Sample Size Calculations for
Completely Randomized Treatment-control Designs

For any specified power $1-\beta_1$, this routine computes the required sample size n for completely randomized designs in which differential expression between n treatment units and n 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.

μ_1: Mean difference in log-expression between treatment and control conditions as postulated under the alternative hypothesis H_1.

σ_d: Anticipated standard deviation of the difference in log-expression between treatment and control conditions. See the example below for the relation between the standard deviation of the difference and the experimental error standard deviation.

$|\mu_1|/\sigma_d$: Standardized statistical difference in gene expression between treatment and control conditions under 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.

Either from previous experiments or from a pilot study, estimate the experimental error standard deviation σ of gene log-expression. The standard deviation of the difference in log-expression between treatment and control conditions is then given by $\sigma_d = \sqrt{2}\sigma$.

Example
To illustrate, suppose σ is anticipated to be 0.40 in a completely randomized experiment. Furthermore, suppose that $\mu_1 = 1.00$, $E(R_0) = 1$, $G_0 = 2000$ and the desired individual power level is 0.90. As $\sigma_d = \sqrt{2}\sigma = \sqrt{2}(0.40) = 0.566$, the ratio $|\mu_1|/\sigma_d = 1.00/0.566 = 1.77$. The required sample size for each treatment group can be seen to be $n = 8$ and, hence, a total of $2n = 2(8) = 16$ experimental units are required.