Examples of Standard Error Adjustment

Obtaining a Statistic Using Both SRS and Complex Survey Methods in SPSS

This resource document will provide you with an example of the analysis of a variable in a complex sample survey dataset using SPSS. A subset of the public-use version of the Early Child Longitudinal Studies ECLS-K rounds one and two data from 1998 accompanies this example, as well as an SPSS syntax file. The stratified probability design of the ECLS-K requires that researchers use statistical software programs that can incorporate multiple weights provided with the data in order to obtain accurate descriptive or inferential statistics.

Research question

This dataset training exercise will answer the research question “Is there a difference in mathematics achievement gain from fall to spring of kindergarten between boys and girls?”

Step 1- Get the data ready for use in SPSS

There are two ways for you to obtain the data for this exercise. You may access a training subset of the ECLS-K Public Use File prepared specifically for this exercise by clicking here, or you may use the ECLS-K Public Use File (PUF) data that is available at

http://nces.ed.gov/ecls/dataproducts.asp.

If you use the training dataset, all of the variables needed for the analysis presented herein will be included in the file. If you choose to access the PUF, extract the following variables from the online data file (also referred to by NCES as an ECB or “electronic code book”):

CHILDID	CHILD IDENTIFICATION NUMBER
C1R4MSCL	C1 RC4 MATH IRT SCALE SCORE (fall)
C2R4MSCL	C2 RC4 MATH IRT SCALE SCORE (spring)
GENDER	GENDER
BYCW0	BASE YEAR CHILD WEIGHT FULL SAMPLE
BYCW1 through C1CW90	BASE YEAR CHILD WEIGHT REPLICATES 1 through 90
BYCWSTR	BASE YEAR CHILD STRATA VARIABLE
BYCWPSU	BASE YEAR CHILD PRIMARY SAMPLING UNIT

Export the data from this ECB to SPSS and be sure to name your file, ‘ECLSK_c1c2_panel_demo’. Finally, download the SPSS syntax file prepared for this exercise by clicking here.

Step 2- Use SPSS to calculate an estimate and accompanying standard error

Start SPSS, then open the training SPSS data file and the corresponding syntax file, 'Practice_SPSS_Analyses.sps'. Below is a screen shot of the typical SPSS Statistics Syntax Editor showing the training syntax file. All the code that is in that syntax file is also commented below.

First, let's explore the descriptive statistics of the training dataset by running the following syntax:

```
DESCRIPTIVES VARIABLES=gender bycw0 c1r4rscl c1r4mscl c2r4rscl c2r4mscl bycwstr
bycwpsu bycw1
  /STATISTICS=MEAN STDDEV MIN MAX.
```

