scorocs®

Prove and improve the good you do.™

How to Transform Regression Coefficients to Integer Weights So Scores Go from 0 to 100

Mark Schreiner

13 August 2022

This note uses a simple example to show how to transform coefficients from a regression on categorical indicators to non-negative integers whose sum for a given case ranges from 0 to a desired maximum (such as 100).

In addition, the lowest value for a given indicator is always linked with a weight of 0, and weights for the other values of the given indicator increase.

Algorithm

A scorecard is based on a regression (e.g., Logit or OLS) which is linear in the parameters and also linear in the indicators:

$$y = \alpha \cdot 1 + \beta \cdot X + \varepsilon.$$

Here, y is the dependent variable, α is the coefficient on the constant 1, β is a vector of estimated coefficients (in general, some negative, some positive, and all with many decimal places), and X is a conformable vector of indicators. Suppose you want the transform the coefficients into scorecard points such that:

- All points are zeroes or positive integers
- The lowest possible score is zero
- The highest possible score is some positive integer h

The K indicators are indexed from 1 to k. Each indicator has k_J categories indexed from 0 to k_J. The indicators may be represented by 0/1 flags or by one variable whose values are integers from 0 to k_J. Either way, the first category (the one coded zero) is assumed to be omitted from the regression, prevent perfect collinearity.

The coefficient on category j of indicator k is β_{kj}. The regression coefficient on the omitted category is zero by definition, so, $\beta_{k0} = 0$ for all k.

For an indicator *k*, the minimum coefficient β_{kmin} is defined as Min(β_{k1}, β_{k2}, . . . , β_{kj}).

The shifted regression coefficient γ_{kj} is defined as $\beta_{kj} - \beta_{kmin}$. While the minimum of β_{kj} could be positive, negative, or zero, the minimum of γ_{kj} is always zero.

Let the maximum shifted coefficient for each category γ_{kmax} be Max(γ_{k1}, γ_{k2}, . . . , γ_{kJ}).

Finally, define the transformed scorecard points p_{kj} for category *j* of indicator *k* as:

$$p_{kj} = \text{Round}\left(\frac{\gamma_{kj} \cdot h}{\sum_{k=1}^{K} \gamma_{k\max}}, 0 \right),$$

where the function Round(*x*, 0) returns the integer nearest to *x*.

Given that only one category is relevant for each indicator when a given case is scored, the score (sum of points) has a minimum of 0 and a maximum of *h*.

Due to the Round(·) function, the maximum may not always be exactly *h*. In that case, the find a category of an indicator that:

- Occurs for a relatively infrequently, and
- Has a relatively high point value

Then adjust the point value for this category up or down as required to make the maximum score be *h*. Without this step, users will question why the maximum score is not exactly *h*, even though it does not, in fact, matter for prediction.

Example

Suppose that there are three indicators ($K = 3$) and that the first has two categories ($k_1 = 2$) and that the second has four ($k_2 = 4$). The desired score range is zero to $h = 100$. The estimated regression equation is:

$$y = -2.314 \cdot 1 + 0 \cdot x_{11} - 2 \cdot x_{12} + 3 \cdot x_{13} + 0 \cdot x_{21} - 3 \cdot x_{22} + 1 \cdot x_{23} + 5 \cdot x_{24}.$$

The regression coefficients (ignoring the coefficient on the constate term) are:

$\beta_{11} = 0$

$\beta_{12} = -2$

$\beta_{13} = 3$

$\beta_{21} = 0$

$\beta_{22} = -3$

$\beta_{23} = 1$

$\beta_{24} = 5$.

The minimum coefficients for each indicator are:

$\beta_{1min} = \text{Min}(0, -2, 3) = -2$,

$\beta_{2min} = \text{Min}(0, -3, 1, 5) = -3$.

The shifted coefficients are then:

$\gamma_{11} = 0 - (-2) = 2$

$\gamma_{12} = -2 - (-2) = 0$

$\gamma_{13} = 3 - (-2) = 5$

$\gamma_{21} = 0 - (-3) = 3$

$\gamma_{22} = -3 - (-3) = 0$

$\gamma_{23} = 1 - (-3) = 4$

$\gamma_{24} = 5 - (-3) = 8$.

