ABSTRACT: Data are often indexed by a set of labels that reflect certain experimental conditions of interest. When these labels have, in addition, some symmetry in their particular structure, the methodology of symmetry studies (Viana, 2008) can be used to facilitate the analysis and interpretation of data. In the present work, the symmetry-related properties derived for the study of data indexed by quaternary sequences in length of three are studied in detail within that context.

KEYWORDS: structured data; permutation groups; canonical projection; sum of squares decomposition.

1 Introduction

Many experimental situations justify the use of a set of symbols to index the data. Consider, for example, the set \( S = \{ a, g, c, t \} \) where symbols \( a \) (adenine), \( g \) (guanine), \( c \) (cytosine) and \( t \) (thymine) are the four DNA nucleotides. If the data are the frequencies with which the sequences in length of three appear in a given genome, then the results \( X = \{ x_{aaa}, x_{aag}, \ldots, x_{ttt} \} \) are indexed by the sixty-four possible concatenations of three of the elements of \( S \). This is an example of structured data and the underlying symmetries imply a partition of the set of indexes that reflect the properties of interest. The appropriate tool to describe these properties is the action of a group into the set of indexes. In this case, the study is extremely enriched by the extensive background available on the action of groups on sets of symbols.

In Section 2 we outline the basic definitions of group theory, starting with the concept of a group, its linear representation, group characters and the canonical projections theorem for finite groups, following Viana (2008), Diaconis (1988) and Serre (1977). In Section 3 we show, specifically for the symmetric groups \( S_3 \) and \( S_4 \) and the dihedral group \( D_4 \), the underlying algebraic properties associated with the canonical projectors obtained when these groups act on a set of symbols. A more complete study of that subject may be found in Souza (2009).
2 Group representations, characters and canonical projections

A pair \((G, \ast)\), where \(G\) is a set and \(\ast\) a closed associative binary operation constitutes a group if \(G\) has an identity and each element in \(G\) has an inverse also in \(G\). It is usual to refer to the group \((G, \ast)\) as the group \(G\). If \(\text{GL}(V)\) are the non-singular linear maps in a vector space \(V\) then any mapping \(\rho : G \rightarrow \text{GL}(V)\) is a linear representation of the group \(G\) if \(\rho(s)\rho(t) = \rho(st)\). The dimension of the representation is the dimension of the vector space \(V\). If \(W\) is a subspace of \(V\) such that, for all \(w \in W\) and all \(s, t \in G\), \(\rho(s)w \in W\), then \(W\) is a called a stable subspace and the homomorphism \(G \rightarrow \text{GL}(W)\) a sub-representation. The zero subspace and the subspace \(W = V\) are the trivial subspaces. A representation \(\rho\) is said irreducible if it admits no sub-representations except the trivial. Every representation is a direct sum of irreducible representations.

In this work we are interested in the groups of permutations of \(n\) objects, the so-called symmetric groups \(S_n\). The importance of symmetric groups is stressed by Cayley theorem stating that “any finite group is isomorphic to a subgroup of a symmetric group”.

Given a finite set \(X\) and a group \(G\), an action of \(G\) on \(X\) is a function \(\varphi : G \times X \rightarrow X\) with the properties:

- \(\varphi(1, x) = x\) for all \(x \in X\)
- \(\varphi(s, \varphi(t, x)) = \varphi(st, x)\) for all \(x \in X\)

and all \(s, t \in G\). If \(X = G\) then the action is called regular.

The trace of a square matrix \(A = (a_{ij})\) is defined as \(\text{tr}(A) = \sum a_{ii}\). Given a linear representation (in dimension of \(n\)) \(\rho : G \rightarrow \text{GL}(V)\), the scalar function \(\chi_\phi(s) = \text{tr}(\rho_s)\) is called the character of the representation, with the properties:

- \(\chi(1) = n\)
- \(\chi(s^{-1}) = \chi(s)\)

(where \(\chi(s)\) indicates, from now on, the complex conjugation) and \(\chi(t) = \chi(st)\).

