Diagrams and Procedures for Partition of Variation

Pierre Legendre

Diagrams describing the partitions of variation of a response data table by two (Fig. 1), three (Fig. 2) and four tables (Fig. 3) of explanatory variables. The fraction names [a] to [p] in the output of \texttt{varpart} function follow the notation in these Venn diagrams, and the diagrams were produced using the \texttt{showvarparts} function.

![Diagram](image)

Figure 1: 3 regression/canonical analyses and 3 subtraction equations are needed to estimate the 4 (= $2^2$) fractions. [a] and [c] can be tested for significance (3 canonical analyses per permutation). Fraction [b] cannot be tested singly.
Figure 2: 7 regression/canonical analyses and 10 subtraction equations are needed to estimate the 8 ($= 2^3$) fractions. [a] to [c] and subsets containing [a] to [c] can be tested for significance (4 canonical analyses per permutation to test [a] to [c]). Fractions [d] to [g] cannot be tested singly.

Figure 3: 15 regression/canonical analyses and 27 subtraction equations are needed to estimate the 16 ($= 2^4$) fractions. [a] to [d] and subsets containing [a] to [d] can be tested for significance (5 canonical analyses per permutation to test [a] to [d]). Fractions [e] to [o] cannot be tested singly.
### Variation partitioning for two explanatory data tables --

Table 1 with m1 explanatory variables, Table 2 with m2 explanatory variables. 
Number of fractions: 4, called [a] ... [d].

γ indicates the 3 regression or canonical analyses that have to be computed.

# Partial canonical analyses are only computed if tests of significance or biplots are needed.

<table>
<thead>
<tr>
<th>Compute</th>
<th>Fitted</th>
<th>Residuals</th>
<th>Derived fractions</th>
<th>Degrees of freedom, numerator of F</th>
</tr>
</thead>
<tbody>
<tr>
<td>γ Y.1</td>
<td>[a+b]</td>
<td>[c+d]</td>
<td>df(a+b) = m1</td>
<td></td>
</tr>
<tr>
<td>γ Y.2</td>
<td>[b+c]</td>
<td>[a+d]</td>
<td>df(b+c) = m2</td>
<td></td>
</tr>
<tr>
<td>γ Y.1,2</td>
<td>[a+b+c]</td>
<td>[d]</td>
<td>df(a+b+c) = 3m ≤ m1m2 (there may be collinearity)</td>
<td></td>
</tr>
<tr>
<td>γ Y.1,2</td>
<td>[a]</td>
<td>[d]</td>
<td>df(a) = m3-m2</td>
<td></td>
</tr>
<tr>
<td>γ Y.2</td>
<td>[c]</td>
<td>[d]</td>
<td>df(c) = m3-m1</td>
<td></td>
</tr>
</tbody>
</table>

Partial analyses

(4) [a] - [a+b+c] - [b+c] 
controlling for 1 table X (5) [c] - [a+b+c] - [a+b] 
(6) [b] - [a+b] - [b+c] - [a+b+c] 
(7) [d] - [a+b+c] - [b] - [c+h] 

The only testable fractions are those that can be obtained directly by regression or canonical analysis.

### Tests of significance --

\[
F(a+b) = (\frac{(a+b)/m1}{(c+d)/(n-1-m1)})
\]

\[
F(b+c) = (\frac{(b+c)/m2}{(a+d)/(n-1-m2)})
\]

\[
F(a+b+c) = (\frac{(a+b+c)/m3}{(b+c)/m2})
\]

\[
F(a) = (\frac{(a)/m3-m2}{(b+c)/m2})
\]

The only testable fractions are those that can be obtained directly by regression or canonical analysis.

### Variation partitioning for three explanatory data tables --

Table 1 with m1 explanatory variables, Table 2 with m2 explanatory variables, Table 3 with m3 explanatory variables. 
Number of fractions: 8, called [a] ... [h].

γ indicates the 7 regression or canonical analyses that have to be computed.

