Effects of Storage on Raffinose Content of Sugar Beets. II. Evaluation of Variance Components for Optimum Sample Size Determination

M. D. FINKNER, R. E. FINKNER AND R. F. OLSON[1]

Breeding for improvement of sugar beets in the past has been largely confined to the factors of yield and percent sucrose. The source and extent of experimental and sampling errors associated with the estimation of these two characters have been studied by Immer (5)[2] and Immer and LeClerg (6). Recently however, new techniques, such as flame spectrophotometry and paper chromatography, have been developed which can be used for the determination of many melassigenic substances for mass operation. Very little information is available regarding the magnitude of experimental and sampling errors associated with the use of these new techniques.

The concept of variance components originated with R. A. Fisher who expressed the additive properties of variances in the first edition of his text "Statistical Methods for Research Workers" (4). Early development of variance components as a working tool was made by Yates and Zacopanay (9). The concept of variance components was used in this study to evaluate sources of variation contributing to the sampling and experimental errors associated with the determination of raffinose content in sugar beets by the means of paper chromatography.

Materials and Methods

The materials used in this study were the same as those described previously (3). There were four varieties, four replications and four storage periods arranged in a split plot design. A total of 35 beets of each variety, in each of the four replications, were topped and stored in the Rocky Ford, Colorado, root cellar on October 24, 1955. Seven beets of each variety within each replication were selected at random and stored together in crates, or a total of 28 beets per crate. The 16 crates, one crate per each replication and for four different storage dates were randomized and stored in the root cellar at a temperature of approximately 4 degrees Centigrade. The remaining beets, seven roots of each of the four varieties of each replication, were bulked together in large onion sacks and all four sacks,

---

[1] Biometrician, New Mexico Agricultural Experiment Station, University Park, New Mexico. Manager Research Station and Research Chemist, respectively, American Crystal Sugar Company, Rocky Ford, Colorado.

[2] Numbers in parentheses refer to literature cited.

each containing a single replication, were buried in the storage pile at Manzanola, Colorado, on the evening of October 25, 1955.

After three weeks the beet samples in the storage pile were taken out, analyzed and compared with samples stored three weeks in the root cellar. Beet samples stored in the root cellar were analyzed at three week intervals. The non-stored check was in storage one day while the experiment was being set up.

Each beet in the storage test was sampled twice by rasping. Each of the two resulting pulp samples was split into half and placed in a plastic sack and frozen. When all beets were rasped the individual samples were taken out of the deep freeze at random and spotted on Whatman filter paper number 1. Determination of raffinose content was similar to the method described by Brown (2) and Serro and Brown (8) and the results are given as percent on dry substance of the beet juice.

Results and Discussion

Table 1 gives the complete analysis of variance and the expected values of mean squares according to Anderson and Bancroft (1) for beets stored in the storage pile compared with beets stored in the root cellar. Variability associated with beets within plots (a sampling error), rasps from the same beet (a sub-sampling error), and determinations (a laboratory error) can be estimated from this analysis. Experimental errors associated with varieties and storage methods can also be estimated from the analysis of variance. Experimental errors will not be considered here however, since it is the usual procedure to base the required number of replications upon some factor (generally yield) which is more variable than a chemical determination.

Table 1.—Analysis of Variance and the Expected Mean Squares of Four Varieties and Two Different Types of Storage for Raffinose Content Expressed as Percent on Dry Substance.

Source of Variation	D.F.	M.S.	Expected Mean Squares
Total	671		
Replications	2[1]	.3237	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2 + 224\sigma_p^2$
Varieties[2]	3	1.5624	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2 + 56\sigma_a^2 + 168\Theta_v$
Error A (Rep. x Var.)	6	.1450	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2 + 56\sigma_a^2$
Storage	1	.1357	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2 + 336\Theta_s$
Varieties x Storage	3	.2996	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2 + 84\Theta_{vs}$
Error B	8	.3307	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2$
Beets within plots	144	.0650	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2$
Rasping from same beet	168	.0141	$\sigma_d^2 + 2\sigma_r^2$
Determinations	336	.0113	σ_d^2

[1] One replication was lost while removing the beets from the storage pile.

[2] Θ_v, Θ_s, and Θ_{vs} refer to the fixed effects of varieties, storage, and variety by storage interaction, respectively.

