

Estandarización: Macro !STD para SPSS Statistics

DIRECT AND INDIRECT STANDARDIZATION

Created 1999.11.08 Last revised 2012.05.09

(c) JM.Domenech

Programmer: R.Sesma

Vancouver reference:

Domenech JM. Macro !STD for SPSS Statistics.

Direct and Indirect Standardization [computer program].

V2012.05.09. Bellaterra: Universitat Autònoma de Barcelona; 2012.

Available from: <http://www.metodo.uab.cat/macros.htm>.

Macro Call (tested for SPSS 18.0 and higher):

```
!STD Type = Type of Standardization
      D - Direct Standardization (rates and ratios)
      I - Indirect Standardization (rates)
      A1 = Variable name with the list of cases in Group 1 by Stratum
      {/T1 = Variable name with the list of person-time in Group 1 by Stratum
      or /N1} = Variable name with the list of population in Group 1 by Stratum
      StdPop = Variable name with the values of the Standard Population
      [StdSize] = Standard Population Size
      Obs: Size of the Observed Population (default)
      Std: Size of the Standard Population
      [Factor] = Multiplying factor for Risk/Rates values. By default, FACTOR=1
      Possible values are powers of 10: 1 10 100 1000 10000 ....
      Str = Variable name with Stratum values
      [StrLabel] = Variable name with Stratum labels (String, up to 7 characters)
      [Split] = Variable name with Split values
      [SplitLabel] = Variable name with Split labels (String, up to 10 characters)
      [Print] = TABLE - Print Stratum tables; default
      NOTABLE - Do NOT print Stratum tables
      [/CL] = Confidence level (%); by default: 95 (50 <= CL < 100)
      [/NST] = Name of the study (Label)
      [HELP] Print Macro call arguments and Samples of calls
```

Parameters for records an additional Group

```
A0 = Variable name with the list of cases in Group 0 by Stratum
{/T0 = Variable name with the list of person-time in Group 0 by Stratum
or /N0} = Variable name with the list of population in Group 0 by Stratum
```

Parameters for compute Comparative Mortality Figure

```
[{/ICS = Crude rate in STANDARD POPULATION (ICS > 0)
or /RCS}] = Crude risk in STANDARD POPULATION (0 < RCS < 1)
```

Notes: Parameters enclosed in Square Brackets [] are optional

Parameters enclosed in Braces { } indicate a choice between elements

Parameters /T1, /T0 and /ICS records RATE DATA.

Parameters /N1, /N0 and /RCS records RISK DATA.

WARNING: The macro does not verify variable names. Wrong variable names

will produce unexpected results.

A new dataset is created, but not saved.

This dataset contains the macro results.

Samples of calls:

!STD HELP.

```
!STD Type=D /A1=Cases.M /N1=Pop.M /A0=Cases.F /N0=Pop.F /StdPop=StdUS /StdSize=Obs
/Str=AgeGroup /StrLabel=AgeGroupL /Split=Year /Factor= 100000 /CL=95
/NST='Childhood Cancers'.
```

```
!STD Type=D /A1=CancSolt /T1=PobSolt /A0=CancCas /T0=PobCas /StdPop=Std
/StdSize=Std /Str=GrEdad /ICS=0.000536371 /CL=95.
```

!! La macro realiza comprobaciones de los parámetros antes de ejecutarla pero **no son exhaustivas** debido a las limitaciones del lenguaje de macros; por ejemplo, si hay un nombre de variable erróneo la macro presenta un malfuncionamiento pero no da ningún aviso.

Por este motivo se evitarán pérdidas de tiempo siendo muy cuidadoso al escribir la sintaxis; si la macro no presenta los resultados solicitados se debe revisar la sintaxis y consultar la ayuda.

!! La macro **no respeta los filtros**: trabaja con todos los casos aunque esté activado un filtro. Para analizar sólo una parte de los casos es necesario utilizar el procedimiento SELECT IF para seleccionar los casos y eliminar el resto de forma permanente. Una vez eliminados se ejecuta la macro.