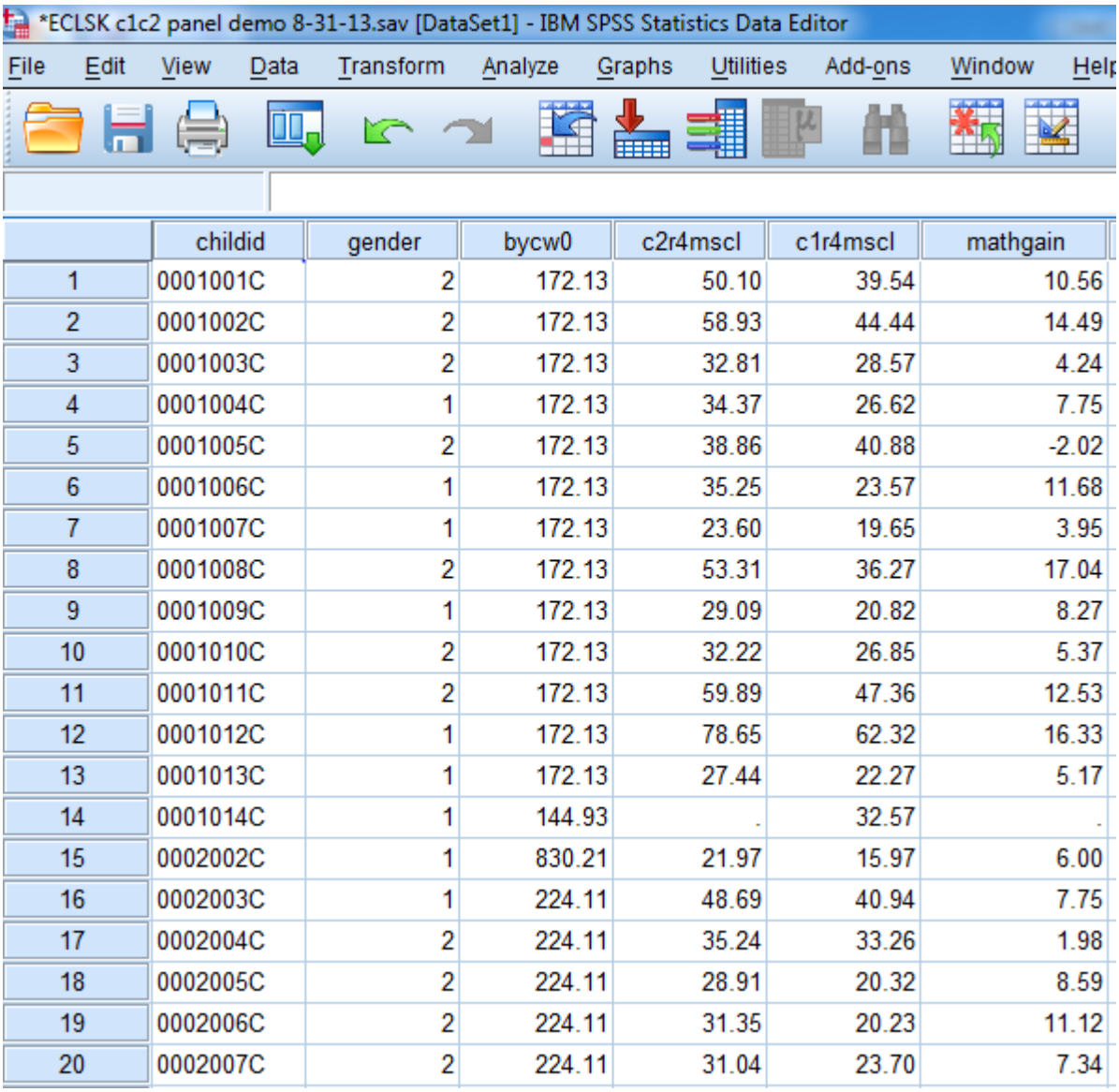
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
gender	21396	1	2	1.49	.500
c1c2 child panel weight full sa	21192	.00	900.00	182.2954	133.11527
c1 rc4 reading irt scale score	17622	21.01	138.51	35.2145	10.19878
c1 rc4 math irt scale score	18636	10.51	115.65	25.9054	9.09918
c2 rc4 reading irt scale score	18937	22.23	156.85	46.4586	14.03521
c2 rc4 math irt scale score	19649	11.57	113.80	36.2733	12.00449
c1c2 c panel wt taylor series s	18211	1	89	51.55	27.038
c1c2 c panl wt taylor ser prim	18211	1	80	5.80	11.974
c1c2 child panel weight replica	21192	.00	1349.80	182.5282	137.52066
Valid N (listwise)	16724				

Next create the math gain score variable that will be used in the analysis.

```
COMPUTE mathgain=c2r4mscl-c1r4mscl.
EXECUTE.
```

Referring back to the SPSS Statistics Data Editor, you can examine the data and check that the computed variable looks as expected. In this screen shot, the variables have been reordered to show the two math scores used to create the math gain score variable alongside of 'mathgain' for the first 20 cases in the training dataset.



	childid	gender	bycw0	c2r4mscl	c1r4mscl	mathgain
1	0001001C	2	172.13	50.10	39.54	10.56
2	0001002C	2	172.13	58.93	44.44	14.49
3	0001003C	2	172.13	32.81	28.57	4.24
4	0001004C	1	172.13	34.37	26.62	7.75
5	0001005C	2	172.13	38.86	40.88	-2.02
6	0001006C	1	172.13	35.25	23.57	11.68
7	0001007C	1	172.13	23.60	19.65	3.95
8	0001008C	2	172.13	53.31	36.27	17.04
9	0001009C	1	172.13	29.09	20.82	8.27
10	0001010C	2	172.13	32.22	26.85	5.37
11	0001011C	2	172.13	59.89	47.36	12.53
12	0001012C	1	172.13	78.65	62.32	16.33
13	0001013C	1	172.13	27.44	22.27	5.17
14	0001014C	1	144.93	.	32.57	.
15	0002002C	1	830.21	21.97	15.97	6.00
16	0002003C	1	224.11	48.69	40.94	7.75
17	0002004C	2	224.11	35.24	33.26	1.98
18	0002005C	2	224.11	28.91	20.32	8.59
19	0002006C	2	224.11	31.35	20.23	11.12
20	0002007C	2	224.11	31.04	23.70	7.34

SPSS data analysis under the different assumptions

For comparison purposes, you will first run the analysis as if this data were SRS, that is, a simple random sample ***with no weight*** adjustments for sampling design or nonresponse. In this first run, you will not apply any weight. In the second run, you will repeat a standard analysis (assuming SRS) ***with the main sampling weight***.

