

Soc 63993, Advanced Social Statistics II
Homework No. 8
Structural Coefficients/R²

Problems 1 and 2. The file *evilstd.sps* will generate the computer runs you need for this problem. Copy it from my web page.

This program contains two examples (which you will treat as problem 1 and problem 2). In each example, two regressions are run. Indicate what each example tells you about the problems that can arise if you focus on R² and on standardized (path) rather than metric (structural) coefficients. Do your best to explain why these problems occur.

Here is a copy of the program:

```
* Evils of standardization:
* This program illustrates problems with the use of standardized
* variables and standardized coefficients.

* Hypothetical data: a sample of 1200 blacks and 1200 whites. Assume
* that equal-sized samples of blacks and whites have been drawn, but
* that in this population, whites outnumber blacks by 8 to 1. Hence,
* blacks have been oversampled by a factor of 8, and when the groups
* are analyzed together, each black should therefore be weighted only
* 1/8 as heavily as each white. In this hypothetical example,
* all individuals conveniently have incomes of $3000, $3500, $4000,
* $4500, $5000, or $5500, hence cases are grouped together.

* Or, as an alternative, assume that 2 populations have been sampled
* from. In one population, whites outnumber blacks by 8 to 1, whereas
* in the other population it is a 50-50 split.

* The vars are:
* Ncases - the number of cases having a particular combination of
* values
* Inc - Income
* Race - 0 = Black, 1 = White
* Inc2 - Income with random error. The change in income conveniently
* works out to be $1100 either way. The variability in income
* increases as a result.

DATA LIST FREE / Ncases Inc Race Inc2.
BEGIN DATA.
200 3000 0 1900
200 3000 0 4100
200 3500 1 2400
200 3500 1 4600
200 4000 0 2900
200 4000 0 5100
200 4500 1 3400
200 4500 1 5600
200 5000 0 3900
200 5000 0 6100
200 5500 1 4400
200 5500 1 6600
END DATA.
```

```

List  Variables = Ncases Inc Race Inc2.

* Compute the correct weight.  We weight each case by Ncases;
* then, weight the Black cases only 1/8 as heavily as each white
* case; then, since sample size has been deflated, inflate the weights
* for all cases by a constant factor so that the total sample size
* is correct.
Compute      WgtRight = Ncases.
If           (Race = 0) WgtRight = Ncases/8.
Compute      WgtRight = WgtRight * 2400/(1200 + 1200/8).

* The incorrect weight ignores the oversampling of blacks, and weights
* all cases equally.
Compute      WgtWrong = Ncases.

*
*           E X A M P L E   1
* What effect does improper weighting of cases have on
* coefficients in a perfectly specified model? .

* First, use the correct weight.
WEIGHT BY WgtRight.
REGRESSION /VARIABLES Inc Race /DESCRIPTIVES DEFAULT
           /DEPENDENT Inc /METHOD ENTER Race.

* Now, repeat the analysis using the incorrect weight.  What differences
* occur?.

Weight by WgtWrong.
REGRESSION /VARIABLES Inc Race /DESCRIPTIVES DEFAULT
           /DEPENDENT Inc /METHOD ENTER Race.

*
*           E X A M P L E   2
* Suppose variability in the dependent variable increases on a random
* basis; that is, the residual variance that cannot be explained by
* other variables goes up.  What effect does an increase in residual
* variance have on metric and standardized coefficients?

* To examine this, we compare Inc and Inc2.  Inc2 adds random variation
* to Inc.

* Use the correct weight.
Weight by WgtRight.
REGRESSION /VARIABLES Inc Race /DESCRIPTIVES DEFAULT
           /DEPENDENT Inc /METHOD ENTER Race/
REGRESSION /VARIABLES Inc2 Race /DESCRIPTIVES DEFAULT
           /DEPENDENT Inc2 /Method enter race.

```

Problem 3. Present a substantive example, real or hypothetical, in which the value of R^2 may be high but inaccurate or misleading, e.g. the model is mis-specified in some way. If possible, cite a real example from the published literature, but otherwise make one up. If hypothetical, try to make the example “reasonable,” i.e. the model might at first glance seem to make sense but on closer examination it really isn’t correct.