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 `varpart` function follow the notation in these Venn diagrams, and the diagrams were produced using the `showvarparts` function.
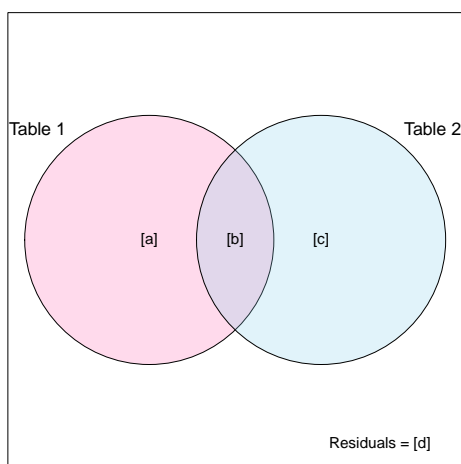

Figure 1: 3 regression/canonical analyses and 3 subtraction equations are needed to estimate the $4 \ (= 2^2)$ fractions.
[a] and [c] can be tested for significance (3 canonical analyses per permutation). Fraction [b] cannot be tested singly.
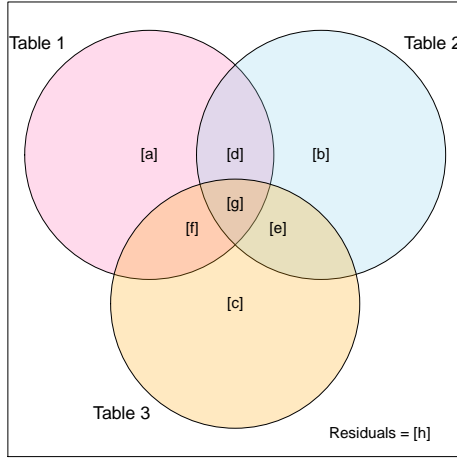
Figure 2: 7 regression/ canonical analyses and 10 subtraction equations are needed to estimate the $8 (= 2^3)$ fractions.
[a] to [c] and subsets containing [a] to [c] can be tested for significance (4 canonical analyses per permutation to test [a] to [c]). Fractions [d] to [g] cannot be tested singly.
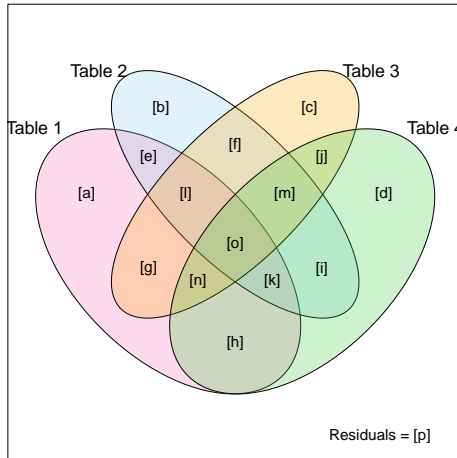

Figure 3: 15 regression/ canonical analyses and 27 subtraction equations are needed to estimate the $16 (= 2^4)$ fractions.
[a] to [d] and subsets containing [a] to [d] can be tested for significance (5 canonical analyses per permutation to test [a] to [d]). Fractions [e] to [o] cannot be tested singly.

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]
√ 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.

Compute	Fitted	Residuals	Derived fractions	Degrees of freedom, numerator of F
√ Y.1	[a+b]	[c+d]	(1)	df(a+b) = m1
√ Y.2	[b+c]	[a+d]	(2)	df(b+c) = m2
√ Y.1,2	[a+b+c]	[d]	(3)	df(a+b+c) = m3 ≤ m1+m2 (there may be collinearity)
# Y.1|2	[a]	[d]		df(a) = m3-m2
# Y.2|1	[c]	[d]		df(c) = m3-m1

Partial analyses	(4) [a] = [a+b+c] - [b+c]	df(a) = m3-m2*
controlling for 1 table X	(5) [c] = [a+b+c] - [a+b]	df(c) = m3-m1*
	(6) [b] = [a+b] + [b+c] - [a+b+c]	df(b) = m1+m2-(m1+m2) = 0
	(7) [d] = residuals = 1 - [a+b+c]	df2(d) = n-1-m3 for denominator of F

* Calculation of d.f. for difference between nested models: see Sokal & Rohlf (1981, 1995) equation 16.14.

