Use of Robust z in Detecting Unstable Items in Item Response Theory Models

Huynh Huynh, University of South Carolina
Patrick Meyer, University of Virginia

The first part of this paper describes the use of the robust $z_R$ statistic to link test forms using the Rasch (or one-parameter logistic) model. The procedure is then extended to the two-parameter and three-parameter logistic and two-parameter partial credit (2PPC) models. A real set of data was used to illustrate the extension. The linking results illustrate the efficacy of the robust $z_R$ vis-à-vis some of the most commonly used processes such as the Stocking and Lord (1983) linking process.

Educational and psychological assessments often require the use of different forms of the same test. These forms usually have a set of common items that are used to link all resulting scores to a common scale. Item response theory (IRT) software such as the Rasch WINSTEPS (Linacre, 2006) and PARSCALE (Muraki & Bock, 1997) have been used for linking purposes. Although there are many ways to accomplish this task, most statewide assessment programs rely on the “anchor test” design for yearly linking. In many situations, this design calls for a separate calibration of each form and then a process to link all these forms together through a set of “stable” common items. The selection of the common items (aka linkers) is usually based on several criteria. During test form construction, potential linkers are chosen to reflect overall test content and the range of item difficulty level. Linkers are typically chosen to be not too easy or too difficult. Once the tests are administered, final linkers are oftentimes restricted to common items which are reasonably stable across the forms to be linked. It may be noted that when a common test form is administered to two different groups of examinees (i.e. all items are treated as linking items), then stability in item parameters is prerequisite for the invariance of the construct under measurement between these two groups.

For technical work based on the Rasch model, the robust $z_R$ statistic (Huynh, 2000; Huynh & Rawls, 2009) has been used widely in large-scale assessment programs (including South Carolina, Arkansas, Maryland, Minnesota, and New Mexico) in detecting items that are unstable (outliers) in yearly linking/equating. The statistic has also been used in studying stability of item parameters across gender and ethnicity groups (Kim & Huynh, 2009). The purpose of this study is to extend the use of the robust $z_R$ statistic to the three-parameter logistic (3PL) and two-parameter partial credit (2PPC) models. A real set of data will be used to illustrate the extension. The linking results also will be used to illustrate the efficacy of the robust $z_R$ vis-à-vis some of the most commonly used processes such as the Stocking and Lord (1983) linking process.

Use of Robust z Statistic in Detecting Outliers

The robust z statistic originates from robust/resistant statistical procedures. (see Hogg, 1979; Huber, 1964; and Huynh, 1982). Huynh took note that most procedures for detecting outliers are based on a “robustification” of the traditional z statistic. Let $D$ be a variable. Traditionally the z statistic is defined as $z = (D - \text{mean})/\text{standard deviation}$. However, both the mean and standard deviation (SD) are influenced by outlying observations. So in order to pinpoint precisely at the outliers, it may be more efficient to look for a z-like statistic that is not affected by the outliers. Let $Md$ and $IQR$ be the median and inter-quartile range. For the
normal distribution, the IQR is equal to 1.35 × SD or SD = 0.74 × IQR or 0.74(IQR). With the quantity 0.74(IQR) emulating the standard deviation, a robust version of the traditional z statistic can be taken as the ratio $z_R = (D - Md)/[0.74(IQR)]$. When the D values come from a normal distribution, the robust $z_R$ statistic follows (asymptotically) a normal distribution with zero mean and unit standard deviation. A level of significance (two-tailed alpha) may be selected and a positive critical mean and unit standard deviation. The South Carolina (SC) linking protocols (2001) calls for two quality indices for Rasch linking: (a) the correlation $CORR$ to be at least .95 and (b) the $RSD$ to be within .9 and 1.1.

The correlation between two estimated values can be treated as a reliability coefficient ($r_{rel}$). Since $r_{rel} < \sqrt{r_{rel}}$, we have $r_{rel} > r_{val}^2$. Since $r_{val}$ is at least 0.97, the other coefficient $r_{rel}$ is at least $(0.97)^2$ or about 0.95. Also according to Huynh, the two bounds for the ratio $RSD$ are the results of the significance test that the population value of this ratio is one. When the standard deviations of $R_1$ and $R_2$ are equal (i.e. when $RSD = 1$), Pearson correlation between the quantities $(R_1 + R_2)$ and $(R_1 - R_2)$ is exactly zero. The traditional t test for the null hypothesis ($H_0$) of zero correlation can be used to check that the hypothesis that $RSD$ is equal to 1. Assuming a correlation of 0.95 between $R_1$ and $R_2$ and at the 5% level of significance, the null hypothesis that true value of $RSD$ is 1 is acceptable if the observed value of $RSD$ is between 0.9 and 1.1.

