COMBINING SCORES FROM DIFFERENT TESTS 2: MULTIPLE REGRESSION EQUATIONS

Contents

If viewing on screen you can click on a contents item above to jump to the page the item is on

In clinical practice on most occasions when a multiple regression formula is used it is almost certainly one provided in a test manual or a journal paper.

Virtually all the examples in the following text involve the use Z-scores as predictors. This makes life easier, but, as you will realise, there are raw score equivalents for all of the formulas discussed below. And some examples will be provided.

1. The simplest case – uncorrelated predictor variables

A problem that frequently arises in clinical practice is the question of the extent to which a mood state or disorder can influence cognitive performance especially on tests involving a speed component.

Suppose that we have a test of mental speed which correlates -.4 with a test of depression, and that the test of mental speed in turn correlates .6 with IQ.

The intelligence and mental speed tests have been administered to a patient who obtains an IQ of 130, and a mental speed score at the 25th percentile. The correlation between this test of mental speed and IQ is .6.

You therefore predict from the IQ that the patient’s Z score on mental speed should be 1.2. The patient’s score looks suspiciously low.

But you have also noticed that the patient seems very depressed, so you administer the Beck depression inventory and discover that on this measure as well, the patient’s score is two standard deviations above the mean. Further you know from a literature search that the correlation between the depression Inventory and the test of mental speed is – 0.4.

You also know that the correlation between the depression Scale and the measure of intelligence is zero.

How can you make use of this information to see whether the poor performance remains abnormal after the effects of depression have been allowed for?

One thing you can do is use a multiple regression equation to predict what the patient's score would be expected to be taking BOTH intelligence and depression into account.

Making use of the extra information.

The easiest form of the multiple regression equation will serve us well here

If we have two ***uncorrelated*** predictors (intelligence and depression in this case, the formula for predicting a third variable with which they are both correlated is .

$$Z_y = r_{ay}Z_a + r_{by}Z_b$$

So in this case the predicted Z score on the test of mental speed will be (.6 x 2.0) + (- .4 x 2.0)

This equals 1.2 – 0.8, which is 0.4.

The standard error of prediction for the combined (intelligence plus depression) variable will be the square root of (1.0 minus the variance accounted for by the two predictor variables)

$$\sigma_{pred} = \sqrt{1-\left(r_{ay}^2 + r_{by}^2\right)}$$

In the present example this will equal the square root of 0.48, which is 0.69

So, now that we have taken depression as well as IQ into account, the predicted Z score on the test of mental speed is 0.4, and 95 percent of people with this IQ and this level of depression would be expected to score in the range + 1.76 to – 0.96. In fact, the patients Z score on the distribution of predicted score is -1.06, so about 14 percent of people with this IQ and this level of depression would perform at a lower on the speed test.

2. More than two uncorrelated predictors

As long as the predictors are uncorrelated the predicted score on the criterion variable *Y* will be given by the formula

$$Zy = r_{x1y}Z_{x1} + r_{x2y}Z_{x2} \ldots. + r_{xn}Z_{xn}$$

3. The more complex case: Correlated predictors

When the predictor variables have correlations between them, the method described above is inappropriate.

The method used is to derive beta weights. These are derived in such a way as to minimise the squared deviations between predicted Z-scores and actual Z-scores on the criterion variable.

These beta weights – symbolised as β_a, β_b, etc, are, for the two predictor variable case:

$$\beta_a = \frac{r_{ay} - r_{by}r_{ab}}{1 - r_{ab}^2}$$

$$\beta_b = \frac{r_{by} - r_{ay}r_{ab}}{1 - r_{ab}^2}$$

To predict Zy we simply use the equation

$$Zy = \beta_a Za + \beta_b Zb$$

Notice that if the correlation between the predictors A and B is zero, then β_a will equal r_{ay}, and β_b will equal r_{by}.

The value of the multiple correlation coefficient (symbolised R) is given by the formula;

$$R = \sqrt{\beta_a r_{ay} + \beta_b r_{by}}$$

Let's work out what R would be if r_{ay} was .5, r_{by} was .6, and r_{ab} was zero. The value of R would be .78.

On the other hand if r_{ab} was .8, the value of R would only be .60.

So an increase in the correlation between valid predictors tends to lower the value of R.

What happens if we reduce the value of r_{by} to zero? If r_{ay} was .5 and r_{ab} was .8, then R would be .83!

So, a variable with no relationship to the criterion, but a strong relationship to a valid predictor, can be entered into the equation and improve predictability.

4. Examples of multiple regression equations

In general, any multiple regression equation used clinically, and involving more than two predictors will be based on an equation provided by a test producer, or one drawn from a journal paper of from a book. The weights will have been worked out for you

All you will need to do is apply them.

