subsections, $\beta = \lambda$ denotes the rate statistic and $\beta = \gamma$ denotes the risk statistic.

Attributable Fraction with Indirect Standardization

The attributable fraction is the fraction of event rate or risk in the exposed population that is attributable to exposure:

$$R_a = \frac{\hat{\beta}_s - \hat{\beta}_r}{\hat{\beta}_s}$$

With a standardized mortality ratio $R_{sm}$, the attributable fraction is estimated by

$$R_a = \frac{R_{sm} - 1}{R_{sm}}$$

The confidence intervals for the attributable fraction can be computed using the confidence intervals for $R_{sm}$. That is, with a confidence interval $(R_l, R_u)$ for $R_{sm}$, the corresponding $R_a$ confidence interval is given by

$$\left( \frac{R_l - 1}{R_l}, \frac{R_u - 1}{R_u} \right)$$
Population Attributable Fraction with Indirect Standardization

The population attributable fraction for a population is the fraction of event rate or risk in a given time period that is attributable to exposure. The population attributable fraction is

\[ R_{pa} = \frac{\hat{\beta}_0 - \hat{\beta}_r}{\hat{\beta}_0} \]

where

\[ \hat{\beta}_0 = \frac{D_s + D_r}{T_s + T_r} \]

is the combined rate in the total population for the rate statistic and where

\[ \hat{\beta}_0 = \frac{D_s + D_r}{N_s + N_r} \]

is the combined risk in the total population for the risk statistic.

Denote \( \rho = D_s/(D_s + D_r) \), the proportion of exposure among events, then \( R_{pa} \) can also be expressed as

\[ R_{pa} = \rho \frac{R_{sm} - 1}{R_{sm}} \]

where \( R_{sm} \) is the standardized mortality ratio.

An approximate confidence interval for the population attributable rate \( R_{pa} \) can be derived by using the complementary log transformation (Greenland 2008, p. 296). That is, with

\[ \mathcal{H} = \log(1 - R_{pa}) \]

a variance estimator for the estimated \( \mathcal{H} \) is given by

\[ \text{Var}(\hat{\mathcal{H}}) = \frac{R_{pa}^2}{(1 - R_{pa})^2} \left( \frac{\hat{V}}{(R_{sm} - 1)^2} + \frac{2}{D_s (R_{sm} - 1)} + \frac{D_r}{D_s (D_s + D_r)} \right) \]

where \( \hat{V} \) is a variance estimate for \( \log(R_{sm}) \).

Mantel-Haenszel Estimation

With Mantel-Haenszel estimation, you specify one study population that consists of subjects who are exposed to a factor and another study population that consists of subjects who are not exposed to the factor. Denote the numbers of events in the exposed and nonexposed study populations by \( D_1 \) and \( D_2 \), respectively.

For the rate estimate, denote the population-times in the two populations by \( T_1 \) and \( T_2 \), respectively. Then the event rates in the two populations can be expressed as the following equations, respectively:

\[ \hat{\lambda}_1 = \frac{D_1}{T_1} \quad \text{and} \quad \hat{\lambda}_2 = \frac{D_2}{T_2} \]

Similarly, for the risk estimate, denote the numbers of observations in the two populations by \( N_1 \) and \( N_2 \), respectively. Then the event risks in the two populations can be expressed as the following equations, respectively:

\[ \hat{\gamma}_1 = \frac{D_1}{N_1} \quad \text{and} \quad \hat{\gamma}_2 = \frac{D_2}{N_2} \]

In the next two subsections, \( \beta = \lambda \) denotes the rate statistic and \( \beta = \gamma \) denotes the risk statistic.
Attributable Fraction with Mantel-Haenszel Estimation

The attributable fraction is the fraction of event rate or risk in the exposed population that is attributable to exposure:

\[ R_a = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\hat{\beta}_1} \]

Denote the rate or risk ratio by \( R = \hat{\beta}_1 / \hat{\beta}_2 \). The attributable fraction is given by

\[ R_a = \frac{R - 1}{R} \]

The confidence intervals for the attributable fraction can be computed using the confidence intervals for the rate or risk ratio \( R \). That is, with a confidence interval \((R_l, R_u)\) for \( R \), the corresponding \( R_a \) confidence interval is given by

\[ \left( \frac{R_l - 1}{R_l}, \frac{R_u - 1}{R_u} \right) \]

For Mantel-Haenszel estimation, you can use the Mantel-Haenszel rate or risk ratio to estimate \( R \).

Population Attributable Fraction with Mantel-Haenszel Estimation

The population attributable fraction for a population is the fraction of event rate or risk in a given time period that is attributable to exposure. The population attributable fraction is

\[ R_{pa} = \frac{\hat{\beta}_0 - \hat{\beta}_2}{\hat{\beta}_0} \]

where

\[ \hat{\beta}_0 = \frac{D_1 + D_2}{T_1 + T_2} \]

is the combined rate in the total population for the rate statistic and where

\[ \hat{\beta}_0 = \frac{D_1 + D_2}{N_1 + N_2} \]

is the combined risk in the total population for the risk statistic.

Denote the proportion of exposure among events as \( \rho = D_1 / (D_1 + D_2) \). Then \( R_{pa} \) can also be expressed as

\[ R_{pa} = \rho \frac{R - 1}{R} \]

where \( R = \hat{\beta}_1 / \hat{\beta}_2 \) is the rate or risk ratio.

An approximate confidence interval for the population attributable rate \( R_{pa} \) can be derived by using the complementary log transformation (Greenland 2008, p. 296). That is, with

\[ H = \log(1 - R_{pa}) \]
a variance estimator for the estimated $\hat{\mathcal{H}}$ is given by

$$\text{Var}(\hat{\mathcal{H}}) = \frac{\mathcal{R}_{pa}^2}{(1 - \mathcal{R}_{pa})^2} \left( \frac{\hat{V}}{(\mathcal{R} - 1)^2} + \frac{2}{\mathcal{D}_1 (\mathcal{R} - 1)} + \frac{\mathcal{D}_2}{\mathcal{D}_1 (\mathcal{D}_1 + \mathcal{D}_2)} \right)$$

where $\hat{V}$ is a variance estimate for $\log(\mathcal{R})$.

For Mantel-Haenszel estimation, you can use the Mantel-Haenszel rate or risk ratio to estimate $\mathcal{R}$.

---

**Applicable Data Sets and Required Variables for Method Specifications**

The METHOD= and DATA= options are required in the STDRATE procedure. The METHOD= option specifies the standardization method, and the DATA= and REFDATA= options specify the study populations and reference population, respectively. You can use the GROUP= option in the POPULATION statement to identify various study populations. Table 109.2 lists applicable data sets for each method.

<table>
<thead>
<tr>
<th>METHOD=</th>
<th>Number of Populations in DATA= Data Set</th>
<th>REFDATA= Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIRECT</td>
<td>1</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>X</td>
</tr>
<tr>
<td>MH</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>INDIRECT</td>
<td>1</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 109.2 lists the required variables for each method.

<table>
<thead>
<tr>
<th>METHOD=</th>
<th>STAT=</th>
<th>DATA= Data Set</th>
<th>REFDATA= Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RATE</td>
<td>RATE</td>
<td>RISK</td>
</tr>
<tr>
<td>DIRECT</td>
<td>RATE</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>RISK</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MH</td>
<td>RATE</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>RISK</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>INDIRECT</td>
<td>RATE</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>RISK</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
The symbol “X” indicates that the variable is either explicitly specified or implicitly available from other variables. For example, when STAT=RATE, the variable RATE is available if the corresponding variables EVENT and TOTAL are specified.

### Applicable Confidence Limits for Rate and Risk Statistics

In the STDRATE procedure, the METHOD= option specifies the standardization method, and the STAT= option specifies either rate or risk for standardization. Table 109.4 lists applicable confidence limits for different methods with standardized rate, rate SMR, standardized risk, and risk SMR.

#### Table 109.4  Applicable Confidence Limits for Standardized Rate and Risk Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>METHOD=</th>
<th>Normal</th>
<th>Lognormal</th>
<th>Gamma</th>
<th>Poisson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>DIRECT</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>MH</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>INDIRECT</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Rate SMR</td>
<td>INDIRECT</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Risk</td>
<td>DIRECT</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MH</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>INDIRECT</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Risk SMR</td>
<td>INDIRECT</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Table 109.5 lists applicable confidence limits for stratum-specific rate, rate SMR, risk, and risk SMR.

