Multiple Regression with WINSTEPS
A Rasch Solution to Regression Confusion

Benjamin D. Wright, University of Chicago

The purpose of Rasch measurement is to build and verify a useful "yardstick" – a stable, portable, reproducible instrument for making linear measures. What makes a yardstick useful is the calibration of its reference points, which mark out a visible linear metric that maintains its spacing as long as it is used in a sensible way. Just like the yardstick in your closet, a useful "yardstick" does not change the distances between its calibration marks from object to object, place to place, or time to time, as long as you apply it as sensibly as you apply the yardstick in your closet.

What follows explains how to use the Rasch measurement program, WINSTEPS (winsteps.com, Linacre, 2000) to solve multiple regression problems in a new way that avoids the sample covariance dependence and missing data problems which interfere with inferential stability.

The data used by Rasch measurement to build yardsticks can originate as any set of ordinal indicators: dichotomies, ratings, partial credits, counts, as well as any already developed metric like those in commerce and science. The way data expressed in the decimal fractions of an existing metric, like inches or mg/Dl, is entered into WINSTEPS is:

1. Recode each decimal fraction X into interval (or log interval) integer Y:

$$Y = M(X-MIN)/(MAX-MIN) + 1/2$$

For Y = 0,9 use M = 0,N <10
For Y= 00,99 use M = 00,NN <100

If an incoming metric is expected to have a ratio effect in the yardstick you are constructing, you can anticipate this by using log X instead of X in the above formula.

2. Your choice of MIN and MAX can be made locally from the smallest incoming value for MIN and the largest incoming value for MAX. Or you can choose values for MIN and MAX which are natural to their originating metric, so long as the values you choose embrace the range of incoming data.

3. Your choice of M depends on how many ordinal integer categories you wish to use for your yardstick construction. In our practice we have not encountered any situation for which M>9 was more informative than M<10, but WINSTEPS does enable you to maintain the linear articulation of your incoming decimal fractions up to 100 steps from 00 to 99 by setting M=99.

4. In order for WINSTEPS to print the original decimal fractions Y of your incoming data next to their recoded integers X, use control variable CFILE= to label each integer category X with its corresponding decimal fraction midpoint Y:

$$X = [(Y-1/2)(MAX-MIN)/M] + MIN$$
(for log interval X = exp X)

This labeling enables you to see the extent to which the scale of decimal fractions X, however linear in X, does not make a linear contribution to your new yardstick. It is often the case that the linear intervals of X do not produce linear separations among their code values Y in the new metric defined by your yardstick.

5. If any of your decimal fraction variables X have useful reference points, such as freezing at 32 degrees or normal body temperature at 98.6 degrees, you can reference your item calibration representations of these variables by pivoting the calibration of the equivalent Y integer step at that reference point.

How to Use WINSTEPS to Solve Multiple Regression Problems

1. Organize your incoming variables into three groups:

Dependent Variables = DV to be predicted by IV
Independent Variables = IV to predict DV's.
Conditional Variables = CV to condition prediction for interaction with other variables like: gender, age, culture, language, wealth . . .

2. Apply one of the following three "regression" formulations:

a. DV positioned on the IV set in terms of a *DV defined variable*.

Anchor objects (persons) at their incoming DV values. This establishes your dependent variable. Then use WINSTEPS to find the best set of IV calibrations for predicting these anchored DV values. This formulation optimizes prediction, but binds IV calibrations to DV sample dependence.

b. DV positioned on IV in terms of an *IV defined variable.*

Apply WINSTEPS to the IV set to find the best variable this IV set can define, independent of any DV. This requires a sequence of stepwise analyses by which members of the IV set are edited until a best possible IV variable has become defined (The steps are listed below).

Because the construction of the IV variable is entirely independent of DV data, this formulation enables the simultaneous evaluation of any number of DV's.

Anchor to the item/step calibrations of this best IV variable and then insert all DV's and use WINSTEPS to show how well these DV's are predicted by the IV just constructed independently of any DV distribution and also of any sample dependent covariance among the IV. This sample free construction of a single variable defined by the IV set optimizes the inferential stability of DV predictions.

c. Middle Ground *Short Cut.*

Combine all DV and IV in one WINSTEPS analysis. The result will fall between formulations (a) and (b). But they will be dominated by (b) to the extent that IV information exceeds DV information.

