Title

> **dstdize** — Direct and indirect standardization

Description	Quick start	Menu	Syntax
Options for dstdize	Options for istdize	Remarks and examples	Stored results
Methods and formulas	Acknowledgments	References	Also see

Description

dstdize produces standardized rates, a weighted average of the stratum-specific rates.

istdize produces indirectly standardized rates that are appropriate when the stratum-specific rates for the population being studied are either unavailable or unreliable.

istdize also calculates a point estimate and exact confidence interval for the study population's standardized mortality ratio (SMR) or the standardized incidence ratio (SIR).

Quick start

Direct standardization

Use reference population saved in mypop.dta to standardize v1 with stratum identifier svar and stratum size v2 for catvar

    dstdize v1 v2 svar, by(catvar) using(mypop)

As above, but with reference population in memory where catvar = 1

    dstdize v1 v2 svar, by(catvar) base(1)

As above, but with reference population in memory where catvar = "nation"

    dstdize v1 v2 svar, by(catvar) base("nation")

Indirect standardization

Use population cases and size saved in cases and pop to standardize study cases v3 and stratum size v4 at each level of svar

    istdize v3 v4 svar using mypop.dta, popvars(cases pop)

As above, but standardize subpopulations identified by levels of catvar

    isdize v3 v4 svar using mypop.dta, popvars(cases pop) by(catvar)

As above, but standardize by population stratum-specific and crude rates saved in srate and crate and display summary of standard population

    isdize v3 v4 svar using mypop.dta, rate(srate crate) by(catvar) print

As above, but indicate that the crude population rate is 0.01

    isdize v3 v4 svar using mypop.dta, rate(srate .01) by(catvar) print

Menu

dstdize

Statistics > Epidemiology and related > Other > Direct standardization

istdize

Statistics > Epidemiology and related > Other > Indirect standardization

Syntax

Direct standardization

> dstdize *charvar popvar stratavars* $\big[\,$*if*$\,\big]$ $\big[\,$*in*$\,\big]$, by(*groupvars*) $\big[\,$*dstdize_options*$\,\big]$

Indirect standardization

> istdize *casevar$_s$ popvar$_s$ stratavars* $\big[\,$*if*$\,\big]$ $\big[\,$*in*$\,\big]$ using *filename*,
>
> $\big\{\,$popvars(*casevar$_p$ popvar$_p$*) $|$ rate(*ratevar$_p$* $\big\{$ # $|$ *crudevar$_p$* $\big\}$) $\big\}$
>
> $\big[\,$*istdize_options*$\,\big]$

charvar is the characteristic to be standardized across different subpopulations identified by *groupvars*.

popvar defines the weights used in standardization.

stratavars defines the strata across which the weights are to be averaged in dstdize. For istdize, *stratavars* defines the strata for which *casevar$_s$* is measured.

casevar$_s$ is the variable name for the study population's number of cases. If by(*groupvars*) is specified, *casevar$_s$* must be constant or missing within each group defined by combinations of *groupvars*.

popvar$_s$ identifies the number of subjects in each strata in the study population.

filename must be a Stata dataset and contain *popvar* and *stratavars*.

dstdize options	Description
Main	
* <u>by</u>(*groupvars*)	study populations
<u>us</u>ing(*filename*)	use standard population from Stata dataset
base(*#* │ *string*)	use standard population from a value of grouping variable
level(*#*)	set confidence level; default is level(95)
Options	
saving(*filename*)	save computed standard population distribution as a Stata dataset
<u>f</u>ormat(%*fmt*)	final summary table display format; default is %10.0g
print	include table summary of standard population in output
nores	suppress storing results in r()

* <u>by</u>(*groupvars*) is required.

istdize options	Description
Main	
* <u>pop</u>vars(*casevar$_p$* *popvar$_p$*)	for standard population, *casevar$_p$* is number of cases and *popvar$_p$* is number of individuals
* <u>rate</u>(*ratevar$_p$* {*#* │ *crudevar$_p$*})	*ratevar$_p$* is stratum-specific rates and *#* or *crudevar$_p$* is the crude case rate value or variable
level(*#*)	set confidence level; default is level(95)
Options	
by(*groupvars*)	variables identifying study populations
<u>f</u>ormat(%*fmt*)	final summary table display format; default is %10.0g
print	include table summary of standard population in output

* Either <u>pop</u>vars(*casevar$_p$* *popvar$_p$*) or <u>rate</u>(*ratevar$_p$* {*#* │ *crudevar$_p$*}) must be specified.

Options for dstdize

Main

by(*groupvars*) is required for the dstdize command; it specifies the variables identifying the study populations. If base() is also specified, there must be only one variable in the by() group. If you do not have a variable for this option, you can generate one by using something like generate newvar=1 and then use newvar as the argument to this option.

using(*filename*) or base(*#* │ *string*) may be used to specify the standard population. You may not specify both options. using(*filename*) supplies the name of a .dta file containing the standard population. The standard population must contain the *popvar* and the *stratavars*. If using() is not specified, the standard population distribution will be obtained from the data. base(*#* │ *string*) lets you specify one of the values of *groupvar*—either a numeric value or a string—to be used as the standard population. If neither base() nor using() is specified, the entire dataset is used to determine an estimate of the standard population.

level(*#*) specifies the confidence level, as a percentage, for a confidence interval of the adjusted rate. The default is level(95) or as set by set level; see [U] **20.7 Specifying the width of confidence intervals**.

```
                    Options
```

saving(*filename*) saves the computed standard population distribution as a Stata dataset that can be
  used in further analyses.

format(%*fmt*) specifies the format in which to display the final summary table. The default is
  %10.0g.

print includes a table summary of the standard population before displaying the study population
  results.

nores suppresses storing results in r(). This option is seldom specified. Some results are stored
  in matrices. If there are more groups than matsize, dstdize will report "matsize too small".
  Then you can either increase matsize or specify nores. The nores option does not change how
  results are calculated but specifies that results need not be left behind for use by other programs.