#### Table 109.5  Applicable Confidence Limits for Strata Rate and Risk Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
</tr>
<tr>
<td>Rate</td>
<td>X</td>
</tr>
<tr>
<td>Rate SMR</td>
<td>X</td>
</tr>
<tr>
<td>Risk</td>
<td>X</td>
</tr>
<tr>
<td>Risk SMR</td>
<td>X</td>
</tr>
</tbody>
</table>
The STDRATE procedure displays the “Standardization Information” table by default. In addition, the procedure also displays the “Standardized Rate Estimates” table (with the default STAT=RATE option in the PROC STDRATE statement) and the “Standardized Risk Estimates” table (with the STAT=RISK option) by default. The rest of this section describes the output tables in alphabetical order.

**Attributable Fraction Estimates**

The “Attributable Fraction Estimates” table displays the following information:

- **Parameter**: attributable rate and population attributable rate for the rate statistic, and attributable risk and population attributable risk for the risk statistic
- **Estimate**: estimate of the parameter
- **Method**: method to construct confidence limits
- **Lower and Upper**: lower and upper confidence limits

**Effect Estimates**

The “Effect Estimates” table displays the following information:

- **Standardized Rate**: directly standardized rates for study populations
- **Standardized Risk**: directly standardized risks for study populations

When EFFECT=RATIO, the table displays the following:

- **Estimate**: the rate or risk ratio estimate
- **Log Ratio**: the logarithm of rate ratio or risk ratio estimate
- **Standard Error**: standard error of the logarithm of the ratio estimate
- **Z**: the standard Z statistic
- **Pr > |Z|**: the p-value for the test

When EFFECT=DIFF, the table displays the following:

- **Estimate**: the rate or risk difference estimate
- **Standard Error**: standard error of the difference estimate
- **Z**: the standard Z statistic
- **Pr > |Z|**: the p-value for the test
Chapter 109: The STDRATE Procedure

Standardization Information

The “Standardization Information” table displays the input data sets, type of statistic to be standardized, standardization method, and number of strata. The table also displays the variance divisor for the risk estimate, and the rate multiplier for the rate estimate. With a rate multiplier $c$, the rates per $c$ population-time units are displayed in the output tables.

Standardized Morbidity/Mortality Ratio

The “Standardized Morbidity/Mortality Ratio” table displays the following information:

- SMR: standardized morbidity/mortality ratio
- Standard Error: standard error for SMR
- Lower and Upper: lower and upper confidence limits for SMR
- Test Statistic: SMR-1, for the test of SMR=1
- Estimate: value of test statistic
- Standard Error: standard error of the estimate
- $Z$: the standard $Z$ statistic
- $Pr > |Z|$: the $p$-value for the test

Standardized Rate Estimates

The “Standardized Rate Estimates” table displays the following information:

- Population: study populations, and reference population for indirect standardization
- Number of Events: number of events in population
- Population-Time: total contributed time in population, for the rate statistic
- Crude Rate: event rate in the population
- Expected Number of Events
- SMR: standardized morbidity/mortality ratio, for indirect standardization
- Standardized Rate: for the rate statistic
- Standard Error: standard error of the standardized estimate of rate
- Confidence Limits: lower and upper confidence limits for standardized estimate
Standardized Risk Estimates

The “Standardized Risk Estimates” table displays the following information:

- Population: study populations, and reference population for indirect standardization
- Number of Events: number of events in population
- Number of Observations: number of observations in population
- Crude Risk: event risk in the population
- Expected Number of Events
- SMR: standardized morbidity/mortality ratio, for indirect standardization
- Standardized Risk
- Standard Error: standard error of the standardized estimate of risk
- Confidence Limits: lower and upper confidence limits for standardized estimate

Strata Effect Estimates

The “Strata Effect Estimates” table displays the following information for each stratum:

- Stratum Index: a sequential stratum identification number
- STRATA variables: the levels of STRATA variables
- Rate: rates for the study populations, for the rate statistic
- Risk: risks for the study populations, for the risk statistic

When EFFECT=DIFF, the table displays the following information for each stratum:

- Estimate: rate or risk difference estimate of the study populations
- Standard Error: the standard error of the difference estimate
- Confidence Limits: confidence limits for the difference estimate

When EFFECT=RATIO, the table displays the following information for each stratum:

- Estimate: rate or risk ratio estimate of the study populations
- Confidence Limits: confidence limits for the ratio estimate
Strata Statistics

For each POPULATION statement, the “Strata Information” table displays the following information for each stratum:

- Stratum Index: a sequential stratum identification number
- STRATA variables: the levels of STRATA variables

If the REFERENCE statement is specified, the table displays the following information for each stratum in the reference population:

- Population-Time Value: population-time for the rate statistic for direct standardization
- Population-Time Proportion: proportion for the population-time
- Number of Observations Value: number of observations for the risk statistic for direct standardization
- Number of Observations Proportion: proportion for the number of observations
- Rate: for the rate statistic for indirect standardization
- Risk: for the risk statistic for indirect standardization

For the rate statistic, the table displays the following information for each stratum in the specified study data set:

- Number of Events
- Population-Time Value
- Population-Time Proportion
- Rate Estimate
- Standard Error: standard error for the rate estimate if the CL=NORMAL suboption is specified in the STATS option in the STRATA statement
- Confidence Limits: confidence limits for the risk estimate if the CL suboption is specified in the STATS option in the STRATA statement
- Expected Number of Events: expected number of events that use the reference population population-time for direct standardization, Mantel-Haenszel weight for Mantel-Haenszel estimation, or reference population rate for indirect standardization
For the risk statistic, the table displays the following information for each stratum in the specified study data set:

- Number of Events
- Number of Observations Value
- Number of Observations Proportion
- Risk
- Standard Error: standard error for the risk estimate, if the CL=NORMAL suboption is specified in the STATS option in the STRATA statement
- Confidence Limits: confidence limits for the risk estimate, if the CL suboption is specified in the STATS option in the STRATA statement
- Expected Number of Events: expected number of events that uses the reference population number of observations for direct standardization, Mantel-Haenszel weight for Mantel-Haenszel estimation, or reference population risk for indirect standardization

**Strata SMR Estimates**

The “Strata SMR Estimates” table displays the following information for each stratum:

- Stratum Index: a sequential stratum identification number
- STRATA variables: the levels of STRATA variables
- Number of Events
- Expected Number of Observations
- SMR Estimate
- Standard Error: standard error for the SMR estimate, if the CL=NORMAL suboption is specified in the SMR option in the STRATA statement
- Confidence Limits: confidence limits for the SMR estimate, if the CL suboption is specified in the SMR option in the STRATA statement

For the rate statistic, the table also displays the following information for each stratum:

- Population-Time
- Reference Rate

For the risk statistic, the table also displays the following information for each stratum:

- Number of Observations
- Reference Risk
### ODS Table Names

PROC STDRATE assigns a name to each table it creates. You must use these names to refer to tables when you use the Output Delivery System (ODS). These names are listed in Table 109.6. For more information about ODS, see Chapter 20, “Using the Output Delivery System.”

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>AttrFraction</td>
<td>Attributable fraction</td>
<td>PROC STDRATE</td>
<td>METHOD=INDIRECT(AF)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PROC STDRATE</td>
<td>METHOD=MH(AF)</td>
</tr>
<tr>
<td>Effect</td>
<td>Effect estimates</td>
<td>PROC STDRATE</td>
<td>EFFECT</td>
</tr>
<tr>
<td>SMR</td>
<td>Standardized morbidity/mortality ratio</td>
<td>PROC STDRATE</td>
<td>METHOD=INDIRECT</td>
</tr>
<tr>
<td>StdInfo</td>
<td>Standardization information</td>
<td>PROC STDRATE</td>
<td></td>
</tr>
<tr>
<td>StdRate</td>
<td>Standardized rate estimates</td>
<td>PROC STDRATE</td>
<td>STAT=RATE</td>
</tr>
<tr>
<td>StdRisk</td>
<td>Standardized risk estimates</td>
<td>PROC STDRATE</td>
<td>STAT=RISK</td>
</tr>
<tr>
<td>StrataEffect</td>
<td>Strata effect estimates</td>
<td>STRATA</td>
<td>EFFECT</td>
</tr>
<tr>
<td>StrataStats</td>
<td>Strata statistics</td>
<td>STRATA</td>
<td>STATS</td>
</tr>
<tr>
<td>StrataSMR</td>
<td>Strata SMR estimates</td>
<td>STRATA</td>
<td>SMR</td>
</tr>
</tbody>
</table>

### Graphics Output

This section describes the use of ODS for creating graphics with the STDRATE procedure. To request these graphs, ODS Graphics must be enabled and you must specify the associated graphics options in the PROC STDRATE statement. For more information about ODS Graphics, see Chapter 21, “Statistical Graphics Using ODS.”

**Strata Distribution Plot**

The PLOTS=DIST option displays the proportion of exposed time or sample size for each stratum in the populations.