Ejemplo 1: Estandarización directa (US Childhood Cancers.xls)

La pestaña *Childhood Cancers* del archivo Excel reproducido a continuación contiene la distribución por sexo y edad de los cánceres diagnosticados en niños y adolescentes (*All Childhood Cancers Combined*) los años 2001, 2005 y 2009 (<http://wonder.cdc.gov/cancernpcr-v2009.html>). La pestaña *US 2000 Std Pop* de este archivo contiene la distribución por grupos de edad de la población estándar de los Estados Unidos de América del año 2000 (<http://www.cdc.gov/nchs/data/statnt/statnt20.pdf>).

El objetivo del estudio es calcular los riesgos estandarizados en cada año y compararlos por sexo.

	A	B	C	D	E
1	Year	Sex	AgeGroup	Cases	Population
2	2001	F	0	431	1783603
3	2001	F	1	1397	6787548
4	2001	F	5	948	8929324
5	2001	F	10	1164	9210928
6	2001	F	15	1787	8895008
7	2001	M	0	473	1863015
8	2001	M	1	1603	7104694
9	2001	M	5	1136	9370681
10	2001	M	10	1310	9677392
11	2001	M	15	2115	9463898
12	2005	F	0	429	1920146
13	2005	F	1	1480	7658048
14	2005	F	5	960	9175135

	A	B	C	D
1	AgeGroup	AgeGroupL	StdPop	
2	0	0	3795	
3	1	1-4	15192	
4	5	5-9	19920	
5	10	10-14	20057	
6	15	15-19	19820	
7				

Archivo: **US Childhood Cancers.xls**

Para realizar este análisis con la macro !STD se necesita una ventana de datos con la siguiente estructura: para cada grupo de edad los datos por sexo y el valor de la población estándar deben estar en la misma fila. Esta estructura se puede repetir si se desean estandarizar más poblaciones, en este estudio se repite para los años 2005 y 2009.

	Year	AgeGroup	Cases.F	Population.F	Cases.M	Population.M	AgeGroupL	StdPop
1	2001	0	431	1783603	473	1863015	0	3795
2	2001	1	1397	6787548	1603	7104694	1-4	15192
3	2001	5	948	8929324	1136	9370681	5-9	19920
4	2001	10	1164	9210928	1310	9677392	10-14	20057
5	2001	15	1787	8895008	2115	9463898	15-19	19820
6	2005	0	429	1920146	505	2012200	0	3795
7	2005	1	1480	7658048	1774	8000072	1-4	15192
8	2005	5	960	9175135	1213	9608892	5-9	19920
9	2005	10	1243	9735388	1337	10220528	10-14	20057
10	2005	15	2133	9825614	2369	10379816	15-19	19820
11	2009	0	452	1991224	543	2082240	0	3795
12	2009	1	1504	7969251	1776	8327776	1-4	15192
13	2009	5	1024	9641038	1282	10083089	5-9	19920
14	2009	10	1247	9337774	1331	9788023	10-14	20057
15	2009	15	2141	10039664	2332	10586169	15-19	19820

Preparación de los datos

```
GET DATA
  /TYPE=XLS
  /FILE='C:\...\US Childhood Cancers.xls'
  /SHEET=name 'Childhood Cancers'
  /CELLRANGE=full
  /READNAMES=on
  /ASSUMEDSTRWIDTH=10.
EXECUTE.
DATASET NAME Childhood_Cancers WINDOW=FRONT.
```

	Year	Sex	AgeGroup	Cases	Population
1	2001	F	0	431	1783603
2	2001	F	1	1397	6787548
3	2001	F	5	948	8929324
4	2001	F	10	1164	9210928
5	2001	F	15	1787	8895008
6	2001	M	0	473	1863015
7	2001	M	1	1603	7104694
8	2001	M	5	1136	9370681
9	2001	M	10	1310	9677392
10	2001	M	15	2115	9463898
11	2005	F	0	429	1920146
12	2005	F	1	1480	7658048

```
SORT CASES BY Year AgeGroup Sex.
CASESTOVARS
  /ID=Year AgeGroup
  /INDEX=Sex
  /GROUPBY=INDEX.
```