Options for istdize

```
                    Main
```

popvars(*casevar$_p$* *popvar$_p$*) or rate(*ratevar$_p$* *#* | *ratevar$_p$* *crudevar$_p$*) must be specified with ist-
  dize. Only one of these two options is allowed. These options are used to describe the standard
  population's data.

  With popvars(*casevar$_p$* *popvar$_p$*), *casevar$_p$* records the number of cases (deaths) for each stratum
  in the standard population, and *popvar$_p$* records the total number of individuals in each stratum
  (individuals at risk).

  With rate(*ratevar$_p$* $\{$*#* | *crudevar$_p$*$\}$), *ratevar$_p$* contains the stratum-specific rates. *#* | *crudevar$_p$*
  specifies the crude case rate either by a variable name or by the crude case rate value. If a crude
  rate variable is used, it must be the same for all observations, although it could be missing for
  some.

level(*#*) specifies the confidence level, as a percentage, for a confidence interval of the adjusted
  rate. The default is level(95) or as set by set level; see [U] **20.7 Specifying the width of
  confidence intervals**.

```
                    Options
```

by(*groupvars*) specifies variables identifying study populations when more than one exists in the
  data. If this option is not specified, the entire study population is treated as one group.

format(%*fmt*) specifies the format in which to display the final summary table. The default is
  %10.0g.

print outputs a table summary of the standard population before displaying the study population
  results.

Remarks and examples   stata.com

Remarks are presented under the following headings:

> *Direct standardization*
> *Indirect standardization*

In epidemiology and other fields, you will often need to compare rates for some characteristic
across different populations. These populations often differ on factors associated with the characteristic
under study; thus directly comparing overall rates may be misleading.

See van Belle et al. (2004, 642–684), Fleiss, Levin, and Paik (2003, chap. 19), or Kirkwood and Sterne (2003, chap. 25) for a discussion of direct and indirect standardization.

Direct standardization

The direct method of adjusting for differences among populations involves computing the overall rates that would result if, instead of having different distributions, all populations had the same standard distribution. The standardized rate is defined as a weighted average of the stratum-specific rates, with the weights taken from the standard distribution. Direct standardization may be applied only when the specific rates for a given population are available.

dstdize generates adjusted summary measures of occurrence, which can be used to compare prevalence, incidence, or mortality rates between populations that may differ on certain characteristics (for example, age, gender, race). These underlying differences may affect the crude prevalence, mortality, or incidence rates.

▷ Example 1

We have data (Rothman 1986, 42) on mortality rates for Sweden and Panama for 1962, and we wish to compare mortality in these two countries:

```
. use http://www.stata-press.com/data/r14/mortality
(1962 Mortality, Sweden & Panama)
. describe
Contains data from http://www.stata-press.com/data/r14/mortality.dta
  obs:            6                         1962 Mortality, Sweden & Panama
 vars:            4                         14 Apr 2014 16:18
 size:           90
```

variable name	storage type	display format	value label	variable label
nation	str6	%9s		Nation
age_category	byte	%9.0g	age_lbl	Age Category
population	float	%10.0gc		Population in Age Category
deaths	float	%9.0gc		Deaths in Age Category

```
Sorted by:
. list, sepby(nation) abbrev(12) divider
```

	nation	age_category	population	deaths
1.	Sweden	0 - 29	3145000	3,523
2.	Sweden	30 - 59	3057000	10,928
3.	Sweden	60+	1294000	59,104
4.	Panama	0 - 29	741,000	3,904
5.	Panama	30 - 59	275,000	1,421
6.	Panama	60+	59,000	2,456

We divide the total number of cases in the population by the population to obtain the *crude rate*:

```
. collapse (sum) pop deaths, by(nation)
. list, abbrev(10) divider
```

	nation	population	deaths
1.	Panama	1075000	7,781
2.	Sweden	7496000	73,555

```
. generate crude = deaths/pop
. list, abbrev(10) divider
```

	nation	population	deaths	crude
1.	Panama	1075000	7,781	.0072381
2.	Sweden	7496000	73,555	.0098126

If we examine the total number of deaths in the two nations, the total crude mortality rate in Sweden is higher than that in Panama. From the original data, we see one possible explanation: Swedes are older than Panamanians, making direct comparison of the mortality rates difficult.

Direct standardization lets us remove the distortion caused by the different age distributions. The adjusted rate is defined as the weighted sum of the crude rates, where the weights are given by the standard distribution. Suppose that we wish to standardize these mortality rates to the following age distribution:

```
. use http://www.stata-press.com/data/r14/1962, clear
(Standard Population Distribution)
. list, abbrev(12) divider
```

	age_category	population
1.	0 - 29	.35
2.	30 - 59	.35
3.	60+	.3

```
. save 1962
file 1962.dta saved
```

If we multiply the above weights for the age strata by the crude rate for the corresponding age category, the sum gives us the standardized rate.

```
. use http://www.stata-press.com/data/r14/mortality
(1962 Mortality, Sweden & Panama)
. generate crude=deaths/pop
. drop pop
. merge m:1 age_cat using 1962
(note: variable age_category was byte, now float to accommodate using data's
       values)
```