**Strata Effect Plot**

The PLOTS=EFFECT option displays the stratum-specific effect measure of rate difference, rate ratio, risk difference, or risk ratio. In addition, the crude effect measure and confidence limits of these stratum-specific effect estimates are also displayed.
Strata Rate Plot

The PLOTS=RATE option displays the stratum-specific rate estimates and their confidence limits of populations. In addition, the overall crude rates of populations are also displayed.

Strata Risk Plot

The PLOTS=RISK option displays the stratum-specific risk estimates and their confidence limits of populations. In addition, the overall crude risks of populations are also displayed.

Strata SMR Plot

The PLOTS=SMR option displays the SMR for each stratum in the populations.

ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, “Statistical Graphics Using ODS.”

Before you create graphs, ODS Graphics must be enabled (for example, by specifying the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section “Enabling and Disabling ODS Graphics” on page 607 in Chapter 21, “Statistical Graphics Using ODS.”

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section “A Primer on ODS Statistical Graphics” on page 606 in Chapter 21, “Statistical Graphics Using ODS.”

PROC STDRATE assigns a name to each graph it creates. You can use these names to refer to the graphs when you use ODS. To request the graph, ODS Graphics must be enabled and you must use the PLOTS option in the PROC STDRATE statement to specify the plot-request indicated in Table 109.7.

<table>
<thead>
<tr>
<th>ODS Graph Name</th>
<th>Plot Description</th>
<th>plot-request in PLOTS Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>StrataDistPlot</td>
<td>Strata proportion of exposed time or sample size</td>
<td>DIST</td>
</tr>
<tr>
<td>StrataEffectPlot</td>
<td>Strata effect measure of rate difference, rate ratio, risk difference, or risk ratio</td>
<td>EFFECT</td>
</tr>
<tr>
<td>StrataRatePlot</td>
<td>Strata rate estimates</td>
<td>RATE</td>
</tr>
<tr>
<td>StrataRiskPlot</td>
<td>Strata risk estimates</td>
<td>RISK</td>
</tr>
<tr>
<td>StrataSMRPlot</td>
<td>Strata SMR of rates or risks</td>
<td>SMR</td>
</tr>
</tbody>
</table>
Chapter 109: The STDRATE Procedure

Examples: STDRATE Procedure

Example 109.1: Comparing Directly Standardized Rates

This example computes directly standardized mortality rates for populations in the states of Alaska and Florida, and then compares these two standardized rates with a rate ratio statistic.

The Alaska data set contains the stratum-specific mortality information in a given period of time for the state of Alaska (Alaska Bureau of Vital Statistics 2000a, b). Variables Sex and Age are the grouping variables that form the strata in the standardization, and variables Death and PYear indicate the number of events and person-years, respectively. The COMMA7. format is specified in the DATA step to input numerical values that contain commas in PYear.

data Alaska;
  State='Alaska';
  input Sex $ Age $ Death PYear:comma7.;
datalines;
  Male  00-14  37 81,205
  Male  15-34  68 93,662
  Male  35-54 206 108,615
  Male  55-74 369 35,139
  Male  75+  556 5,491
  Female 00-14  78 77,203
  Female 15-34 181 85,412
  Female 35-54 395 100,386
  Female 55-74 555 32,118
  Female 75+  479 7,701
;

The Florida data set contains the corresponding stratum-specific mortality information for the state of Florida (Florida Department of Health 2000, 2013). Variables Sex and Age are the grouping variables that form the strata in the standardization, and variables Death and PYear indicate the number of events and person-years, respectively.

data Florida;
  State='Florida';
  input Sex $ Age $ Death:comma6. PYear:comma9.;
datalines;
  Male  00-14 1,189 1,505,889
  Male  15-34 2,962 1,972,157
  Male  35-54 10,279 2,197,912
  Male  55-74 26,354 1,383,533
  Male  75+  42,443 554,632
  Female 00-14 1,312 1,445,831
  Female 15-34 1,870 1,870,430
  Female 35-54 5,630 2,246,737
  Female 55-74 18,309 1,612,270
  Female 75+ 11,056 868,838
;
Example 109.1: Comparing Directly Standardized Rates

The TwoStates data set contains the data sets Alaska and Florida:

```latex
\begin{verbatim}
data TwoStates;
    length State $ 7.;
    set Alaska Florida;
run;
\end{verbatim}
```

The US data set contains the corresponding stratum-specific person-years information for the United States (US Bureau of the Census 2011). Variables Sex and Age are the grouping variables that form the strata in the standardization, and the variable PYear indicates the person-years.

```latex
\begin{verbatim}
data US;
    input Sex $ Age $ PYear:comma10.;
datalines;
   Male  00-14  30,854,207
   Male  15-34  40,199,647
   Male  35-54  40,945,028
   Male  55-74  19,948,630
   Male  75+    6,106,351
   Female 00-14 29,399,168
   Female 15-34 38,876,268
   Female 35-54 41,881,451
   Female 55-74 22,717,040
   Female 75+  10,494,416
;
\end{verbatim}
```

The following statements invoke the STDRATE procedure and compute the direct standardized rates for the states of Florida and Alaska by using the United States as the reference population. The DATA= option names the data set for the study populations, and the REFDATA= option names the data set for the reference population.

```latex
\begin{verbatim}
ods graphics on;
proc stdrate data=TwoStates
    refdata=US
    method=direct
    stat=rate(mult=1000)
    effect
    plots(only)=(dist effect)
    ;
    population group=State event=Death total=PYear;
    reference total=PYear;
    strata Sex Age / effect;
run;
\end{verbatim}
```

The METHOD=DIRECT option requests direct standardization, and the STAT=RATE option specifies the rate statistic for standardization. With the EFFECT option, the procedure computes the rate effect between the study populations with the default rate ratio statistics.
The “Standardization Information” table in Output 109.1.1 displays the standardization information.

**Output 109.1.1 Standardization Information**

<table>
<thead>
<tr>
<th>Standardization Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Group Variable</td>
</tr>
<tr>
<td>Reference Data Set</td>
</tr>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Statistic</td>
</tr>
<tr>
<td>Number of Strata</td>
</tr>
<tr>
<td>Rate Multiplier</td>
</tr>
</tbody>
</table>

With ODS Graphics enabled, the PLOTS(ONLY)=(DIST EFFECT) option displays the strata distribution plot and the strata effect plot, but does not display the default strata rate plot.

The strata distribution plot displays proportions for stratum-specific person-years in the study populations and reference population, as shown in Output 109.1.2.

**Output 109.1.2 Strata Distribution Plot**
The EFFECT option in the STRATA statement and the STAT=RATE option request that the “Strata Rate Effect Estimates” table in Output 109.1.3 display the stratum-specific rate effect statistics between the two study populations. The default EFFECT=RATIO in the PROC STDRATE statement requests that the stratum-specific rate ratio statistics be displayed.

### Output 109.1.3 Strata Effect Estimates

<table>
<thead>
<tr>
<th>Stratum Index</th>
<th>Sex</th>
<th>Age</th>
<th>Alaska</th>
<th>Florida</th>
<th>Rate Ratio</th>
<th>95% Lognormal Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Female</td>
<td>00-14</td>
<td>1.010</td>
<td>0.6266</td>
<td>1.61231</td>
<td>1.27940 2.03185</td>
</tr>
<tr>
<td>2</td>
<td>Female</td>
<td>15-34</td>
<td>2.119</td>
<td>0.6597</td>
<td>3.21208</td>
<td>2.74812 3.75437</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>35-54</td>
<td>3.935</td>
<td>2.5059</td>
<td>1.57025</td>
<td>1.41795 1.73889</td>
</tr>
<tr>
<td>4</td>
<td>Female</td>
<td>55-74</td>
<td>17.280</td>
<td>11.3560</td>
<td>1.52166</td>
<td>1.39844 1.65574</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>75+</td>
<td>62.200</td>
<td>61.5638</td>
<td>1.01033</td>
<td>0.92341 1.10542</td>
</tr>
<tr>
<td>6</td>
<td>Male</td>
<td>00-14</td>
<td>0.456</td>
<td>0.7896</td>
<td>0.57707</td>
<td>0.41604 0.80044</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>15-34</td>
<td>0.726</td>
<td>1.5019</td>
<td>0.48339</td>
<td>0.38010 0.61476</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>35-54</td>
<td>1.897</td>
<td>4.6767</td>
<td>0.40554</td>
<td>0.35330 0.46552</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>55-74</td>
<td>10.501</td>
<td>19.0483</td>
<td>0.55129</td>
<td>0.49746 0.61094</td>
</tr>
<tr>
<td>10</td>
<td>Male</td>
<td>75+</td>
<td>101.257</td>
<td>76.5246</td>
<td>1.32319</td>
<td>1.21699 1.43866</td>
</tr>
</tbody>
</table>

The “Strata Rate Effect Estimates” table shows that except for the age group 75+, Alaska has lower mortality rates for male groups and higher mortality rates for female groups than Florida. For age group 75+, Alaska has higher mortality rates than Florida for both male and female groups.
With ODS Graphics enabled and two study populations, the PLOTS=EFFECT option displays the stratum-specific effect measures and their associated confidence limits, as shown in Output 109.1.4. The STAT=RATE option and the default EFFECT=RATIO option request that the strata rate ratios be displayed. By default, confidence limits are generated with 95\% confidence level. This plot displays the stratum-specific rate ratios in the “Strata Rate Effect Estimates” table in Output 109.1.3.