The maximum shifted coefficients for each indicator are:

$\gamma_{1max} = \text{Max}(2, 0, 5) = 5$,

$\gamma_{2max} = \text{Max}(3, 0, 4, 8) = 8$.

Note that $\gamma_{1max} + \gamma_{2max} = 13$.

The transformed scorecard points p_{kj} are then:

$$p_{11} = \text{Round}(2 \div 13 \cdot 100, 0) = 15,$$
$$p_{12} = \text{Round}(0 \div 13 \cdot 100, 0) = 0,$$
$$p_{13} = \text{Round}(5 \div 13 \cdot 100, 0) = 38,$$
$$p_{21} = \text{Round}(3 \div 13 \cdot 100, 0) = 23,$$
$$p_{22} = \text{Round}(0 \div 13 \cdot 100, 0) = 0,$$
$$p_{23} = \text{Round}(4 \div 13 \cdot 100, 0) = 31,$$
$$p_{24} = \text{Round}(8 \div 13 \cdot 100, 0) = 62.$$

The minimum score is when $x_1 = 2$ and $x_2 = 2$, for a score of zero.

The maximum score is when $x_1 = 3$ and $x_2 = 4$, for a score of 100.

Results

Non-specialists probably will not understand a scorecard presented like this:

Variable		Estimate
Intercept	1	-5.5784
id_dist_ind	1	0.6154
id_dist_ind	2	1.2459
id_dist_ind	3	1.7190
id_dist_ind	4	2.0310
id_dist_ind	5	2.4196
aga_0099	1	0.7031
aga_0099	2	1.1111
aga_0099	3	1.6074
aga_0099	4	2.4801
aga_0099	5	3.4177
aga_0099	6	3.9685
aga_0099	7	5.4029
res_wall	1	0.4054
res_wall	2	0.9182
res_cook	1	0.3209
res_toil	1	0.3603
ass_matt	1	0.0941
ass_matt	2	0.3631
ass_matt	3	0.6463
ass_tv	1	1.1942
ass_cell	1	0.8911
ass_cell	2	1.4228
ag_ry	1	0.4936

They are more likely to understand a scorecard presented like this:

Question	Response	Points
1. In which district does the household live? (*record without asking*)	A. Kenema, or Pujehun	0
	B. Tonkolili, or Bombali	5
	C. Freetown (western area, rural), Port Loko, Bo, Moyamba, Karene, Falaba, or Koinadugu	9
	D. Kono, or Bonthe	13
	E. Freetown	15
	F. Kailahun, or Kambia	18
2. How many members does the household have?	A. Nine or more	0
	B. Eight	5
	C. Seven	8
	D. Six	12
	E. Five	19
	F. Four	26
	G. Three	30
	H. One or two	41
3. What is the main construction material of the outside walls of the household's dwelling? (*By observation; ask if in doubt*)	A. Wattle and mud, or mud bricks	0
	B. Mud bricks plastered with cement, or other	3
	C. Cement blocks, wooden boards, or corrugated iron/zinc sheets	7
4. Does the household cook mostly with charcoal, cooking gas, or electricity?	A. No	0
	B. Yes	2
5. What kind of toilet does the household use?	A. None (bush, field, waterside), hanging toilet/hanging latrine, or composting toilet	0
	B. Pit latrine without slab (open pit) or with slab, ventilated improved pit latrine (VIP), flush (to pit latrine, septic tank, or piped sewer system)	3
6. How many mattresses does the household have?	A. None	0
	B. One	1
	C. Two	3
	D. Three or more	5
7. Does the household have a television?	A. No	0
	B. Yes	9
8. How many mobile phones does the household have?	A. None	0
	B. One	7
	C. Two or more	11
9. In the past 12 months, did the household grow rice or cassava for its own consumption?	A. Yes	0
	B. No	4