Result	# of obs.	
not matched	0	
matched	6	(_merge==3)

```
. list, sepby(age_category) abbrev(12)
```

	nation	age_category	deaths	crude	population	_merge
1.	Sweden	0 - 29	3,523	.0011202	.35	matched (3)
2.	Panama	0 - 29	3,904	.0052686	.35	matched (3)
3.	Panama	30 - 59	1,421	.0051673	.35	matched (3)
4.	Sweden	30 - 59	10,928	.0035747	.35	matched (3)
5.	Sweden	60+	59,104	.0456754	.3	matched (3)
6.	Panama	60+	2,456	.0416271	.3	matched (3)

```
. generate product = crude*pop
. by nation, sort: egen adj_rate = sum(product)
. drop _merge
. list, sepby(nation)
```

	nation	age_ca~y	deaths	crude	popula~n	product	adj_rate
1.	Panama	30 - 59	1,421	.0051673	.35	.0018085	.0161407
2.	Panama	0 - 29	3,904	.0052686	.35	.001844	.0161407
3.	Panama	60+	2,456	.0416271	.3	.0124881	.0161407
4.	Sweden	30 - 59	10,928	.0035747	.35	.0012512	.0153459
5.	Sweden	60+	59,104	.0456754	.3	.0137026	.0153459
6.	Sweden	0 - 29	3,523	.0011202	.35	.0003921	.0153459

Comparing the standardized rates indicates that the Swedes have a slightly lower mortality rate.

To perform the above analysis with dstdize, type

```
. use http://www.stata-press.com/data/r14/mortality, clear
(1962 Mortality, Sweden & Panama)
. dstdize deaths pop age_cat, by(nation) using(1962)
```

-> nation= Panama

Stratum	Pop.	Cases	Unadjusted Pop. Dist.	Stratum Rate[s]	Std. Pop. Dst[P]	s*P
0 - 29	741000	3904	0.689	0.0053	0.350	0.0018
30 - 59	275000	1421	0.256	0.0052	0.350	0.0018
60+	59000	2456	0.055	0.0416	0.300	0.0125

Totals:	1075000	7781	Adjusted Cases:	17351.2
			Crude Rate:	0.0072
			Adjusted Rate:	0.0161
		95% Conf. Interval:	[0.0156, 0.0166]	

-> nation= Sweden

Stratum	Pop.	Cases	Unadjusted Pop. Dist.	Stratum Rate[s]	Std. Pop. Dst[P]	s*P
0 - 29	3145000	3523	0.420	0.0011	0.350	0.0004
30 - 59	3057000	10928	0.408	0.0036	0.350	0.0013
60+	1294000	59104	0.173	0.0457	0.300	0.0137

Totals:	7496000	73555	Adjusted Cases:	115032.5
			Crude Rate:	0.0098
			Adjusted Rate:	0.0153
		95% Conf. Interval:	[0.0152, 0.0155]	

Summary of Study Populations:

nation	N	Crude	Adj_Rate	Confidence Interval
Panama	1075000	0.007238	0.016141	[0.015645, 0.016637]
Sweden	7496000	0.009813	0.015346	[0.015235, 0.015457]

The summary table above lets us make a quick inspection of the results within the study populations, and the detail tables give the behavior among the strata within the study populations.

◁

▷ Example 2

We have individual-level data on persons in four cities over several years. Included in the data is a variable indicating whether the person has high blood pressure, together with information on the person's age, sex, and race. We wish to obtain standardized high blood pressure rates for each city for 1990 and 1992, using, as the standard, the age, sex, and race distribution of the four cities and two years combined.

Our dataset contains

```
. use http://www.stata-press.com/data/r14/hbp

. describe
Contains data from http://www.stata-press.com/data/r14/hbp.dta
  obs:          1,130
  vars:             7                          21 Feb 2014 06:42
  size:        19,210
```

variable name	storage type	display format	value label	variable label
id	str10	%10s		Record identification number
city	byte	%8.0g		
year	int	%8.0g		
sex	byte	%8.0g	sexfmt	
age_group	byte	%8.0g	agefmt	
race	byte	%8.0g	racefmt	
hbp	byte	%8.0g	yn	high blood pressure

```
Sorted by:
```

The dstdize command is designed to work with aggregate data but will work with individual-level data only if we create a variable recording the population represented by each observation. For individual-level data, this is one:

```
. generate pop = 1
```

On the next page, we specify print to obtain a listing of the standard population and level(90) to request 90% rather than 95% confidence intervals. Typing if year==1990 | year==1992 restricts the data to the two years for both summary tables and the standard population.

```
. dstdize hbp pop age race sex if year==1990 | year==1992, by(city year) print
> level(90)
```

	Stratum	Pop.	Dist.
	—————Standard Population—————		
15 - 19	Black Female	35	0.077
15 - 19	Black Male	44	0.097
15 - 19	Hispanic Female	5	0.011
15 - 19	Hispanic Male	10	0.022
15 - 19	White Female	7	0.015
15 - 19	White Male	5	0.011
20 - 24	Black Female	43	0.095
20 - 24	Black Male	67	0.147
20 - 24	Hispanic Female	14	0.031
20 - 24	Hispanic Male	13	0.029
20 - 24	White Female	4	0.009
20 - 24	White Male	21	0.046
25 - 29	Black Female	17	0.037
25 - 29	Black Male	44	0.097
25 - 29	Hispanic Female	7	0.015
25 - 29	Hispanic Male	13	0.029
25 - 29	White Female	9	0.020
25 - 29	White Male	16	0.035
30 - 34	Black Female	16	0.035
30 - 34	Black Male	32	0.070
30 - 34	Hispanic Female	2	0.004
30 - 34	Hispanic Male	3	0.007
30 - 34	White Female	5	0.011
30 - 34	White Male	23	0.051

```
Total:                        455
(6 observations excluded because of missing values)
```

```
-> city year= 1 1990
```