	Year	AgeGroup	Cases.F	Population.F	Cases.M	Population.M
1	2001	0	431	1783603	473	1863015
2	2001	1	1397	6787548	1603	7104694
3	2001	5	948	8929324	1136	9370681
4	2001	10	1164	9210928	1310	9677392
5	2001	15	1787	8895008	2115	9463898
6	2005	0	429	1920146	505	2012200
7	2005	1	1480	7658048	1774	8000072
8	2005	5	960	9175135	1213	9608892
9	2005	10	1243	9735388	1337	10220528
10	2005	15	2133	9825614	2369	10379816
11	2009	0	452	1991224	543	2082240
12	2009	1	1504	7969251	1776	8327776
13	2009	5	1024	9641038	1282	10083089
14	2009	10	1247	9337774	1331	9788023
15	2009	15	2141	10039664	2332	10586169

```
GET DATA /TYPE=XLS /FILE='C:\...\US Childhood Cancers.xls' /SHEET=name 'US 2000 Std Pop'
  /CELLRANGE=full /READNAMES=on /ASSUMEDSTRWIDTH=10.
EXECUTE.
DATASET NAME US_2000_StdPop WINDOW=FRONT.
SORT CASES BY AgeGroup(A).
```

```
DATASET ACTIVATE Childhood_Cancers.
SORT CASES BY AgeGroup(A).
MATCH FILES /FILE=* /TABLE='US_2000_StdPop' /BY AgeGroup.
EXECUTE.
DATASET CLOSE US_2000_StdPop.
SORT CASES BY Year(A) AgeGroup(A).
SAVE OUTFILE='C:\...\US Childhood Cancers.sav'.
```

	Year	AgeGroup	Cases.F	Population.F	Cases.M	Population.M	AgeGroupL	StdPop
1	2001	0	431	1783603	473	1863015	0	3795
2	2001	1	1397	6787548	1603	7104694	1-4	15192
3	2001	5	948	8929324	1136	9370681	5-9	19920
4	2001	10	1164	9210928	1310	9677392	10-14	20057
5	2001	15	1787	8895008	2115	9463898	15-19	19820
6	2005	0	429	1920146	505	2012200	0	3795
7	2005	1	1480	7658048	1774	8000072	1-4	15192
8	2005	5	960	9175135	1213	9608892	5-9	19920
9	2005	10	1243	9735388	1337	10220528	10-14	20057
10	2005	15	2133	9825614	2369	10379816	15-19	19820
11	2009	0	452	1991224	543	2082240	0	3795
12	2009	1	1504	7969251	1776	8327776	1-4	15192
13	2009	5	1024	9641038	1282	10083089	5-9	19920
14	2009	10	1247	9337774	1331	9788023	10-14	20057
15	2009	15	2141	10039664	2332	10586169	15-19	19820

La sintaxis de la página anterior crea la matriz de datos con la estructura adecuada para la macro.

El primer GET lee los datos de la pestaña '**Childhood Cancers**' del Excel y con el procedimiento DATASET NAME se le asigna a la ventana de datos el nombre **Childhood_Cancers** (este nombre no puede tener espacios en blanco).

En segundo lugar, CASETOVAR reestructura la matriz de datos para situar ambos sexos en la misma fila. Observe que la variable de índice es **Sex** (con valores F y M) y después de la reestructuración las variables "Cases" y "Population" pasan a ser "Cases.F", "Cases.M" y "Population.F", "Population.M". Este paso sólo se debe realizar sólo si los datos de sexo masculino y femenino del archivo Excel están en registros diferentes.

En tercer lugar se debe añadir a cada registro el correspondiente valor de la población estándar con el procedimiento MATCH FILES. Para ello, primero, con un GET se lee los datos de la pestaña '**US 2000 Std Pop**' del Excel y con el procedimiento DATASET NAME se le asigna a la nueva ventana de datos el nombre **US_2000_StdPop**. Seguidamente se ordenan los registros de esta ventana por la variable de enlace (match) **AgeGroup**.

A continuación se activa la ventana de datos **Childhood_Cancers** con **DATASET ACTIVATE** y con **MATCH FILES** se añaden a esta ventana de datos las variables AgeGroupL y StdPop de la tabla de claves situada en la ventana **US_2000_StdPop**. Finalmente el procedimiento **DATASET CLOSE** cierra la ventana **US_2000_StdPop** y **SORT CASES** ordena la ventana de datos a estandarizar por la variable Year y dentro de cada año por AgeGroup.

