Weighting and Adjusting Weights: An Example[1]

The following pages show the impact of weighting data in SPSS with regression and descriptive statistics. For more information about weights and sampling, please see:

http://staff.washington.edu/glynn/samples_weights.pdf

It may also be helpful to read **"Statistical Significance: A non-mathematical explanation"**, which you will find at:

http://staff.washington.edu/glynn/statistical_difference.pdf

The entire program and the data that were used for this example can be found at:

http://staff.washington.edu/glynn/adjweight2.sps

The program creates a small data set (10 cases) with two variables. The regression procedure is used on this data set showing a positive but statistically insignificant relationship between the two variables. A weight of 100,000 created and applied. This makes each case represent 100,000 cases. The regression procedure is run. The coefficients are the same, but the standard errors are reduced, the T values are increased, and the relationship now appears to be strongly significant. This apparently statistical significance is false for the actual sample size of 10, and is an artifact of weighting. Because SPSS does not "normalize" weights for regression, the standard errors and T values are calculated using an N of 1,000,000 (10*100,000 = 1,000,000) instead of the actual N of 10.

To avoid this serious problem, weights must be adjusted so that the weighted N is the same as the unweighted N. The mean of weights should be 1. For information about how adjust weights, see:

Adjusting or Normalizing Weights "On the Fly" in SPSS

http://staff.washington.edu/glynn/adjspss.pdf

Adjusting, or Normalizing Weights "On the Fly" in SAS

http://staff.washington.edu/glynn/adjustsa.pdf

[1]Prepared by Patty Glynn, University of Washington. 3/26/04 C:\all\help\helpnew\adjweight2.wpd

```
** adjweight2.sps .
** Demonstrate impact of failing to adjust weights in SPSS with regression .
list var = all .

List
       y       x1

     .00     4.00
    1.00     8.00
    2.00     8.00
    3.00     5.00
    4.00     3.00
    5.00     4.00
    6.00     3.00
    7.00     7.00
    8.00     9.00
    9.00     9.00

Number of cases read:  10    Number of cases listed:  10

** Run a regression with weight off .
** There is a positive but statistically insignificant relationship
** between Y and X1.
weight off  .
```

title 'Regression and Descriptives with weight off' .

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT y
  /METHOD=ENTER x1 .
```

Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.315(a)	.099	-.014	3.04822

a Predictors: (Constant), x1

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.167	1	8.167	.879	.376(a)
	Residual	74.333	8	9.292		
	Total	82.500	9			

a Predictors: (Constant), x1
b Dependent Variable: y

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.167	2.669		.812	.440
	x1	.389	.415	.315	.938	.376

a Dependent Variable: y

title 'Regression and Descriptives with weight off' .

desc var = Y X1 .

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
y	10	.00	9.00	4.5000	3.02765
x1	10	3.00	9.00	6.0000	2.44949
Valid N (listwise)	10				

```
** Create a fake weight for this example.
** For this simple example, each case represents 100,000 cases .

compute fakewght = 100000 .
** Turn weight on .
weight by fakewght .
```

title 'Regression and Descriptives with weight on' .

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT y
  /METHOD=ENTER x1 .
```

Regression

Variables Entered/Removed(b)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.315(a)	.099	.099	2.72642

a Predictors: (Constant), x1

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	816666.667	1	816666.667	109865.251	.000(a)
	Residual	7433333.333	999998	7.433		
	Total	8250000.000	999999			

a Predictors: (Constant), x1
b Dependent Variable: y

title 'Regression and Descriptives with weight on' .

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.167	.008		287.010	.000
	x1	.389	.001	.315	331.459	.000

a Dependent Variable: y

desc var = Y X1 .

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
y	1000000	.00	9.00	4.5000	2.87228
x1	1000000	3.00	9.00	6.0000	2.32379
Valid N (listwise)	1000000				

```
** Normalize weights.
weight off .
compute adwgt = fakewght / 100000 .
** Weight by the adjusted weight .
weight by adwgt .
```

title 'Regression and Descriptives, weighting by adjusted weights' .

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT y
  /METHOD=ENTER x1 .
```

title 'Regression and Descriptives, weighting by adjusted weights' .

Regression

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.315(a)	.099	-.014	3.04822

a Predictors: (Constant), x1

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	8.167	1	8.167	.879	.376(a)
	Residual	74.333	8	9.292		
	Total	82.500	9			

a Predictors: (Constant), x1
b Dependent Variable: y

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.167	2.669		.812	.440
	x1	.389	.415	.315	.938	.376

a Dependent Variable: y

desc var = Y X1 .

Descriptives

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
y	10	.00	9.00	4.5000	3.02765
x1	10	3.00	9.00	6.0000	2.44949
Valid N (listwise)	10				