Normalization method for Eleonore's data

Jacopo Cirrone
Computer Science Department
Courant Institute, New York University
New York, NY
Email: cirrone@courant.nyu.edu

I. Description

We have multiple batches containing different ecotypes. And each ecotype consists of 10 plants.
A normalization method is needed because different batches are grown in different days, hence trait values might be affected by other factors (i.e. batch effect). That's why each batch includes Columbia ecotype, which is a reference standard.

Let $r + 1$ be the number of ecotype within a batch. ($(r + 1)th$ represents Columbia).

Let n be the number of batches.

Let t_{xyz} be the median (across all of 10 plants) of trait x of ecotype y in batch z.

Let $tCol_{xz}$ be the median (across all of 10 plants) of trait x of columbia in batch z.

To normalize t_{xyz} we will:

- compute $\overline{tCol_x}$ which is the median of $tCol_{xi}, \forall i$ with $1 \leq i \leq n$ (median of Columbia across all batches).
- compute scaling factor for trait x in batch z

$$scalFact_{xz} = \frac{\overline{tCol_x}}{tCol_{xz}}$$

- compute normalized value of trait x for each ecotype y in batch z

$$\hat{t}_{xyz} = (scalFact_{xz}) * t_{xyz}$$
$$\forall y, 1 \leq y \leq r$$