La ventana de datos a estandarizar se guarda con el nombre '**US Childhood Cancers.sav**'.

Parámetros de la macro

Para realizar la estandarización directa se ejecutará la macro !STD con los siguientes parámetros:

```
!STD Type=D /A1=Cases.M /N1=Population.M /A0=Cases.F /N0=Population.F /StdPop=StdPop  
/StdSize=Obs /Str=AgeGroup /StrLabel=AgeGroupL /Split=Year /Factor= 100000 /CL=95  
/NST='Childhood Cancers'
```

- El parámetro Type contiene el tipo de estandarización (D=Directa).
- Los parámetros A1 y N1 contienen los nombres de las variables con el número de casos de cáncer y con la subpoblación de sexo masculino.
- Los parámetros A0 y N0 contienen los nombres de las variables con el número de casos de cáncer y con la subpoblación de sexo femenino. Si en lugar de número de sujetos de la población se dispusiera de número de personas-tiempo los parámetros N1 y N0 se sustituyen por T1 y T0.
- El parámetro StdPop contiene el nombre de la variable con la población de referencia.
- El parámetro StdSize sirve para controlar el tamaño de la población estándar que se imprime en los resultados. La opción de defecto, StdSize=Obs, indica que el tamaño de la población estándar se debe escalar al tamaño de la proporción observada; la ventaja de esta opción es que se puede ver el efecto de la estandarización comparando los casos observados con los esperados.
- El parámetro Str contiene el nombre de la variable con los estratos. Esta variable debe ser numérica.
- El parámetro StrLabel es opcional y contiene el nombre de la variable con las etiquetas de los estratos.
- El parámetro Split sólo es necesario si la matriz de datos contiene más de una población a estandarizar. En este estudio hay tres poblaciones que corresponden a los años 2001, 2005 y 2009. Split=Year indica que la variable Year contiene el identificador de cada población a estandarizar; esta variable debe ser numérica.
- El parámetro Factor= 100000 indica que los riesgos se deben calcular por cada 100000 sujetos. Este parámetro debe ser múltiplo de 10 y por defecto vale 1.
- El parámetro CL=95 indica que estimar los índices con una confianza del 95%; si se omite este parámetro se utiliza una confianza del 95%.
- El parámetro Nst es optativo y contiene el nombre del estudio escrito entre comillas.

Ejecución de la macro

Macro !STD Type=D /A1=Cases.M /N1=Population.M /A0=Cases.F /N0=Population.F /StdPop=StdPop DIRECT STANDARDIZATION STUDY: Childhood Cancers GROUP 1: Population.M GROUP 0: Population.F Year= 2001 GROUP 1 ----- Observed ----- Standard Expected Stratum Cases Population Risk(*) Population Cases Cases Population Risk(*) Population Cases ----- 0 473 1863015 25,3889529 1805384 458,37 431 1783603 24,1645703 1715149 414,46 1-4 1603 7104694 22,5625481 7227245 1630,65 1397 6787548 20,5818066 6866021 1413,15 5-9 1136 9370681 12,1229183 9476483 1148,83 948 8929324 10,6167051 9002839 955,80 10-14 1310 9677392 13,5367049 9541657 1291,63 1164 9210928 12,6371632 9064757 1145,53 15-19 2115 9463898 22,3480853 9428910 2107,18 1787 8895008 20,0899201 8957645 1799,58 ----- TOTAL: 6637 37479680 17,7082622 37479680 6636,65 5727 35606411 16,0841821 35606411 5728,53 ----- Year= 2005 GROUP 1 ----- Observed ----- Standard Expected Stratum Cases Population Risk(*) Population Cases Cases Population Risk(*) Population Cases ----- 0 505 2012200 25,0969089 1937457 486,24 429 1920146 22,3420511 1845589 412,34 1-4 1774 8000072 22,1748004 7755955 1719,87 1480 7658048 19,3260737 7388192 1427,85 5-9 1213 9608892 12,6237239 10169735 1283,80 960 9175135 10,4630613 9687519 1013,61 10-14 1337 10220528 13,0815159 10239678 1339,51 1243 9735388 12,7678527 9754145 1245,39 15-19 2369 10379816 22,8231406 10118683 2309,40 2133 9825614 21,7085670 9638887 2092,46 ----- TOTAL: 7198 40221508 17,8958979 40221508 7138,81 6245 38314331 16,2993842 38314331 6191,66 ----- Year= 2009 GROUP 1 ----- Observed ----- Standard Expected Stratum Cases Population Risk(*) Population Cases Cases Population Risk(*) Population Cases ----- 0 543 2082240 26,0776856 1968565 513,36 452 1991224 22,6996059 1877604 426,21 1-4 1776 8327776 21,3262220 7880483 1680,61 1504 7969251 18,8725390 7516351 1418,53 5-9 1282 10083089 12,7143577 10333019 1313,78 1024 9641038 10,6212630 9855563 1046,79 10-14 1331 9788023 13,5982517 10404084 1414,77 1247 9337774 13,3543605 9923345 1325,20 15-19 2332 10586169 22,0287434 10281146 2264,81 2141 10039664 21,3254149 9806088 2091,19 ----- TOTAL: 7264 40867297 17,7746035 40867297 7187,32 6368 38978951 16,3370225 38978951 6307,91 ----- 													
<div>continúa ...</div>													

