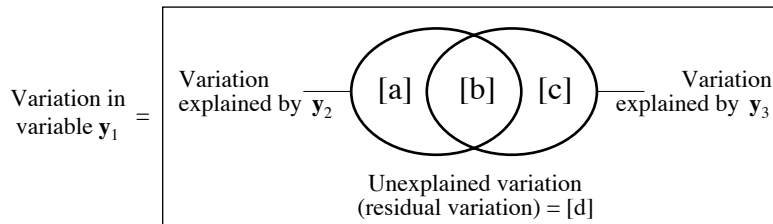


Variation partitioning

Box 4.1

Variation partitioning, which is described in detail in Subsection 10.3.5, provides a general framework to illustrate the similarities and differences between the coefficient of multiple determination and the partial and semipartial correlation coefficients, as well as the corresponding F -statistics.

Three variables only, y_1 , y_2 , and y_3 , are considered in this example. In the following Venn diagram, the rectangle represents the total sum of squares of variable y_1 :



In the multiple regression of y_1 on y_2 and y_3 , $\hat{y}_1 = b_0 + b_2 y_2 + b_3 y_3$ (this is an application of eq. 10.15), the coefficient of multiple determination, which is the square of the coefficient of multiple correlation, is:

$$R_{1,23}^2 = \frac{[a + b + c]}{[a + b + c + d]} \quad \text{with} \quad F = \frac{[a + b + c]/2}{[d]/(n-3)}$$

The partial correlation of y_1 with y_2 while controlling for the effect of y_3 is:

$$r_{12.3} = \sqrt{\frac{[a]}{[a + d]}} \quad \text{with} \quad F = \frac{[a]/1}{[d]/(n-3)}$$

The semipartial correlation of y_1 with y_2 in the presence of y_3 is:

$$r_{1(2.3)} = \sqrt{\frac{[a]}{[a + b + c + d]}} \quad \text{with} \quad F = \frac{[a]/1}{[d]/(n-3)}$$

The coefficients of partial and semipartial correlation receive the same sign as the corresponding coefficient of partial regression.

The test of a partial regression coefficient, b_2 or b_3 , is the same (i.e. it has the same F -statistic) as the test of the corresponding partial correlation coefficient, $r_{12.3}$ or $r_{13.2}$. The F -statistic is always the ratio of two *independent* portions of the variation of y_1 , each one divided by its degrees of freedom; see eqs. 4.39 and 4.40.