	Stratum	Pop.	Cases	Unadjusted Pop. Dist.	Stratum Rate[s]	Std. Pop. Dst[P]	s*P
15 - 19	Black Female	6	2	0.128	0.3333	0.077	0.0256
15 - 19	Black Male	6	0	0.128	0.0000	0.097	0.0000
15 - 19	Hispanic Male	1	0	0.021	0.0000	0.022	0.0000
20 - 24	Black Female	3	0	0.064	0.0000	0.095	0.0000
20 - 24	Black Male	11	0	0.234	0.0000	0.147	0.0000
25 - 29	Black Female	4	0	0.085	0.0000	0.037	0.0000
25 - 29	Black Male	6	1	0.128	0.1667	0.097	0.0161
25 - 29	Hispanic Female	2	0	0.043	0.0000	0.015	0.0000
25 - 29	White Female	1	0	0.021	0.0000	0.020	0.0000
30 - 34	Black Female	1	0	0.021	0.0000	0.035	0.0000
30 - 34	Black Male	6	0	0.128	0.0000	0.070	0.0000

```
Totals:                       47      3    Adjusted Cases:      2.0
                                               Crude Rate:   0.0638
                                            Adjusted Rate:   0.0418
                           90% Conf. Interval: [0.0074, 0.0761]
```

```
-> city year= 1 1992
```

	Stratum		Pop.	Cases	Unadjusted Pop. Dist.	Stratum Rate[s]	Std. Pop. Dst[P]	s*P
15 - 19	Black	Female	3	0	0.054	0.0000	0.077	0.0000
15 - 19	Black	Male	9	0	0.161	0.0000	0.097	0.0000
15 - 19	Hispanic	Male	1	0	0.018	0.0000	0.022	0.0000
20 - 24	Black	Female	7	0	0.125	0.0000	0.095	0.0000
20 - 24	Black	Male	9	0	0.161	0.0000	0.147	0.0000
20 - 24	Hispanic	Female	1	0	0.018	0.0000	0.031	0.0000
25 - 29	Black	Female	2	0	0.036	0.0000	0.037	0.0000
25 - 29	Black	Male	11	1	0.196	0.0909	0.097	0.0088
25 - 29	Hispanic	Male	1	0	0.018	0.0000	0.029	0.0000
30 - 34	Black	Female	7	0	0.125	0.0000	0.035	0.0000
30 - 34	Black	Male	4	0	0.071	0.0000	0.070	0.0000
30 - 34	White	Female	1	0	0.018	0.0000	0.011	0.0000

```
Totals:                       56        1     Adjusted Cases:        0.5
                                                  Crude Rate:    0.0179
                                               Adjusted Rate:    0.0088
                              90% Conf. Interval: [0.0000, 0.0226]
```

```
-> city year= 2 1990
```

	Stratum		Pop.	Cases	Unadjusted Pop. Dist.	Stratum Rate[s]	Std. Pop. Dst[P]	s*P
15 - 19	Black	Female	5	0	0.078	0.0000	0.077	0.0000
15 - 19	Black	Male	7	1	0.109	0.1429	0.097	0.0138
15 - 19	Hispanic	Male	1	0	0.016	0.0000	0.022	0.0000
20 - 24	Black	Female	7	1	0.109	0.1429	0.095	0.0135
20 - 24	Black	Male	8	0	0.125	0.0000	0.147	0.0000
20 - 24	Hispanic	Female	5	0	0.078	0.0000	0.031	0.0000
20 - 24	Hispanic	Male	2	0	0.031	0.0000	0.029	0.0000
20 - 24	White	Male	2	0	0.031	0.0000	0.046	0.0000
25 - 29	Black	Female	3	0	0.047	0.0000	0.037	0.0000
25 - 29	Black	Male	9	0	0.141	0.0000	0.097	0.0000
25 - 29	Hispanic	Female	2	0	0.031	0.0000	0.015	0.0000
25 - 29	White	Female	1	0	0.016	0.0000	0.020	0.0000
25 - 29	White	Male	2	1	0.031	0.5000	0.035	0.0176
30 - 34	Black	Female	1	0	0.016	0.0000	0.035	0.0000
30 - 34	Black	Male	5	0	0.078	0.0000	0.070	0.0000
30 - 34	Hispanic	Female	2	0	0.031	0.0000	0.004	0.0000
30 - 34	White	Female	1	0	0.016	0.0000	0.011	0.0000
30 - 34	White	Male	1	0	0.016	0.0000	0.051	0.0000

```
Totals:                       64        3     Adjusted Cases:        2.9
                                                  Crude Rate:    0.0469
                                               Adjusted Rate:    0.0449
                              90% Conf. Interval: [0.0091, 0.0807]
```

```
-> city year= 2 1992
```

