3.7 MSstatsTMT: Details of parameter estimation and testing for differential abundance

Pairwise comparisons between conditions Model-based testing for differentially abundant proteins between pairs of conditions is carried out through a contrast of the condition means. Denote $\beta = (\beta_1, \beta_2, \ldots, \beta_C)$ the parameters associated with the terms *Condition* in Supplementary Equation S1-5, and $l = (l_1, l_2, \ldots, l_C)$ a vector of coefficients where $\sum l_c = 0$. A contrast is defined as $l^T\beta = \sum_c l_c\beta_c$. For example, a pairwise comparison testing proteins for differential abundance between Condition 1 and Condition 2 can be expressed as a contrast with $l = (1, -1, 0, \ldots, 0)$. We are interested in the null hypothesis $l^T\beta = \sum l_c\beta_c = 0$.

We estimate the parameters of the model using restricted maximum likelihood [7, 8, 9] to obtain $\hat{\beta}$, the contrast $l^T\hat{\beta} = \sum_c l_c\hat{\beta}_c$, and the corresponding t-statistic [10]

$$t = \frac{l^T\hat{\beta}}{\sqrt{ls^2\hat{V}l^T}} \tag{6}$$

Here s^2 is the estimate of σ^2 in Supplementary Equation S1, and $\hat{V}$ is the unscaled variance-covariance matrix of $\hat{\beta}$, such that $s^2\hat{V}$ is the variance-covariance matrix of $\hat{\beta}$, and $\sqrt{ls^2\hat{V}l^T}$ is the standard error of the contrast. The estimates s^2 and $\hat{V}$ are obtained by restricted maximum likelihood. The matrix $\hat{V}$ is a function of estimates $\hat{\sigma}^2_M$ for the random effect of mixtures, $\hat{\sigma}^2_T$ for the random effect of technical replicates, and $\hat{\sigma}^2_S$ for the random effect of subject. Therefore, the standard error of the contrast takes into account both technical variance σ^2_M, σ^2_T, σ^2 and biological variance σ^2_S.

In general and unbalanced designs, the degrees of freedom of the t-statistic in Supplementary Equation S6 are derived by Satterthwaite approximation [11]

$$df = \frac{2(s^2l\hat{V}l^T)^2}{[\text{VAR}(s^2l\hat{V}l^T)]}. \tag{7}$$

The calculation of $[\text{VAR}(s^2l\hat{V}l^T)]$ is described in [12, 10]. In unbalanced design, different contrasts of a same protein may have different degrees of freedom.

Empirical Bayes variance moderation When the number of biological replicates in each condition is small, we adopt Empirical Bayes moderation from *limma* [13]. Briefly, we assume that the estimate of the error variance σ^2 in Supplementary Equation S1 s^2 follows a scaled chi-square distribution with ν degrees of freedom

$$s^2|\sigma^2 \sim \frac{\sigma^2}{\nu}\chi^2_\nu, \tag{8}$$

The variance σ^2 of each protein is in turn assumed to follow a scaled inverse chi-square prior distribution with prior degrees of freedom d_0 and prior variance s^2_0

$$\frac{1}{\sigma^2} \sim \frac{1}{d_0s^2_0}\chi^2_{d_0}. \tag{9}$$

The degree of freedom ν is estimated as implemented in the R package lmerTest [12]. The parameters d_0 and s^2_0 are estimated from the distribution of the observed s^2 of all the proteins using an Empirical Bayes approach as implemented in the R package *limma* [14]. Finally, the posterior variance estimate is incorporated into the residual variance of each protein

$$\tilde{s}^2 = \frac{s^2_0d_0 + s^2\nu}{d_0 + \nu}. \tag{10}$$

Then, the moderated t-statistic for the contrast l becomes

$$\tilde{t} = \frac{l^T\hat{\beta}}{\sqrt{l\tilde{s}^2\hat{V}l^T}}, \tag{11}$$

with $df + d_0$ degrees of freedom.