

Soc 195B 04W

EXAMPLES OF DIRECT STANDARDIZATION

This annotated Chip output file illustrates the use of the “Standardize” command using a file named “status98_freq.chp” that has the same information as “status98.chp.” (The file named “status98.chp” is formatted in such a way that the standardization command does not work with it, but other commands do work with it.)

I have placed a self-extracting file named “status98_freq.sfx” on the course web site. It may be found in the “Index of Course Materials.”

1. To begin this session I opened a log file named “status98_21oct02.log.”

2. I next opened the file “status98_freq.chp.”

```
Status98 (Status attainment model; source: '96 & '98 GSS; Full-time workers
only)
N = 2195
```

3. Here is the result of requesting the tabulation of respondent’s income by father’s occupational prestige. Clearly there is an association between the two variables.

Income/Papres

	Low	Medium	High	All
\$35K+	32.8	34.0	44.2	37.3
\$17.5K-\$	43.1	42.6	39.8	41.7
<\$17.5K	24.1	23.4	16.1	21.0
100%=	677	721	797	N = 2195

4. I next issued commands to percentage respondent’s income by father’s occupational prestige, controlling for respondent’s education. This makes for a very large table, consisting of four conditional subtables:

Ed = <12yrs

	Low	Medium	High	All
\$35K+	20.3	14.7	7.7	16.0
\$17.5K-\$	40.5	41.3	42.3	41.1
<\$17.5K	39.2	44.0	50.0	42.9
100%=	74	75	26	N = 175

Examples of Direct Standardization

Ed = 12yrs

	Low	Medium	High	All
\$35K+	23.5	21.7	29.3	24.3
\$17.5K-\$	47.8	49.3	45.9	47.8
<\$17.5K	28.7	29.0	24.8	27.9
100%=	268	221	157	N = 646

Ed = 13-15yrs

	Low	Medium	High	All
\$35K+	32.3	26.4	38.6	32.6
\$17.5K-\$	45.2	49.1	42.6	45.5
<\$17.5K	22.6	24.5	18.8	21.9
100%=	155	163	176	N = 494

Ed = 16+yrs

	Low	Medium	High	All
\$35K+	52.2	54.6	53.9	53.8
\$17.5K-\$	35.6	33.2	36.3	35.2
<\$17.5K	12.2	12.2	9.8	11.0
100%=	180	262	438	N = 880

5. It is difficult to summarize, much less interpret such a complex table, yet it is highly desirable to do so. Fortunately, there are tools to deal with this situation. The simplest of these is known as “direct standardization.” We will henceforth refer to this tool as “standardization,” since we will not be using variants such as “indirect standardization.”

We will use standardization to form a particular type of weighted average over conditional subtables, so that we can summarize controlled associations between pairs of variables. Before explaining the calculations, I want to illustrate how you can use Chip to produce standardized results.

Although you can’t see it in this transcript file, after extracting the income by father’s occupational prestige table within each educational level (i.e., controlling for education), I issued the “standardize” command. That produced the following output:

```
Status98 (Status attainment model; source: '96 & '98 GSS; Full-time workers
only)
File
Causal order:
  Ed -> Inco* -> Papr* -> Regi* -> Race* -> Sex* -> Age* -> Marr* -> Sibs*
  4   x   3   x   3   x   4   x   2   x   2   x   3   x   3   x   3
N = 2197
```

6. The above output should be read as telling us that until we tell it otherwise, Chip will standardize all associations for educations (note the lack of an asterisk next to “Ed”). (Notice that, unfortunately, the *N* is 2,197

and not 2,195. This is due to internal rounding errors in Chip. In the mathematics of standardization, the *N* stays exactly the same.)

Still, Chip will do nothing until you give it the next command. The command is not printed in the log file, but it was simply “frequency,” which produced the following table of frequencies standardized for education:

	Low	Medium	High	100%=	
\$35K+	254	248	317	820	
\$17.5K-\$	282	303	331	917	
<\$17.5K	147	163	149	461	
100%=	684	715	798	N =	2197

7. I presented these frequencies so that you could see that the number of respondents remains the same under standardization. The number of individuals in each father’s occupational prestige category also remains the same.

Next I asked for “percent down”:

	Low	Medium	High	All	
\$35K+	37.2	34.7	39.8	37.3	
\$17.5K-\$	41.2	42.4	41.5	41.7	
<\$17.5K	21.6	22.9	18.7	21.0	
100%=	684	715	798	N =	2197

8. The above percentage table summarizes the association between respondent’s income and father’s occupational prestige, controlling respondent’s education. It shows that there is virtually no association.