	Stratum	Pop.	Cases	Unadjusted Pop. Dist.	Stratum Rate[s]	Std. Pop. Dst[P]	s*P
15 - 19 Black	Female	1	0	0.015	0.0000	0.077	0.0000
15 - 19 Black	Male	5	0	0.075	0.0000	0.097	0.0000
15 - 19 Hispanic	Female	3	0	0.045	0.0000	0.011	0.0000
15 - 19 Hispanic	Male	1	0	0.015	0.0000	0.022	0.0000
15 - 19 White	Male	1	0	0.015	0.0000	0.011	0.0000
20 - 24 Black	Female	8	0	0.119	0.0000	0.095	0.0000
20 - 24 Black	Male	11	0	0.164	0.0000	0.147	0.0000
20 - 24 Hispanic	Female	6	0	0.090	0.0000	0.031	0.0000
20 - 24 Hispanic	Male	4	2	0.060	0.5000	0.029	0.0143
20 - 24 White	Female	1	0	0.015	0.0000	0.009	0.0000
20 - 24 White	Male	2	0	0.030	0.0000	0.046	0.0000
25 - 29 Black	Female	2	0	0.030	0.0000	0.037	0.0000
25 - 29 Black	Male	3	0	0.045	0.0000	0.097	0.0000
25 - 29 Hispanic	Female	2	0	0.030	0.0000	0.015	0.0000
25 - 29 Hispanic	Male	4	0	0.060	0.0000	0.029	0.0000
25 - 29 White	Female	4	0	0.060	0.0000	0.020	0.0000
25 - 29 White	Male	2	0	0.030	0.0000	0.035	0.0000
30 - 34 Black	Female	1	0	0.015	0.0000	0.035	0.0000
30 - 34 Black	Male	2	0	0.030	0.0000	0.070	0.0000
30 - 34 Hispanic	Male	1	0	0.015	0.0000	0.007	0.0000
30 - 34 White	Female	2	0	0.030	0.0000	0.011	0.0000
30 - 34 White	Male	1	0	0.015	0.0000	0.051	0.0000

```
Totals:                          67        2     Adjusted Cases:       1.0
                                                    Crude Rate:      0.0299
                                                 Adjusted Rate:      0.0143
                                 90% Conf. Interval: [0.0025, 0.0260]
```

```
-> city year= 3 1990
```

	Stratum	Pop.	Cases	Unadjusted Pop. Dist.	Stratum Rate[s]	Std. Pop. Dst[P]	s*P
15 - 19 Black	Female	3	0	0.043	0.0000	0.077	0.0000
15 - 19 Black	Male	1	0	0.014	0.0000	0.097	0.0000
15 - 19 Hispanic	Female	1	0	0.014	0.0000	0.011	0.0000
15 - 19 White	Female	3	0	0.043	0.0000	0.015	0.0000
15 - 19 White	Male	1	0	0.014	0.0000	0.011	0.0000
20 - 24 Black	Female	1	0	0.014	0.0000	0.095	0.0000
20 - 24 Black	Male	9	0	0.130	0.0000	0.147	0.0000
20 - 24 Hispanic	Male	3	0	0.043	0.0000	0.029	0.0000
20 - 24 White	Female	2	0	0.029	0.0000	0.009	0.0000
20 - 24 White	Male	8	1	0.116	0.1250	0.046	0.0058
25 - 29 Black	Female	1	0	0.014	0.0000	0.037	0.0000
25 - 29 Black	Male	8	3	0.116	0.3750	0.097	0.0363
25 - 29 Hispanic	Male	4	0	0.058	0.0000	0.029	0.0000
25 - 29 White	Female	1	0	0.014	0.0000	0.020	0.0000
25 - 29 White	Male	6	0	0.087	0.0000	0.035	0.0000
30 - 34 Black	Male	6	2	0.087	0.3333	0.070	0.0234
30 - 34 White	Male	11	5	0.159	0.4545	0.051	0.0230

```
Totals:                          69       11     Adjusted Cases:       6.1
                                                    Crude Rate:      0.1594
                                                 Adjusted Rate:      0.0885
                                 90% Conf. Interval: [0.0501, 0.1268]
```

```
-> city year= 3 1992
                                          ————Unadjusted————  Std.
                                             Pop.  Stratum  Pop.
                     Stratum      Pop.   Cases  Dist. Rate[s] Dst[P]   s*P
```

Stratum			Pop.	Cases	Pop. Dist.	Stratum Rate[s]	Std. Pop. Dst[P]	s*P
15 - 19	Black	Female	2	0	0.054	0.0000	0.077	0.0000
15 - 19	Hispanic	Male	3	0	0.081	0.0000	0.022	0.0000
15 - 19	White	Female	2	0	0.054	0.0000	0.015	0.0000
15 - 19	White	Male	1	0	0.027	0.0000	0.011	0.0000
20 - 24	Black	Male	3	0	0.081	0.0000	0.147	0.0000
20 - 24	Hispanic	Female	1	0	0.027	0.0000	0.031	0.0000
20 - 24	Hispanic	Male	3	0	0.081	0.0000	0.029	0.0000
20 - 24	White	Female	1	0	0.027	0.0000	0.009	0.0000
20 - 24	White	Male	6	1	0.162	0.1667	0.046	0.0077
25 - 29	Hispanic	Male	1	0	0.027	0.0000	0.029	0.0000
25 - 29	White	Male	5	1	0.135	0.2000	0.035	0.0070
30 - 34	Black	Male	1	0	0.027	0.0000	0.070	0.0000
30 - 34	White	Male	8	5	0.216	0.6250	0.051	0.0316

```
Totals:                             37       7    Adjusted Cases:     1.7
                                                      Crude Rate:   0.1892
                                                   Adjusted Rate:   0.0463
                                  90% Conf. Interval: [0.0253, 0.0674]
```

```
-> city year= 5 1990
                                          ————Unadjusted————  Std.
                                             Pop.  Stratum  Pop.
                     Stratum      Pop.   Cases  Dist. Rate[s] Dst[P]   s*P
```

