HOW TO ....
A PRACTICAL GUIDE TO PSYCHOMETRICS

B. Predicting one variable from another

Contents

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

1.   Formulas for the correlation coefficient

The basic formula for the product moment correlation coefficient is:

$$r_{xy} = \frac{\sum xy}{N\sigma_x \sigma_y}$$

If we are using Z-scores, this becomes:

$$r_{xy} = \frac{\sum Z_x Z_y}{N}$$

2.   How to Predict using Z-scores

To predict a Z-score on one test from the Z-score on another, use the formula

$$\hat{Z}_y = r_{xy}Z_x$$

Example: Suppose somebody obtains a Verbal IQ of 130 on a test with a mean of 100 and a standard deviation of 15. What would be the predicted Z-score on a test of Reading Comprehension if this correlate .7 with verbal IQ?

Answer:

$$((130 - 100)/15) \times 0.7 = 1.4$$

3.   Predict using test scores

To predict score on test Y from test X, the regression formula is:

$$\hat{Y} = r_{xy}\frac{\sigma_y}{\sigma_x}(X - M_x) + M_y$$

Example. Somebody obtains an IQ of 145 on a test with a mean of 100 and a standards deviation of 15. The correlation between this test and a test drawing ability (which has a mean of 50 and a standard deviation of 10) is 0.45. What is the predicted score on the drawing test?

Answer.

$$\hat{Y} = 0.45\frac{10}{15}(145 - 100) + 50 = 63.5$$

4.   How to Calculate error of prediction

Scores predicted by regression equations will not always be exactly correct. There will be differences between scores on Test Y predicted from Test X and the scores actually obtained on Test Y. The actual scores will be distributed around the predicted score with a standard deviation equal to $\sqrt{1-r_{xy}^2}$ in Z-scores, or $\sigma_y\sqrt{1-r_{xy}^2}$ if raw scores are used

5.   How to Test the significance of the difference between a predicted and an obtained score

It is probably easier to use Z-scores here. The formula is:

$$Z_{diff} = \frac{Z_y - r_{xy}Z_x}{\sqrt{1-r_{xy}^2}}$$

Example. Jack gets a Verbal IQ of 115 and a Performance IQ of 95. What proportion of people would get a difference as large or larger than this between predicted and actual Performance IQ, supposing that the correlation between Verbal and Performance IQs is .8?

Answer

$$Z_{diff} \text{ will equal } \frac{1-(-0.33)}{\sqrt{1-0.64}} = \frac{1.33}{0.60} = 2.22$$

Consulting Tables for the Normal Distribution we find that only about 1.5 percent of people with that Verbal IQ would be expected to get a Performance IQ as low as this.

6.   How to: Given an (as yet) unvalidated predictor U whose correlation with a validated predictor V is known, estimate the range within which the correlation between A and the criterion (C) is likely to lie

This somewhat convoluted title refers to the following problem. People sometimes argue that because Test V is quite highly correlated with some criterion C, and because Test U is correlated with Test V, then Test U is likely to be correlated with criterion C as well.

(Logicians will smell a fallacy brewing here!)

Obviously the only way to find the actual correlation between U and C is to gather data and do the sums.

However it is possible to calculate the range within which the correlation between U and C would lie from the correlation between V and C and the correlation between U and V.

We can work out the upper and lower limits of the range within which r_{uc} is likely to lie

The formula is:

$$r_{uc_{(upper)}} = r_{vc}r_{uv} + \sqrt{r_{vc}^2 r_{uv}^2 - r_{vc}^2 - r_{uv}^2 + 1}$$

$$r_{uc_{(lower)}} = r_{vc}r_{uv} - \sqrt{r_{vc}^2 r_{uv}^2 - r_{vc}^2 - r_{uv}^2 + 1}$$

The results of the calculation can sometimes surprise.

To make life easier use the table on the next page.

Maximum and minimum correlations

ruv	0.90		0.80		0.70		0.60		0.50	
rvc	*Max. r*	*Min. r*	*Max. r*	*Min. r*	*Max. r*	*Min. r*	*Max. r*	*Min. r*	*Max. r*	*Min. r*
0.90	1.00	0.62	0.98	0.46	0.94	0.32	0.89	0.19	0.83	0.07
0.80	0.98	0.46	1.00	0.28	0.99	0.13	0.96	0.00	0.92	-0.12
0.70	0.94	0.32	0.99	0.13	1.00	-0.02	0.99	-0.15	0.97	-0.27
0.60	0.89	0.19	0.96	0.00	0.99	-0.15	1.00	-0.28	0.99	-0.39
0.50	0.83	0.07	0.92	-0.12	0.97	-0.27	0.99	-0.39	1.00	-0.50
0.40	0.76	-0.04	0.87	-0.23	0.93	-0.37	0.97	-0.49	0.99	-0.59
0.30	0.69	-0.15	0.81	-0.33	0.89	-0.47	0.94	-0.58	0.98	-0.68
0.20	0.61	-0.25	0.75	-0.43	0.84	-0.56	0.90	-0.66	0.95	-0.75
0.10	0.52	-0.34	0.68	-0.52	0.78	-0.64	0.86	-0.74	0.91	-0.81

C = criterion

V = validated test, i.e., ctest whose correlation with the criterion is known

U = unvalidated test, i.e., test whose correlation with the criterion is unknown

rvc = correlation between validated test and criterion

ruv = the correlation between the validated and the unvalidated tests

Values in the body of the table give the upper and lower limits for the correlation between the unvalidated test and the criterion

Example: A test of auditory immediate recall (V) correlates 0.8 (rvc) with a test of auditory delayed recall (C).
A test of word knowledge (U) correlates 0.7 with the test of immediate recall.(ruv)
What is the range within which the correlation between word knowledge and delayed recall will lie?

We find the column for ruv = 0.7, and the row for rvc = 0.8
The cell where the row and column intersect gives us the range of possible values for ruc, which is between .13 and .99