Tests of significance --

F(a+b) = ([a+b]/m1)/([c+d]/(n-1-m1))
F(b+c) = ([b+c]/m2)/([a+d]/(n-1-m2))
F(a+b+c) = ([a+b+c]/m3)/([d]/(n-1-m3))

F(a) = ([a]/(m3-m2))/([d]/(n-1-m3))
F(c) = ([c]/(m3-m1))/([d]/(n-1-m3))

The only testable fractions are those that can be obtained directly by regression or canonical analysis.
The non-testable fraction is [b]. That fraction cannot 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, Table3 with m3 explanatory variables
Number of fractions: 8, called [a] ... [h]
√ 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.

Compute	Fitted	Residuals	Derived fractions	Degrees of freedom, numerator of F
Direct canonical analysis				
√ Y.1	[a+d+f+g]	[b+c+e+h]	(1)	df(a+d+f+g) = m1
√ Y.2	[b+d+e+g]	[a+c+f+h]	(2)	df(b+d+e+g) = m2
√ Y.3	[c+e+f+g]	[a+b+d+h]	(3)	df(c+e+f+g) = m3
√ Y.1,2	[a+b+d+e+f+g]	[c+h]	(4)	df(a+b+d+e+f+g) = m4 ≤ m1+m2 (collinearity?)
√ Y.1,3	[a+c+d+e+f+g]	[b+h]	(5)	df(a+c+d+e+f+g) = m5 ≤ m1+m3 (collinearity?)
√ Y.2,3	[b+c+d+e+f+g]	[a+h]	(6)	df(b+c+d+e+f+g) = m6 ≤ m2+m3 (collinearity?)
√ Y.1,2,3	[a+b+c+d+e+f+g]	[h]	(7)	df(a+b+c+d+e+f+g) = m7 ≤ m1+m2+m3 (collinearity?)
# Y.1|2	[a+f]	[c+h]		df(a+f) = m4-m2
# Y.1|3	[a+d]	[b+h]		df(a+d) = m5-m3
# Y.2|1	[b+e]	[c+h]		df(b+e) = m4-m1
# Y.2|3	[b+d]	[a+h]		df(b+d) = m6-m3
# Y.3|1	[c+e]	[b+h]		df(c+e) = m5-m1
# Y.3|2	[c+f]	[a+h]		df(c+f) = m6-m2
# Y.1|2,3	[a]	[h]		df(a) = m7-m6
# Y.2|1,3	[b]	[h]		df(b) = m7-m5
# Y.3|1,2	[c]	[h]		df(c) = m7-m4

Partial analyses	(8) [a] = [a+b+c+d+e+f+g] - [b+c+d+e+f+g]	df(a) = m7-m6
controlling for two tables X	(9) [b] = [a+b+c+d+e+f+g] - [a+c+d+e+f+g]	df(b) = m7-m5
	(10) [c] = [a+b+c+d+e+f+g] - [a+b+d+e+f+g]	df(c) = m7-m4

controlling for one table X	(11) [a+d] = [a+c+d+e+f+g] - [c+e+f+g]	df(a+d) = m5-m3
	(12) [a+f] = [a+b+d+e+f+g] - [b+d+e+g]	df(a+f) = m4-m2
	(13) [b+d] = [b+c+d+e+f+g] - [c+e+f+g]	df(b+d) = m6-m3
	(14) [b+e] = [a+b+d+e+f+g] - [a+d+f+g]	df(b+e) = m4-m1
	(15) [c+e] = [a+c+d+e+f+g] - [a+d+f+g]	df(c+e) = m5-m1
	(16) [c+f] = [b+c+d+e+f+g] - [b+d+e+g]	df(c+f) = m6-m2

