The *t test* for testing the significance of the difference between two estimates of the same person from two subtests composed of mutually exclusive sets of items

David Andrich
The University of Western Australia

Suppose that from responses from a group of persons to all items of the test that the item parameter estimates for all items are obtained from these responses using the Rasch model (dichotomous or polytomous), and that there are two, well defined, subtests which have no common items.

One approach to checking whether or not there is a violation of the model with respect to the two subtests is to test the significance of the difference between estimates of each person on the two subtests.

Often this is considered a test of whether the two subtests assess the same dimension.

Let $(\hat{\beta}_{n1}, \hat{\beta}_{n2}); (\sigma_{n1}, \sigma_{n2})$ be the pairs of estimates and pairs of standard errors of their estimates from subtests 1 and 2 respectively. We may write

$$\begin{aligned} \hat{\beta}_{n1} &= \beta_{n1} + e_{n1} \\ \hat{\beta}_{n2} &= \beta_{n2} + e_{n2}. \end{aligned} \tag{1}$$

Then in forming a *t test* to check the null hypothesis of $\beta_{n1} - \beta_{n2} = 0$ for each person, we may write

$$t = \frac{\hat{\beta}_{n1} - \hat{\beta}_{n2}}{\sqrt{\sigma_{n1}^2 + \sigma_{n2}^2}}, \tag{2}$$

where the degrees of freedom are (1, F). If the sample size is relatively large, then F is also large, and t approximates a standard, normal deviate. Then if its absolute value is greater than 1.96, it can be said that this difference is significant at the 5% level of confidence.

The question that may arise is whether this is an *independent* or a *dependent t test*. The suggestion that it might be a dependent *t test* arises from the observation that the same person is involved in the comparison.

However, the dependent *t test* arises in a different context: specifically, it arises in the comparison between two *means* on the same variable when measures on the two variables are related. One example occurs when the *same* people are assessed before and after a treatment of some kind and the null hypothesis is that there is no effect of the treatment. A second example is when twins are assessed on the same variable, and the study of differences between the means of the twins takes account of the correlation between the measures of the twins. In the case of a *t test* for dependent samples it is necessary to be able to match the scores on a-priori grounds, as in the above examples.

There are two ways to calculate the t value for dependent samples. Both begin with the following formulation.

The first way is to simply form differences between the two values. Let

$$\hat{\beta}_{n1} = \beta_n + d_{n1} + e_{n1};\ \hat{\beta}_{n2} = \beta_n + d_{n2} + e_{n2} \tag{3}$$

where β_n is the common location of person n on the two measures, (d_{n1}, d_{n2}) are unique effects of each person n on the two measures, and (e_{n1}, e_{n2}) are the respective errors of measurement. Then taking the difference between $\hat{\beta}_{n1}$ and $\hat{\beta}_{n2}$
gives

$$\begin{aligned} \hat{\beta}_{n1} - \hat{\beta}_{n2} &= \beta_n - \beta_n + d_{n1} - d_{n2} + e_{n1} - e_{n2} \\ &= d_{n1} - d_{n2} + e_{n1} - e_{n2} \\ &= D_n + E_n. \end{aligned} \tag{4}$$

where $D_n = d_{n1} - d_{n2};\ E_n = e_{n1} - e_{n2}$.

On the assumption that E$[E]$ = 0, the null hypothesis that $\bar{\beta}_{n1} - \bar{\beta}_{n2} = \bar{d}_1 - \bar{d}_2 = 0$ is identical to the null hypothesis that $\bar{D} = 0$. The variance of $\bar{D}$ is given by

$$V[\bar{D}] = \frac{V[D]}{N}$$

where N is the number of persons. Then the *t test* value is given by

$$t = \frac{\bar{D}}{\sqrt{V[\bar{D}]}}, \tag{5}$$

with degrees of freedom 1 and $N-1$. With large N, t again approximates a standard, normal deviate.

In the above subtraction of the two estimates for the same person, the source of the correlation of the two values, the common β_n across persons, has been subtracted out, which in turn accounts for the dependence between β_{n1} and β_{n2}.