We list some important results involving character theory and irreducible representations:

- If \(\chi\) and \(\psi\) are characters of non isomorphic irreducible representations of a group \(G\) then \((\chi | \psi) = 1\) and \((\chi | \psi) = 0\), where \((\cdot | \cdot)\) is the usual scalar product \(\frac{1}{|G|} \sum_{s \in G} \chi(s) \psi(s)^*\). It means that the characters of irreducible representations form an orthonormal system.
- The number of irreducible representations of a group \(G\) (up to isomorphism) is the number of conjugacy classes. Recall that the equivalence relation \(t' = sts^{-1}\) \((t, t', s \in G)\) partitions \(G\) into conjugacy classes.
- The character \(r_G\) of the regular representation of a group \(G\) of order \(g\) is given by \(r_G(1) = g\) and \(r_G(s) = 0\) if \(s \neq 1\).
- (Canonical Decomposition Theorem) Let \(\rho : G \rightarrow \text{GL}(V)\) be a representation of a group \(G\) and \(V = U_1 \oplus \ldots \oplus U_m\) a decomposition (direct sum) of the space \(V\) in irreducible subspaces. Let \(\chi_1, \ldots, \chi_h\) be the characters (with \(n_1, \ldots, n_h\) the
corresponding degrees) of the distinct irreducible representations $\rho_1, \ldots, \rho_h$ ($\rho_i : G \rightarrow \text{GL}(W_i)$). Let $V_i$ be the direct sum of those $U_j$ which are isomorphic to $W_i$. The decomposition $V = V_1 \oplus \cdots \oplus V_h$ is the so-called canonical decomposition. It does not depend on the initially decomposition in irreducible representations and the (canonical) projector $P_i$ of $V$ into $V_i$ is given by the matrix

$$P_i = \frac{n_i}{g} \sum_{s \in G} \chi_i^s(v) \rho_s.$$ 

The matrices $P_i$ have some important properties: $P_i^2 = P_i$ (they are projectors), $P_i P_j = P_j P_i = 0$ for $i \neq j$ (they are algebraically orthogonal) and $P_1 + \cdots + P_h = I$ (they add to the identity). If $X = (X_1, \ldots, X_h)$ is a vector data in $V$, with $\mu = (\mu_1, \ldots, \mu_h)$ the vector of means, the projectors $P_i$ decompose the sum of squares $\sum_{j=1}^h X_j^2$ into orthogonal parts $\sum_{j=1}^h X_j^2 P_i X_j$, which are fundamental in the application of the canonical decomposition to the analysis of continuous univariate normal data. Each quadratic form $(X_j^2 P_i X_j)$ carries an algebraic property, expressed by $\mu_j P_i \mu = 0$, that can be tested. Note that, if each $X_j$ can be considered to have normal distribution, these algebraic properties are the null hypotheses to be tested in the usual analyses of variance framework.

Usually, the first canonical projector $P_1$ corresponds to the space of constants, so that it has always the form of a $n \times n$ matrix with all components equal to $1/n$, where $n$ is the dimension of the representation. We write $P_1 = A_n$ and $I_n - A_n = Q_n$. These are called the standard projectors or the standard decomposition and they are the only non-null projectors obtained when the symmetric group $S_n$ acts naturally in the set $\{1, 2, \ldots, n\}$.

### 3 The action of a group on a set of symbols

The set $V = \{aaa, aag, aag, \ldots, ttt\}$ of indexes considered in the introduction, may be seen as the set $V = \{f : L \rightarrow S\}$ of functions from $L = \{1, 2, 3\}$ to $S = \{a, g, c, t\}$. Given a group $G$, we may consider the left action (position symmetries) $f \sigma^{-1}$ of subgroups of the symmetric group $S_3$ on the set of positions $L$ and the right action (symbol symmetries) $\sigma f$ of subgroups of $S_3$ on $S$. Each of these actions gives a partition of $V$ in subsets (orbits) where the group acts transitively. Each orbit is taken, then, as a set of indexes for a space of representation of the group, where the canonical projectors are constructed and corresponding analytic properties (when formulated as null hypotheses) are determined.