Fractions estimated	(17) [d] = [a+d] - [a]	df(d) = m1-m1 = 0
by subtraction	(18) [e] = [b+e] - [b]	df(e) = m2-m2 = 0
(cannot be tested)	(19) [f] = [c+f] - [c]	df(f) = m3-m3 = 0
	(20) [g] = [a+b+c+d+e+f+g]-[a+d]-[b+e]-[c+f]	df(g) = (m1+m2+m3)-m1-m2-m3 = 0
	or [g] = [a+d+f+g] - [a] - [d] - [f]	df(g) = m1-m1-0-0 = 0
	(21) [h] = residuals = 1 - [a+b+c+d+e+f+g]	df2(h) = n-1-m7 for denominator of F

Tests of significance --

F(a+d+f+g) = ([a+d+f+g]/m1)/([b+c+e+h]/(n-1-m1))
F(b+d+e+g) = ([b+d+e+g]/m2)/([a+c+f+h]/(n-1-m2))
F(c+e+f+g) = ([c+e+f+g]/m3)/([a+b+d+h]/(n-1-m3))
F(a+b+d+e+f+g) = ([a+b+d+e+f+g]/m4)/([c+h]/(n-1-m4))
F(a+c+d+e+f+g) = ([a+c+d+e+f+g]/m5)/([b+h]/(n-1-m5))
F(b+c+d+e+f+g) = ([b+c+d+e+f+g]/m6)/([a+h]/(n-1-m6))
F(a+b+c+d+e+f+g) = ([a+b+c+d+e+f+g]/m7)/([h]/(n-1-m7))

F(a) = ([a]/(m7-m6))/([h]/(n-1-m7))
F(b) = ([b]/(m7-m5))/([h]/(n-1-m7))
F(c) = ([c]/(m7-m4))/([h]/(n-1-m7))
F(a+d) = ([a+d]/(m5-m3))/([b+h]/(n-1-m5))
F(a+f) = ([a+f]/(m4-m2))/([c+h]/(n-1-m4))
F(b+d) = ([b+d]/(m6-m3))/([a+h]/(n-1-m6))
F(b+e) = ([b+e]/(m4-m1))/([c+h]/(n-1-m4))
F(c+e) = ([c+e]/(m5-m1))/([b+h]/(n-1-m5))
F(c+f) = ([c+f]/(m6-m2))/([a+h]/(n-1-m6))

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, Table3 with m3 variables, Table4 with m4 variables
Number of fractions: 16, called [a] ... [p].
√ indicates the 15 regression or canonical analyses that have to be computed.

Compute     Fitted                     Residuals              Derived fractions                                          Degrees of freedom