**Output 109.1.4** Strata Effect Measure Plot

![Strata Rate Ratios with 95\% Lognormal Confidence Limits](image)
The “Directly Standardized Rate Estimates” table in Output 109.1.5 displays directly standardized rates and related statistics.

### Output 109.1.5 Directly Standardized Rate Estimates

<table>
<thead>
<tr>
<th>State</th>
<th>Observed Events</th>
<th>Population-Time</th>
<th>Crude Rate</th>
<th>Expected Events</th>
<th>Population-Time</th>
<th>Standard Rate</th>
<th>95% Normal Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alaska</td>
<td>2924</td>
<td>626932</td>
<td>4.664</td>
<td>2270876</td>
<td>281422206</td>
<td>8.0693</td>
<td>0.1643 7.7472 8.3913</td>
</tr>
<tr>
<td>Florida</td>
<td>162795</td>
<td>15658229</td>
<td>10.3968</td>
<td>2176572</td>
<td>281422206</td>
<td>7.7342</td>
<td>0.0195 7.6959 7.7725</td>
</tr>
</tbody>
</table>

The MULT=1000 suboption in the STAT=RATE option requests that rates per 1,000 person-years be displayed. The table shows that although the crude rate in the Florida population (10.3968) is higher than the crude rate in the Alaska population (4.664), the resulting standardized rate in the Florida population (7.7342) is lower than the crude rate in the Alaska population (8.0693).

The EFFECT option requests that the “Rate Effect Estimates” table in Output 109.1.6 display the log rate ratio statistics of the two directly standardized rates by default.

### Output 109.1.6 Effect Estimates

| State | Rate Ratio | Lognormal Confidence Limits | Log Rate Ratio | Standard Error | Z Pr > |Z| |
|-------|------------|-----------------------------|----------------|----------------|--------|---|
| Alaska| 8.0693     | 7.7342                      | 1.0433         | 1.00220        | 1.08614| 0.0205 2.07 0.0387 |
| Florida| 7.7342     | 1.0433                      | 1.00220        | 1.08614        | 0.0205 2.07 0.0387 |

The table shows that with a log rate ratio statistic 1.0433, the resulting p-value is 0.0387, which indicates that the death rate is significantly higher in Alaska than in Florida at the 5% significance level.
Example 109.2: Computing Mantel-Haenszel Risk Estimation

This example uses Mantel-Haenszel method to estimate the effect of household smoking on respiratory symptoms of school children, after adjusting for the effects of the student’s grade and household pets.

Suppose that the School data set contains the stratum-specific numbers of cases of respiratory symptoms in a given school year for a school district. Variables Pet and Grade are the grouping variables that form the strata in the standardization, and the variable Smoking identifies students who have smokers in their households. The variables Case and Student indicate the number of cases with respiratory symptoms and the total number of students, respectively.

```plaintext
data School;
  input Smoking $ Pet $ Grade $ Case Student;
  datalines;
  Yes Yes K-1 109 807
  Yes Yes 2-3 106 791
  Yes Yes 4-5 112 868
  Yes No K-1 168 1329
  Yes No 2-3 162 1337
  Yes No 4-5 183 1594
  No Yes K-1 284 2403
  No Yes 2-3 266 2237
  No Yes 4-5 273 2279
  No No K-1 414 3398
  No No 2-3 372 3251
  No No 4-5 382 3270
;
```

The following statements invoke the STDRATE procedure and compute the Mantel-Haenszel rate difference statistic between students with household smokers and students without household smokers:

```plaintext
ods graphics on;
proc stdrate data=School
  method=mh
  stat=risk
  effect=diff
  plots=all;
  population group=Smoking event=Case total=Student;
  strata Pet Grade / order=data stats(cl=none) effect;
run;
```

The ORDER=DATA option in the STRATA statement sorts the strata by the order of appearance in the input data set.
The “Standardization Information” table in Output 109.2.1 displays the standardization information.

**Output 109.2.1** Standardization Information

<table>
<thead>
<tr>
<th>Standardization Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>WORK.SCHOOL</td>
</tr>
<tr>
<td>Group Variable</td>
</tr>
<tr>
<td>Smoking</td>
</tr>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Mantel-Haenszel</td>
</tr>
<tr>
<td>Statistic</td>
</tr>
<tr>
<td>Risk</td>
</tr>
<tr>
<td>Number of Strata</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

The STATS option in the STRATA statement requests that the STDRA TE procedure display a “Mantel-Haenszel Standardized Strata Statistics” table for study populations, as shown in Output 109.2.2. The table displays the strata information and the expected number of events in each stratum. The Expected Events column shows the expected number of events when the Mantel-Haenszel weights are applied to the corresponding stratum-specific risks in the study populations. The CL=NONE suboption requests that confidence limits for strata risks not be displayed.

**Output 109.2.2** Mantel-Haenszel Standardized Strata Statistics

<table>
<thead>
<tr>
<th>Mantel-Haenszel Standardized Strata Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Population</td>
</tr>
<tr>
<td>Number of Observations</td>
</tr>
<tr>
<td>Smoking</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>No</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>Yes</td>
</tr>
</tbody>
</table>

With ODS Graphics enabled, the PLOTS=ALL option displays all appropriate plots. With the METHOD=MH and STAT=RISK options, these plots include the strata distribution plot, strata risk plot, and strata effect plot.
The strata distribution plot displays proportions for stratum-specific numbers of students in the study populations, as shown in Output 109.2.3.

Output 109.2.3 Strata Distribution Plot
The strata risk plot displays stratum-specific risk estimates with confidence limits in the study populations, as shown in Output 109.2.4. This plot displays stratum-specific risk estimates in the “Mantel-Haenszel Standardized Strata Statistics” table in Output 109.2.2. In addition, the overall crude risks for the two study populations are also displayed. By default, strata levels are displayed on the vertical axis.

Output 109.2.4 Strata Risk Plot
The EFFECT option in the STRATA statement requests that the “Strata Risk Effect Estimates” table be displayed, as shown in Output 109.2.5. The EFFECT=DIFF option in the PROC STDRATE statement requests that strata risk differences be displayed.

Output 109.2.5 Strata Effect Estimates

<table>
<thead>
<tr>
<th>Stratum Index</th>
<th>Pet</th>
<th>Grade</th>
<th>No</th>
<th>Yes</th>
<th>Risk Difference</th>
<th>Standard Error</th>
<th>95% Normal Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yes</td>
<td>K-1</td>
<td>0.11819</td>
<td>0.13507</td>
<td>-.016883</td>
<td>.013716</td>
<td>-.043766 .010001</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>2-3</td>
<td>0.11891</td>
<td>0.13401</td>
<td>-.015098</td>
<td>.013912</td>
<td>-.042366 .012169</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>4-5</td>
<td>0.11979</td>
<td>0.12903</td>
<td>-.009243</td>
<td>.013257</td>
<td>-.035225 .016740</td>
</tr>
<tr>
<td>4</td>
<td>No</td>
<td>K-1</td>
<td>0.12184</td>
<td>0.12641</td>
<td>-.004574</td>
<td>.010704</td>
<td>-.025554 .016405</td>
</tr>
<tr>
<td>5</td>
<td>No</td>
<td>2-3</td>
<td>0.11443</td>
<td>0.12117</td>
<td>-.006740</td>
<td>.010527</td>
<td>-.027373 .013892</td>
</tr>
<tr>
<td>6</td>
<td>No</td>
<td>4-5</td>
<td>0.11682</td>
<td>0.11481</td>
<td>0.002014</td>
<td>.009762</td>
<td>-.017120 .021148</td>
</tr>
</tbody>
</table>

The “Strata Risk Effect Estimates” table shows that for the stratum of students without household pets in Grade 4–5, the risk is higher for students without household smokers than for students with household smokers. For all other strata, the risk is lower for students without household smokers than for students with household smokers. The difference is not significant in each stratum because the null value 0 is between the lower and upper confidence limits.
With ODS Graphics enabled, the PLOTS=EFFECT option displays the plot with the stratum-specific risk effect measures and their associated confidence limits, as shown in Output 109.2.6. The EFFECT=DIFF option requests that the risk difference be displayed. By default, confidence limits are generated with 95% confidence level. This plot displays the stratum-specific risk differences in the “Strata Risk Effect Estimates” table in Output 109.2.5.