Estimates of the components of variance associated with determinations (σ_d^2), raspings (σ_r^2) and beets (σ_b^2) were computed from the expected mean squares of Table 1.

$$
\begin{aligned}
&(1) \quad \sigma_d^2 \doteq .0113 \\
&(2) \quad \sigma_d^2 + 2\sigma_r^2 \doteq .0141 \\
&\qquad 2\sigma_r^2 \doteq .0141 - \sigma_d^2 \\
&\qquad \sigma_r^2 \doteq \frac{.0141 - .0113}{2} \\
&\qquad \sigma_r^2 \doteq .0014 \\
&(3) \quad \sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 \doteq .0650 \\
&\qquad 4\sigma_b^2 \doteq .0650 - (\sigma_d^2 + 2\sigma_r^2) \\
&\qquad \sigma_b^2 \doteq \frac{.0650 - .0141}{4} \\
&\qquad \sigma_b^2 \doteq .0127
\end{aligned}
$$

Since the sampling scheme for raffinose determination involved the random selection of seven beets per plot, two raspings per beet and two determinations per rasping, the variance of a sample mean on a per determination basis was the mean square for beets within plots divided by the total number of determinations per plot. Also since variance components have additive properties, the variance of a sample mean can be equated to the variance components associated with beets, raspings and determinations. Therefore

$$
\begin{aligned}
V_s &\doteq \frac{\sigma_d^2}{b\,r\,d} + \frac{\sigma_r^2}{b\,r} + \frac{\sigma_b^2}{b} \\
&\doteq \frac{.0113}{28} + \frac{.0014}{14} + \frac{.0127}{7} \\
&\doteq .00232
\end{aligned}
$$

where V_s is the variance of a sample mean and b, r, and d refer to the number of beets, raspings, and determinations respectively.

Table 2 shows the value of the estimated variance and the percent of the total sampling variability attributed to each of the various components under study.

Table 2.—Estimates of Variance Components Associated with Field Sampling of Beets for Raffinose Determination.

Variance Components		Value	Percent
Determinations	(σ_d^2)	.0113	44.5
Raspings	(σ_r^2)	.0014	5.5
Beets	(σ_b^2)	.0127	50.0
Total[1]		.0254	100.0

---

[1] Does not include plot variability for either main plots or sub-plots.

The component which exhibited the greatest portion of the total sampling variability was associated with beets within plots. This was to be expected as the varieties used in this investigation were extremely heterozygous. Also, the seven beets were selected at random from each plot allowing for the possible effect of some intra-plot soil variation. The only other component of significance was that associated with the laboratory determinations. This may be interpreted as the failure of the laboratory technique to be completely reproducible.

The analysis of variance and expected values of mean squares for the second phase of the investigation are shown in Table 3. Again only the sampling variability is of particular interest. The sampling scheme for the second phase was identical to that of the first phase although many more plots were involved.

Table 3.—Analysis of Variance and the Expected Mean Squares of Four Varieties and Four Different Storage Dates for Raffinose Content Expressed as Percent on Dry Substance.

Source of Variation	D.F.	M.S.	Expected Mean Squares
Total	1791		
Replications	3	.8649	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2 + 448\sigma_p^2$
Varieties	3	4.6281	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2 + 112\sigma_a^2 + 448\Theta_v$
Error A (Rep. x Var.)	9	.2872	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2 + 112\sigma_a^2$
Storage	3	6.7044	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2 + 448\Theta_s$
Varieties x Storage	9	.2812	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2 + 112\Theta_{vs}$
Error B	36	.1082	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2 + 28\sigma_e^2$
Beets within plots	384	.1156	$\sigma_d^2 + 2\sigma_r^2 + 4\sigma_b^2$
Rasping from same beet	448	.0347	$\sigma_d^2 + 2\sigma_r^2$
Determinations	896	.0149	σ_d^2

[1] Θ_v, Θ_s, and Θ_{vs} refer to the fixed effects of varieties, storage, and variety by storage interaction, respectively.

Estimates of the variance components, computed as previously shown, and the percent of the total sampling variability attributed to each are listed in Table 4.

Table 4.—Estimates of Variance Components Associated with Field Sampling of Beets for Raffinose Determination.

Variance Components		Value	Percent
Determinations	(σ_d^2)	.0149	33.1
Raspings	(σ_r^2)	.0099	22.0
Beets	(σ_b^2)	.0202	44.9
Total[1]		.0450	100.0

[1] Does not include plot variability for either main plots or sub-plots.

Again the greatest variance was from beets within plots which contributed about 45% of the total sampling error. Following it was the component for laboratory determinations. Although the component of variance associated with the raspings was somewhat larger in this case, contributing 22 percent of the total sampling error, it was not significant at the .05 level of probability.

The variance of a sample mean for this phase was estimated as follows:

$$
\begin{aligned}
V_s &\doteq \frac{\sigma_d^2}{b\,r\,d} + \frac{\sigma_r^2}{b\,r} + \frac{\sigma_b^2}{b}. \\
&\doteq \frac{.0149}{28} + \frac{.0099}{14} + \frac{.0202}{7} \\
&\doteq .00413
\end{aligned}
$$

The variance of a sample mean can be reduced to any value desirable by merely increasing the sampling intensity. Disregarding cost the most efficient method of reducing the variance of a sample mean would be to increase the number of beets per plot. Not only was the beet-to-beet variability the greatest of the three sampling components, but the number of beets appeared in the denominator of all three components.