Direct canonical analysis
√ Y.1       [a+e+g+h+k+l+n+o]          [b+c+d+f+i+j+m+p] (1)                                                             df(a+e+g+h+k+l+n+o) = m1
√ Y.2       [b+e+f+i+k+l+m+o]          [a+c+d+g+h+j+n+p] (2)                                                             df(b+e+f+i+k+l+m+o) = m2
√ Y.3       [c+f+g+j+l+m+n+o]          [a+b+d+e+h+i+k+p] (3)                                                             df(c+f+g+j+l+m+n+o) = m3
√ Y.4       [d+h+i+j+k+m+n+o]          [a+b+c+e+f+g+l+p] (4)                                                             df(d+h+i+j+k+m+n+o) = m4
√ Y.1,2     [a+b+e+f+g+h+i+k+l+m+n+o]        [c+d+p] (5)                                                                 df(a+b+e+f+g+h+i+k+l+m+n+o) = m5  ≤ m1+m2
√ Y.1,3     [a+c+e+f+g+h+j+k+l+m+n+o]        [b+d+i+p] (6)                                                              df(a+c+e+f+g+h+j+k+l+m+n+o) = m6  ≤ m1+m3
√ Y.1,4     [a+d+e+g+h+i+j+k+l+m+n+o]        [b+c+f+p] (7)                                                              df(a+d+e+g+h+i+j+k+l+m+n+o) = m7  ≤ m1+m4
√ Y.2,3     [b+c+e+f+g+i+j+k+l+m+n+o]        [a+d+h+p] (8)                                                              df(b+c+e+f+g+i+j+k+l+m+n+o) = m8  ≤ m2+m3
√ Y.2,4     [b+d+e+f+h+i+j+k+l+m+n+o]        [a+c+g+p] (9)                                                              df(b+d+e+f+h+i+j+k+l+m+n+o) = m9  ≤ m2+m4
√ Y.3,4     [c+d+f+g+h+i+j+k+l+m+n+o]        [a+b+e+p] (10)                                                             df(c+d+f+g+h+i+j+k+l+m+n+o) = m10 ≤ m3+m4
√ Y.1,2,3   [a+b+c+e+f+g+h+i+j+k+l+m+n+o]       [d+p] (11)                                                             df(a+b+c+e+f+g+h+i+j+k+l+m+n+o) = m11 ≤ m1+m2+m3
√ Y.1,2,4   [a+b+d+e+f+g+h+i+j+k+l+m+n+o]       [c+p] (12)                                                             df(a+b+d+e+f+g+h+i+j+k+l+m+n+o) = m12 ≤ m1+m2+m4
√ Y.1,3,4   [a+c+d+e+f+g+h+i+j+k+l+m+n+o]       [b+p] (13)                                                             df(a+c+d+e+f+g+h+i+j+k+l+m+n+o) = m13 ≤ m1+m3+m4
√ Y.2,3,4   [b+c+d+e+f+g+h+i+j+k+l+m+n+o]       [a+p] (14)                                                             df(b+c+d+e+f+g+h+i+j+k+l+m+n+o) = m14 ≤ m2+m3+m4
√ Y.1,2,3,4 [a+b+c+d+e+f+g+h+i+j+k+l+m+n+o]       [p] (15)                                                            df(a+b+c+d+e+f+g+h+i+j+k+l+m+n+o) = m15 ≤ m1+m2+m3+m4

Partial analyses
controlling for one table X                         (16) [a+g+h+n] = [a+b+e+f+g+h+i+k+l+m+n+o] - [b+e+f+i+k+l+m+o]      df(a+g+h+n) = m5 - m2
                                                    (17) [a+e+h+k] = [a+c+e+f+g+h+j+k+l+m+n+o] - [c+f+g+j+l+m+n+o]      df(a+e+h+k) = m6 - m3
                                                    (18) [a+e+g+l] = [a+d+e+g+h+i+j+k+l+m+n+o] - [d+h+i+j+k+m+n+o]      df(a+e+g+l) = m7 - m4

                                                    (19) [b+f+i+m] = [a+b+e+f+g+h+i+k+l+m+n+o] - [a+e+g+h+k+l+n+o]      df(b+f+i+m) = m5 - m1
                                                    (20) [b+e+i+k] = [b+c+e+f+g+i+j+k+l+m+n+o] - [c+f+g+j+l+m+n+o]      df(b+e+i+k) = m8 - m3
                                                    (21) [b+e+f+l] = [b+d+e+f+h+i+j+k+l+m+n+o] - [d+h+i+j+k+m+n+o]      df(b+e+f+l) = m9 - m4

                                                    (22) [c+f+j+m] = [a+c+e+f+g+h+j+k+l+m+n+o] - [a+e+g+h+k+l+n+o]      df(a) = m6 - m1
                                                    (23) [c+g+j+n] = [b+c+e+f+g+i+j+k+l+m+n+o] - [b+e+f+i+k+l+m+o]      df(a) = m8 - m2
                                                    (24) [c+f+g+l] = [c+d+f+g+h+i+j+k+l+m+n+o] - [d+h+i+j+k+m+n+o]      df(a) = m10 - m4