3. Two ways to introduce CV variables.

a. Several Separate Analyses.

For CV's with few categories, repeat Step 2 for each CV sub-group. Compare maps.

b. Sequence of Composite Analyses

Include CVs in each analysis and use person separations, fit statistics and residual analyses to expose the extent to which each CV interferes (or helps).

How to Construct a Best IV Variable

1. Item Polarity: Examine the correlations between item responses and person measures in the Item Misfit Table to identify and correct all negative relations by reversing their scoring.

2. Category Articulation: Examine the Rating Scales Structure Table to identify noisy and uninformative categories that you can improve by rescoring these categories.

3. Item Dimensionality: Examine the Item Principal Component Analysis of Response Residuals Table to find out whether there is a secondary item dimension large enough or meaningful enough to isolate.

4. Person Dimensionality: If the relative size or item content of the first item residual factor interests you, examine the Person Principal Component Analysis of Response Residuals Table to identify and evaluate the effect of this secondary dimension on person measures.

5. Variable Sharpening: Reexamine the Item Misfit Table to evaluate the effects on person separation (in the Summary Table) of deleting items with large infit mean squares (e.g.>1.3) in order to find the most efficient definition of your IV variable.

How WINSTEPS Improves on Multiple Regression MR

1. MR arithmetic and stochastic interpretation depends on normally distributed continuous linear data.

WINSTEPS accepts discrete ordinal data of any distribution and constructs linear continuous measures from them. Every analysis of raw ordinal observations requires this step to prepare for linear statistical analysis.

2. MR is vulnerable to missing data.

When rows and columns are connected, WINSTEPS conjoint additivity corrects for missing data automatically.

3. MR posits a single dimensioned DV to which the IV, whatever their own dimensions, must produce a co-linear contribution.

WINSTEPS extracts the best possible single linear dimension, which the data support and estimates continuous linear measures, standard errors and fit statistics on this dimensions for all item, step and person parameters.

4. MR regression coefficients and multiple R's are hard to interpret because they defy visualization.

WINSTEPS constructs linear measures, qualified by errors and fit statistics and reports them on linear MAPs which show, in complete detail, the positional relationship between all values of the DV in terms of all values of the IV. The resulting positional relationships are complete, easy to see and easy to interpret.

A Gout Application of WINSTEPS Regression

Table 1 shows three panels from the application of WINSTEPS Rasch regression to the medical data discussed in the article, "Rasch Measurement Instead of Regression" by Wright, Perkins and Dorsey.

The top panel lists, on the right, the eight definitions of the medical items analyzed. The next column to the left, "SCORE CORR." are response/measure correlations. For the three anchored blood measures which define the independent variable, IV, these correlations correspond to standardized regression coefficients. For the five dependent variables, DV, listed below, they correspond to the usual multiple regression prediction correlations.

Next to the left are two columns of mean square fit statistics. When these mean squares are near 1.00, they document a valid relationship among the three anchored blood chemistry IV's: Uric Acid, Urea Nitrogen and Creatinine. They also validate or invalidate the regressions on the three blood chemistry IV of the five DV's: Gout, Hypertension, Diuretics, Kidney Stones and Diabetes.

Gout does best with a prediction correlation of .61, closely followed by hypertension. Both correlations and outfit mean squares expose the failure of this three blood chemistry IV to predict kidney stones or diabetes.

The middle panel of Table 1 illustrates the IV linear coding of chemical metrics mg/dl onto 10 category integer scales, 0 to 9. It also shows the pivot marking for each blood chemistry at the Merck Manual step from "normal" to "high". The middle panel also lists the distribution of the 96 patients across the 10 levels for each IV and the average measure at each category of the new medical variable which the three blood chemistries were found to define. Since the chemical mg/dl metrics are evenly represented by the 10 categories, the non-linear distributions of these average measures is noteworthy. This non-linearity shows that the medical implications of increases in these blood chemistries are not collinear with increases in their chemical metric mg/dl. The clinical implications of a particular increment in mg/dl varies with mg/dl level. This irregularity muddies clinical evaluations of blood chemistry changes. The WINSTEPS analysis makes the specifics of this non-linearity evident and provides, instead, a new medical metric, which is linear in its clinical implications.