For example, take the formula that Crawford and Allan (1997) put forward to predict IQ from demographic variables. The formula is:

Predicted Full Scale IQ = 87.14 + (5.21 x class) + (1.78 x years of education) + (0.18 x age)

(Crawford, J.R, and Allan, K.M. (1977) Estimating pre-morbid IQ with demographic variables: Regression equations derived from a UK sample. *The Clinical Neuropsychologist*, **11**, 192 – 197)

Several formulas have been provided by Sullivan, R., Senior, G., and Hennessy, M. (2000) Australian Age-Education and Premorbid Cognitive/Intellectual Estimates for the WAIS-III *Poster Presented at the 6th Annual Conference of the APS College of Clinical Neuropsychologists Hunter Valley, NSW, Australia October 12 – October 15, 2000*

The paper can be found on-line at:

http://www.usq.edu.au/users/senior/Posters/SullivanPoster.htm

The equations to predict Verbal IQ on the WAIS are shown below. But, as is always the case, you should read the paper before using them. The authors will have given details of sample and of cautions relating to the use of the equations. The values to use for the different variables are listed after the table.

Equations for predicting Verbal IQ *from* Sullivan, Senior and Hennessey (2000)				
Equation	**N**	**r**	**R^2**	**SE estimate**
85.51 + 5.0 (educ) + 0.2 (age) – 2.87(sex)	229	.546	.298	9.36
101.33 – 0.58(NARTerr) + 4.18(educ)	129	.635	.403	8.44
110.51 – 0.48(AUSNARTerr) + 2.97(educ) – 3.01(sex)	214	.660	.463	8.50
59.27 + 0.68(STW) + 3.75(educ) + 0.11(age) – 2.7(sex)	229	.604	.364	8.92
45.65 + 0.58(WRAT3) + 3.5(educ) + 0.11(age) – 3.24(sex)	229	.652	.425	8.49
54.90 + 1.38(SHIPvoc) + 3.66(educ) –0.18(age)	128	.658	.433	8.28

Where:

Age	= age in years
Sex	1 = Male, 2 = Female
Education	1 = less than 9 years
	2 = 9 to 10 years
	3 = 11 to 12 years
	4 = 13 to 15 iears
	5 = 16 or more years
NARTerr	= NART Error Score
AUSNARTerr	= AUSNART Error Score
STW	= Spot–the–Word raw Score
WRAT 3	= WRAT 3 Reading raw score
SHIPvoc	= Shipley Vocabulary raw score

As an example, let's suppose we wanted to predict the Verbal IQ of a 30 year old woman with 12 years of education.

The first equation in the table could be used here.

The values would be:

$$VIQ = 85.51 + (5 \text{ x}3) + (0.2 \text{ x } 30) - (2.87 \text{ x } 2) = 101$$

IQs for 30 year old women with 12 years of education would thus be expected to have a mean of 100.77 and a standard deviation equal to the Standard Error of Estimate, which is given in the table as 9.36.

5. Do it yourself

Alternatively you might have gathered data yourself, which you would presumably analyse using statistics software. The resulting output would give you the weights to apply to your predictors.

(If you do not have a program for multiple regression analysis, **Smith's Statistics Package**, which is free, and which has versions for both Windows and Mac OS, can be downloaded from:

http://www.economics.pomona.edu/StatSite/framepg.htm

Choose 'Smith's Statistical Package' at the bottom of the Home Page)

An example might help here

Suppose you had been interested in the relationship of the intensity of headache pain to anxiety and depression. You had gathered the data in the table below. These data consist of scores on a test of anxiety, a test of depression and a rating of the intensity of the pain experienced.

Anxiety	Depression	Pain
18.00	10.00	2.00
20.00	17.00	1.00
21.00	16.00	3.00
21.00	18.00	6.00
23.00	20.00	5.00
23.00	17.00	6.00
25.00	28.00	10.00
30.00	25.00	8.00

You use computer software to carry out a multiple regression analysis of your data, and somewhere in the output tables you find the following information.

	Unstandardized Coefficients	Standardized Coefficients
	B	Beta
(Constant)	- 5.824	
Anxiety	.175	.210
Depression	.371	.678

The Unstandardised Coefficients Column gives you the constant and weights you need to make raw score predictions.

In the current example we can predict Pain rating by the formula:

Pain Rating = - 5.824 + .175 times Anxiety Score +. 371 times Depression Score

So if a new person arrives and scores 10 for Anxiety and 20 for Depression, the predicted pain Score would be:

$$- 5.624 + (.175 \times 10) + (.371 \times 20) = 3.55$$

The program will also give you the Standard Error of Estimate which in this case is 1.835, which is of course, the standard deviation of the distribution of actual Pain Scores around the predicted Pain Score.

The standardised coefficients column gives the Beta weights for use in predicting from Z-scores. So the Z-score multiple regression equation for predicting pain from Anxiety and Depression will be:

$$Z_{pain} = (.210 \text{ x } Z_{anxiety}) + (.678\ Z_{depression})$$

Thus somebody obtaining a Z score of -1 on the Anxiety Test, and a Z-score of 1.5 on the Depression Test would have a Pain Z-score equal to:

$$(-1 \text{ x } .210) + (1.5 \text{ x } .678) = .807$$

This is, of course, only an example. You would not normally consider conducting a multiple regression analysis with so few cases.

Finally your statistics program will probably allow you to visually inspect the closeness of fit between you predicted and obtained scores.

Here is a diagram showing the relationship between predicted and actual pain scores.