To complete the correct analysis using SPSS, you would then conduct a third run using one of the analytic options presented within Step 3 to calculate appropriate standard errors that will give you more useful and accurate results when conducting significance testing or in creating confidence intervals in subsequent analysis steps.

First, we will calculate simple descriptive statistics, the average math score gain of all children and then again, by gender.

```
EXAMINE VARIABLES=mathgain BY gender
  /PLOT NONE
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.
```

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	gender	N	Percent	N	Percent	N	Percent
mathgain	male	9000	82.2%	1950	17.8%	10950	100.0%
	female	8702	83.3%	1744	16.7%	10446	100.0%

Descriptives

	gender			Statistic	Std. Error
mathgain	male	Mean		10.5263	.07527
		95% Confidence Interval for Mean	Lower Bound	10.3787	
			Upper Bound	10.6738	
		5% Trimmed Mean		10.1652	
		Median		9.5750	
		Variance		50.996	
		Std. Deviation		7.14117	
	female	Mean		10.1763	.06836
		95% Confidence Interval for Mean	Lower Bound	10.0423	
			Upper Bound	10.3103	
		5% Trimmed Mean		9.9048	
		Median		9.4350	
		Variance		40.667	
		Std. Deviation		6.37705	

*The Descriptives table shown here has been truncated to fit the page. T-Test results not shown.

The output above indicates that the average math score gain for boys is estimated as 10.53 with a standard error of 0.075. The average math score for girls is estimated as 10.18 with a standard error of 0.068. The answer to our main question about whether the difference of 0.35 in the gain scores of boys and girls depends on the accuracy of the mean gain scores and of these standard errors. If you run a t-test on these data, it will indicate that the difference is statistically significant.

```
T-TEST GROUPS=gender(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=mathgain
  /CRITERIA=CI(.95).
```

However, the method shown above of estimating the average gain scores is misleading. Even in SRS analyses, when we have a main sampling weight, we must apply it.

Statistical Analysis of NCES Datasets Employing a Complex Sample Design > Examples > Slide 11 of 13

Next, run the same analysis using the main sampling weight, bycw0.

```
WEIGHT BY bycw0.
EXAMINE VARIABLES=mathgain BY gender
  /PLOT NONE
  /STATISTICS DESCRIPTIVES
  /CINTERVAL 95
  /MISSING LISTWISE
  /NOTOTAL.

T-TEST GROUPS=gender(1 2)
  /MISSING=ANALYSIS
  /VARIABLES=mathgain
  /CRITERIA=CI(.95).
```

Case Processing Summary

		Cases					
		Valid		Missing		Total	
	gender	N	Percent	N	Percent	N	Percent
mathgain	male	1959828	98.4%	31440	1.6%	1991268	100.0%
	female	1844869	98.7%	23601	1.3%	1868471	100.0%

Descriptives

	gender			Statistic	Std. Error
mathgain	male	Mean		10.3680	.00507
		95% Confidence Interval for Mean	Lower Bound	10.3581	
			Upper Bound	10.3779	
		5% Trimmed Mean		10.0124	
		Median		9.4500	
		Variance		50.300	
		Std. Deviation		7.09223	
	female	Mean		10.1539	.00471
		95% Confidence Interval for Mean	Lower Bound	10.1446	
			Upper Bound	10.1631	
		5% Trimmed Mean		9.8805	
		Median		9.4100	
		Variance		40.962	
		Std. Deviation		6.40013	

*The Descriptives table shown here has been truncated to fit the page. T-Test results not shown.

The weighted analysis above shows that the average estimated mathematics score gain for boys has been reduced to 10.37 (with a decrease in the standard error from 0.075 to 0.005) as compared to the unweighted estimate. The estimate for girls is reduced slightly to 10.15, and the standard error has decreased from 0.068 to 0.005. Thus the gender difference is now 0.21, but a t-test run using the full sample weight gives a significance result similar to the unweighted version run previously.

This method is also misleading because the complex sample design has not been taken into account to calculate the correct standard errors that accompany the properly weighted estimates.

Step 3- Consider your next analytic steps within SPSS or another software package

To complete the correct analysis using SPSS, analysts have two options for taking the complex sampling designs of NCES surveys into account: the SPSS Complex Samples add-on or the SPSS R plugin. It is important to note that these analysis features are not included in the SPSS Statistics Core system.

Users who purchase the Complex Samples add-on will be able to use PSU and strata variables included in NCES datasets to estimate variance using Taylor series linearization methods exclusively. Accordingly, both PSU and strata variables are provided within the training dataset for your analytic purposes.

Users of NCES data who wish to use replication techniques to conduct their analysis may either (a) use the SPSS R plugin which will allow you to access the R survey package, or (b) consult a resource document detailing another statistical package that will incorporate multiple weights provided within the dataset in order to obtain accurate descriptive or inferential statistics and their corresponding standard errors.