**Output 109.2.6 Strata Risk Plot**

![Strata Risk Differences with 95% Normal Confidence Limits](image)
The “Mantel-Haenszel Standardized Risk Estimates” table in Output 109.2.7 displays the Mantel-Haenszel standardized risks and related statistics.

**Output 109.2.7**  Standardized Risk Estimates (Mantel-Haenszel Estimation)

<table>
<thead>
<tr>
<th>Smoking</th>
<th>Observed Events</th>
<th>Number of Observations</th>
<th>Crude Risk</th>
<th>Expected Events</th>
<th>Weight</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>95% Normal Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>1991</td>
<td>16838</td>
<td>0.1182</td>
<td>566.172</td>
<td>4791.43</td>
<td>0.1182</td>
<td>0.00250</td>
<td>0.1133 0.1231</td>
</tr>
<tr>
<td>Yes</td>
<td>840</td>
<td>6726</td>
<td>0.1249</td>
<td>599.602</td>
<td>4791.43</td>
<td>0.1251</td>
<td>0.00404</td>
<td>0.1172 0.1331</td>
</tr>
</tbody>
</table>

The EFFECT=DIFF option requests that the “Risk Effect Estimates” table display the risk difference statistic for the two directly standardized risks, as shown in Output 109.2.8.

**Output 109.2.8**  Mantel-Haenszel Effect Estimates

| Smoking | Risk Effect Estimates | 95% Normal Confidence Limits | Standard Error | Z Pr > |Z| |
|---------|-----------------------|-----------------------------|----------------|---------|---|
| No      | 0.1182                | -0.00698                    | 0.002330       | 0.00475 | -1.47 | 0.1418 |
| Yes     | 0.1251                | -0.016284                   | 0.002330       | 0.00475 | -1.47 | 0.1418 |

The table shows that although the standardized risk for students without household smokes is lower than the standardized risk for students with household smokes, the difference (–0.00698) is not significant at the 5% significance level, ($p$-value 0.1418).
Example 109.3: Computing Attributable Fraction Estimates

This example computes the excess event risk fraction that is attributable to a specific chemical exposure for workers in a factory.

Suppose that the Factory data set contains the stratum-specific event information for exposure to a specific chemical agent. The variable Age is the grouping variable that forms the strata. The variables Event_E and Count_E indicate the number of events and number of workers for workers with the specific chemical exposure, respectively. The variables Event_NE and Count_NE indicate the number of events and number of workers for workers without the specific chemical exposure, respectively.

```
data Factory;
   input Age $ Event_E Count_E Event_NE Count_NE;
datalines;
   20-29 31 352 143 2626
   30-39 57 486 392 4124
   40-49 62 538 459 4662
   50-59 50 455 337 3622
   60-69 38 322 199 2155
   70+ 9 68 35 414
;
```

The following statements invoke the STDRATE procedure and compute the attributable risk and population attributable risk for the chemical exposure:

```
ods graphics on;
proc stdrate data=Factory
   refdata=Factory
   method=indirect(af)
   stat=risk
   plots(stratum=horizontal)
;
population event=Event_E total=Count_E;
reference event=Event_NE total=Count_NE;
strata Age / stats;
run;
```

The “Standardization Information” table in Output 109.3.1 displays the standardization information.

Output 109.3.1 Standardization Information

<table>
<thead>
<tr>
<th>Standardization Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Reference Data Set</td>
</tr>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Statistic</td>
</tr>
<tr>
<td>Number of Strata</td>
</tr>
</tbody>
</table>
The STATS option in the STRATA statement requests that the “Indirectly Standardized Strata Statistics” table in Output 109.3.2 display the strata information and the expected number of events at each stratum. The Expected Events column shows the expected numbers of events when the stratum-specific risks in the reference data set are applied to the corresponding numbers of workers in the study data set.

### Output 109.3.2 Strata Information (Indirect Standardization)

The STDRATE Procedure

<table>
<thead>
<tr>
<th>Stratum Index</th>
<th>Age</th>
<th>Observed Events</th>
<th>Number of Observations</th>
<th>Proportion</th>
<th>Crude Risk</th>
<th>Standard Error</th>
<th>95% Normal Confidence Limits</th>
<th>Value</th>
<th>Proportion</th>
<th>Crude Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20-29</td>
<td>31</td>
<td>352</td>
<td>0.1585</td>
<td>0.088068</td>
<td>0.015105</td>
<td>0.038463 0.117673</td>
<td>2626</td>
<td>0.1492</td>
<td>0.05446</td>
</tr>
<tr>
<td>2</td>
<td>30-39</td>
<td>57</td>
<td>486</td>
<td>0.2188</td>
<td>0.117284</td>
<td>0.014595</td>
<td>0.088678 0.145890</td>
<td>4124</td>
<td>0.2343</td>
<td>0.09505</td>
</tr>
<tr>
<td>3</td>
<td>40-49</td>
<td>62</td>
<td>538</td>
<td>0.2422</td>
<td>0.115242</td>
<td>0.013767</td>
<td>0.088260 0.142224</td>
<td>4662</td>
<td>0.2648</td>
<td>0.09846</td>
</tr>
<tr>
<td>4</td>
<td>50-59</td>
<td>50</td>
<td>455</td>
<td>0.2049</td>
<td>0.109890</td>
<td>0.014662</td>
<td>0.081153 0.138627</td>
<td>3622</td>
<td>0.2058</td>
<td>0.09304</td>
</tr>
<tr>
<td>5</td>
<td>60-69</td>
<td>38</td>
<td>322</td>
<td>0.1450</td>
<td>0.118012</td>
<td>0.017979</td>
<td>0.082774 0.153251</td>
<td>2155</td>
<td>0.1224</td>
<td>0.09234</td>
</tr>
<tr>
<td>6</td>
<td>70+</td>
<td>9</td>
<td>68</td>
<td>0.0306</td>
<td>0.132353</td>
<td>0.041095</td>
<td>0.051809 0.212897</td>
<td>414</td>
<td>0.0235</td>
<td>0.08454</td>
</tr>
</tbody>
</table>

With ODS Graphics enabled and the specified STAT=RISK option, the default PLOTS=RISK option displays the stratum-specific risk estimates in the study and reference populations, as shown in Output 109.3.3. The STRATUM=HORIZONTAL global option in the PLOTS option displays the strata information on the horizontal axis. The plot displays the stratum-specific risk estimates in the “Indirectly Standardized Strata Statistics” table in Output 109.3.2. In addition, confidence limits for the risk estimates in the study population and the overall crude risks for the two populations are also displayed.
The METHOD=INDIRECT option requests that the “Standardized Morbidity/Mortality Ratio” table in Output 109.3.4 display the SMR, its 95% confidence limits, and the test for the null hypothesis $H_0 : \text{SMR} = 1$.

### Output 109.3.4 Standardized Morbidity/Mortality Ratio

<table>
<thead>
<tr>
<th>Standardized Morbidity/Mortality Ratio</th>
<th>95% Normal Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observed Events</td>
<td>Expected Events</td>
</tr>
<tr>
<td>-----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>247</td>
<td>196.151</td>
</tr>
</tbody>
</table>

The “Standardized Morbidity/Mortality Ratio” table shows that SMR=1.259, the 95% confidence limits do not contain the null value SMR=1, and the null hypothesis of SMR=1 is rejected at $\alpha = 0.05$ level from the normal test.
The “Indirectly Standardized Risk Estimates” table in Output 109.3.5 displays the standardized risks and related statistics.

**Output 109.3.5** Standardized Risks (Indirect Standardization)

<table>
<thead>
<tr>
<th>Study Population</th>
<th>Indirectly Standardized Risk Estimates</th>
<th>Standardized Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed Events</td>
<td>Number of Observations</td>
</tr>
<tr>
<td></td>
<td>247</td>
<td>2221</td>
</tr>
</tbody>
</table>

The AF suboption in the METHOD=INDIRECT option requests that the “Attributable Fraction Estimates” table display the attributable risk and population attributable risk, as shown in Output 109.3.6

**Output 109.3.6** Attributable Fraction Estimates

<table>
<thead>
<tr>
<th>Attributable Fraction Estimates</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attributable Risk</td>
<td>0.20587 0.10013 0.28937</td>
</tr>
<tr>
<td>Population Attributable Risk</td>
<td>0.02806 0.01159 0.04426</td>
</tr>
</tbody>
</table>

The attributable risk fraction 0.206 indicates that 20.6% of all events in the chemical exposure group are attributed to the chemical exposure, and the population attributable risk fraction 0.028 indicates that about 2.8% of all events in the total population are attributed to the chemical exposure.