When both criteria for “acceptable linking results” are satisfied, then all items are treated as “stable” and all potential linkers are used in the linking process. However, if one of the criteria does not hold, then some very unstable items will be deleted, starting with items with largest robust $z_R$ statistic. (The $z_R$ statistics are computed only once.) This process is stopped when either the criteria are met or 20% of potential linkers have already been deleted. Huynh noted that, in assessment situations like the South Carolina Basic Skills Assessment Program (BSAP), each test has several strands (subtests), each with about five items, and it would not be desirable to delete more than one item (20%) for each subtest.

### Extension of Methodology Based on Robust $z$ to 3PL Models

Let $a_i$ and $b_i$ be the slope and location of a given item in the 3PL model for one group of students (first calibration). Let $a_2$ and $b_2$ be the same parameters for another group of students (second calibration). Linking the second set of item parameters to the first set requires determination of two constants $A$ and $B$ that set the following two transformations:

$$a_2 = Aa_1,$$

$$b_2 = \frac{(b_1 - B)}{A}.$$

Note that these transformation equations do not involve the pseudo-chance (“c”) item parameter. When the two sets of calibrated item parameters are perfectly linked, the two constants $A$ and $B$ are identical for all linkers. However, due to sampling variations, the
equations usually do not hold for all these items. So a
statistical process has to be used to find to constants $A$
and $B$ that fit (in some statistical sense) the calibrated
parameters of all items. Once the fitting constants $A$
and $B$ are found, then can be used to find the estimated
parameters of items what are not part of the linking
items.

There are a variety of methods to fit the linking
constants $A$ and $B$ to the calibrated item data of the
linkers. Among them are the Mean-Mean, Mean-Sigma,
Haebra, and Stocking-Lord methods (Haebaer, 1980;
Loyd & Hoover, 1980; Marco, 1977; Stocking & Lord,
1983). A new method that is based on robust $z_R$
statistics is described next.

The general process of the robust $z_R$ method is
described as follows. Equation 1 can be written as:

$$\log(A) = \log(a_1) - \log(a_1).$$

By writing $R_1 = \log(a_1)$ and $R_2 = \log(a_2)$, it can be seen that the robust $z_R$
procedure described for the Rasch model can be extended to
identify items that are unstable in the “log (a)” (slope)
parameter and needs to be deleted in the linking process.

The value of $\log(A)$ can then be taken as the mean of the
discrepancy $\log(a_2) - \log(a_1)$ of all surviving linkers. The
value of $A$ will then be computed via the formula $A = \exp[\log(A)]$. Applying this $A$ value to Equation 2 for all
linkers that have survived so far, Equation 2 now
becomes $B = Ab_2 - b_1$. By writing $R_1 = b_1$ and $R_2 =
Ab_2$, it can be seen again that the robust $z_R$
procedure described for the Rasch model can be extended to
identify items that are unstable in the “b” (location)
parameter and needs to be deleted in the linking process.

Those that are stable in location may then be used to find
the linking constant $B$. This constant is the mean of the
discrepancy $(Ab_2 - b_1)$ taking over all surviving linking
items.

As noted at the beginning of this section of the
paper, the linking constants $A$ and $B$ are defined
theoretically using the slope (“a”) and location (“b”) item
parameters. The robust $z_R$ does not rely on the
pseudo-chance item parameter (“c”), and conduct the
data analysis sequentially, first with the slope (“a”) and
then follow this up with the location (“b”) parameter. It
may be noted that the “c” parameters are hard to estimate,
especially when the sample size is small. In fact,
in a number of situations, the “c” parameter has to be
fixed at a certain value or certain range to allow to
calibration process to converge. So it seems to make
sense to set aside the “c” parameter in fitting the linking
constants $A$ and $B$ to the item data. It may also be noted
that there are more sampling fluctuations for the slope
(“a”) than for the location (“b”) estimates. This seems to
justify the use of the robust $z_R$ for the slope parameter
first.