# Partial canonical analyses are only computed if tests of significance or biplots are needed.

<table>
<thead>
<tr>
<th>Compute</th>
<th>Fitted</th>
<th>Residuals</th>
<th>Derived fractions</th>
<th>Degrees of freedom, numerator of F</th>
</tr>
</thead>
</table>
| Direct canonical analysis

| γ Y.1   | [a+dfg] | [b+ce+h] | df(a+dfg) = m1    |                                 |
| γ Y.2   | [b+deg] | [c+ef]   | df(b+deg) = m2    |                                 |
| γ Y.3   | [c+efg] | [a+b+dh] | df(c+efg) = m3    |                                 |
| γ Y.1,2 | [a+bd+efg] | [c+ef] | df(a+bd+efg) = m4 ≤ m1m2 (collinearity?) |                                 |
| γ Y.1,3 | [a+ce+dfg] | [b+eh] | df(a+ce+dfg) = m5 ≤ m1m3 (collinearity?) |                                 |
| γ Y.2,3 | [b+ce+dfg] | [a+eh] | df(b+ce+dfg) = m6 ≤ m2m3 (collinearity?) |                                 |
| γ Y.1,2,3 | [a+b+ce+dfg] | [h] | df(a+b+ce+dfg) = m7 ≤ m1m2m3 (collinearity?) |                                 |
| γ Y.1,2 | [a] | [h] | df(a) = m5-m3     |                                 |
| γ Y.1,3 | [a+b] | [c+h] | df(a+b) = m5-m1   |                                 |
| γ Y.2,3 | [b+eh] | [c+h] | df(b+eh) = m6-m3  |                                 |
| γ Y.310 | [a]  | [h]  | df(a) = m5-m1     |                                 |
| γ Y.1,11 | [b] | [h]  | df(b) = m7-m5     |                                 |
| γ Y.312 | [c]  | [h]  | df(c) = m7-m4     |                                 |

Partial analyses

(8) [a] - [a+b+c+de+fg] - [b+ce+dfg] 
controlling for two tables X

(9) [b] - [a+b+c+de+fg] - [a+c+de+fg] 
(10) [c] - [a+b+c+de+fg] - [a+b+de+fg] 

controlling for one table X

(11) [a] = [a+ce+dfg] - [c+efg] 
(12) [a+b] = [a+ef] - [b+ef] 
(13) [b+d] = [b+ef] - [c+ef] 
(14) [b+e] = [a+bd+efg] - [a+ef] 
(15) [c+e] = [a+ce+dfg] - [a+dfg] 
(16) [c+f] = [b+ce+dfg] - [b+efg] 

Fracions estimated

(17) [d] = [a+d] - [a] 
by subtraction

(18) [e] = [b+e] - [b] 
(cannot be tested)

(19) [f] = [c+f] - [c] 
(20) [g] = [b+ce+dfg] - [a+d] - [b+e] - [c+f] 
(21) [h] = residuals = 1 - [a+b+ce+dfg] 

| Tests of significance --

\[
F(a+b+dfg) = (\frac{(a+b+dfg)/m1}{(c+d+ef)/m1-m2})
\]

\[
F(b+d+eg) = (\frac{(b+d+eg)/m2}{(a+c+f)/m2-m3})
\]

\[
F(c+ef) = (\frac{(c+ef)/m3}{(a+b+df)/m3-m4})
\]

\[
F(a+b+dfg) = (\frac{(a+b+dfg)/m4}{(b+c+df)/m4-m5})
\]

\[
F(a+b+ce+dfg) = (\frac{(a+b+ce+dfg)/m5}{(b+ce+df)/m5-m6})
\]

\[
F(a+b+c+de+fg) = (\frac{(a+b+c+de+fg)/m6}{(a+c+d+ef)/m6-m7})
\]

\[
F(a) = (\frac{(a)/m7-m6}{(b+c)/m6-m5})
\]

\[
F(b) = (\frac{(b)/m5-m4}{(a+c+df)/m5-m6})
\]

\[
F(c) = (\frac{(c)/m6-m5}{(a+b+df)/m6-m4})
\]

\[
F(a+d) = (\frac{(a+d)/m3-m2}{(b+c+e)/m3-m4})
\]

\[
F(a+f) = (\frac{(a+f)/m4-m3}{(b+c+d)/m4-m5})
\]

\[
F(b+d) = (\frac{(b+d)/m6-m5}{(a+c+df)/m6-m4})
\]

\[
F(b+e) = (\frac{(b+e)/m4-m3}{(a+c+d)/m4-m5})
\]

\[
F(c+e) = (\frac{(c+e)/m5-m4}{(a+b+df)/m5-m6})
\]

\[
F(c+f) = (\frac{(c+f)/m6-m5}{(a+b+ef)/m6-m4})
\]