To illustrate the left action of $S_3$, consider the genome of the HIV type 1 virus (available at the National Center for Biotechnology Information, under accession number
K02013) and the experiment in which we count the occurrences of each DNA word in a given orbit, along six consecutive subregions, each one in length of 900 Bps. The purpose is the study of the analytic properties associated with the canonical projectors, formulated as parametric null hypotheses.

In what follows we always suppose that the first canonical projector \( P_1 \) corresponds to the space of constants and its interpretation will be omitted.

3.1 The symmetric group \( S_3 \) and its left action on the set \( V \)

We write \( S_3 = \{1,(123),(132),(23),(12),(13)\} \). Its left action on the set \( V \) results in four orbits in size of one, of the form \( \{aaa\} \), four orbits in size of six, of the form \( \{(agc),(gca),(cag),(gac),(cga),(acg)\} \) and twelve orbits in size of three, of the form \( \{(aag),(aga),(gag)\} \).

As the size one orbits are of no interest we consider only the following two cases:

3.1.1 The left action of \( S_3 \) on the transitive parts of size three. It gives the standard projectors

\[
A_3 = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad Q_3 = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}.
\]

Considering a data vector \( X = (X_1, X_2, X_3) \), the resulting decomposition is given by:

\[
X'X = X'A + X'QX \\
\sum x_i^2 = 3\bar{x}^2 + (X_i - \bar{X})^2.
\]

The null hypotheses associated to \( Q_3 \) is

\[ H_0 : \mu_1 = \mu_2 = \mu_3. \]

The following tables (Tables 1, 2A and 2B) are the counting data for each orbit in six contiguous regions of the genome and the corresponding analysis of variance (where SS stands for sum of squares and DF stands for degree of freedom). The asymptotic normal distribution of counting data and their independence (e.g. Diaconis, 1988) are assumed in this preliminary analysis of the data. To better understand the data table consider the column correspondent to orbit \( \{aag,aga,gaa\} \) and note that the first six numbers \( (34,26,27,37,29,23) \) are the frequencies of word \( aag \) at the six regions \( R_1, R_2, R_3, R_4, R_5 \) and \( R_6 \). The next six numbers \( (29,34,32,33,24,24) \) are the frequencies of word \( aga \), and so on.
Table 1 - Counting data in each orbit generated by the left action of $S_3$ on the transitive parts of size three in six contiguous regions of the genome

<table>
<thead>
<tr>
<th>Orbit</th>
<th>aag</th>
<th>aac</th>
<th>aat</th>
<th>agg</th>
<th>acc</th>
<th>att</th>
<th>ggc</th>
<th>ggt</th>
<th>gtc</th>
<th>ctt</th>
<th>ctt</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_1$</td>
<td>34</td>
<td>15</td>
<td>21</td>
<td>29</td>
<td>15</td>
<td>20</td>
<td>13</td>
<td>7</td>
<td>13</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>$R_2$</td>
<td>26</td>
<td>16</td>
<td>30</td>
<td>16</td>
<td>12</td>
<td>23</td>
<td>6</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>$R_3$</td>
<td>27</td>
<td>20</td>
<td>34</td>
<td>19</td>
<td>13</td>
<td>20</td>
<td>9</td>
<td>5</td>
<td>7</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>$R_4$</td>
<td>37</td>
<td>13</td>
<td>26</td>
<td>25</td>
<td>3</td>
<td>23</td>
<td>12</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>$R_5$</td>
<td>29</td>
<td>14</td>
<td>22</td>
<td>22</td>
<td>8</td>
<td>12</td>
<td>12</td>
<td>6</td>
<td>13</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>$R_6$</td>
<td>23</td>
<td>11</td>
<td>34</td>
<td>13</td>
<td>14</td>
<td>17</td>
<td>7</td>
<td>12</td>
<td>8</td>
<td>10</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 2A - Analysis of variance for the counting data in each orbit generated by the left action of $S_3$ on the transitive parts of size three in six contiguous regions of the genome