The bottom panel of Table 1 sums up the diagnostic implications of these analyses. The multiple regression prediction correlations, repeated from above, show that Gout at .61 is better predicted than hypertension at .51. Far more useful, however, are the measurement positions of each diagnostic indicator. The gout indicators, rounded to 48 for "No Gout" and 59 for "Gout", mark the positions on the three blood chemistry yardstick where the odds for the presence of gout shift from 1/2 at 48 to 2/1 at 59. The hypertension indicators, rounded to 49 and 59, provide a similar interpretation with respect to hypertension. At the bottom we see again, in metric form, the futility of trying to predict kidney stones or diabetes from these three blood chemistries.

When the 12 unit distance (59.44 - 47.70 = 11.74) between the gout indicators is compared to the 10 unit distance (58.83 – 48.93 = 9.90) between the hypertension indicators, we see the metric implications of their .61 > .51 multiple correlation difference. The ratio of those distances, 11.74/9.90 = 1.19, measures how much better this yardstick predicts gout than hypertension. Similar comparisons can be made among all five dependent variables.

The piece de resistance for clinical interpretation, however, is displayed in Figure 2 of "Rasch Measurement Instead of Regression" by Wright, Perkins and Dorsey. In that Figure, the position of any patient measure with respect to the "No Gout" and "Gout" indicators makes the clinical interpretation of the measure obvious. See that discussion of Figure 2 to appreciate the clinical advantage of WINSTEPS Rasch measurement "regression" analysis.

Table 1. WINSTEPS MULTIPLE REGRESSION Results

RAW SCORE	COUNT	MEASURE	ERROR	INFIT MNSQ	OUTFIT MNSQ	SCORE CORR.	ITEMS	
408	96	60.3A	.7	.83	.82	.83	URIC ACID	INDEPENDENT
325	96	62.7A	.8	.70	.75	.72	UREA NITROGEN	VARIABLES
181	96	55.3A	.8	.89	1.07	.62	CREATININE	
48	96	53.7	1.9	.80	.88	.61	GOUT	DEPENDENT
45	96	55.2	1.9	.91	1.08	.51	HyperTense	VARIABLES
22	96	66.0	2.1	1.01	.81	.39	Diuretic	Successful
6	96	79.5	3.5	1.19	3.33	-.03	KidneyStone	Unsuccessful
9	96	75.7	2.9	1.19	4.78	-.06	Diabetes	

Table 1. (continued)

SCORE VALUE	DATA COUNT	%	AVERAGE MEASURE	ITEM	DIAGNOSTIC MEASURES	
0	1	1	29.12	URIC ACID	2.1 mg/dl	
1	4	4	34.91		3.3	
2	11	11	42.12		4.5	
3	20	21	48.66		5.7	
4	21	22	55.76		6.9 normal	
5	12	13	55.28		8.1 high	55
6	14	15	61.31		9.3	
7	11	11	64.51		10.5	
8	1	1	64.67		11.7	
9	1	1	70.53		12.9	
0	1	1	29.12	UREA	0 mg/dl	INDEPENDENT
1	1	1	36.80	NITROGEN	5	
2	21	22	45.80		10	
3	41	43	52.40		15 normal	
4	18	19	57.91		20 high	58
5	6	6	62.29		25	
6	3	3	66.61		30	
7	2	2	69.44		35	
8	2	2	69.02		40	
9	1	1	73.50		45	
0	8	8	38.95	CREATININE	0.7 mg/dl	VARIABLES
1	39	41	50.63		1.0 normal	
2	28	29	55.14		1.3 high	55
3	13	14	60.61		1.6	
4	3	3	60.12		1.9	
5	1	1	67.51		2.2	
6	1	1	61.88		2.5	
7	1	1	69.75		2.8	
8	1	1	73.50		3.1	
9	1	1	71.37		3.4	
0	48	50	47.70	GOUT	DIAGNOSIS	
1	48	50	59.44	1/0= +12	R= +.61	59
0	51	53	48.93	HYPERTENSE	DIAGNOSIS	DEPENDENT
1	45	47	58.83	1/0= +10	R= +.51	59
0	74	77	51.50	DIURETIC	DIAGNOSIS	VARIABLES
1	22	23	60.52	1/0= +9	R= +.39	61
0	90	94	53.65	KidneyStone	No Diagnosis	
1	6	6	52.42	1/0= -1	R= -.03	
0	87	91	53.76	Diabetes	No Diagnosis	
1	9	9	51.77	1/0= -2	R= -.06	