Chapter 23
Fit of Responses to the Polytomous Rasch Model

The Fit-Residual

The residual of the response x_{ni} of each person n for each item i is simply

$$x_{ni} - E[x_{ni}], \tag{23.1}$$

where

$$E[x_{ni}] = \sum_{x=0}^{m_i} x \ \Pr\{x_{ni}\}, \tag{23.2}$$

$$\Pr\{x_{ni}\} = \frac{1}{\gamma_{ni}} e^{-\sum_{k=0}^{x} \tau_{ki} + x(\beta_n - \delta_i)}, \quad \tau_{0i} \equiv 0 \tag{23.3}$$

and $\gamma_{ni} = \sum_{x=0}^{m_i} e^{-\sum_{k=0}^{x} \tau_{ki} + x(\beta_n - \delta_i)}$ is a normalizing factor which ensures that the probabilities of Eq. (23.2) sum to 1.

To obtain $E[x_{ni}]$, the estimates $\left(\hat{\beta}_n, \ \hat{\delta}_i, \ \hat{\tau}_{ki}\right)$ are placed into Eq. (23.3) and in turn into Eq. (23.2). Because we insert estimates $\left(\hat{\beta}_n, \ \hat{\delta}_i, \ \hat{\tau}_{ki}\right)$, $E[x_{ni}]$ of Eq. (23.2) could be written with a 'hat' as $\hat{E}[x_{ni}]$. However, we generally do not do that, understanding that in the context, we have inserted the estimates. The residual itself is a difference. To assess whether the magnitude is large or not, it is referenced to its standard deviation. Therefore, the *standardized residual*

$$z_{ni} = \frac{x_{ni} - E[x_{ni}]}{\sqrt{V[x_{ni}]}} \tag{23.4}$$

is formed where $V[x_{ni}] = E\left[x_{ni}^2\right] - (E[x_{ni}])^2$ is the variance of x_{ni}, $E[x_{ni}^2] = \sum_{x=0}^{m_i} x^2 \Pr\{x_{ni}\}$.

D. Andrich and I. Marais, *A Course in Rasch Measurement Theory*,
Springer Texts in Education, https://doi.org/10.1007/978-981-13-7496-8_23

The theoretical mean over an imagined infinite number of replications is zero, $E[z_{ni}] = 0$. If known values of $\left(\hat{\beta}_n,\ \hat{\delta}_i,\ \hat{\tau}_{ki}\right)$, which did not come from the estimates from the data, were used in obtaining $E[x_{ni}]$ and $V[x_{ni}]$, then the variance $V[z_{ni}] = 1$. This is just the variance of standardized scores. However, because we are using the estimates of the parameters from the same data as we are using to form the residuals, the variance of the residuals will be less than 1, and it will be the degrees of freedom. Say the degrees of freedom are $f_{ni} < 1$, which we use shortly.

Deriving the Fit-Residual for the Persons

Because the sum of the residuals will be close to zero, no matter how good the fit, to obtain a magnitude of the residual, it is first squared. From Eq. (23.4),

$$Y_n^2 = \sum_{i=1}^{I} z_{ni}^2. \tag{23.5}$$

Y_n^2 itself has an expected value given by $E\left[Y_n^2\right] = E\left[\sum_i Y_{ni}^2\right] = \sum_i E\left[Y_{ni}^2\right]$.

In the case where no parameters are estimated, $E\left[Y_{ni}^2\right] = 1$ so that $\sum_i E\left[Y_{ni}^2\right] = 1$. However, in the case where degrees of freedom are lost through estimation, $E[Y_{ni}^2] = f_{ni}$. We estimate the degrees of freedom by subtracting the effective number of parameters estimated from the data and then apportioning this to each person–item combination and then summing over all the items.

$$E\left[Y_n^2\right] = \sum_i f_{ni} = f_n = I\frac{(N-1)(I-1)-(m-1)}{NI},$$

that is

$$f_n = [(N-1)(I-1)-(m-1)]/N, \tag{23.6}$$

where f_n are the degrees of freedom associated with each person n.

Then the residual $Y_n^2 - E[Y_n^2]$, which has an expected value of 0, can be used to test the fit of the responses of person n. In order to make a formal test of fit, this residual can be standardized by calculating the variance of Y_n^2. This can be obtained from $V\left[Y_n^2\right] = V\left[\sum_i Y_{ni}^2\right] = \sum_i V\left[Y_{ni}^2\right]$.

The test statistic T_{n1} can then take the form

$$T_{n1} = \frac{Y_n^2 - E\left[Y_n^2\right]}{\sqrt{V\left[Y_n^2\right]}}. \tag{23.7}$$

A transformation of T_{n1}, see Eq. (23.9), which makes the distribution more symmetrical, can be made. This is done simply by first forming the mean square ratio

$$Y_n^2/f_n, \tag{23.8}$$

which has an expected value of 1, and then taking its natural logarithm.

Because Eq. (23.8) is in the ratio form,
if $Y_n^2/f_n = c$ and $Y_n^2 > f_n$, then c > *0* and

$$\log\left(Y_n^2/f_n\right) = \log\ c = C > 0.$$

If $Y_n^2 < f_n$ by a symmetrical amount in the ratio, that is $Y_n^2/f_n = 1/c$, then

$$\log\left(Y_n^2/f_n\right) = -\log\ c = -C.$$

For example, if Y_n^2 is 3 times its expected value of f_n, then

$$\log\left(Y_n^2/f_n\right) = \log\ 3 = 1.099.$$

And if Y_n^2 is 1/3 times its expected value of f_n, then

$$\log\left(Y_n^2/f_n\right) = -\log\ 3 = -1.099.$$

Furthermore, when $Y_n^2 = f_n$, then $\log\left(Y_n^2/f_n\right) = 0$.