                                                    (25) [d+i+j+m] = [a+d+e+g+h+i+j+k+l+m+n+o] - [a+e+g+h+k+l+n+o]      df(a) = m7 - m1
                                                    (26) [d+h+j+n] = [b+d+e+f+h+i+j+k+l+m+n+o] - [b+e+f+i+k+l+m+o]      df(a) = m9 - m2
                                                    (27) [d+h+i+k] = [c+d+f+g+h+i+j+k+l+m+n+o] - [c+f+g+j+l+m+n+o]      df(a) = m10 - m3

controlling for two tables X                        (28) [a+e] = [a+c+d+e+f+g+h+i+j+k+l+m+n+o]-[c+d+f+g+h+i+j+k+l+m+n+o]   df(a+e) = m13 - m10
                                                    (29) [a+g] = [a+b+d+e+f+g+h+i+j+k+l+m+n+o]-[b+d+e+f+h+i+j+k+l+m+n+o]   df(a+g) = m12 - m9
                                                    (30) [a+h] = [a+b+c+e+f+g+h+i+j+k+l+m+n+o]-[b+c+e+f+g+i+j+k+l+m+n+o]   df(a+h) = m11 - m8

                                                    (31) [b+e] = [b+c+d+e+f+g+h+i+k+l+m+n+o]-[c+d+f+g+h+i+j+k+l+m+n+o]     df(b+e) = m14 - m10
                                                    (32) [b+f] = [a+b+d+e+f+g+h+i+k+l+m+n+o]-[a+d+e+g+h+i+j+k+l+m+n+o]     df(b+f) = m12 - m7
                                                    (33) [b+i] = [a+b+c+e+f+g+h+i+k+l+m+n+o]-[a+c+e+f+g+h+j+k+l+m+n+o]     df(b+i) = m11 - m6

                                                    (34) [c+f] = [a+c+d+e+f+g+h+i+j+k+l+m+n+o]-[a+d+e+g+h+i+j+k+l+m+n+o]   df(c+f) = m13 - m7
                                                    (35) [c+g] = [a+b+c+e+f+g+h+i+j+k+l+m+n+o]-[a+b+e+f+g+h+i+k+l+m+n+o]   df(c+g) = m12 - m9
                                                    (36) [c+j] = [a+b+c+e+f+g+h+i+j+k+l+m+n+o]-[a+b+e+f+g+h+i+k+l+m+n+o]   df(c+j) = m11 - m5

                                                    (37) [d+h] = [b+c+d+e+f+g+h+i+j+k+l+m+n+o]-[b+c+e+f+g+i+j+k+l+m+n+o]   df(d+h) = m14 - m8
                                                    (38) [d+i] = [a+c+d+e+f+g+h+i+j+k+l+m+n+o]-[a+c+e+f+g+h+j+k+l+m+n+o]   df(d+i) = m13 - m6
                                                    (39) [d+j] = [a+b+d+e+f+g+h+i+j+k+l+m+n+o]-[a+b+e+f+g+h+i+k+l+m+n+o]   df(d+j) = m12 - m5

controlling for three tables X                      (40) [a] = [a+b+c+d+e+f+g+h+i+j+k+l+m+n+o] - [b+c+d+e+f+g+h+i+j+k+l+m+n+o]   df(a) = m15 - m14
                                                    (41) [b] = [a+b+c+d+e+f+g+h+i+j+k+l+m+n+o] - [a+c+d+e+f+g+h+i+j+k+l+m+n+o]   df(b) = m15 - m13
                                                    (42) [c] = [a+b+c+d+e+f+g+h+i+j+k+l+m+n+o] - [a+b+d+e+f+g+h+i+j+k+l+m+n+o]   df(c) = m15 - m12
                                                    (43) [d] = [a+b+c+d+e+f+g+h+i+j+k+l+m+n+o] - [a+b+c+e+f+g+h+i+j+k+l+m+n+o]   df(d) = m15 - m11