<table>
<thead>
<tr>
<th>Orbit</th>
<th>aag</th>
<th>aac</th>
<th>aat</th>
<th>agg</th>
<th>acc</th>
<th>att</th>
<th>ggc</th>
<th>ggt</th>
<th>gtc</th>
<th>ctt</th>
<th>ctt</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_1$</td>
<td>41</td>
<td>22</td>
<td>15</td>
<td>26</td>
<td>20</td>
<td>10</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>$R_2$</td>
<td>37</td>
<td>28</td>
<td>21</td>
<td>24</td>
<td>23</td>
<td>22</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>$R_3$</td>
<td>37</td>
<td>26</td>
<td>24</td>
<td>25</td>
<td>15</td>
<td>19</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>$R_4$</td>
<td>33</td>
<td>35</td>
<td>23</td>
<td>25</td>
<td>15</td>
<td>19</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>$R_5$</td>
<td>34</td>
<td>24</td>
<td>30</td>
<td>22</td>
<td>16</td>
<td>15</td>
<td>3</td>
<td>8</td>
<td>0</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>$R_6$</td>
<td>30</td>
<td>27</td>
<td>30</td>
<td>16</td>
<td>18</td>
<td>12</td>
<td>2</td>
<td>17</td>
<td>1</td>
<td>16</td>
<td>6</td>
</tr>
</tbody>
</table>

**Analysis of Variance**

<table>
<thead>
<tr>
<th>Source</th>
<th>Degree of Freedom</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS treat</td>
<td>15</td>
<td>454,00</td>
<td>30,27</td>
<td>3,48</td>
<td>0.057</td>
</tr>
<tr>
<td>SS res</td>
<td>15</td>
<td>314,50</td>
<td>20,97</td>
<td>2.39</td>
<td>0.125</td>
</tr>
<tr>
<td>SS tot</td>
<td>17</td>
<td>768,50</td>
<td>44,63</td>
<td>5,06</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 2B - Analysis of variance with the counting data for each orbit generated by the left action of $S_3$ on the transitive parts of size three in six contiguous regions of the genome

<table>
<thead>
<tr>
<th>Orbit:</th>
<th>g g c</th>
<th>g g t</th>
<th>g c c</th>
<th>G t t</th>
<th>c c t</th>
<th>c t t</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS treat = $X'(Q_3 \odot A_3)X$</td>
<td>294,78</td>
<td>395,11</td>
<td>313,44</td>
<td>32,33</td>
<td>24,33</td>
<td>1,44</td>
</tr>
<tr>
<td>DF treat = $tr(Q_3 \odot A_3)$</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>SS res = $X'(I_3 \odot Q_3)X$</td>
<td>55,00</td>
<td>151,33</td>
<td>44,83</td>
<td>273,67</td>
<td>176,17</td>
<td>177,67</td>
</tr>
<tr>
<td>DF res = $tr(I_3 \odot Q_3)$</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>SS tot = $X'(I_{15} - A_3 \odot A_3)X$</td>
<td>349,78</td>
<td>546,44</td>
<td>358,28</td>
<td>306,00</td>
<td>200,50</td>
<td>179,11</td>
</tr>
<tr>
<td>DF tot = $tr(I_{15} - A_3 \odot A_3)$</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>$F_0$</td>
<td>40,20</td>
<td>19,58</td>
<td>52,43</td>
<td>0,89</td>
<td>0,40</td>
<td>0,06</td>
</tr>
<tr>
<td>P-value</td>
<td>0,00</td>
<td>0,0001</td>
<td>0,00</td>
<td>0,4328</td>
<td>0,3789</td>
<td>0,9411</td>
</tr>
</tbody>
</table>

From Tables 2A and 2B it is evident that orbits such as att, gtt, cct, and ctt show a markedly different frequency distribution, relative to the remaining orbits, and should then find an explanation within the molecular biology context. As indicated earlier, the determination of what properties might be of contextual interest is precisely the objective of the symmetry studies. Clearly, in this first example, such property (equality of counts or uniform distribution) is expected because it is the consequence of the standard reduction. Under different symmetries, however, the properties of interest are not evident. The methodology will give then a precise description of what these properties are.