**Extension to Mixed-Format 2PPC Items**

Although this expression was developed for 3PL binary
(multiple-choice) items, it can also be applied to partial
credit items that follow a two-parameter partial credit
(2PPC) model (Muraki, 1992, 1993). It may be noted that
a 2PPC item with $k$ score categories can be explicitly
defined as (k-1) conditional 2PL binary items. The $j$-th
conditional binary item is the “item” that has only ($j$-1) and
$j$ as scores. (see Masters, 1982, for a formulation of
conditional items in the Rasch model.) All conditional
binary items have identical slope. The location parameter
of each conditional item is often called “threshold”
parameters in IRT software such as PARSCALE
(Muraki & Bock, 1997).

As for a 3PL model, for a set of 2PPC linking items,
the robust $z_R$ procedure starts with detecting items that
are unstable along the slope dimension and need to be
set aside in the linking process. Then the robust $z_R$
procedure is applied to detect any threshold that is
unstable along the location dimension and needs to be
deleted. Theoretically only thresholds that are unstable
need to be deleted from the calibration process.

As noted at the beginning of this section of the
paper, the linking constants $A$ and $B$ are defined
theoretically using the slope (“a”) and location (“b”) item
parameters. The robust $z_R$ does not rely on the
pseudo-chance item parameter (“c”), and conduct the
data analysis sequentially, first with the slope (“a”) and
then follow this up with the location (“b”) parameter. It
may be noted that the “c” parameters are hard to estimate,
especially when the sample size is small. In fact,
in a number of situations, the “c” parameter has to be
fixed at a certain value or certain range to allow to
calibration process to converge. So it seems to make
sense to set aside the “c” parameter in fitting the linking

**An Illustration**

The performance of the robust $z$ procedure will be
illustrated using a set of archival data from a large-scale
state assessment program. The data came from the
administration of a math test to 5th grade students in 2006. The math test has 40 multiple-choice items and two constructed response items (with four score categories 0, 1, 2, and 3). Group 1 (N = 4045) is comprised of one-third of all male students who had free or reduced priced lunch. Group 2 is comprised of one-third of all female students who had to pay their lunches (N = 3692). As can be seen from Table 1, these two groups differ considerably in terms of ability.

Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th>Student Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>4045</td>
<td>21.09</td>
<td>8.22</td>
<td>.86</td>
</tr>
<tr>
<td>Group 2</td>
<td>3692</td>
<td>28.67</td>
<td>8.30</td>
<td>.87</td>
</tr>
</tbody>
</table>

PARSCALE was used to estimate the item parameters for each group separately. Each calibration set the mean at zero and the standard deviation at one. The items parameters \((a_1, a_2; b_1, b_2; c_1, c_2)\) for Group 1 and Group 2 are reported in the Appendix A. Items with ID from 1 to 40 are for the 40 multiple-choice items. Data listed for the codes 41A, 41B, and 41C are the common slope and threshold parameter for the three non-zero scores 1, 2, and 3 of the first CR items. Similarly, data listed for the codes 42A, 42B, and 42C are the common slope and threshold parameter for the three non-zero scores 1, 2, and 3 of the second CR items.

The first set of robust \(z\) statistics were first computed for the difference \(A_{\text{diff}} = \log(a_1) - \log(a_2)\). Using the cut score 1.96 for the robust \(z\), two items (ID = 26 with \(z_R = 4.261\); and 38 with \(z_R = -2.88\)) were found to be “unstable” along the slope dimension. Based on the remaining 44 data items, the linking constant (on the log scale) is -0.19609, which is transformed back to \(A = 0.822\) on the original slope scale.

The second set of robust \(z\) statistics were computed for the difference \(B_{\text{diff}} = B_2 - b_1\) using the 44 “stable” items. Using also the cut score 1.96 for the robust \(z\), six items were found to be “unstable” along the location dimension. They are listed as follows: ID = 17 (\(z_R = 2.624\)); ID = 21 (\(z_R = -2.37\)); ID = 28 (\(z_R = -2.58\)); ID = 33 (\(z_R = 2.399\)); ID = 35 (\(z_R = 3.924\)); and ID = 44 (\(z_R = 1.987\)). Based on the remaining “stable” items, the linking constant was found to be \(B = 1.072\).