It was also apparent that it was unnecessary to rasp each beet more than once. Therefore, if one should double the number of beets per plot, rasp each beet once and run duplicate determinations, the total number of chemical determinations would be the same as before. The variance of a sample mean, however, would be as follows:

$$
\begin{aligned}
V_s &\doteq \frac{.0149}{28} + \frac{.0099}{14} + \frac{.0202}{14} \\
&\doteq .00268
\end{aligned}
$$

This would reduce the variance of a sample mean 35 percent without requiring any additional chemical determinations. Considering the same sampling pattern for the first phase of the experiment the variance would have been .00141, a reduction of 39 percent.

A primary objective in the statistical design of such a sampling procedure is to mininize the cost of obtaining the sample estimate if the desired degree of precision is fixed, or conversely, to maximize the precision of the estimate obtained from a given amount of expenditure including personnel, time and equipment. If the cost per sampling unit is known or can be accurately estimated, and if the cost is proportional to the number of units sampled at each level, the optimum allocation of the sampling units can be determined. Marcuse (7) has developed cost functions for a three stage sampling pattern such as the one described in this paper. Therefore the optimum allocation of sampling units for a minimum cost subject to the constraint that the allowable amount of variance is preassigned is as follows:

$$b' = \frac{\sigma_b}{v} \; \frac{\sum\limits_{i=b,r,d} (\sigma_i \sqrt{c_i})}{\sqrt{c_b}}$$

$$r' = \frac{\sigma_r}{\sigma_b} \sqrt{\frac{c_b}{c_r}}$$

$$d' = \frac{\sigma_d}{\sigma_r} \sqrt{\frac{c_r}{c_d}}$$

where v is the preassigned variance, σ_b, σ_r, and σ_d are the standard deviations associated with beets, raspings, and determinations respectively; and c_b, c_r, and c_d are the unit costs for beets, raspings, and determinations respectively.

Likewise Marcuse (7) has shown that the optimum allocation when the total amount of cost is fixed is:

$$b'' = \frac{\sigma_b}{\sum\limits_{i=b,r,d} (\sigma_i \sqrt{c_i})} \; \frac{c}{\sqrt{c_b}}$$

$$r'' = \frac{\sigma_r}{\sigma_b} \sqrt{\frac{c_b}{c_r}}$$

$$d'' = \frac{\sigma_d}{\sigma_r} \sqrt{\frac{c_r}{c_d}}$$

where c is the total allowable cost and the other notations are the same as above.

Cost information for the sampling procedure of these experiments was obtained from actual expenditures for labor and materials. (Expenditures such as capital outlay, taxes, etc. were not included). The cost per beet for harvesting, labeling and storing was approximately twenty cents (\$0.20). The cost for a single rasping was determined to be about ten cents (\$0.10). The cost per single determination of raffinose including calculations necessary to obtain a proper value based on dry substance was approximately twenty cents (\$0.20).

Since the sampling procedures for both of the experiments reported in this paper were identical, the best available estimates of the variance components were the weighted averages of the components from the two experiments. The proper weighing factor was the degrees of freedom associated with each component.

Therefore estimates of the variance components for the three sampling stages are:

$$s_d^2 = .0139 \quad (s_d = .1179)$$
$$s_r^2 = .0076 \quad (s_r = .0872)$$
$$s_b^2 = .0182 \quad (s_b = .1349)$$

Using these estimates along with the cost information that $c_d = .20$, $c_r = .10$, and $c_b = .20$ optimum allocation of sampling units can be obtained. Assuming that a variance of a sample mean of .0025 ($s_x = .05$) will provide adequate precision, the sampling units for each stage were computed as follows:

$$b' = \frac{s_b}{V} \frac{\sum\limits_{i = b, r, d} (s_i \sqrt{c_i})}{\sqrt{c_b}}$$

$$= \frac{.1349}{.0025} \left[\frac{(.1349)\ (.4472) + (.0872)\ (.3162) + (.1179)\ (.4472)}{.4472} \right]$$

$$= \left[\frac{.1349}{.0025}\right] \left[\frac{.1406}{.4472}\right]$$

$$= \frac{.01897}{.00112}$$

$$= 16.9$$

$$r' = \frac{s_r}{s_b} \sqrt{\frac{c_b}{c_r}}$$

$$= \left[\frac{.0872}{.1349}\right] \left[\frac{.4472}{.3162}\right]$$

$$= \frac{.03900}{.04266}$$

$$= 0.91$$

$$d' = \frac{s_d}{s_r} \sqrt{\frac{c_r}{c_d}}$$

$$= \left[\frac{.1179}{.0872}\right] \left[\frac{.3162}{.4472}\right]$$

$$= \frac{.03728}{.03900}$$

$$= 0.96$$

Obviously, the sampling units must be whole numbers. Therefore, the optimum sampling procedure is 17 beets per plot, one rasping per beet, and one chemical determination per rasping. This would be accomplished at a total cost of $8.50 per plot which is minimum for these conditions.