Fractions estimated by subtraction                  (44) [e] = [a+e] - [a]                                             df(e) = m1-m1 = 0
(cannot be tested)                                  (45) [f] = [b+f] - [b]                                             df(f) = m2-m2 = 0
                                                    (46) [g] = [a+g] - [a]                                             df(g) = m1-m1 = 0
                                                    (47) [h] = [a+h] - [a]                                             df(h) = m1-m1 = 0
                                                    (48) [i] = [b+i] - [b]                                             df(i) = m2-m2 = 0
                                                    (49) [j] = [c+j] - [c]                                             df(j) = m3-m3 = 0
                                                    (50) [k] = [a+e+h+k] - [a+e] - [h]                                 df(k) = m1-m1-0 = 0
                                                    (51) [l] = [a+e+g+l] - [a+e] - [g]                                 df(l) = m1-m1-0 = 0
                                                    (52) [m] = [b+f+i+m] - [b+f] - [i]                                 df(m) = m2-m2-0 = 0
                                                    (53) [n] = [a+g+h+n] - [a+g] - [h]                                 df(n) = m1-m1-0 = 0
                                                    (54) [o] = [a+e+g+h+k+l+n+o] - [a+e+h+k] - [g] - [l] - [n]         df(o) = m1-m1-0-0-0 = 0
                                                    (55) [p] = residuals = 1 - [a+b+c+d+e+f+g+h+i+j+k+l+m+n+o]         df2(p) = n-1-m15
```

Tests of significance --

F(a+e+g+h+k+l+n+o) = ([a+e+g+h+k+l+n+o]/m1)/([b+c+d+f+i+j+m+p]/(n-1-m1))
F(b+e+f+i+k+l+m+o) = ([b+e+f+i+k+l+m+o]/m2)/([a+c+d+g+h+j+n+p]/(n-1-m2))
F(c+f+g+j+l+m+n+o) = ([c+f+g+j+l+m+n+o]/m3)/([a+b+d+e+h+i+k+p]/(n-1-m3))
F(d+h+i+j+k+m+n+o) = ([d+h+i+j+k+m+n+o]/m4)/([a+b+c+e+f+g+l+p]/(n-1-m4))
F(a+b+e+f+g+h+i+k+l+m+n+o) = ([a+b+e+f+g+h+i+k+l+m+n+o]/m5)/([c+d+j+p]/(n-1-m5))
F(a+c+e+f+g+h+j+k+l+m+n+o) = ([a+c+e+f+g+h+j+k+l+m+n+o]/m6)/([b+d+i+p]/(n-1-m6))
F(a+d+e+g+h+i+j+k+l+m+n+o) = ([a+d+e+g+h+i+j+k+l+m+n+o]/m7)/([b+c+f+p]/(n-1-m7))
F(b+c+e+f+g+i+j+k+l+m+n+o) = ([b+c+e+f+g+i+j+k+l+m+n+o]/m8)/([a+d+h+p]/(n-1-m8))
F(b+d+e+f+h+i+j+k+l+m+n+o) = ([b+d+e+f+h+i+j+k+l+m+n+o]/m9)/([a+c+g+p]/(n-1-m9))
F(c+d+f+g+h+i+j+k+l+m+n+o) = ([c+d+f+g+h+i+j+k+l+m+n+o]/m10)/([a+b+e+p]/(n-1-m10))
F(a+b+c+e+f+g+h+i+j+k+l+m+n+o) = ([a+b+c+e+f+g+h+i+j+k+l+m+n+o]/m11)/([d+p]/(n-1-m11))
F(a+b+d+e+f+g+h+i+j+k+l+m+n+o) = ([a+b+d+e+f+g+h+i+j+k+l+m+n+o]/m12)/([c+p]/(n-1-m12))
F(a+c+d+e+f+g+h+i+j+k+l+m+n+o) = ([a+c+d+e+f+g+h+i+j+k+l+m+n+o]/m13)/([b+p]/(n-1-m13))
F(b+c+d+e+f+g+h+i+j+k+l+m+n+o) = ([b+c+d+e+f+g+h+i+j+k+l+m+n+o]/m14)/([a+p]/(n-1-m14))
F(a+b+c+d+e+f+g+h+i+j+k+l+m+n+o) = ([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) = ([a+g+h+n]/(m5-m2))/([c+d+j+p]/(n-1-m5))
For the other fractions controlling for one table X, the F-statistics are constructed in the same way

F(a+e) = ([a+e]/(m13-m10))/([b+p]/(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) = ([a]/(m15-m14))/([p]/(n-1-m15))
F(b) = ([b]/(m15-m13))/([p]/(n-1-m15))
F(c) = ([c]/(m15-m12))/([p]/(n-1-m15))
F(d) = ([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.