The Attributable fraction can also be computed by using Mantel-Haenszel method.

Suppose that the Factory1 data set contains the stratum-specific event information for exposure to a specific chemical agent. The variable Age is the grouping variable that forms the strata, and the variable Exposure identifies workers with chemical exposure. The variables Event and Count indicate the number of events and number of workers, respectively.
data Factory1;
  input Exposure $ Age $ Event Count;
datalines;
Yes 20-29 31 352
Yes 30-39 57 486
Yes 40-49 62 538
Yes 50-59 50 455
Yes 60-69 38 322
Yes 70+ 9 68
No 20-29 143 2626
No 30-39 392 4124
No 40-49 459 4662
No 50-59 337 3622
No 60-69 199 2155
No 70+ 35 414;

The following statements invoke the STDRATE procedure and compute the attributable risk and population attributable risk for the chemical exposure:

```
proc stdrate data=Factory1
  method=mh(af)
  stat=risk
  effect;
  population group(order=data exposed='Yes')=Exposure
    event=Event total=Count;
  strata Age;
run;
```

The GROUP=EXPOSURE option specifies the variable Exposure, whose values identify the various populations. The ORDER= suboption specifies the order in which the values of Exposure are to be displayed, and the EXPOSED= option identifies the exposed group in the derivation of the attributable fraction.

The “Standardization Information” table in Output 109.3.7 displays the standardization information.

**Output 109.3.7** Standardization Information

<table>
<thead>
<tr>
<th>Standardization Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Group Variable</td>
</tr>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Statistic</td>
</tr>
<tr>
<td>Number of Strata</td>
</tr>
</tbody>
</table>
The “Mantel-Haenszel Standardized Risk Estimates” table in Output 109.3.8 displays the Mantel-Haenszel standardized risks and related statistics.

**Output 109.3.8** Standardized Risk Estimates (Mantel-Haenszel Estimation)

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Observed Events</th>
<th>Number of Observations</th>
<th>Crude Risk</th>
<th>Expected Events</th>
<th>Weight</th>
<th>Standard Error</th>
<th>Standard Error</th>
<th>95% Normal Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>247</td>
<td>2221</td>
<td>0.1112</td>
<td>219.122</td>
<td>1970.26</td>
<td>0.1112</td>
<td>0.00667</td>
<td>0.0981 0.1243</td>
</tr>
<tr>
<td>No</td>
<td>1565</td>
<td>17603</td>
<td>0.0889</td>
<td>174.134</td>
<td>1970.26</td>
<td>0.0884</td>
<td>0.00214</td>
<td>0.0842 0.0926</td>
</tr>
</tbody>
</table>

The EFFECT option requests that the “Risk Effect Estimates” table display the risk ratio statistic for the two directly standardized risks, as shown in Output 109.3.9.

**Output 109.3.9** Mantel-Haenszel Effect Estimates

<table>
<thead>
<tr>
<th>Exposure</th>
<th>Risk Ratio</th>
<th>95% Lognormal Confidence Limits</th>
<th>Log Risk Ratio</th>
<th>Standard Error</th>
<th>Z</th>
<th>Pr &gt;</th>
<th>Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>0.1112</td>
<td>1.2584 1.10851 1.42845</td>
<td>0.2298</td>
<td>0.0647</td>
<td>3.55</td>
<td>0.0004</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>0.0884</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The AF suboption in the METHOD=MH option requests that the “Attributable Fraction Estimates” table display the attributable risk and population attributable risk, as shown in Output 109.3.10

**Output 109.3.10** Attributable Fraction Estimates

<table>
<thead>
<tr>
<th>Attributable Fraction Estimates Exposed = Yes</th>
<th>95% Confidence Limits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Estimate</td>
</tr>
<tr>
<td>Attributable Risk</td>
<td>0.20531</td>
</tr>
<tr>
<td>Population Attributable Risk</td>
<td>0.02799</td>
</tr>
</tbody>
</table>
Similar to the results of using the SMR estimates, the attributable risk fraction (0.205) indicates that 20.5\% of all events in the chemical exposure group are attributed to the chemical exposure, and the population attributable risk fraction (0.028) indicates that about 2.8\% of all events in the total population are attributed to the chemical exposure.

**Example 109.4: Displaying SMR Results from BY Groups**

This example illustrates the use of ODS OUTPUT statement to save standardized mortality ratios for different causes and to display these statistics together in a table and in a plot.

The Florida_Cs data set contains the stratum-specific mortality information for stomach cancer and skin cancer in year 2000 for the state of Florida (Florida Department of Health 2000, 2013). The variable Age is the grouping variable that forms the strata in the standardization. The variables Event_C16, Event_C43, and PYear identify the number of events for stomach cancer, the number of events for skin cancer, and the person-years, respectively. The COMMA9. format is specified in the DATA step to input numerical values that contain commas in PYear.

```plaintext
data Florida_Cs;
  input Age $1-5 Event_C16 Event_C43 PYear: comma9.;
datalines;
00-04  0  0  953,785
05-14  0  0  1,997,935
15-24  0  4  1,885,014
25-34  1 14  1,957,573
35-44 19 43  2,356,649
45-54 64 72  2,088,000
55-64 114 70  1,548,371
65-74 201 126  1,447,432
75-84 294 136  1,087,524
85+  136  73  335,944;
```

The following statements construct and list the mortality information by cancer cause:

```plaintext
data Florida_Cs;
  set Florida_Cs;
  Cause='Stomach';  Event=Event_C16;  output;
  Cause='Skin';    Event=Event_C43;  output;
  drop Event_C16 Event_C43;
run;

proc sort data=Florida_Cs;
  by Cause;
run;

proc print data=Florida_Cs;
  var Cause Age Event PYear;
run;
```
Chapter 109: The STDRATE Procedure

Output 109.4.1 Florida Data

<table>
<thead>
<tr>
<th>Obs</th>
<th>Cause</th>
<th>Age</th>
<th>Event</th>
<th>PYear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Skin</td>
<td>00-04</td>
<td>0</td>
<td>953785</td>
</tr>
<tr>
<td>2</td>
<td>Skin</td>
<td>05-14</td>
<td>0</td>
<td>1997935</td>
</tr>
<tr>
<td>3</td>
<td>Skin</td>
<td>15-24</td>
<td>4</td>
<td>1885014</td>
</tr>
<tr>
<td>4</td>
<td>Skin</td>
<td>25-34</td>
<td>14</td>
<td>1957573</td>
</tr>
<tr>
<td>5</td>
<td>Skin</td>
<td>35-44</td>
<td>43</td>
<td>2356649</td>
</tr>
<tr>
<td>6</td>
<td>Skin</td>
<td>45-54</td>
<td>72</td>
<td>2088000</td>
</tr>
<tr>
<td>7</td>
<td>Skin</td>
<td>55-64</td>
<td>70</td>
<td>1548371</td>
</tr>
<tr>
<td>8</td>
<td>Skin</td>
<td>65-74</td>
<td>126</td>
<td>1447432</td>
</tr>
<tr>
<td>9</td>
<td>Skin</td>
<td>75-84</td>
<td>136</td>
<td>1087524</td>
</tr>
<tr>
<td>10</td>
<td>Skin</td>
<td>85+</td>
<td>73</td>
<td>335944</td>
</tr>
<tr>
<td>11</td>
<td>Stomach</td>
<td>00-04</td>
<td>0</td>
<td>953785</td>
</tr>
<tr>
<td>12</td>
<td>Stomach</td>
<td>05-14</td>
<td>0</td>
<td>1997935</td>
</tr>
<tr>
<td>13</td>
<td>Stomach</td>
<td>15-24</td>
<td>0</td>
<td>1885014</td>
</tr>
<tr>
<td>14</td>
<td>Stomach</td>
<td>25-34</td>
<td>1</td>
<td>1957573</td>
</tr>
<tr>
<td>15</td>
<td>Stomach</td>
<td>35-44</td>
<td>19</td>
<td>2356649</td>
</tr>
<tr>
<td>16</td>
<td>Stomach</td>
<td>45-54</td>
<td>64</td>
<td>2088000</td>
</tr>
<tr>
<td>17</td>
<td>Stomach</td>
<td>55-64</td>
<td>114</td>
<td>1548371</td>
</tr>
<tr>
<td>18</td>
<td>Stomach</td>
<td>65-74</td>
<td>201</td>
<td>1447432</td>
</tr>
<tr>
<td>19</td>
<td>Stomach</td>
<td>75-84</td>
<td>294</td>
<td>1087524</td>
</tr>
<tr>
<td>20</td>
<td>Stomach</td>
<td>85+</td>
<td>136</td>
<td>335944</td>
</tr>
</tbody>
</table>

