Comparing Responses to Dichotomous Attributes in Single-Case Designs

Paul R. Yarnold, Ph.D.
Optimal Data Analysis, LLC

Imagine a single case is assessed on two or more binary attributes over a series of measurements or testing sessions. For example, a chronic pain patient might rate their pain and fatigue each evening before sleep, as “worse than average” or “better than average”. Alternatively, imagine following two or more different corporate stocks over a given time period, keeping a record at the end of every trading day regarding whether the closing price was higher or lower than on the prior day. Both examples are single-case series, and in both examples an interesting question is whether the binary responses to the different attributes differ. Does the patient have more bad pain days than bad fatigue days? Does one stock close higher more often than another stock? This article discusses how to accomplish such analysis and address such questions using UniODA. The example selected to illustrate the method provides clear evidence that ipsative standardization is critical in obtaining meaningful results in the analysis of serial data.

A common practice across disciplines is the conversion of an ordered measure into a binary measure on the basis of an investigator-defined split of the scale into dichotomous categories. Common splitting methods used, for example, to create binary class variables (used as the dependent variable in discriminant analysis) or attributes (used as independent variables in logistic regression analysis) are: mean-based splits (higher or lower than the mean); median-based splits (higher or lower than the median); quartile-based splits (exclude middle quartiles); and z-score-based splits (exclude observations with absolute z<1). Fraught with perils such as errors of omission—missing true effects by missing appropriate splits, and of commission—finding spurious effects using splits that induce paradoxical confounding, this practice nevertheless remains widely used.¹ A recent discovery suggests use of statistically unmotivated splits is further complicated by the differential use of physically identical response scales, for example Likert-type scales, by a single individual rating different phenomena.² Said another way, the rater’s perception of phenomena interacts with the rater’s interpretation of and use of the scales response options. The example used presently demonstrates this latter peril.

Data were obtained from a log of 297 consecutive ratings of pain and fatigue made by a patient with fibromyalgia.² Symptoms were rated on an 11-point Likert-Type scale ranging
from 0 (“not at all bothersome”) to 10 (“extremely bothersome”). Analysis is conducted to assess if the patient rated these symptoms similarly: that is, were symptoms experienced with comparable intensity or did ratings of one symptom exceed ratings of the other symptom—in which case the former is described as being dominant or primary versus the latter. Descriptive data for these 297 pain and fatigue ratings are given elsewhere.

Analysis Involving Raw Data

For exposition, symptom data are split using the mean- or median-based methods that happen to be the same presently (this is not an uncommon occurrence). These are among the more frequently-used splitting methods because they offer the advantage of maximizing sample size and statistical power. Therefore, the binary pain attribute is dummy-coded as “0” if the raw rating is less than 6 (average or less severe) and as “1” otherwise (more severe than average). As is common practice, the same strategy was used to dummy-code the binary fatigue attribute.

As is done for similar analysis involving ordered attributes, the analysis begins by constructing a new data set (new.txt) with 2n observations (2*297=594), each of which forms a row in new.txt. First, copy all 297 binary pain ratings (0 or 1) to new.txt (which at this point has 297 rows). Next, beneath these, copy all 297 binary fatigue ratings (new.txt now has 594 rows). Add the number “0” (an arbitrary class category dummy-code) delimited by a space to the beginning of each of the first (top) set of 297 rows, and then likewise add the number “1” (arbitrary class dummy-code) delimited using a space to the beginning of each of the second (bottom) set of 297 rows. Using new.txt as input, the post hoc hypothesis that the patient had a different proportion of bad pain days than of bad fatigue days across the series of 297 entries is tested running the following UniODA code (control commands are indicated in red):

```
OPEN new.txt;
OUTPUT example.out;
VARS class rating;
CLASS class;
ATTR rating;
CATEGORICAL rating;
GO;
```

The CATEGORICAL (CAT) command is used to indicate categorical attributes. Monte Carlo simulation is not run to estimate p because UniODA software computes and provides exact p for binary applications. The resulting UniODA model was: if rating=0 (average or below) then predict attribute=0 (pain); otherwise predict fatigue. This model indicates that fatigue ratings were greater than pain ratings (p<0.0001). The confusion table for the model is given in Table 1: ESS and ESP are both 44.8, reflecting moderate strength.

<table>
<thead>
<tr>
<th>Predicted Symptom</th>
<th>Actual Symptom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td>213</td>
</tr>
<tr>
<td>Fatigue</td>
<td>84</td>
</tr>
</tbody>
</table>

71.7%

<table>
<thead>
<tr>
<th>Predicted Symptom</th>
<th>Actual Symptom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pain</td>
<td>80</td>
</tr>
<tr>
<td>Fatigue</td>
<td>217</td>
</tr>
</tbody>
</table>

73.1%

Analysis Involving Ipsatively Standardized Data

Ipsative standardization of raw data into z-score form is done using the mean and SD computed for data from an observation (not a sample of observations), and it is appropriate for analysis of serial data as a means of eliminating variability attributable to “base-rate” differences between observations that essentially introduce noise into the data. Analyses performed on raw
data are thus repeated here after pain and fatigue ratings have been ipsatively standardized. Thus, ipsatively standardized pain scores are coded as “0” if $z_{\text{pain}} \leq 0$, and as “1” otherwise, and the ipsatively standardized fatigue ratings are coded in a likewise manner.

The analysis begins as before: construct a new data set (new.txt) having $2n$ observations (here, 594), each forming a row in new.txt. Copy all 297 $z_{\text{pain}}$–based binary codes to new.txt (at this point new.txt has 297 rows), then beneath these copy all 297 $z_{\text{fatigue}}$–based binary codes (new.txt now has 594 rows). Add the number “0” (an arbitrary class category dummy-code) delimited by a space to the beginning of each of the first (top) set of 297 rows, and then likewise add the number “1” (arbitrary dummy-code) delimited by a space to the beginning of each of the second (bottom) set of 297 rows. Using new.txt as input, the post hoc hypothesis that the patient had a different proportion of bad pain days than of bad fatigue days across the series of 297 entries is tested running the same UniODA code used for raw score analysis with the following changes:

- **VARS** class Zrating;
- **ATTR** Zrating;
- **CAT** Zrating;

In stark contrast to findings based on raw data, the results for ipsatively standardized data indicate no statistical support for the post hoc hypothesis ($p < 0.19$), and model ESS and ESP were both extremely weak at less than 5.8. The confusion table is shown in Table 2. Which of these results reflects the truth? In the end this case calls for selecting between the definition of an average rating made by any researcher who uses statistically unmotivated numerical thresholds that define psychological conditions to all patients’ data; or the individual patients who are making the ratings to describe their cognitive and emotional experience.

Table 2: Confusion Table for UniODA Model Discriminating Binary Pain and Fatigue Ratings Based on Split of Ipsatively Standardized Data

<table>
<thead>
<tr>
<th>Predicted Symptom</th>
<th>Pain</th>
<th>Fatigue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Pain</td>
<td>149</td>
<td>148</td>
</tr>
<tr>
<td>Symptom Fatigue</td>
<td>132</td>
<td>165</td>
</tr>
<tr>
<td></td>
<td>53.0%</td>
<td>52.7%</td>
</tr>
</tbody>
</table>

References


Author Notes

E-mail: Journal@OptimalDataAnalysis.com.

Mail: Optimal Data Analysis, LLC
1220 Rosecrans St., #330
San Diego, CA 9210

ODA Blog: http://odajournal.wordpress.com