Stratum			Pop.	Cases	Pop. Dist.	Stratum Rate[s]	Std. Pop. Dst[P]	s*P
15 - 19	Black	Female	9	0	0.196	0.0000	0.077	0.0000
15 - 19	Black	Male	7	0	0.152	0.0000	0.097	0.0000
15 - 19	Hispanic	Male	1	0	0.022	0.0000	0.022	0.0000
15 - 19	White	Male	1	0	0.022	0.0000	0.011	0.0000
20 - 24	Black	Female	4	0	0.087	0.0000	0.095	0.0000
20 - 24	Black	Male	6	0	0.130	0.0000	0.147	0.0000
20 - 24	Hispanic	Female	1	0	0.022	0.0000	0.031	0.0000
25 - 29	Black	Female	3	1	0.065	0.3333	0.037	0.0125
25 - 29	Black	Male	5	0	0.109	0.0000	0.097	0.0000
25 - 29	Hispanic	Female	1	0	0.022	0.0000	0.015	0.0000
25 - 29	White	Female	2	1	0.043	0.5000	0.020	0.0099
30 - 34	Black	Female	2	0	0.043	0.0000	0.035	0.0000
30 - 34	Black	Male	3	0	0.065	0.0000	0.070	0.0000
30 - 34	White	Male	1	0	0.022	0.0000	0.051	0.0000

```
Totals:                             46       2    Adjusted Cases:     1.0
                                                      Crude Rate:   0.0435
                                                   Adjusted Rate:   0.0223
                                  90% Conf. Interval: [0.0020, 0.0426]
```

```
-> city year= 5 1992
                                           ——Unadjusted——   Std.
                                            Pop.  Stratum   Pop.
                       Stratum     Pop.   Cases Dist. Rate[s] Dst[P]   s*P

  15 - 19     Black   Female        6        0  0.087 0.0000  0.077 0.0000
  15 - 19     Black     Male        9        0  0.130 0.0000  0.097 0.0000
  15 - 19 Hispanic   Female         1        0  0.014 0.0000  0.011 0.0000
  15 - 19 Hispanic     Male         2        0  0.029 0.0000  0.022 0.0000
  15 - 19     White   Female        2        0  0.029 0.0000  0.015 0.0000
  15 - 19     White     Male        1        0  0.014 0.0000  0.011 0.0000
  20 - 24     Black   Female       13        0  0.188 0.0000  0.095 0.0000
  20 - 24     Black     Male       10        0  0.145 0.0000  0.147 0.0000
  20 - 24 Hispanic     Male         1        0  0.014 0.0000  0.029 0.0000
  20 - 24     White     Male         3        0  0.043 0.0000  0.046 0.0000
  25 - 29     Black   Female        2        0  0.029 0.0000  0.037 0.0000
  25 - 29     Black     Male         2        0  0.029 0.0000  0.097 0.0000
  25 - 29 Hispanic     Male         3        0  0.043 0.0000  0.029 0.0000
  25 - 29     White     Male         1        0  0.014 0.0000  0.035 0.0000
  30 - 34     Black   Female        4        0  0.058 0.0000  0.035 0.0000
  30 - 34     Black     Male         5        0  0.072 0.0000  0.070 0.0000
  30 - 34 Hispanic     Male         2        0  0.029 0.0000  0.007 0.0000
  30 - 34     White   Female        1        0  0.014 0.0000  0.011 0.0000
  30 - 34     White     Male         1        1  0.014 1.0000  0.051 0.0505

Totals:                            69        1    Adjusted Cases:    3.5
                                                      Crude Rate:  0.0145
                                                   Adjusted Rate:  0.0505
                                    90% Conf. Interval: [0.0505, 0.0505]


Summary of Study Populations:
    city
    year          N     Crude     Adj_Rate      Confidence Interval

       1
    1990          47  0.063830    0.041758   [  0.007427,    0.076089]
       1
    1992          56  0.017857    0.008791   [  0.000000,    0.022579]
       2
    1990          64  0.046875    0.044898   [  0.009072,    0.080724]
       2
    1992          67  0.029851    0.014286   [  0.002537,    0.026035]
       3
    1990          69  0.159420    0.088453   [  0.050093,    0.126813]
       3
    1992          37  0.189189    0.046319   [  0.025271,    0.067366]
       5
    1990          46  0.043478    0.022344   [  0.002044,    0.042644]
       5
    1992          69  0.014493    0.050549   [  0.050549,    0.050549]
```

◁

Indirect standardization

Standardization of rates can be performed via the indirect method whenever the stratum-specific rates are either unknown or unreliable. If the stratum-specific rates are known, the direct standardization method is preferred.

To apply the indirect method, you must have the following information:

• The observed number of cases in each population to be standardized, O. For example, if death rates in two states are being standardized using the U.S. death rate for the same period, you must know the total number of deaths in each state.

• The distribution across the various strata for the population being studied, $n_1, \ldots, n_k$. If you are standardizing the death rate in the two states, adjusting for age, you must know the number of individuals in each of the k age groups.

• The stratum-specific rates for the standard population, $p_1, \ldots, p_k$. For example, you must have the U.S. death rate for each stratum (age group).

• The crude rate of the standard population, C. For example, you must have the U.S. mortality rate for the year.

The indirect adjusted rate is then

$$R_{\text{indirect}} = C\frac{O}{E}$$

where E is the expected number of cases (deaths) in each population. See *Methods and formulas* for a more detailed description of calculations.

▷ Example 3

This example is borrowed from Kahn and Sempos (1989, 95–105). We want to compare 1970 mortality rates in California and Maine, adjusting for age. Although we have age-specific population counts for the two states, we lack age-specific death rates. Direct standardization is not feasible here. We can use the U.S. population census data for the same year to produce indirectly standardized rates for these two states.

