Power Calculations for
Completely Randomized Treatment-Control Designs

This routine computes the individual power value $1-\beta_1$ for a completely randomized design with n treatment units and n control units ($2n$ units in total). This power value is the expected fraction of truly differentially expressed genes that will be correctly declared as differentially expressed by the tests.

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.

$\psi_1 = n\left(|\mu_1|/\sigma_d\right)^2$: The non-centrality parameter for the design.

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. Consider a completely randomized treatment-control design involving $G_0 = 5000$ undifferentially expressed genes. The investigator wishes to control the mean number of false positives at $E(R_0) = 2$ and to detect a two-fold differential expression between the treatment and control conditions. The two-fold difference represents a value of $|\mu_1| = \log_2(2.0) = 1.000$ on a log-2 scale. The experimental error standard deviation is anticipated to be about $\sigma = 0.40$ on a log-2 scale. The standard deviation of the difference in log-expression between treatment and control conditions is therefore given by $\sigma_d = \sqrt{2}\sigma = \sqrt{2}(0.40) = 0.5657$. Thus, the ratio $|\mu_1|/\sigma_d$ equals $1.000/0.5657 = 1.768$. Eight replications are to be used so $n = 8$. For these specifications, the non-centrality parameter equals $\psi_1 = 8(1.768)^2 = 25.0$. The computer routine gives an individual power level of $1-\beta_1 = 0.93$ for these inputs. Thus, about 93 percent of genes that exhibit a two-fold differential expression between treatment and control (whether up- or down-regulated) are expected to be discovered with this study design.