After another transformation which can be found in the *Further Reading*, we reach the ratio

$$T_{n2} = \frac{\log\left(Y_n^2/f_n\right)}{\sqrt{V\left(Y_n^2/f_n\right)}} = \frac{f_n\left(\log Y_n^2 - \log f_n\right)}{\sqrt{V\left(Y_n^2\right)}}. \tag{23.9}$$

This is a more symmetrical distribution than the one in Eq. (23.7) with $E\,[T_{n2}] = 0$ and $V\,[T_{n2}] = 1$. The proper shape of this distribution is not known but it should be close to a normal distribution. If Y_n^2 were a χ^2 distribution on 20 or more degrees of freedom, then the logarithmic transformation would convert a 2.5% one-tailed test for a normal distribution ($T_{n2} \cong 2$) to a 1% one-tailed test on the original χ^2 distribution. For this reason, a T_{n2} value of $|T_{n2}| = 2$ is taken as a general critical value for fit of a person to the model if $f_n > 20$. The effect of the variance of Y_n^2 not being $2f_n$, as would be the case if it were actually distributed as χ^2, is not known, but on the basis of simulations, the particular statistic seems to work very well.

Other similar-based statistics reported in the literature (Wright & Stone, 1979; Wright & Masters, 1982) use a different weighting procedure to account for the fact that $V\left[Y_{ni}^2\right]$ is a function of $(\beta_n - \delta_i)$ and to make Y_n^2 symmetrical.

If the value from Eq. (23.9) is large in magnitude and negative, then the response profile of the person is very Guttman-like. For example, in proficiency assessment, easy items are answered correctly and difficult items incorrectly. On the other hand, if it is large in magnitude and positive, then the person's response profile is erratic relative to the difficulties of the items.

Of course, the earlier advice to use fit statistics in context, and not absolutely, holds here as well. Thus, a key to interpreting this statistic is not simply to use an absolute value, such as +2.5 or −2.5, but to order the persons by this fit statistic and see how the values change and if there are some persons at either extreme who are very different from those a little less extreme—that is, see if there are small differences between them or if there is a big jump in values. If there is, then these are the persons whose profiles would be of most concern.

Deriving the Fit-Residual for the Items

The difference, $x_{ni} - E[x_{ni}]$, can also be standardized for each item and summed over the persons attempting the item. Beginning with Eq. (23.4), by taking the summation over persons, equivalent to Eq. (23.5), gives for an item

$$Y_i^2 = \sum_{n=1}^{N} z_{ni}^2. \tag{23.10}$$

Transforming these squared terms produces the item-based test statistic T_{i1}, equivalent to Eq. (23.7) for the persons.

$$T_{i1} = \frac{Y_i^2 - E\left[Y_i^2\right]}{\sqrt{V\left[Y_i^2\right]}}. \tag{23.11}$$

As with the T_{n1} statistic for the persons, the distribution of T_{i1} is also not symmetrical. Therefore, using $\pm 2\, T_{i1}$, say, as an approximation to the 95% confidence interval in a normal distribution would be misleading. The same logarithmic transformation described for the person case can also be made to T_{i1}, which makes the distribution more symmetrical.

The degrees of freedom for each item are approximated by

$$E\left[Y_i^2\right] = \sum_n f_{ni} = f_i = N\frac{(N-1)(I-1)-(m-1)}{NI}.$$

That is,

$$f_i = \frac{(N-1)(I-1)-(m-1)}{I}. \tag{23.12}$$

Then the fit statistic is given by

$$T_{i2} = \frac{\log\left(Y_i^2/f\right)_i}{\sqrt{V\left(Y_i^2/f_i\right)}} = \frac{f_n\left(\log Y_i^2 - \log f_i\right)}{\sqrt{V\left(Y_i^2\right)}}. \tag{23.13}$$

As with the fit-residual for persons, if the value of Eq. (23.13) is large in magnitude and negative, then the response profile for the item is very Guttman-like. For example, in the case of assessment of proficiency, the less able tend to answer the item incorrectly and the more able correctly. In relation to the item characteristic curve, a negative value which is large in magnitude implies that the observed proportions in the class intervals will be steeper than the curve.

If the value of Eq. (23.13) is large in magnitude and positive, then the response profile for the item is very non-Guttman-like. For example, in the case of assessment of proficiency, it is not the case that the less able tend to answer the item incorrectly and the more able correctly. In relation to the item characteristic curve, a positive value which is large in magnitude implies that the observed proportions in the class intervals will be less steep than the curve.

References

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA Press.

Wright, B. D., & Stone, M. H. (1979). The measurement model. *Best test design: Rasch measurement* (pp. 1–17). Chicago: MESA Press.

Further Reading

Andrich, D., Sheridan, B. E., & Luo, G. (2018). *RUMM2030: Rasch unidimensional models for measurement. Interpreting RUMM2030 Part III estimation and statistical techniques* (pp. 15–25). Perth, Western Australia: RUMM Laboratory.