

**Key\_function:** myest is used for half-min imputation, myest\_mi is used for multiple imputation.

**Simulation:**

- data\_gen.r is used for generating the data
- Missing\_only.R is used for the design that data are generated from the Gaussian distribution and only contain missing values (Table 1, S1)
- Measurement\_error\_only.R is used for the design that data are generated from the Gaussian distribution and only contain measurement error (Table 2)
- Missing\_measurement.R is used for the design that data are generated from the Gaussian distribution and contain both measurement error and missing values (Table 3, S2, S3)
- Empirical\_data\_distribution.R is used for the design that data are generated from the empirical data distribution (Table 4, S6)
- Missing\_measurement\_multiple\_datasets.R is used for the design of multiple outcomes (Table S5)
- Oracle.R is used for the results of Oracle results (Table S4).

**Real\_Data\_analysis:** Codes of data analysis for four platforms: GC-MS, LC-MS, Lipidizer, NMR.

- imputation.R: impute the datasets using multiple imputation method: min-imputation, multiple imputation
- method\_MI.R: the algorithm for multiple imputation datasets
- method\_min.R: the algorithm for min imputation datasets
- single.R: Summary the results for single outcome (either breast cancer or colorectal cancer)
- both.R: Summary the results for both outcomes (common features for breast cancer and colorectal cancers)