From the U.S. census, the standard population for this example was entered into Stata and saved in popkahn.dta.

```
. use http://www.stata-press.com/data/r14/popkahn, clear

. list age pop deaths rate, sep(4)
```

	age	population	deaths	rate
1.	<15	57,900,000	103,062	.00178
2.	15–24	35,441,000	45,261	.00128
3.	25–34	24,907,000	39,193	.00157
4.	35–44	23,088,000	72,617	.00315
5.	45–54	23,220,000	169,517	.0073
6.	55–64	18,590,000	308,373	.01659
7.	65–74	12,436,000	445,531	.03583
8.	75+	7,630,000	736,758	.09656

The standard population contains for each age stratum the total number of individuals (pop) and both the age-specific mortality rate (rate) and the number of deaths. The standard population need not contain all three. If we have only the age-specific mortality rate, we can use the rate(*ratevar_p crudevar_p*) or rate(*ratevar_p* #) option, where *crudevar_p* refers to the variable containing the total population's crude death rate or # is the total population's crude death rate.

Now let's look at the states' data (study population):

```
. use http://www.stata-press.com/data/r14/kahn
. list, sep(4)
```

	state	age	populat~n	death	st	death_~e
1.	California	<15	5,524,000	166,285	1	.0016
2.	California	15-24	3,558,000	166,285	1	.0013
3.	California	25-34	2,677,000	166,285	1	.0015
4.	California	35-44	2,359,000	166,285	1	.0028
5.	California	45-54	2,330,000	166,285	1	.0067
6.	California	55-64	1,704,000	166,285	1	.0154
7.	California	65-74	1,105,000	166,285	1	.0328
8.	California	75+	696,000	166,285	1	.0917
9.	Maine	<15	286,000	11,051	2	.0019
10.	Maine	15-24	168,000	.	2	.0011
11.	Maine	25-34	110,000	.	2	.0014
12.	Maine	35-44	109,000	.	2	.0029
13.	Maine	45-54	110,000	.	2	.0069
14.	Maine	55-64	94,000	.	2	.0173
15.	Maine	65-74	69,000	.	2	.039
16.	Maine	75+	46,000	.	2	.1041

For each state, the number of individuals in each stratum (age group) is contained in the pop variable. The death variable is the total number of deaths observed in the state during the year. It must have the same value for all observations in the group, as for California, or it could be missing in all but one observation per group, as for Maine.

To match these two datasets, the strata variables must have the same name in both datasets and ideally the same levels. If a level is missing from either dataset, that level will not be included in the standardization.

With kahn.dta in memory, we now execute the command. We will use the print option to obtain the standard population's summary table, and because we have both the standard population's age-specific count and deaths, we will specify the popvars(*casevar_p popvar_p*) option. Or, we could specify the rate(rate 0.00945) option because we know that 0.00945 is the U.S. crude death rate for 1970.

```
. istdize death pop age using http://www.stata-press.com/data/r14/popkahn,
> by(state) pop(deaths pop) print
```

```
————Standard Population————
```

Stratum	Rate
<15	0.00178
15-24	0.00128
25-34	0.00157
35-44	0.00315
45-54	0.00730
55-64	0.01659
65-74	0.03583
75+	0.09656

Standard population's crude rate: 0.00945

-> state= California

Indirect Standardization

Stratum	Standard Population Rate	Observed Population	Cases Expected
<15	0.0018	5,524,000	9,832.72
15-24	0.0013	3,558,000	4,543.85
25-34	0.0016	2,677,000	4,212.46
35-44	0.0031	2,359,000	7,419.59
45-54	0.0073	2,330,000	17,010.10
55-64	0.0166	1,704,000	28,266.14
65-74	0.0358	1,105,000	39,587.63
75+	0.0966	696,000	67,206.23
Totals:		19953000	178078.73

Observed Cases: 166,285
SMR (Obs/Exp):   0.93
SMR exact 95% Conf. Interval: [0.9293, 0.9383]
Crude Rate: 0.0083
Adjusted Rate: 0.0088
95% Conf. Interval: [0.0088, 0.0089]

-> state= Maine

Indirect Standardization

Stratum	Standard Population Rate	Observed Population	Cases Expected
<15	0.0018	286,000	509.08
15-24	0.0013	168,000	214.55
25-34	0.0016	110,000	173.09
35-44	0.0031	109,000	342.83
45-54	0.0073	110,000	803.05
55-64	0.0166	94,000	1,559.28
65-74	0.0358	69,000	2,471.99
75+	0.0966	46,000	4,441.79
Totals:		992000	10515.67

Observed Cases: 11,051
SMR (Obs/Exp):   1.05
SMR exact 95% Conf. Interval: [1.0314, 1.0707]
Crude Rate: 0.0111
Adjusted Rate: 0.0099
95% Conf. Interval: [0.0097, 0.0101]

```
Summary of Study Populations (Rates):
                Cases
     state   Observed      Crude   Adj_Rate  Confidence Interval
  ─────────────────────────────────────────────────────────────
  California   166,285   0.008334   0.008824  [0.008782, 0.008866]
       Maine    11,051   0.011140   0.009931  [0.009747, 0.010118]
Summary of Study Populations (SMR):
                Cases       Cases                      Exact
     state   Observed    Expected    SMR     Confidence Interval
  ─────────────────────────────────────────────────────────────
  California   166,285  178,078.73   0.934    [0.929290, 0.938271]
       Maine    11,051   10,515.67   1.051    [1.031405, 1.070688]
```

◁

Stored results

dstdize stores the following in r():

Scalars
  r(k)   number of populations

Macros
  r(by)   variable names specified in by()
  r(c#)   values of r(by) for #th group