The only testable fractions are those that can be obtained directly by regression or canonical analysis.
Variation partitioning for four explanatory data tables --

Table 1 with m1 variables, Table 2 with m2 variables, Table 3 and Table 4 each with m4 variables

Indicates the 15 regression or canonical analyses that have to be computed.

<table>
<thead>
<tr>
<th>Direct canonical analysis</th>
<th>Residuals</th>
<th>Derived fractions</th>
<th>Degrees of freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y.1 [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o]</td>
<td>df(a) = m1 - m2</td>
<td>df(a) = m1 - m2</td>
<td>m1</td>
</tr>
<tr>
<td>Y.2 [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o]</td>
<td>df(a) = m1 - m2</td>
<td>df(a) = m1 - m2</td>
<td>m1</td>
</tr>
<tr>
<td>Y.3 [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o]</td>
<td>df(a) = m1 - m2</td>
<td>df(a) = m1 - m2</td>
<td>m1</td>
</tr>
<tr>
<td>Y.4 [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o]</td>
<td>df(a) = m1 - m2</td>
<td>df(a) = m1 - m2</td>
<td>m1</td>
</tr>
</tbody>
</table>

Partial analyses controlling for one table X

| (16) [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o] | df(a) = m1 - m2 | df(a) = m1 - m2 | m1 |
| (17) [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o] | df(a) = m1 - m2 | df(a) = m1 - m2 | m1 |
| (18) [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o] | df(a) = m1 - m2 | df(a) = m1 - m2 | m1 |

Partial analyses controlling for two tables X

| (28) [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o] | df(a) = m1 - m2 | df(a) = m1 - m2 | m1 |
| (29) [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o] | df(a) = m1 - m2 | df(a) = m1 - m2 | m1 |
| (30) [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o] | df(a) = m1 - m2 | df(a) = m1 - m2 | m1 |

Partial analyses controlling for three tables X

| (48) [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o] | df(a) = m1 - m2 | df(a) = m1 - m2 | m1 |
| (49) [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o] | df(a) = m1 - m2 | df(a) = m1 - m2 | m1 |
| (50) [ao] = [b] + [c] + [d] + [e] + [f] + [g] + [h] + [i] + [j] + [k] + [l] + [m] + [n] + [o] | df(a) = m1 - m2 | df(a) = m1 - m2 | m1 |

Fractions estimated by subtraction (cannot be tested)

| (44) [e] = [a] - [f] | df(e) = m1 - m2 = 0 |
| (45) [f] = [a] + [b] + [c] + [d] | df(f) = m1 - m2 = 0 |
| (46) [g] = [a] - [b] + [c] + [d] | df(g) = m1 - m2 = 0 |
| (47) [h] = [a] - [b] + [c] + [d] | df(h) = m1 - m2 = 0 |
| (48) [i] = [a] + [b] + [c] + [d] | df(i) = m1 - m2 = 0 |
| (49) [j] = [a] - [b] + [c] + [d] | df(j) = m1 - m2 = 0 |
| (50) [k] = [a] + [b] + [c] + [d] | df(k) = m1 - m2 = 0 |
| (51) [l] = [a] + [b] + [c] + [d] | df(l) = m1 - m2 = 0 |
| (52) [m] = [a] + [b] + [c] + [d] | df(m) = m1 - m2 = 0 |
| (53) [n] = [a] + [b] + [c] + [d] | df(n) = m1 - m2 = 0 |
| (54) [o] = [a] + [b] + [c] + [d] | df(o) = m1 - m2 = 0 |
| (55) [p] = residuals - 1 = [a] + [b] + [c] + [d] | df(p) = m1 - m2 = 0 |
Tests of significance --