### 3.1.2 The left action of $S_3$ on the transitive parts of size six

The following projectors (and corresponding null hypotheses) are obtained

$$P_2 = \frac{1}{6} \begin{bmatrix} 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} = Q_2 \odot A_3$$

$$H_0: \mu_1 + \mu_2 + \mu_3 = \mu_4 + \mu_5 + \mu_6$$

$$P_3 = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & -1 & -1 \\ 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & -1 & -1 & 2 \end{bmatrix} = I_2 \odot Q_3$$

$$H_0: \begin{cases} \mu_1 = \mu_2 = \mu_3 \\ \mu_4 = \mu_5 = \mu_6 \end{cases}$$
The counting data for each orbit and the analysis of variance:

Table 3 - Counting data in each orbit generated by the left action of $S_3$ on the transitive parts of size six in six contiguous regions of the genome.

<table>
<thead>
<tr>
<th>Orbit</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
<th>$R_5$</th>
<th>$R_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>agc</td>
<td>23</td>
<td>11</td>
<td>18</td>
<td>18</td>
<td>28</td>
<td>13</td>
</tr>
<tr>
<td>cag</td>
<td>18</td>
<td>12</td>
<td>22</td>
<td>21</td>
<td>24</td>
<td>16</td>
</tr>
<tr>
<td>gca</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>acg</td>
<td>33</td>
<td>30</td>
<td>27</td>
<td>34</td>
<td>22</td>
<td>24</td>
</tr>
<tr>
<td>cga</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>gac</td>
<td>16</td>
<td>15</td>
<td>10</td>
<td>11</td>
<td>17</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Orbit</th>
<th>$R_1$</th>
<th>$R_2$</th>
<th>$R_3$</th>
<th>$R_4$</th>
<th>$R_5$</th>
<th>$R_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>act</td>
<td>12</td>
<td>16</td>
<td>11</td>
<td>8</td>
<td>18</td>
<td>11</td>
</tr>
<tr>
<td>tac</td>
<td>13</td>
<td>7</td>
<td>14</td>
<td>10</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>cta</td>
<td>16</td>
<td>13</td>
<td>16</td>
<td>13</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>atc</td>
<td>9</td>
<td>17</td>
<td>10</td>
<td>15</td>
<td>10</td>
<td>17</td>
</tr>
<tr>
<td>tca</td>
<td>7</td>
<td>13</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>cat</td>
<td>9</td>
<td>18</td>
<td>14</td>
<td>18</td>
<td>21</td>
<td>16</td>
</tr>
</tbody>
</table>

Again here, we observe a contrasting counting distribution among the orbits, and this would find its explanation within the context of the virus and its eventual strategies (e.g., Doi (1991)).

Table 4 - Analysis of variance for the counting data in each orbit generated by the left action of $S_3$ on the transitive parts of size six in six contiguous regions of the genome.

<table>
<thead>
<tr>
<th>Orbit:</th>
<th>a g c</th>
<th>a g t</th>
<th>a c t</th>
<th>g c t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SS_3 = \chi (P \circ A_n) \chi$</td>
<td>2433.78</td>
<td>210.25</td>
<td>10.02</td>
<td>336.11</td>
</tr>
<tr>
<td>$DF_3 = tr(P \circ A_n)$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$SS_3 = \chi (P \circ A_n) \chi$</td>
<td>906.11</td>
<td>250.22</td>
<td>91.44</td>
<td>10.88</td>
</tr>
<tr>
<td>$DF_3 = tr(P \circ A_n)$</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>$SSres = \chi (I \circ Q_n) \chi$</td>
<td>492.67</td>
<td>892.50</td>
<td>324.83</td>
<td>164.00</td>
</tr>
<tr>
<td>$DFres = tr(I \circ Q_n)$</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>$F_0.2$</td>
<td>148.20</td>
<td>7.07</td>
<td>0.92</td>
<td>61.48</td>
</tr>
<tr>
<td>$F_0.4$</td>
<td>13.79</td>
<td>2.10</td>
<td>2.11</td>
<td>4.70</td>
</tr>
<tr>
<td>P-value</td>
<td>0.000</td>
<td>0.105</td>
<td>0.104</td>
<td>0.005</td>
</tr>
</tbody>
</table>
Observe that $SS_2$ and $SS_3$ correspond, respectively, to projectors $P_2$ and $P_3$.