Appendix B lists a sample SAS program for the robust \(z\) analysis.

For illustration purposes, the software STUIRT was also used to find the linking constants \(A\) and \(B\) for the same set of data. Table 2 reports the results of various linking methods.

Table 2: Results of six linking processes

<table>
<thead>
<tr>
<th>Method</th>
<th>(A)</th>
<th>(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean/Mean</td>
<td>0.823</td>
<td>1.165</td>
</tr>
<tr>
<td>Mean/Sigma</td>
<td>0.770</td>
<td>1.130</td>
</tr>
<tr>
<td>Haebara</td>
<td>0.802</td>
<td>1.036</td>
</tr>
<tr>
<td>Stocking-Lord</td>
<td>0.823</td>
<td>1.078</td>
</tr>
<tr>
<td>Robust z</td>
<td>0.822</td>
<td>1.072</td>
</tr>
</tbody>
</table>

It is interesting to see that the robust \(z\) results are almost identical to those obtained from the Stocking-Lord procedure, yet the robust \(z\) method is computationally much simpler than the Stocking-Lord procedure. This finding applies only to this data set and should be considered as suggestive of an interesting possibility that needs further research.

Educational Importance

IRT models are widely used linking educational assessment. The robust \(z\) statistic has proved useful in detecting unstable items in the Rasch model. The statistic is intuitively appealing and simple to use. This paper extends the robust \(z\) to the 3PL binary items and 2PPC partial credit items. An illustration based on real data indicates that the linking constants \(A\) and \(B\) obtained from the robust \(z\) procedure are strikingly similar to those obtained from the well-known Stocking-Lord procedure. This observation applies only to this data set and should be considered as suggestive of an interesting possibility that needs further research.