The US_Cs data set contains the corresponding stratum-specific mortality information for the United States (Miniño et al. 2002; US Bureau of the Census 2011). The variable Age is the grouping variable that forms the strata in the standardization. The variables Event_C16, Event_C43, and PYear identify the number of events for stomach cancer, the number of events for skin cancer, and the person-years, respectively.

data US_Cs;
  input Age $1-5 Event_C16 Event_C43 PYear:comma10.;
datalines;
00-04  0  0  19,175,798
05-14  1  1  41,077,577
15-24  14 41 39,183,891
25-34 124 186 39,892,024
35-44 484 626 45,148,527
45-54 1097 1199 37,677,952
55-64 1804 1303 24,274,684
65-74 3054 1637 18,390,986
75-84 3833 1624 12,361,180
85+ 2234 803 4,239,587
;

The following statements construct and list the mortality information by cancer cause:

data US_Cs;
  set US_Cs;
  Cause='Stomach'; Event=Event_C16; output;
  Cause='Skin'; Event=Event_C43; output;
  drop Event_C16 Event_C43;
run;
Example 109.4: Displaying SMR Results from BY Groups

```
proc sort data=US_Cs;
  by Cause;
run;

proc print data=US_Cs;
  var Cause Age Event PYear;
run;
```

Output 109.4.2 lists the mortality information by cancer cause.

### Output 109.4.2 Florida Data

<table>
<thead>
<tr>
<th>Obs</th>
<th>Cause</th>
<th>Age</th>
<th>Event</th>
<th>PYear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Skin</td>
<td>00-04</td>
<td>0</td>
<td>19175798</td>
</tr>
<tr>
<td>2</td>
<td>Skin</td>
<td>05-14</td>
<td>1</td>
<td>41077577</td>
</tr>
<tr>
<td>3</td>
<td>Skin</td>
<td>15-24</td>
<td>41</td>
<td>39183891</td>
</tr>
<tr>
<td>4</td>
<td>Skin</td>
<td>25-34</td>
<td>186</td>
<td>39892024</td>
</tr>
<tr>
<td>5</td>
<td>Skin</td>
<td>35-44</td>
<td>626</td>
<td>45148527</td>
</tr>
<tr>
<td>6</td>
<td>Skin</td>
<td>45-54</td>
<td>1199</td>
<td>37677952</td>
</tr>
<tr>
<td>7</td>
<td>Stomach</td>
<td>55-64</td>
<td>1303</td>
<td>24274684</td>
</tr>
<tr>
<td>8</td>
<td>Stomach</td>
<td>65-74</td>
<td>1637</td>
<td>18390986</td>
</tr>
<tr>
<td>9</td>
<td>Stomach</td>
<td>75-84</td>
<td>1624</td>
<td>12361180</td>
</tr>
<tr>
<td>10</td>
<td>Stomach</td>
<td>85+</td>
<td>803</td>
<td>4239587</td>
</tr>
<tr>
<td>11</td>
<td>Stomach</td>
<td>00-04</td>
<td>0</td>
<td>19175798</td>
</tr>
<tr>
<td>12</td>
<td>Stomach</td>
<td>05-14</td>
<td>1</td>
<td>41077577</td>
</tr>
<tr>
<td>13</td>
<td>Stomach</td>
<td>15-24</td>
<td>14</td>
<td>39183891</td>
</tr>
<tr>
<td>14</td>
<td>Stomach</td>
<td>25-34</td>
<td>124</td>
<td>39892024</td>
</tr>
<tr>
<td>15</td>
<td>Stomach</td>
<td>35-44</td>
<td>484</td>
<td>45148527</td>
</tr>
<tr>
<td>16</td>
<td>Stomach</td>
<td>45-54</td>
<td>1097</td>
<td>37677952</td>
</tr>
<tr>
<td>17</td>
<td>Stomach</td>
<td>55-64</td>
<td>1804</td>
<td>24274684</td>
</tr>
<tr>
<td>18</td>
<td>Stomach</td>
<td>65-74</td>
<td>3054</td>
<td>18390986</td>
</tr>
<tr>
<td>19</td>
<td>Stomach</td>
<td>75-84</td>
<td>3833</td>
<td>12361180</td>
</tr>
<tr>
<td>20</td>
<td>Stomach</td>
<td>85+</td>
<td>2234</td>
<td>4239587</td>
</tr>
</tbody>
</table>

The following statements invoke the STDRATE procedure and request indirect standardization to compute the skin and stomach SMR estimates for the state of Florida. The BY statement requests separate analyses of causes that are defined by the Cause variable.

```
ods graphics on;
ods select StdInfo StrataSmrPlot Smr;
proc stdrate data=Florida_Cs refdata=US_Cs
  stat=rate
  method=indirect
  plots=smr
;
  population event=Event total=PYear;
  reference event=Event total=PYear;
  strata Age;
  by Cause;
  ods output smr=Smr_Cs;
run;
```

Only the tables and plots that are specified in the ODS SELECT statement are displayed.
The STDINFO option in the ODS SELECT statement requests that the “Standardization Information” table display the standardization information for the first BY group, skin cancer, as shown in Output 109.4.3

**Output 109.4.3** Standardization Information

The **STDRATE** Procedure

Causes=Skin

<table>
<thead>
<tr>
<th>Standardization Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Reference Data Set</td>
</tr>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Statistic</td>
</tr>
<tr>
<td>Number of Strata</td>
</tr>
<tr>
<td>Rate Multiplier</td>
</tr>
</tbody>
</table>

The STRATASMRPLOT option in the ODS SELECT statement requests that the strata SMR plot display stratum-specific SMR estimates for skin cancer with confidence limits, as shown in Output 109.4.4.

**Output 109.4.4** Strata SMR Plot

Strata SMR Estimates with 95% Normal Confidence Limits
The SMR option in the ODS SELECT statement requests that the “Standardized Morbidity/Mortality Ratio” table display the SMR, its confidence limits, and the test for the null hypothesis $H_0 : SMR = 1$ for skin cancer, as shown in Output 109.4.5. With the default ALPHA=0.05, 95% confidence limits are constructed.

**Output 109.4.5 Standardized Morbidity/Mortality Ratio**

<table>
<thead>
<tr>
<th>Cause=Skin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardized Morbidity/Mortality Ratio</td>
</tr>
<tr>
<td>Observed Events</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>538</td>
</tr>
</tbody>
</table>

Similarly, the “Standardization Information” table in Output 109.4.6 displays the standardization information for the second BY group, stomach cancer.

**Output 109.4.6 Standardization Information**

**The STDRATE Procedure**

<table>
<thead>
<tr>
<th>Cause=Stomach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardization Information</td>
</tr>
<tr>
<td>Data Set</td>
</tr>
<tr>
<td>Reference Data Set</td>
</tr>
<tr>
<td>Method</td>
</tr>
<tr>
<td>Statistic</td>
</tr>
<tr>
<td>Number of Strata</td>
</tr>
<tr>
<td>Rate Multiplier</td>
</tr>
</tbody>
</table>
The “Strata SMR Plot” displays stratum-specific SMR estimates with confidence limits for stomach cancer, as shown in Output 109.4.7.

**Output 109.4.7**  Strata SMR Plot

The “Standardized Morbidity/Mortality Ratio” table displays the SMR, its confidence limits, and the test for the null hypothesis $H_0 : \text{SMR} = 1$ for stomach cancer, as shown in Output 109.4.8.