In the following cases we describe the analytic properties of interest without reporting the analysis of the counting data, which is available from Souza, D.J. (2009).

3.2 The symmetric group $S_4$ and its right action on the set $V$

We order $S_4$ as:

$$S_4 = \{1, (12)(34), (13)(24), (14)(23), (243), (142), (123), (134), (234), (132), (143), (124), (34), (12), (1423), (1324), (23), (1342), (1243), (14), (24), (1432), (13), (1234) \}.$$  

Its right action on the set $V$ results in one orbit in size of four, of the form

$$\{(aa)(gg)(cc)(mm)\}$$

three orbits in size of twelve,


and one orbit in size of twenty four


3.2.1 The right action of $S_4$ on the transitive parts of size four. We obtain the standard projectors $A_4$ and $Q_4$ with the null hypotheses $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$ associated to $Q_4$

3.2.2 The right action of $S_4$ on the transitive parts of size twelve. We obtain the projectors
Note that both the null hypotheses associated to $P_3$ and $P_4$ are easily evaluated from $P_3 + P_4$:

\[
\begin{bmatrix}
3 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & 3 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & -1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \quad \begin{bmatrix}
\mu_1 = \mu_2 = \mu_3 = \mu_4 \\
\mu_5 = \mu_6 = \mu_7 = \mu_8 \\
\mu_9 = \mu_{10} = \mu_{11} = \mu_{12}
\end{bmatrix}
\]

and:

\[
\begin{bmatrix}
2 & -2 & 0 & 0 & -1 & 1 & 1 & -1 & 1 & 1 & -1 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix} \quad \begin{bmatrix}
\mu_1 = \mu_2 = \mu_3 = \mu_4 \\
\mu_5 = \mu_6 = \mu_7 = \mu_8 \\
\mu_9 = \mu_{10} = \mu_{11} = \mu_{12}
\end{bmatrix}
\]
3.2.3 The right action of $S_4$ on the transitive parts of size twenty four. It gives the standard projectors $A_{24}$ and $Q_{24}$ with the null hypotheses $H_0 : \mu_1 = \mu_2 = \ldots = \mu_{24}$ associated to $Q_{24}$.

3.3 The dihedral group $D_4$ and its right action on the set $V$

If we associate $D_4$ with the symmetries of a square centered at the origin of $R^2$, then we may write it as $D_4 = \{ r_1, r_2, r_3, t_1, t_2, t_3, t_4 \}$, where:

$\{ r_1, r_2, r_3, r_4 \} = \{ 1, (13)(24), (1234), (1432) \}$ are rotations and

$\{ t_1, t_2, t_3, t_4 \} = \{ (24), (13), (12)(34), (14)(23) \}$ are reflections.

![Figure 1 - Symmetries of $D_4$ associated to a square centered at the origin of $R^2$.](image)

The right action of $D_4$ on the set of labels $V$ results in four orbits in size of four,

$\{ (aaa), (ggg), (ccc), (ttt) \}$, $\{ (aac), (cca), (ttg), (ggt) \}$

$\{ (aca), (cac), (tgg), (gtg) \}$, $\{ (acc), (caa), (tgg), (ggt) \}$

and six orbits in size of eight

$\{ (aag)(ctc)(tan)(gge)(ntc)(gga)(ccg)(aat) \}$,

$\{ (aga)(etc)(tat)(gcg)(ttc)(gag)(cgc)(ata) \}$,

$\{ (agg)(cct)(taa)(gcc)(ttg)(gag)(ccg)(att) \}$,

$\{ (age)(cta)(tag)(get)(tgc)(get)(gat)(cga)(atc) \}$,

$\{ (agt)(cgt)(tac)(gca)(tcg)(gca)(cgt)(atg) \}$,

3.3.1 The right action of $D_4$ on the transitive parts of size four. It gives the following projectors, with corresponding analytic properties $H_0$:

$$P_2 = \frac{1}{4} \begin{bmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{bmatrix} = Q_2 \oplus A_2 \quad H_0: \mu_1 + \mu_2 = \mu_3 + \mu_4$$

$$P_1 = \frac{1}{2} \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix} = \mathbb{I}_4 \oplus Q_2 \quad H_0: \begin{cases} \mu_1 = \mu_2 \\ \mu_3 = \mu_4 \end{cases}$$

