

## Simplistic Statistics for Control of Polytomous Items

Several issues ago, I discussed estimating the logit difficulties with a simple *pair* algorithm, although this is viewed with distain in some quarters because it's only least squares and does not involve maximum likelihood or Bayesian estimation. It begins with counting the number of times item *a* is correct and item *b* incorrect and vice versa; then converting the counts to log odds; and finally computing the logit estimates for dichotomous items as the row averages, if the data are sufficiently well behaved. If the data aren't sufficiently well behaved, it could involve solving some simultaneous equations instead of taking the simple average.

This machinery was readily adaptable to include polytomous items by translating the items scores, 0 to  $m_i$ , into the corresponding  $m_i + 1$  Guttman patterns. That is, a five-point item has six possible scores 0 to 5 and six Guttman patterns (00000, 10000, 11000, 11100, 11110, and 11111). Treating these just like five more dichotomous items allows us to use the exactly the same algorithm to compute the logit difficulty (aka, *threshold*) estimates. (The constraints on the allowable patterns means there will always be some missing log odds and the row averages never work; polytomous items will always require solving the simultaneous equations but the computer doesn't much care.)

While the pair algorithm leads to some straightforward statistics of its own for controlling the model, my focus is always on the simple person-item residuals because the symmetry leads naturally to statistics for monitoring the person's behavior as well as the items performance. For dichotomous items, the basic residual is  $y_{ni} = x_{ni} - p_{ni}$ , which can be interpreted as the person's deviation from the item's p-value. The basic residual can be manipulated and massaged any number of ways; for example, a squared standardized residual  $z_{ni}^2 = (x_{ni} - p_{ni})^2 / [p_{ni}(1 - p_{ni})]$ , or  $(1 - p) / p$  if  $x_{ni} = 1$  or  $p / (1 - p)$  if  $x_{ni} = 0$ , which can be interpreted as the *odds against* the response.

A logical extension to polytomous items (Ludlow, 1982) would be, for the basic residual,  $y_{ni} = x_{ni} - E(x_{ni})$  and, for the standardized residual,  $z_{ni} = (x_{ni} - E(x_{ni})) / \sqrt{Var(x_{ni})}$ , where  $x_{ni}$  is the observed item score from 0 to  $m_i$ , where  $m_i$  is greater than one. The interpretation for the basic form is now the deviation in item score (which is the same as the deviation in p-value when  $m_i$  is one.) The interpretation for  $z^2$  is messier. This approach has been used extensively for the past thirty plus years, although not exploited as fully as it might be<sup>1</sup>. And there is an alternative that salvages much of the dichotomous machinery. And we have made dichotomous items out of the polytomous scores already.

We're back in the world of 1's and 0's; or maybe we never left. All thresholds are dichotomies where you either pass, succeed, endorse, or do whatever you need to get by or you don't. We have an observed value  $x = 0$  or  $1$ , an expected value  $p = (0, 1)$ , and any form of residual we like,  $y$  or  $z$ . The following table shows the residuals for the six Guttman patterns, based on a person with logit ability equal zero and a five-point item with nicely spaced thresholds (-2, -1, 0, 1, 2). Because the thresholds are symmetric and the person is centered on them, there is a lot of repetition. Values in bold font are the ones that changed from the preceding panel.

---

<sup>1</sup> As with dichotomous items, we tend to sum over items (occasionally people) to get *Infit* or *Outfit* and proceed merrily on our way trusting everything is fine.

Category	1	2	3	4	5	Sum	
Threshold	-2.0	-1.0	0.0	1.0	2.0		
$P(r=k \beta=0)$	0.13	0.35	0.35	0.13	0.02	1.0*	
$p(x=1 \beta=0)$	0.88	0.73	0.50	0.27	0.12		
$x$	0	0	0	0	0	0	
$y$	-0.9	-0.7	-0.5	-0.3	-0.1	6.3	Squared of Squares
$y^2$	0.8	0.5	0.3	0.1	0.0	1.6	
$z$	-2.7	-1.6	-1.0	-0.6	-0.4	40.2	
$z^2$	7.4	2.7	1.0	0.4	0.1	11.6	

$x$	1	0	0	0	0	1	
$y$	<b>0.1</b>	-0.7	-0.5	-0.3	-0.1	2.3	Squared of Squares
$y^2$	<b>0.0</b>	0.5	0.3	0.1	0.0	0.9	
$z$	<b>0.4</b>	-1.6	-1.0	-0.6	-0.4	10.6	
$z^2$	<b>0.1</b>	2.7	1.0	0.4	0.1	4.4	

$x$	1	1	0	0	0	2	
$y$	0.1	<b>0.3</b>	-0.5	-0.3	-0.1	0.3	Squared of Squares
$y^2$	0.0	<b>0.1</b>	0.3	0.1	0.0	0.4	
$z$	0.4	<b>0.6</b>	-1.0	-0.6	-0.4	1.0	
$z^2$	0.1	<b>0.4</b>	1.0	0.4	0.1	2.0	