**Output 109.4.8** Standardized Morbidity/Mortality Ratio

<table>
<thead>
<tr>
<th>Cause=Stomach</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standardized Morbidity/Mortality Ratio</td>
</tr>
<tr>
<td>95% Normal Confidence Limits</td>
</tr>
<tr>
<td>Z Pr &gt;</td>
</tr>
<tr>
<td>Observed Events</td>
</tr>
<tr>
<td>829</td>
</tr>
</tbody>
</table>
The ODS OUTPUT SMR=SMR_CS statement requests that the “Standardized Morbidity/Mortality Ratio” tables for the two cancer causes be saved in the data set Smr_Cs. The following statements display the selected output variables for the data set:

``` Sas
proc print data=Smr_Cs;
 var Cause ObservedEvents ExpectedEvents Smr SmrLcl SmrUcl;
 run;
```

### Output 109.4.9 SMR Results from BY Groups

<table>
<thead>
<tr>
<th>Obs</th>
<th>Cause</th>
<th>ObservedEvents</th>
<th>ExpectedEvents</th>
<th>Smr</th>
<th>SmrLcl</th>
<th>SmrUcl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Skin</td>
<td>538</td>
<td>528.726</td>
<td>1.0175</td>
<td>0.9316</td>
<td>1.1035</td>
</tr>
<tr>
<td>2</td>
<td>Stomach</td>
<td>829</td>
<td>962.537</td>
<td>0.8613</td>
<td>0.8026</td>
<td>0.9199</td>
</tr>
</tbody>
</table>

The table in Output 109.4.9 shows that the study population (state of Florida) has a higher skin cancer rate and a lower stomach cancer rate than the reference population (United States), but only the lower stomach cancer rate is significant because its corresponding SMR upper confidence limit (0.9199) is less than 1.

The following statements display the standardized morbidity/mortality ratios for the two causes in a plot:

``` Sas
proc sgplot data=Smr_Cs;
 scatter y=Cause x=Smr / group=Cause;
 highlow y=Cause high=SmrUcl low=SmrLcl / highcap=serif lowcap=serif;
 yaxis type=discrete;
 xaxis label="SMR";
 refline 1 / axis=x transparency=0.5; 
 run;
```
Alternatively, you can also use the following statements to obtain separate analyses for the two cancer causes, and then to display these standardized mortality ratios together in a table and in a plot:

```sas
/*----- Perform Separate Analyses for Different Causes ------*/
ods graphics on;
ods select StdInfo StrataSmrPlot Smr;
proc stdrate data=Florida_Cs refdata=US_Cs
  stat=rate
  method=indirect
  plots=smr
;
  population event=Event_C43 total=PYear;
  reference event=Event_C43 total=PYear;
  strata Age;
ods output smr=Smr_c43;
run;
```
ods select StdInfo StrataSmrPlot Smr;
proc stdrate data=Florida_Cs refdata=US_Cs
   stat=rate
   method=indirect
   plots=smr
;
   population event=Event_C16 total=PYear;
   reference event=Event_C16 total=PYear;
   strata Age;
ods output smr=Smr_c16;
run;

/**------------ Combine SMRs -------------*/
data Smr_C43;
   set Smr_C43;
   length Cause $ 7.;
   Cause='Skin';
run;

data Smr_C16;
   set Smr_C16;
   length Cause $ 7.;
   Cause='Stomach';
run;

data Smr Cs;
   set Smr_C43 Smr_C16;
run;

/**------- Display the Cause-Specific SMRs --------*/
proc print data=Smr Cs;
   var Cause ObservedEvents ExpectedEvents Smr SmrLcl SmrUcl;
run;

proc sgplot data=Smr Cs;
   scatter y=Cause x=Smr / group=Cause;
   highlow y=Cause high=SmrUcl low=SmrLcl / highcap=serif lowcap=serif;
   yaxis type=discrete;
   xaxis label="SMR";
   reline 1 / axis=x transparency=0.5;
run;

References


Subject Index

AF
  STDRATE procedure, 8884
attributable fraction
  STDRATE procedure, 8884
attributable fraction estimates
  STDRATE procedure, 8889

confidence limits
  STDRATE procedure, 8864
directly standardized rate
  STDRATE procedure, 8876
distribution plot
  STDRATE procedure, 8894
effect estimates
  STDRATE procedure, 8899
effect measure plot
  STDRATE procedure, 8894
indirect standardization
  STDRATE procedure, 8881
Mantel-Haenszel effect estimation
  STDRATE procedure, 8879
method specifications
  STDRATE procedure, 8887
population attributable fraction
  STDRATE procedure, 8884
rate
  STDRATE procedure, 8872
risk
  STDRATE procedure, 8874
significance level
  STDRATE procedure, 8864
SMR
  STDRATE procedure, 8881
SMR plot
  STDRATE procedure, 8895
standardization information
  STDRATE procedure, 8890
standardized morbidity ratio
  STDRATE procedure, 8890
standardized mortality ratio
  STDRATE procedure, 8890
standardized rate
  STDRATE procedure, 8853
standardized rate estimates
  STDRATE procedure, 8890
standardized risk estimates
  STDRATE procedure, 8891
STDRATE procedure
  AF, 8884
  attributable fraction, 8884
  attributable fraction estimates, 8889
  confidence limits, 8864
  directly standardized rate, 8876
  distribution plot, 8894
  effect estimates, 8889
  effect measure plot, 8894
  indirect standardization, 8881
  introductory example, 8855
  Mantel-Haenszel effect estimation, 8879
  method specifications, 8887
  ODS Graphics names, 8895
  ODS table names, 8894
  population attributable fraction, 8884
  rate, 8872
  risk, 8874
  significance level, 8864
  SMR, 8881
  SMR plot, 8895
  standardization information, 8890
  standardized morbidity ratio, 8890
  standardized mortality ratio, 8890
  standardized rate estimates, 8890
  standardized risk estimates, 8891
  strata effect estimates, 8891
  strata information, 8892
  strata rate plot, 8895
  strata risk plot, 8895
  strata SMR estimates, 8893
  syntax, 8863
  table output, 8889
  strata effect estimates
    STDRATE procedure, 8891
    strata information
      STDRATE procedure, 8892
    strata rate plot
      STDRATE procedure, 8895
    strata risk plot
      STDRATE procedure, 8895
    strata SMR estimates
      STDRATE procedure, 8893
Syntax Index

ALPHA= option
    PROC STDRATE statement, 8864

BY statement
    STDRATE procedure, 8867

CL= option
    PROC STDRATE statement, 8864

DATA option
    PROC STDRATE statement, 8865

EFFECT option
    PROC STDRATE statement, 8865
    STRATA statement (STDRATE), 8870

EVENT= option
    POPULATION statement (STDRATE), 8868
    REFERENCE statement (STDRATE), 8869

EXPOSED= option
    POPULATION statement (STDRATE), 8868

GROUP= option
    POPULATION statement (STDRATE), 8868

METHOD= option
    PROC STDRATE statement, 8865

MISSING option
    STRATA statement (STDRATE), 8870

MULT= option
    POPULATION statement (STDRATE), 8868
    PROC STDRATE statement, 8867
    REFERENCE statement (STDRATE), 8869

ORDER= option
    POPULATION statement (STDRATE), 8868
    STRATA statement (STDRATE), 8870

PLOTS option
    PROC STDRATE statement, 8866

POPEVENT= option
    POPULATION statement (STDRATE), 8868

POPULATION statement
    STDRATE procedure, 8868
    PROC STDRATE statement, see STDRATE procedure

RATE= option
    POPULATION statement (STDRATE), 8868
    REFERENCE statement (STDRATE), 8869

REFDATA option
    PROC STDRATE statement, 8867

REFERENCE statement
    STDRATE procedure, 8869

REFEVENT= option
    REFERENCE statement (STDRATE), 8869

REFRATE= option
    REFERENCE statement (STDRATE), 8869

RISK= option
    POPULATION statement (STDRATE), 8868
    REFERENCE statement (STDRATE), 8869

SMR option
    STRATA statement (STDRATE), 8871

STAT= option
    PROC STDRATE statement, 8867

STATS option
    STRATA statement (STDRATE), 8871

STDRATE procedure, BY statement, 8867

STDRATE procedure, POPULATION statement, 8868
    EVENT= option, 8868
    EXPOSED= option, 8868
    GROUP= option, 8868
    ORDER= option, 8868
    POPEVENT= option, 8868
    RATE= option, 8868
    RISK= option, 8868
    TOTAL= option, 8869

STDRATE procedure, PROC STDRATE statement, 8864
    ALPHA= option, 8864
    CL option, 8864
    DATA= option, 8865
    EFFECT option, 8865
    MULT= option, 8865
    METHOD= option, 8865
    PLOTS option, 8866
    REFDATA= option, 8867
    STAT= option, 8867

STDRATE procedure, REFERENCE statement, 8869
    EVENT= option, 8869
    MULT= option, 8869
    RATE= option, 8869
    REFEVENT= option, 8869
    REFRATE= option, 8869
    REFRISK= option, 8869
    REFTOTAL= option, 8870
    RISK= option, 8869
    TOTAL= option, 8869
STDRATE procedure, STRATA statement, 8870
  EFFECT option, 8870
  INFO option, 8871
  MISSING option, 8870
  ORDER= option, 8870
  SMR option, 8871

STRATA statement
  STDRATE procedure, 8870

TOTAL= option
  POPULATION statement (STDRATE), 8869
  REFERENCE statement (STDRATE), 8869, 8870