A second way of obtaining the same value is to simply obtain the means $\bar{\beta}_{n1}, \bar{\beta}_{n2}$ and write

$$t = \frac{\bar{\hat{\beta}}_{n1} - \bar{\hat{\beta}}_{n2}}{\sqrt{V[\bar{\hat{\beta}}_{n1} - \bar{\hat{\beta}}_{n2}]}}. \qquad (6)$$

Then $V[\bar{\hat{\beta}}_{n1} - \bar{\hat{\beta}}_{n2}] = \frac{V[\hat{\beta}_{n1} - \hat{\beta}_{n2}]}{N^2}$ (7)

where $V[\hat{\beta}_{n1} - \hat{\beta}_{n2}] = V[\hat{\beta}_1] + V[\hat{\beta}_2] - 2COV[\hat{\beta}_1, \hat{\beta}_2]$. (8)

The subtraction of the covariance term $2COV[\hat{\beta}_1, \hat{\beta}_2]$ from $V[\hat{\beta}_1] + V[\hat{\beta}_2]$ in Eq. (8) accounts for any covariance which results from the common persons in the two assessments on the same variable.

The calculation from Eq. (6) using Eqs. (7) and (8) gives an identical t value to that in Eq. (5).

In the case that $COV[\hat{\beta}_1, \hat{\beta}_2]$=0, both ways of obtaining the t value reduce to being numerically equivalent to a *t test* for independent samples in which $V[\hat{\beta}_{n1} - \hat{\beta}_{n2}] = V[\hat{\beta}_1] + V[\hat{\beta}_2]$. However, the formulae above cannot be used unless there is a natural *matching* of assessments.

The above analysis of the *t test* for dependent sample is standard. In this note we call this simply the *dependent* sample *t test* to emphasise that it accounts for the dependence (correlation) *among the persons* between their two measures.

In the above example of the comparison between means of groups of persons, the need to use a dependent sample t test arises from the use of the variance of persons to obtain an estimate of the variance of the difference between the means. In the example of the comparison between the two estimates of the same person in Eq. (2), no such use of a common term is involved. The standard error of the estimate is given from two different sets of items.

However, if the two person estimates had items in common, then the independence would be violated. A formula exists for the extreme case when an estimate on the whole test is to be compared with an estimate on the subtest.

Let $(\hat{\beta}_{nT}, \hat{\beta}_{nS}); (\sigma_{nT}, \sigma_{nS})$ be estimates and standard errors of these estimates for the whole test and a subset of items from the test. In this case the standard errors are not independent.

From a theorem by Erling Andersen (2002), Eq. (2) becomes

$$t = \frac{\hat{\beta}_{nT} - \hat{\beta}_{nS}}{\sqrt{\sigma_{nS}^2 - \sigma_{nT}^2}}. \qquad (9)$$

Note the unusual *subtraction* of variances in the denominator in Eq. (9). Andersen derived this formula in the case of conditional estimates of item parameters in the Rasch model, and applied it to a comparison of item estimates from a whole sample and a subsample of persons. Eq. (9) takes the principle of this formula for the case of person estimates from a whole set of items and a subset of items.

In the specific case of RUMM software, RUMM2020 or RUMM2030, therefore, and irrespective of what it might be called to draw attention to the fact that *two* estimates are being compared (for example the word *pair* might be used),

(i) the *t test* to assess whether the two *estimates of each individual* from two different subtests (with mutually exclusive items) is significantly different, is a standard *t test*.

However, it is stressed that this *t test* is correct only if the two subtests have mutually exclusive items.

(ii) The *t test* to assess whether the pair of *means* of the estimates from the two different subtests (with mutually exclusive items) is significantly different, is a dependent sample *t test*.

Reference

Andersen, E.B. (2002). Residual Diagrams Based on a Remarkably Simple Result Concerning the Variances of Maximum Likelihood Estimators *Journal of Educational and Behavioral Statistics, 27 (1), 19–30.*

t