Under some conditions it would be desirable to determine the optimum sampling scheme for a fixed cost. Suppose it has been determined that sampling costs must not exceed $5.00 per plot.

Computations for optimum allocations of sampling units which will provide a minimum variance were as follows:

$$b'' = \frac{s_b}{\sum\limits_{i=b,r,d} (s_i \sqrt{c_i})} \qquad \frac{C}{\sqrt{c_b}}$$

$$= \left[\frac{.1349}{(.1349)\ (.4472) + (.0872)\ (.3162) + (.1179)\ (.4472)}\right]\left[\frac{5.00}{.4472}\right]$$

$$= \left[\frac{.1349}{.1406}\right]\left[\frac{5.00}{.4472}\right]$$

$$= \frac{.6745}{.0629}$$

$$= 10.7$$

$$r'' = \frac{s_r}{s_b} \sqrt{\frac{c_b}{c_r}}$$

$$= \left[\frac{.0872}{.1349}\right]\left[\frac{.4472}{.3162}\right]$$

$$= \frac{.03900}{.04266}$$

$$= 0.91$$

$$d'' = \frac{s_d}{s_r} \sqrt{\frac{c_r}{c_d}}$$

$$= \left[\frac{.1179}{.0872}\right]\left[\frac{.3162}{.4472}\right]$$

$$= \frac{.03728}{.03900}$$

$$= 0.96$$

Again each of the sampling intensities must be rounded to a whole number. Since r″ and d″ must be rounded to one, it was necessary to round b″ to 10 in order to satisfy the requirement that the cost not exceed \$5.00 per plot. The minimum variance for this sampling pattern is .0040.

It is worthy of note that, except for the first level (beets), the optimum combination of the number of sampling units is independent of the given degree of precision or the fixed total cost. Therefore, when planning an experiment using nested sampling in three stages one needs to be concerned with the given cost or precision only in selecting the number of primary sampling units.

Summary

The concept of variance components was used to evaluate sources of variation contributing to sampling errors associated with the determination of raffinose content in sugar beets by means of paper chromatography. Two experiments were involved, each of which utilized identical three stage nested sampling procedures. Weighted averages of the estimated components of variance were .0182, 0076, and .0139 for beets, raspings, and chemical determinations respectively.

Actual costs were determined to be twenty cents (\$0.20) per beet, ten cents (\$0.10) per rasping, and twenty cents (\$0.20) per chemical determination. Using cost functions developed by Marcuse (7) the optimum allocations of sampling units for a preassigned variance of .0025 is 17 beets per plot, one rasping per beet and one chemical determination per rasping. The total cost resulting from this sampling procedure is \$8.50 per plot. Assuming that the total cost per plot must not exceed \$5.00 the optimum allocation of sampling units is 10 beets, one rasping, and one determination. The variance of the sample mean would then be .0040.

Literature Cited

(1) ANDERSON, R. L., and BANCROFT, T. A. 1952. Statistical theory in research. 1st Ed. McGraw-Hill, New York.

(2) BROWN, R. J. 1952. Quantitative determination of raffinose in mother beets and raw beet juices by paper chromatography. Anal. Chem. 24:384.

(3) FINKNER, R. E. and BAUSERMAN, H. M. 1956. Breeding of sugar beets with reference to sodium, sucrose, and raffinose content. Jour. Amer. Soc. Sugar Beet Tech. IX (2):170-177.

(4) FISHER, R. A. 1925. Statistical methods for research workers. 1st Ed. Oliver and Boyd, Ltd. Edinburgh and London.

(5) IMMER,, F. R. 1932. A study of sampling technic with sugar beets. Jour. of Agr. Res. Vol. 44 (8) :633-647.

(6) IMMER, F. R. and LECLERG, E. L. 1936. Errors of routine analysis for percentage of sucrose and apparent purity coefficient with sugar beets taken from field experments. Jour. Agr. Res. Vol. 52 (7) :505-515.

(7) MARCUSE, S. 1949. Optimum allocation and variance components in nested sampling with an application to chemical analysis. Biometrics 5:189-206.

(8) SERRO, R. F. and BROWN, R. J. 1954. Improved chromatographic method for analysis of sugar beet products. Anal. Chem. 26:890.

(9) YATES, F., and ZACOPANAY, I. 1935. The estimation of the efficiency of sampling with special reference to sampling for yield in agriculture experiments. Jour. Agr. Sci. 25:545-577.