$x$	1	1	1	0	0	3	
$y$	0.1	0.3	<b>0.5</b>	-0.3	-0.1	0.3	Squared of Squares
$y^2$	0.0	0.1	<b>0.3</b>	0.1	0.0	0.4	
$z$	0.4	0.6	<b>1.0</b>	-0.6	-0.4	1.0	
$z^2$	0.1	0.4	<b>1.0</b>	0.4	0.1	2.0	

$x$	1	1	1	1	0	4	
$y$	0.1	0.3	0.5	<b>0.7</b>	-0.1	2.3	Squared of Squares
$y^2$	0.0	0.1	0.3	<b>0.5</b>	0.0	0.9	
$z$	0.4	0.6	1.0	<b>1.6</b>	-0.4	10.6	
$z^2$	0.1	0.4	1.0	<b>2.7</b>	0.1	4.4	

$x$	1	1	1	1	1	5	
$y$	0.1	0.3	0.5	0.7	<b>0.9</b>	6.3	Squared of Squares
$y^2$	0.0	0.1	0.3	0.5	<b>0.8</b>	1.6	
$z$	0.4	0.6	1.0	1.6	<b>2.7</b>	40.2	
$z^2$	0.1	0.4	1.0	2.7	<b>7.4</b>	11.6	

\*Probabilities sum to one when we include category  $k=0$ .

Not surprisingly, for a person with a true expected response of 2.5, we are surprised when the person's response was *zero* or *five* ( $\sum_{k=1}^{m_i} z_{vik}^2 = 11.6$ ); less surprised by responses of *one* or *four* ( $\sum_{k=1}^{m_i} z_{vik}^2 = 4.4$ ); and quite happy with responses of *two* or *three* ( $\sum_{k=1}^{m_i} z_{vik}^2 = 2.0$ ). We would feel pretty much the same looking at  $[\sum z]^2$  or almost any other number in the sum column. Not surprisingly, when we look at the numbers for each category, we are surprised when the person is stopped by a low valued threshold (e.g., the first panel, first column) or not stopped by a high valued (the last panel, last column.)

That's what happens with nicely spaced thresholds targeted on the person. If the annoying happens and some thresholds are reversed, the effects on these calculations are less dramatic than one might expect or hope. For example, with thresholds of (-2, -1, 1, 0, 2), the sum of  $z^2$  for the six Guttman patterns are (11.6, 4.4, 2.0, 4.4, 4.4, and 11.6). Comparing those to the table above, only the fourth value (response  $x=3$ ) is changed at all (4.4 instead of 2.0.) How that would present itself in real data depends on who the people are and how they are distributed. The relevant panel is below; the others are unchanged.

Category	1	2	3	4	5	Sum	
Threshold	-2.0	-1.0	1.0	0.0	2.0		
$P(r=k \beta=0)$	0.17	0.45	0.17	0.17	0.02	1.0*	
$p(x=1 \beta=0)$	0.88	0.73	0.27	0.50	0.12		
$x$	1	1	1	0	0	3	
$y$	0.1	0.3	0.7	-0.5	-0.1	0.3	Squared of Squares
$y^2$	0.0	0.1	0.5	0.3	0.0	0.9	
$z$	0.4	0.6	1.6	-1.0	-0.4	1.6	Squared of Squares
$z^2$	0.1	0.4	2.7	1.0	0.1	4.4	

\*Probabilities sum to one when we include category  $k=0$ .

While there is nothing in the mathematics of the model that says the thresholds must be ordered, it makes the categories, which are ordered, a little puzzling. We are somewhat surprised ( $z^2=2.7$ ) that the person passed the third threshold but at the same time thought the person had a good chance ( $y=-0.5$ ) of passing the fourth.

Reversing the last two thresholds (-2, -1, 0, 2, 1) gives similar results; in this case, only the calculations for response  $x=4$  changes.

Category	1	2	3	4	5	Sum	
Threshold	-2.0	-1.0	0.0	2.0	1.0		
$P(r=k \beta=0)$	0.14	0.38	0.38	0.05	0.02	1.0*	
$P^*(x=1 \beta=0)$	0.88	0.73	0.50	0.12	0.27		
$x$	1	1	1	1	0	4	
$y$	0.1	0.3	0.5	0.9	-0.3	2.3	Squared of Squares
$y^2$	0.0	0.1	0.3	0.8	0.1	1.2	
$z$	0.4	0.6	1.0	2.7	-0.6	16.7	Squared of Squares
$z^2$	0.1	0.4	1.0	7.4	0.4	9.3	

\*Probabilities sum to one when we include category  $k=0$ .

This discussion has been more about the person than the item. Given estimates of the person's logit ability and the item's thresholds, we can say relatively intelligent things about what we think of the person's score on the item; we are surprised if *difficult* thresholds are *passed* or *easy* thresholds are *missed*. Whether or not any of this is visible in the item statistics depends on whether or not there are sufficient numbers of people behaving oddly.

Whether or not the disordered thresholds affects the item mean squares depends on how the item is targeted and the distribution of abilities. Estimation of the threshold logits is still not affected by the ability distribution, which keeps us comfortably in the Rasch family, even if we are a little puzzled.