9. Now let’s consider what happens to the income by education association, controlling father’s occupational prestige. First, let’s remind ourselves of the uncontrolled association. To obtain that, I need to “mouse” through “standard|restore”. Then I issue the commands for the uncontrolled table.

Income/Ed

	<12yrs	12yrs	13-15yrs	16+yrs	All	
\$35K+	16.0	24.3	32.6	53.8	37.3	
\$17.5K-\$	41.1	47.8	45.5	35.2	41.7	
<\$17.5K	42.9	27.9	21.9	11.0	21.0	
100%=	175	646	494	880	N =	2195

10. Here comes the table controlling for father's occupational prestige:

Papres = Low

	<12yrs	12yrs	13-15yrs	16+yrs	All
\$35K+	20.3	23.5	32.3	52.2	32.8
\$17.5K-\$	40.5	47.8	45.2	35.6	43.1
<\$17.5K	39.2	28.7	22.6	12.2	24.1
100%=	74	268	155	180	N = 677

Papres = Medium

	<12yrs	12yrs	13-15yrs	16+yrs	All
\$35K+	14.7	21.7	26.4	54.6	34.0
\$17.5K-\$	41.3	49.3	49.1	33.2	42.6
<\$17.5K	44.0	29.0	24.5	12.2	23.4
100%=	75	221	163	262	N = 721

Papres = High

	<12yrs	12yrs	13-15yrs	16+yrs	All
\$35K+	7.7	29.3	38.6	53.9	44.2
\$17.5K-\$	42.3	45.9	42.6	36.3	39.8
<\$17.5K	50.0	24.8	18.8	9.8	16.1
100%=	26	157	176	438	N = 797

11. Again, such a table is difficult to summarize and interpret. We standardize on father's occupational prestige.

Status98 (Status attainment model; source: '96 & '98 GSS; Full-time workers only)

File

Causal order:

Papre -> Inco* -> Ed* -> Regi* -> Race* -> Sex* -> Age* -> Marr* -> Sibs*
 3 x 3 x 4 x 4 x 2 x 2 x 3 x 3 x 3

N = 2195

12. Next, we percentage the table that Chip has been holding in memory for us:

	<12yrs	12yrs	13-15yrs	16+yrs	All
\$35K+	17.8	25.3	32.6	52.6	37.3
\$17.5K-\$	40.9	47.5	45.4	35.6	41.7
<\$17.5K	41.3	27.2	22.0	11.8	21.0
100%=	173	645	496	881	N = 2195

13. The standardized table can and should be compared to the original, uncontrolled table:

Income/Ed

	<12yrs	12yrs	13-15yrs	16+yrs	All
\$35K+	16.0	24.3	32.6	53.8	37.3
\$17.5K-\$	41.1	47.8	45.5	35.2	41.7
<\$17.5K	42.9	27.9	21.9	11.0	21.0
100%=	175	646	494	880	N = 2195

14. Standardization becomes even more helpful when we want to control for more than one variable. Consider the income by education association, controlling for race and region.

To obtain the 4-way tabulation, I need to “restore” the data to their pre-standardized state. Then I issue the appropriate table, control, and percentaging commands:

Race = White

Region = Northeast

	<12yrs	12yrs	13-15yrs	16+yrs	All
\$35K+	21.2	29.0	43.2	53.1	41.2
\$17.5K-\$	39.4	50.0	45.9	35.2	42.0
<\$17.5K	39.4	21.0	10.8	11.7	16.8
100%=	33	100	74	145	N = 352

Race = White

Region = Midwest

	<12yrs	12yrs	13-15yrs	16+yrs	All
\$35K+	12.5	22.7	33.6	55.6	37.2
\$17.5K-\$	45.8	49.7	45.7	34.3	42.8
<\$17.5K	41.7	27.6	20.7	10.1	20.0
100%=	24	181	116	198	N = 519

Race = White

Region = South

	<12yrs	12yrs	13-15yrs	16+yrs	All
\$35K+	13.4	24.1	25.3	52.4	33.9
\$17.5K-\$	46.3	49.2	47.9	37.8	44.4
<\$17.5K	40.3	26.6	26.7	9.8	21.7
100%=	67	199	146	246	N = 658

