Power Calculations for Matched-Pairs Designs

This routine computes the individual power value $1 - \beta_1$ for a matched-pairs design having n treatment units and n matched control units. 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 matched treatment and control units.

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

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

Example . Consider a matched-pairs 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 standard deviation of the log-expression differences between matched treatment and control units is anticipated from a similar previous study to be $\sigma_d = 0.4243$ on a log-2 scale. Thus, the ratio $|\mu_1|/\sigma_d$ equals $1.000/0.4243 = 2.357$. Four matched pairs are to be used so $n = 4$. For these specifications, the non-centrality parameter equals

$$\psi_1 = 4(2.357)^2 = 22.22$$

The computer routine gives an individual power level of $1-\beta_1 = 0.88$ for these inputs. Thus, about 88 percent of genes that exhibit a two-fold differential expression between matched treatment and control units (whether up- or down-regulated) are expected to be discovered with this study design.