3.3.1 The right action of $D_4$ on the transitive parts of size eight. It gives the regular projector:

$$P_2 = \frac{1}{8} \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad H_0: \mu_1 + \mu_2 + \mu_3 + \mu_4 = \mu_1 + \mu_2 + \mu_3 + \mu_4$$

$$P_3 = \frac{1}{8} \begin{bmatrix} 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad H_0: \mu_1 + \mu_2 + \mu_3 + \mu_4 = \mu_1 + \mu_2 + \mu_3 + \mu_4$$

$$P_4 = \frac{1}{8} \begin{bmatrix} 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad H_0: \mu_1 + \mu_2 + \mu_3 + \mu_4 = \mu_1 + \mu_2 + \mu_3 + \mu_4$$
Observe that with this ordering of the group, the null hypotheses are easily obtained by simple inspection of each projections matrix. It is worth to observe that, if the data vector is written as 

\[ X = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) = (x_1, x_2, 0, 0, 0, 0, 0, 0) + (0, 0, x_3, x_4, 0, 0, 0, 0) \]

\[ (0, 0, 0, 0, x_5, x_6, 0, 0) + (0, 0, 0, 0, x_7, x_8, 0, 0) \]

\[ = x_{r_1} + x_{r_2} + x_{r_3} + x_{r_4} \]

then projectors \( P_2, P_3 \) and \( P_4 \) compare the sum of pairs of these components and \( P_4 \) contrasts inside each component. Additions of these projectors also result in interesting null hypotheses:

\[ P_2 + P_3 = \frac{1}{4} \]

\[
\begin{bmatrix}
1 & 1 & 0 & 0 & -1 & -1 & 0 & 0 \\
1 & 1 & 0 & 0 & -1 & -1 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & -1 \\
-1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\
-1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & -1 & -1 & 0 & 0 & 0 & 1 \\
0 & 0 & -1 & -1 & 0 & 0 & 0 & 1 
\end{bmatrix}
\]

\[ H_0 : \begin{cases} 
\mu_{r_1} = \mu_{r_2} \text{ and } \\
\mu_{r_3} = \mu_{r_4} \text{ and } \\
\mu_{r_1} = \mu_{r_2} \text{ and } \\
\mu_{r_3} = \mu_{r_4} 
\end{cases} \]

\[ P_2 + P_3 = \frac{1}{4} \]

\[
\begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 \\
1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\
0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 \\
0 & 0 & -1 & -1 & 1 & 1 & 0 & 0 \\
-1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 \\
-1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 
\end{bmatrix}
\]

\[ H_0 : \begin{cases} 
\mu_{r_1} + \mu_{r_2} = \mu_{r_3} + \mu_{r_4} \text{ and } \\
\mu_{r_3} + \mu_{r_4} = \mu_{r_1} + \mu_{r_2} 
\end{cases} \]
4 Conclusions

The methodology of symmetry studies, whose main tools are the canonical projectors constructed by the junction of the action of groups on sets of labels and character theory of irreducible representations, showed to be a simple and systematic method to access the data-analytic properties inherent to symmetry-related labels for experimental data.


RESUMO: Pode ocorrer que dados experimentais sejam indexados por um conjunto de índices que refletem alguma particularidade de interesse do experimento. Quando, além disso, esses índices têm alguma simetria em sua estrutura, a metodologia de estudos de simetria (Symmetry studies – Viana 2008) pode ser utilizada para facilitar a análise e interpretação dos dados. Nesse contexto, o presente trabalho apresenta um estudo detalhado de propriedades relacionadas às simetrias em dados indexados por sequências de tamanho três em um conjunto de quatro símbolos.

PALAVRAS-CHAVE: Dados estruturados; grupos de permutação; projeções canônicas; decomposição da soma de quadrados.

References


Received in 02.07.2009.

Approved after revised in 15.03.2010.