Macro ID_power User Guide

9 May 2022

This macro implements prospective calculation of identification power for discovery analyses for genes or other features that are associated with a clinical outcome or state. The identification power for a feature with a particular degree of association is defined as the probability of identifying that feature as associated with outcome in an analysis controlling the false discovery rate (FDR). In addition to the acceptable FDR and the true degree of association the individual feature, the identification power depends on the proportion of truly non-associated (true null) features in the list of candidates and, among the truly associated features, the distribution of the degree of association. The calculation is described in detail in Crager (2012).

Macro ID_power is called as follows:

```
%id_power(
    /* Input Specification */  indsn=, fdr=, method=,
                    trueassoc=, se_est=, altmed=,
                    alt75pctl=, p0=, id=,
    /* Output Specification */  idpower=idpower, outdsn=
        );
```

The macro parameters are described in Table 1.

Table 1. Macro ID_power Parameters

Parameter	Type	Required?	Default Value	Description
indsn	$	Yes	(at temporary library)	(Libname reference and) file name containing input data set.
fdr	$	Yes	—	Name of input data set variable containing false discovery rate.
method	$	No	storey	FDR-controlling method to be used in the study. Default is Storey's method. Specify method = benjamini or method = bh for the Benjamini-Hochberg method.
trueassoc	$	Yes	—	Name of input data set variable that contains the hypothetical true degree of association of the individual gene.
se_est	$	Yes	—	Name of input data set variable that contains the expected standard error of estimate of degree of association. This can in general be determined based on the proposed sample size or number of events in the study.
altmed	$	Yes	—	Name of input data set variable that contains the median true absolute degree of association for the non-null gene population.
alt75pctl	$	Yes	—	Name of input data set variable that contains the 75th percentile of the true absolute degree of association for the non-null gene population.
p0	$	Yes	—	Input data set variable containing the hypothetical proportion of truly null association genes. Typical values range from 0.6 to 0.9. Values less than 0.5 are uncommon.
id	$	No	—	Optional list of variables used to identify records in the output data set.
idpower	$	No	—	Name of output data set variable that will contain the identification power.
outdsn	$	Yes	—	Output data set name.

Reference

Crager MR (2012). Prospective calculation of identification power for individual genes in analyses controlling the false discovery rate. *Genetic Epidemiology* **36**:839–847.