Examples of Direct Standardization

Race = White
Region = West

	<12yrs	12yrs	13-15yrs	16+yrs	All	
\$35K+	18.2	29.8	42.2	55.9	45.7	
\$17.5K-\$	36.4	44.0	38.9	32.4	36.4	
<\$17.5K	45.5	26.2	18.9	11.7	17.9	
100%=	22	84	90	222	N =	418

Race = Black
Region = Northeast

	<12yrs	12yrs	13-15yrs	16+yrs	All	
\$35K+	33.3	18.8	27.3	16.7	21.4	
\$17.5K-\$	33.3	50.0	45.5	66.7	52.4	
<\$17.5K	33.3	31.3	27.3	16.7	26.2	
100%=	3	16	11	12	N =	42

Race = Black
Region = Midwest

	<12yrs	12yrs	13-15yrs	16+yrs	All	
\$35K+	33.3	11.1	22.7	66.7	32.8	
\$17.5K-\$	16.7	50.0	45.5	22.2	37.5	
<\$17.5K	50.0	38.9	31.8	11.1	29.7	
100%=	6	18	22	18	N =	64

Race = Black
Region = South

	<12yrs	12yrs	13-15yrs	16+yrs	All	
\$35K+	11.1	21.4	22.6	42.4	25.8	
\$17.5K-\$	27.8	33.3	48.4	42.4	38.7	
<\$17.5K	61.1	45.2	29.0	15.2	35.5	
100%=	18	42	31	33	N =	124

Race = Black
Region = West

	<12yrs	12yrs	13-15yrs	16+yrs	All	
\$35K+	.0	.0	.0	83.3	27.8	
\$17.5K-\$	100.0	50.0	75.0	.0	44.4	
<\$17.5K	.0	50.0	25.0	16.7	27.8	
100%=	2	6	4	6	N =	18

15. Next I standardize on race and region.

Status98 (Status attainment model; source: '96 & '98 GSS; Full-time workers only)

File

Causal order:

Race -> Regio -> Inco* -> Ed* -> Sex* -> Age* -> Marr* -> Papr* -> Sibs*
 2 x 4 x 3 x 4 x 2 x 3 x 3 x 3 x 3

N = 2195

16. And then percentage the standardized table.

	<12yrs	12yrs	13-15yrs	16+yrs	All
\$35K+	16.9	25.0	32.7	53.0	37.3
\$17.5K-\$	40.8	47.5	45.8	35.4	41.7
<\$17.5K	42.3	27.5	21.5	11.6	21.0
100%=	174	645	492	883	N = 2195

17. Now let's look at the income/race association—uncontrolled:

Income/Race

	White	Black	All
\$35K+	38.6	27.0	37.3
\$17.5K-\$	41.8	41.1	41.7
<\$17.5K	19.6	31.9	21.0
100%=	1947	248	N = 2195

18. Controlling for education:

Ed = <12yrs

	White	Black	All
\$35K+	15.8	17.2	16.0
\$17.5K-\$	43.2	31.0	41.1
<\$17.5K	41.1	51.7	42.9
100%=	146	29	N = 175

Ed = 12yrs

	White	Black	All
\$35K+	25.4	17.1	24.3
\$17.5K-\$	48.8	41.5	47.8
<\$17.5K	25.9	41.5	27.9
100%=	564	82	N = 646

Examples of Direct Standardization

Ed = 13-15yrs

	White	Black	All
\$35K+	34.3	22.1	32.6
\$17.5K-\$	45.1	48.5	45.5
<\$17.5K	20.7	29.4	21.9
100%=	426	68	N = 494

Ed = 16+yrs

	White	Black	All
\$35K+	54.3	47.8	53.8
\$17.5K-\$	35.0	37.7	35.2
<\$17.5K	10.7	14.5	11.0
100%=	811	69	N = 880

19. Standardizing on education:

Status98 (Status attainment model; source: '96 & '98 GSS; Full-time workers only)

File

Causal order:

Ed -> Inco* -> Race* -> Regi* -> Sex* -> Age* -> Marr* -> Papr* -> Sibs*
 4 x 3 x 2 x 4 x 2 x 3 x 3 x 3 x 3

N = 2195

20. Percentaging the standardized association:

	White	Black	All
\$35K+	38.2	30.0	37.3
\$17.5K-\$	41.8	40.9	41.7
<\$17.5K	19.9	29.1	21.0
100%=	1951	244	N = 2195