

## Sample Size Calculations for Matched-Pairs Designs

This routine computes the sample size  $n$  required to achieve a specified power level  $1 - \beta_1$  for a matched-pairs design in which differential expression between  $n$  treatment units and  $n$  matched control units is of interest. The total number of experimental units for the study is  $2n$ .

The following list summarizes notation for items used in the computation.

$E(R_0)$ : Mean number of false positives.

$\mu_1$ : Mean difference in log-expression between treatment and control units as postulated under the alternative hypothesis  $H_1$ .

$\sigma_d$ : Anticipated standard deviation of the difference in log-expression between matched treatment and control units.

$|\mu_1|/\sigma_d$ : Standardized statistical difference in gene expression between matched treatment and control units under the alternative hypothesis  $H_1$ .

$1 - \beta_1$ : Specified power level for an individual gene, which represents the expected proportion of differentially expressed genes that will be declared as such by the tests.

$G_0$ : Anticipated number of genes in the experiment that are *not* differentially expressed.

**Example.** Consider a study where it is anticipated that  $G_0 = 2000$  genes will not be differentially expressed. The investigator wishes to control the mean number of false positives at  $E(R_0) = 1$  and to detect a two-fold difference between treatment and control conditions with an individual power level of 0.90. Previous studies by the investigator may suggest that the standard deviation of gene expression differences in matched pairs will be about  $\sigma_d = 0.50$  on a log-2 scale. The two-fold difference represents a value of  $\log_2(2) = 1.00$  for  $|\mu_1|$  on a log-2 scale. Thus, the ratio  $|\mu_1|/\sigma_d$  equals  $1.00/0.50 = 2.0$ . The computer routine shows  $n = 6$  pairs of treatment and control conditions are required in the study.