References


## APPENDIX A

Item Calibration Data for Group 1 and Group 2

<table>
<thead>
<tr>
<th>ID</th>
<th>a1</th>
<th>a2</th>
<th>b1</th>
<th>b2</th>
<th>c1</th>
<th>c2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.729</td>
<td>0.650</td>
<td>1.585</td>
<td>0.676</td>
<td>0.134</td>
<td>0.110</td>
</tr>
<tr>
<td>2</td>
<td>0.846</td>
<td>0.782</td>
<td>0.635</td>
<td>-0.525</td>
<td>0.304</td>
<td>0.316</td>
</tr>
<tr>
<td>3</td>
<td>0.909</td>
<td>0.816</td>
<td>-0.378</td>
<td>-1.749</td>
<td>0.267</td>
<td>0.161</td>
</tr>
<tr>
<td>4</td>
<td>0.818</td>
<td>0.787</td>
<td>-0.100</td>
<td>-1.092</td>
<td>0.176</td>
<td>0.149</td>
</tr>
<tr>
<td>5</td>
<td>0.742</td>
<td>0.611</td>
<td>-0.195</td>
<td>-1.619</td>
<td>0.215</td>
<td>0.145</td>
</tr>
<tr>
<td>6</td>
<td>0.890</td>
<td>0.888</td>
<td>0.749</td>
<td>-0.406</td>
<td>0.194</td>
<td>0.200</td>
</tr>
<tr>
<td>7</td>
<td>1.741</td>
<td>1.192</td>
<td>1.246</td>
<td>-0.132</td>
<td>0.267</td>
<td>0.243</td>
</tr>
<tr>
<td>8</td>
<td>0.907</td>
<td>0.589</td>
<td>1.016</td>
<td>0.006</td>
<td>0.159</td>
<td>0.059</td>
</tr>
<tr>
<td>9</td>
<td>1.487</td>
<td>1.211</td>
<td>-0.234</td>
<td>-1.352</td>
<td>0.095</td>
<td>0.081</td>
</tr>
<tr>
<td>10</td>
<td>1.228</td>
<td>0.742</td>
<td>0.537</td>
<td>-0.872</td>
<td>0.197</td>
<td>0.075</td>
</tr>
<tr>
<td>11</td>
<td>0.672</td>
<td>0.526</td>
<td>0.070</td>
<td>-1.242</td>
<td>0.089</td>
<td>0.028</td>
</tr>
<tr>
<td>12</td>
<td>1.007</td>
<td>0.690</td>
<td>1.985</td>
<td>0.873</td>
<td>0.272</td>
<td>0.267</td>
</tr>
<tr>
<td>13</td>
<td>1.016</td>
<td>0.996</td>
<td>1.101</td>
<td>0.239</td>
<td>0.229</td>
<td>0.242</td>
</tr>
<tr>
<td>14</td>
<td>0.776</td>
<td>0.816</td>
<td>-0.742</td>
<td>-2.038</td>
<td>0.159</td>
<td>0.189</td>
</tr>
<tr>
<td>15</td>
<td>0.921</td>
<td>0.781</td>
<td>0.463</td>
<td>-0.487</td>
<td>0.162</td>
<td>0.184</td>
</tr>
<tr>
<td>16</td>
<td>0.550</td>
<td>0.507</td>
<td>-0.060</td>
<td>-1.372</td>
<td>0.100</td>
<td>0.121</td>
</tr>
<tr>
<td>17</td>
<td>0.624</td>
<td>0.378</td>
<td>0.477</td>
<td>-1.492</td>
<td>0.259</td>
<td>0.000</td>
</tr>
<tr>
<td>18</td>
<td>0.984</td>
<td>0.976</td>
<td>1.084</td>
<td>0.214</td>
<td>0.167</td>
<td>0.170</td>
</tr>
<tr>
<td>19</td>
<td>0.506</td>
<td>0.473</td>
<td>-2.340</td>
<td>-4.537</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>20</td>
<td>0.594</td>
<td>0.364</td>
<td>1.068</td>
<td>0.220</td>
<td>0.242</td>
<td>0.151</td>
</tr>
<tr>
<td>21</td>
<td>0.687</td>
<td>0.585</td>
<td>-0.055</td>
<td>-0.686</td>
<td>0.323</td>
<td>0.383</td>
</tr>
<tr>
<td>22</td>
<td>0.541</td>
<td>0.566</td>
<td>-1.045</td>
<td>-2.394</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>23</td>
<td>0.691</td>
<td>0.511</td>
<td>1.859</td>
<td>0.747</td>
<td>0.196</td>
<td>0.195</td>
</tr>
<tr>
<td>24</td>
<td>0.843</td>
<td>0.718</td>
<td>0.645</td>
<td>-0.467</td>
<td>0.189</td>
<td>0.177</td>
</tr>
<tr>
<td>25</td>
<td>0.530</td>
<td>0.354</td>
<td>-0.689</td>
<td>-3.629</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>26</td>
<td>0.462</td>
<td>1.080</td>
<td>-2.583</td>
<td>-5.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>27</td>
<td>1.007</td>
<td>0.840</td>
<td>1.922</td>
<td>0.927</td>
<td>0.334</td>
<td>0.352</td>
</tr>
<tr>
<td>28</td>
<td>0.825</td>
<td>0.865</td>
<td>0.709</td>
<td>0.305</td>
<td>0.538</td>
<td>0.647</td>
</tr>
<tr>
<td>29</td>
<td>0.608</td>
<td>0.528</td>
<td>0.499</td>
<td>-0.839</td>
<td>0.125</td>
<td>0.116</td>
</tr>
<tr>
<td>30</td>
<td>1.177</td>
<td>0.814</td>
<td>1.973</td>
<td>1.270</td>
<td>0.511</td>
<td>0.501</td>
</tr>
<tr>
<td>31</td>
<td>0.900</td>
<td>0.555</td>
<td>0.104</td>
<td>-1.618</td>
<td>0.192</td>
<td>0.000</td>
</tr>
<tr>
<td>32</td>
<td>0.861</td>
<td>0.701</td>
<td>0.809</td>
<td>-0.091</td>
<td>0.353</td>
<td>0.286</td>
</tr>
<tr>
<td>33</td>
<td>0.843</td>
<td>0.530</td>
<td>0.640</td>
<td>-1.228</td>
<td>0.103</td>
<td>0.000</td>
</tr>
<tr>
<td>34</td>
<td>1.404</td>
<td>1.220</td>
<td>0.247</td>
<td>-1.019</td>
<td>0.241</td>
<td>0.248</td>
</tr>
<tr>
<td>35</td>
<td>0.446</td>
<td>0.344</td>
<td>0.820</td>
<td>-1.453</td>
<td>0.245</td>
<td>0.064</td>
</tr>
<tr>
<td>36</td>
<td>1.014</td>
<td>0.966</td>
<td>1.837</td>
<td>1.090</td>
<td>0.118</td>
<td>0.150</td>
</tr>
<tr>
<td>37</td>
<td>1.632</td>
<td>1.044</td>
<td>2.129</td>
<td>1.743</td>
<td>0.155</td>
<td>0.126</td>
</tr>
<tr>
<td>38</td>
<td>0.831</td>
<td>0.358</td>
<td>1.012</td>
<td>-1.436</td>
<td>0.132</td>
<td>0.000</td>
</tr>
<tr>
<td>39</td>
<td>1.560</td>
<td>1.192</td>
<td>1.774</td>
<td>1.024</td>
<td>0.215</td>
<td>0.187</td>
</tr>
<tr>
<td>40</td>
<td>0.798</td>
<td>0.615</td>
<td>0.095</td>
<td>-1.358</td>
<td>0.148</td>
<td>0.007</td>
</tr>
<tr>
<td>41A</td>
<td>0.561</td>
<td>0.486</td>
<td>0.784</td>
<td>-0.539</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>41B</td>
<td>0.561</td>
<td>0.486</td>
<td>-0.113</td>
<td>-1.489</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>41C</td>
<td>0.561</td>
<td>0.486</td>
<td>1.166</td>
<td>-0.052</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>42A</td>
<td>0.745</td>
<td>0.737</td>
<td>3.687</td>
<td>2.599</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>42B</td>
<td>0.745</td>
<td>0.737</td>
<td>2.506</td>
<td>1.250</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>42C</td>
<td>0.745</td>
<td>0.737</td>
<td>-0.001</td>
<td>-1.209</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
APPENDIX B