\[ F(ae+gh+kl+mn+o) = \frac{[ae+gh+kl+mn+o]/m1}{[bc+de+fi+j+mp]/(n-1-m1)} \]
\[ F(b+e+fi+kl+mn+o) = \frac{[b+e+fi+kl+mn+o]/m2}{[arc+de+gh+ji+mp]/(n-1-m2)} \]
\[ F(c+fg+js+il+mn+o) = \frac{[c+fg+js+il+mn+o]/m3}{[ax+b+d+ei+h+sp]/(n-1-m3)} \]
\[ F(d+hs+ij+kl+mn+o) = \frac{[d+hs+ij+kl+mn+o]/m4}{[a+bc+ef+gh+sp]/(n-1-m4)} \]
\[ F(a+ce+fgh+jisk+kl+mn+o) = \frac{[ace+fgh+jisk+kl+mn+o]/m5}{[bcde+fi+jsp]/(n-1-m5)} \]
\[ F(a+b+e+fi+kl+mn+o) = \frac{[a+b+e+fi+kl+mn+o]/m6}{[arc+de+gh+ji+mp]/(n-1-m6)} \]
\[ F(b+c+e+fi+kl+mn+o) = \frac{[b+c+e+fi+kl+mn+o]/m7}{[bcde+fi+jsp]/(n-1-m7)} \]
\[ F(bd+ef+hs+ij+kl+mn+o) = \frac{[bd+ef+hs+ij+kl+mn+o]/m8}{[ax+b+d+ei+h+sp]/(n-1-m8)} \]
\[ F(bc+ef+gis+jih+kl+mn+o) = \frac{[bc+ef+gis+jih+kl+mn+o]/m9}{[ace+fgh+jisk+kl+mn+o]/m10} \]
\[ F(bc+ef+gis+jih+kl+mn+o) = \frac{[bc+ef+gis+jih+kl+mn+o]/m10}{[ae+gh+kl+mn+o]/(m5-m2)} \]
\[ F(ac+de+ef+gh+hi+js+kl+mn+o) = \frac{[ac+de+ef+gh+hi+js+kl+mn+o]/m11}{[bcde+fi+jsp]/(n-1-m11)} \]
\[ F(ac+de+ef+gh+hi+js+kl+mn+o) = \frac{[ac+de+ef+gh+hi+js+kl+mn+o]/m12}{[bcde+fi+jsp]/(n-1-m12)} \]
\[ F(ac+de+ef+gh+hi+js+kl+mn+o) = \frac{[ac+de+ef+gh+hi+js+kl+mn+o]/m12}{[bcde+fi+jsp]/(n-1-m13)} \]
\[ F(bac+de+ef+gh+hi+js+kl+mn+o) = \frac{[bac+de+ef+gh+hi+js+kl+mn+o]/m13}{[bcde+fi+jsp]/(n-1-m14)} \]
\[ F(bac+de+ef+gh+hi+js+kl+mn+o) = \frac{[bac+de+ef+gh+hi+js+kl+mn+o]/m13}{[bcde+fi+jsp]/(n-1-m15)} \]
\[ F(a+b+c+d+e+f+g+h+i+j+k+l+m+n+o) = \frac{[a+b+c+d+e+f+g+h+i+j+k+l+m+n+o]/m15}{[p]/(n-1-m15)} \]
\[ F(a+g+h+n) = \frac{[a+g+h+n]/(m5-m2)}{[bcde+fi+jsp]/(n-1-m5)} \]

For the other fractions controlling for one table \( X \), the F-statistics are constructed in the same way

\[ F(ae) = \frac{[ae]/(m1-m10)}{[bcde+fi+jsp]/(n-1-m13)} \]

For the other fractions controlling for two tables \( X \), the F-statistics are constructed in the same way

Fractions controlling for three tables \( X \):
\[ F(a) = \frac{[a]/(m15-m14)}{[p]/(n-1-m15)} \]
\[ F(b) = \frac{[b]/(m15-m13)}{[p]/(n-1-m15)} \]
\[ F(c) = \frac{[c]/(m15-m12)}{[p]/(n-1-m15)} \]
\[ F(d) = \frac{[d]/(m15-m11)}{[p]/(n-1-m15)} \]

Other fractions combining elementary fractions \([a]\) to \([o]\) can be calculated, but cannot be tested because they cannot be obtained by regression.

----------