... continuación

Year		Crude Estimate	Direct Standardized Estimate	95% Confidence Interval	
2001					
Risk(*)	Group 1	17,7082622	17,7073337	17,2812844	18,1333831
	Group 0	16,0841821	16,0884686	15,6717822	16,5051551
Risk Difference(*)		1,6240800	1,6188651	1,0229241	2,2148060
Risk Ratio		1,10097	1,10062	1,06239	1,14023

Year		Crude Estimate	Direct Standardized Estimate	95% Confidence Interval	
2005					
Risk(*)	Group 1	17,8958979	17,7487501	17,3385084	18,1589918
	Group 0	16,2993842	16,1601670	15,7592127	16,5611213
Risk Difference(*)		1,5965137	1,5885831	1,0149434	2,1622229
Risk Ratio		1,09795	1,09830	1,06168	1,13618

Year		Crude Estimate	Direct Standardized Estimate	95% Confidence Interval	
2009					
Risk(*)	Group 1	17,7746035	17,5869798	17,1821464	17,9918132
	Group 0	16,3370225	16,1828583	15,7850329	16,5806836
Risk Difference(*)		1,4375810	1,4041215	,8365344	1,9717086
Risk Ratio		1,08800	1,08677	1,05078	1,12399

(*) per 100000

La macro genera una nueva ventana de datos @ResultsStdD con los riesgos estandarizados y sus intervalos de confianza, y también con las diferencias y razones de riesgos de las subpoblaciones de sexo.

	Year	CR1	DS1	lbDS1	ubDS1	CR0	DS0	lbDS0	ubDS0	CRD	SRD	lbSRD	ubSRD	CRR	SRR	lbSRR	ubSRR
1	2001	17,71	17,71	17,28	18,13	16,08	16,09	15,67	16,51	1,62	1,62	1,02	2,21	1,10	1,10	1,06	1,14
2	2005	17,90	17,75	17,34	18,16	16,30	16,16	15,76	16,56	1,60	1,59	1,01	2,16	1,10	1,10	1,06	1,14
3	2009	17,77	17,59	17,18	17,99	16,34	16,18	15,79	16,58	1,44	1,40	,84	1,97	1,09	1,09	1,05	1,12

Si para posteriores análisis sólo interesa guardar los riesgos crudos y estandarizados, la siguiente sintaxis permite eliminar las variables no deseadas de la ventana de datos de resultados.

```
DATASET ACTIVATE @ResultsStdD.
SAVE OUTFILE='C:\...\US Childhood Cancers Standardized.sav'
/DROP=lbDS1 ubDS1 lbDS0 ubDS0 CRD SRD lbSRD ubSRD CRR SRR lbSRR ubSRR.
GET FILE='C:\...\US Childhood Cancers Standardized.sav'.
```

	Year	CR1	DS1	CR0	DS0
1	2001	17,7083	17,7073	16,0842	16,0885
2	2005	17,8959	17,7488	16,2994	16,1602
3	2009	17,7746	17,5870	16,3370	16,1829