Sample SAS Program for Robust z Analysis for Slope and Location Parameters

```sas
%let CV = 1.96;
Data raw;
Input   ID   $ a1  a2  b1  b2 ;
alog = log(a1);
a2log = log(a2);
Adiff = a2log-alog;
Kode  = 1;
Cards;
* Datelines are here;
;
proc univariate data=raw noprint ; var Adiff; output out=summary Median= MD Mean=Mean Qrange=IQR;
data new1; set summary; kode=1;
data new2; merge raw new1 by kode;
data new3; set new2; zrobust=(Adiff-MD)/(0.74*IQR);
   code ="A-unstable";
   if abs(zrobust) le &CV then code ="A-stable";
proc print data=new3 noobs; var ID a1 a2 allog a2log Adiff zrobust code;
   format a1 a2 allog a2log Adiff zrobust 5.3;
run;

data new4; set new3; if code="A-stable";
proc means data=new4 noprint; var Adiff; output out=LC1 mean=AlogCon;
proc print data=LC1; var AlogCon;
data LC2; set LC1; Kode=1;
   Acon=exp(AlogCon);
proc print data=LC2; var Acon;

data new5; merge new4 LC2 by Kode;
data new6; set new5; Bdiff=b1-Acon*b2;
   keep ID Kode a1 a2 b1 b2 Bdiff;
   proc univariate data=new6 noprint ; var Bdiff; output out=summary2 Median= MD Mean=Mean
   Qrange=IQR;
data summary3; set summary2; kode=1;
data new8; set new6; if code="B-stable";
   proc print data=new8 noobs; var ID b1 b2 Bdiff zrobust code;
   format b1 b2 Bdiff zrobust 5.3;
data new10; set new9; if code="B-stable";
   proc means data=new10; var Bdiff;
run;
```

Citation

Authors

Huynh Huynh
College of Education
University of South Carolina
Columbia, SC 29208.
Email: hhuynh [at] mailbox.sc.edu.

J. Patrick Meyer
Curry School of Education
University of Virginia
P.O. Box 400277
405 Emmet Street South
Charlottesville, VA 22904.
Email: meyerjp [at] virginia.edu