Sample Size Determintation

Xiaoyi Wang

Since the goal is to do a linear regression, we choose the 'pwr.f2.test': a test for the general linear model. The entire process used the 'pwr.f2.test' function from the 'pwr' package in R.

We can perform a multiple regression with gene expression as the dependent variable and concentration, cell age, treatment, cell type and media as independent variables.

The null hypothesis is that none of the independent variables explain any of the variability in gene expression. This would mean their regression coefficients are statistically indistinguishable from 0. The alternative is that at least one of the coefficients is not 0. This is tested with an F test. We can estimate the sample size for this test using the 'pwr.f2.test' function.

The F test has numerator and denominator degrees of freedom. The numerator degrees of freedom, u, is the number of coefficients you'll have in your model (minus the intercept). In this case, $u = 5$. The denominator degrees of freedom, v, is the number of error degrees of freedom: $v = n - u - 1$. This implies $n = v + u + 1$.

The effect size, f2, is $R^2/(1 - R^2)$, where R^2 is the coefficient of determination. In this case R^2 is 0.1, hence f2 is $0.1^2/(1 - 0.1^2)$.

We also know that we want a power of 0.9 and a significance level of 0.05. So that we apply these arguments in function 'pwr.f2.test', and can obtain the v, which is 147.8645, rounded to 148.

Recall $n = v + u + 1$, therefore we need $148 + 5 + 1 = 154$ samples in total.