Matrices
  r(se) $1\times k$ vector of standard errors of adjusted rates
  r(ub_adj) $1\times k$ vector of upper bounds of confidence intervals for adjusted rates
  r(lb_adj) $1\times k$ vector of lower bounds of confidence intervals for adjusted rates
  r(Nobs) $1\times k$ vector of number of observations
  r(crude) $1\times k$ vector of crude rates (*)
  r(adj) $1\times k$ vector of adjusted rates (*)

(*) If, in a group, the number of observations is 0, then 9 is stored for the corresponding crude and adjusted rates.

istdize stores the following in r():

Scalars
  r(k)   number of populations

Macros
  r(by)   variable names specified in by()
  r(c#)   values of r(by) for #th group

Matrices
  r(cases_obs) $1\times k$ vector of number of observed cases
  r(cases_exp) $1\times k$ vector of number of expected cases
  r(ub_adj) $1\times k$ vector of upper bounds of confidence intervals for adjusted rates
  r(lb_adj) $1\times k$ vector of lower bounds of confidence intervals for adjusted rates
  r(crude) $1\times k$ vector of crude rates
  r(adj) $1\times k$ vector of adjusted rates
  r(smr) $1\times k$ vector of SMRs
  r(ub_smr) $1\times k$ vector of upper bounds of confidence intervals for SMRs
  r(lb_smr) $1\times k$ vector of lower bounds of confidence intervals for SMRs

Methods and formulas

The directly standardized rate, S_R, is defined by

$$S_R = \frac{\sum_{i=1}^{k} w_i R_i}{\sum_{i=1}^{k} w_i}$$

(Rothman 1986, 44), where R_i is the stratum-specific rate in stratum i and w_i is the weight for stratum i derived from the standard population.

If n_i is the population of stratum i, the standard error, $se(S_R)$, in stratified sampling for proportions (ignoring the finite population correction) is

$$se(S_R) = \frac{1}{\sum w_i} \sqrt{\sum_{i=1}^{k} \frac{w_i{}^2 R_i (1 - R_i)}{n_i}}$$

(Cochran 1977, 108), from which the confidence intervals are calculated.

For indirect standardization, define O as the observed number of cases in each population to be standardized; $n_1, \ldots, n_k$ as the distribution across the various strata for the population being studied; $R_1, \ldots, R_k$ as the stratum-specific rates for the standard population; and C as the crude rate of the standard population. The expected number of cases (deaths), E, in each population is obtained by applying the standard population stratum-specific rates, $R_1, \ldots, R_k$, to the study populations:

$$E = \sum_{i=1}^{k} n_i R_i$$

The indirectly adjusted rate is then

$$R_{\text{indirect}} = C \frac{O}{E}$$

and O/E is the study population's SMR if death is the event of interest or the SIR for studies of disease (or other) incidence.

The exact confidence interval is calculated for each estimated SMR by assuming a Poisson process as described in Breslow and Day (1987, 69–71). These intervals are obtained by first calculating the upper and lower bounds for the confidence interval of the Poisson-distributed observed events, O—say, L and U, respectively—and then computing $SMR_L = L/E$ and $SMR_U = U/E$.

Acknowledgments

References

Breslow, N. E., and N. E. Day. 1987. *Statistical Methods in Cancer Research: Vol. 2—The Design and Analysis of Cohort Studies.* Lyon: IARC.

Cleves, M. A. 1998. sg80: Indirect standardization. *Stata Technical Bulletin* 42: 43–47. Reprinted in *Stata Technical Bulletin Reprints*, vol. 7, pp. 224–228. College Station, TX: Stata Press.

Cochran, W. G. 1977. *Sampling Techniques.* 3rd ed. New York: Wiley.

Consonni, D. 2012. A command to calculate age-standardized rates with efficient interval estimation. *Stata Journal* 12: 688–701.

Fleiss, J. L., B. Levin, and M. C. Paik. 2003. *Statistical Methods for Rates and Proportions.* 3rd ed. New York: Wiley.

Forthofer, R. N., and E. S. Lee. 1995. *Introduction to Biostatistics: A Guide to Design, Analysis, and Discovery.* New York: Academic Press.

Juul, S., and M. Frydenberg. 2014. *An Introduction to Stata for Health Researchers.* 4th ed. College Station, TX: Stata Press.

Kahn, H. A., and C. T. Sempos. 1989. *Statistical Methods in Epidemiology.* New York: Oxford University Press.

Kirkwood, B. R., and J. A. C. Sterne. 2003. *Essential Medical Statistics.* 2nd ed. Malden, MA: Blackwell.

McGuire, T. J., and J. A. Harrison. 1994. sbe11: Direct standardization. *Stata Technical Bulletin* 21: 5–9. Reprinted in *Stata Technical Bulletin Reprints*, vol. 4, pp. 88–94. College Station, TX: Stata Press.

Pagano, M., and K. Gauvreau. 2000. *Principles of Biostatistics.* 2nd ed. Belmont, CA: Duxbury.

Rothman, K. J. 1986. *Modern Epidemiology.* Boston: Little, Brown.

van Belle, G., L. D. Fisher, P. J. Heagerty, and T. S. Lumley. 2004. *Biostatistics: A Methodology for the Health Sciences.* 2nd ed. New York: Wiley.

Wang, D. 2000. sbe40: Modeling mortality data using the Lee–Carter model. *Stata Technical Bulletin* 57: 15–17. Reprinted in *Stata Technical Bulletin Reprints*, vol. 10, pp. 118–121. College Station, TX: Stata Press.

Also see

[R] **epitab** — Tables for epidemiologists

[SVY] **direct standardization** — Direct standardization of means, proportions, and ratios