

Advanced Mill Database Creation

Edward Souza, SWQL, February 22, 2011

Purpose

To summarize cultivar milling and baking data into a reference for use in adjusting other data sets of experimental lines.

Background

Each group germplasm from breeders that is evaluated by the laboratory should have check cultivars included in the group. As part of the validation process of the data, those checks should be compared to prior performance of the check in quality evaluations. Environmentally adjusted scores are created for each line evaluated in the laboratory. This is calculated based on the difference between an historical average of the cultivar's performance and the observed value within the trial. The Advanced Mill Database is the tabular form of historical averages for cultivars evaluated at the laboratory using the modified Quadrumat flour mill.

Compilation of Data

Databases of previous year's performance are compiled and stored on the SWQL server in the folder marked 'Quality Scores'. Begin with the most recent and add to it all new advanced groups evaluated in the laboratory since the last compilation. The dataset as of 2/22/2011 bridges evaluations with the current and earlier sugar-snap cookie method. Samples evaluated with the earlier method are noted in a method column with 'Old'. All current data should be marked as 'New' for the method column as we are using the revised sugar-snap cookie method exclusively, since 2009. Edit previous entries of new cultivar releases in the past year for consistent naming. Review all naming for consistency as company names can and do change each year.

Analysis of Data

Due to the unbalanced nature of the data, we can only generate approximate means.

1. The current model used includes 'year' of testing and 'cult' or cultivar name as the independent variables to generate a least squares mean.
2. Any entries that appear fewer than seven times in the database should not be analyzed.
3. Mean values for each cultivar as well as the number of observations contributing to the mean are tabled for use in adjustments. Milling, baking, and softness equivalent scores are included in the table and graded.
4. Cookie data has a more complex model; it includes a third variable of 'method' to specify the different baking methods. The new method produces cookies approximately 0.6 cm greater in diameter than the earlier method.
5. When cookie means are entered into the database table, the average value should be increased by half the difference between the two methods so that the value appearing in the table is close to the values produced under the current baking method.
6. Regression models for generating scores are created by modeling the mean scores against the mean milling and baking data for the cultivars in the Advanced Mill Database. Using the means in the table, SAS should solve the regression models:

$$\text{Milling score} = \text{Intercept} + b * \text{Milling yield}$$

Baking score = Intercept + b1*Cookie diameter + b2*Sucrose SRC + b3*Softness equivalent
Softness equivalent score = Intercept + b*Softness equivalent

Milling and baking data analysis uses the models in the templates sheets marked 'Adjustment Factor' sheet to calculate the observed scores for the checks. The observed scores are compared to the historical averages for the checks in the Advanced Mill Database, and the difference is used to calculate the bias for adjusting the score of all the experimental lines to accommodate the environmental influences on the trial.

Annotation of the Set

The data sheets should note averages of all entries and standard errors for a specified number of observations. The raw data used to generate the means should be annotated for date of last entry and any missing or unusual information in the group. Backups should be made on the local computer and server. All technicians should be informed of the conversion to the new database. In the database and in files a date should be noted for conversion to the new database. Each group processed should be marked at the bottom with the version of the database used for analysis.