SEED SAMPLING FOR CERTIFICATION INSPECTION
I. SOME RESULTS ON SAMPLING AND LABORATORY ANALYSIS TECHNIQUES FOR BIRDSFOOT TREFOIL SEED

G. F. Atkinson, W. R. Blischke, R. H. Bradley,
and C. L. Hart

BU-112-M October, 1959

1. Introduction

In the sampling of lots of Birdsfoot Trefoil seed for purity analysis in New York, one-half pound samples are drawn from the lot and submitted to the seed-testing laboratory for analysis. At the seed laboratory a two gram subsample is drawn from the half-pound sample submitted. The seeds are then sorted and, among other things, percents by weight of pure seed, other crop seed, weed seed, and inert matter, are computed for the two gram subsample.

Experimental errors present in the various sampling and laboratory techniques have caused concern to the seedsmen as well as to the seed producers. In particular, this investigation was initiated because of such variation in sampling and analyzing Birdsfoot Trefoil seed to determine whether or not it meets the requirements for seed certification.

The experiment was designed to study several instruments available for sampling bags of seed, two laboratory instruments used for subsampling, variation among samples, and subsample sizes.

2. Description of the Experiment

Ten half-pound samples were drawn from a lot of 1730 pounds of Empire Birdsfoot Trefoil seed at Alton L. Culver and Sons, Trumansburg, New York, with each of the following four probes:

Probe 1: 6 inch V-type, $\frac{1}{2}$" in diameter at open end with $2\frac{3}{4}$" long and $\frac{3}{16}$" to $\frac{5}{16}$" wide slot.

Probe 2: 6 inch straight type, $\frac{1}{4}$" in diameter with $\frac{15}{16}$" x $\frac{3}{16}$" slot.

Probe 3: 12 inch long straight type, $\frac{1}{4}$" in diameter with $\frac{13}{16}$" x $\frac{1}{8}$" slot.

Probe 4: 30 inch long sleeve-type probe with 9 equally spaced $1\frac{7}{8}$" x $\frac{1}{4}$" slots.

Thus, in all, forty one-half-pound samples were drawn.

At the seed laboratory each sample was subsampled four times (with replacement), one 2-gram and one 5-gram subsample being taken with each of two subsampling instruments, a Gamet divider and a Boerner mixer. Each subsample was analyzed and percents pure seed, other crop seed, weed seed, and inert matter recorded.

3. Results and Conclusions

On the basis of this experiment we wish to select a probe, a subsample size, and a subsampling device for use in sampling Birdsfoot Trefoil seed for certification. In each case the statistical analysis is done separately for pure seed, other crop seed, and weed seed.

Two factors are important in making a decision in each case -- bias and sample variation. To check for bias, tests of hypotheses about population means are made. If it can be concluded in each case that the population means are the same, that is, that the various sampling and subsampling devices result in estimates of the same population mean, it will be assumed that in all estimates the bias is negligible. (It is highly unlikely that all estimate the population mean with exactly the same amount of bias.)

The null hypothesis that the four probes estimate the same population mean is tested by means of an F-test in an analysis of variance on totals over the four subsamples. The results are presented in Table 1.

Table 1

		Pure seed		Other crop		Weed seeds	
Source	df	MS	F	MS	F	MS	F
Probes	3	.3382	1.06	.5560	2.42	.0358	<1
Error	36	.3136		.2298		.0634	

Since $F_{.05}(3,36\ df)=2.86$, we conclude that at the 5% level of significance there is for each variable no evidence for rejecting the null hypothesis that the four probe means are the same.

The null hypothesis that there are no differences among the subsample means is tested separately for each probe as follows. Since homogeneity of variance is not a reasonable assumption, analysis of variance is not appropriate. Hence, the non-parametric "Sign Test" is used (Reference 3, section 5.8). For each probe the ten samples are tested as replicates and the Sign Test is used on totals over subsample sizes to test for differences between dividers, on totals over dividers to test for differences between sizes and on differences between dividers for each subsample size to test for a size by divider interaction. The sample Chi-Square values (each with 1 df) are given in Table 2.

Table 2

Variable	Probe	χ^2 Dividers	χ^2 Subsample Sizes	χ^2 Interaction
Pure Seed	1	1.60	.00	.00
	2	.00	.40	.40
	3	.40	.40	.40
	4	.40	2.50	6.40
Other Crop Seed	1	.40	.40	.40
	2	.00	.40	.00
	3	.40	.40	3.60
	4	.10	.40	3.60
Weed Seed	1	6.40	.40	.40
	2	.40	1.60	1.60
	3	6.40	1.60	.00
	4	.00	.40	.00

Since $\chi^2_{.05}(1\ df)=3.841$, three of the sample χ^2's are significant at the 5% level. However, if the null hypothesis were true in all 36 cases, then on the average in .05(36)=1.80 cases the sample Chi-Square would exceed 3.841. A Chi-Square test may be used to determine whether the observed value of three significant Chi-Square is compatible with the expected number, 1.8, namely

$$\chi^2 = \frac{[33-36(.95)]^2}{36(.95)} + \frac{[3-36(.05)]^2}{36(.05)}$$

$$= 0.84$$

$$< 3.841 .$$

It is not unreasonable to assume no differences among the means in all cases.

Under these circumstances, probe, divider, and subsample size are selected on the basis of the sample variances:

a) Selection of probe. The four within probe variances are compared for each method of subsampling. These variances and their rank in each case are given in Table 3.

Table 3

Variable	Probe	Gamet Divider				Boerner Mixer			
		2 gm		5 gm		2 gm		5 gm	
		Var.	Rank	Var.	Rank	Var.	Rank	Var.	Rank
Pure Seed	1	.0693	3	.0158	2	.1312	1	.0327	2
	2	.1048	1	.0366	1	.1041	2	.0367	1
	3	.0880	2	.0138	3	.0593	4	.0221	3
	4	.0399	4	.0113	4	.0854	3	.0190	4
Other Crop Seed	1	.0560	3	.0094	4	.1321	1	.0312	2
	2	.0628	1	.0232	1	.0477	3	.0451	1
	3	.0605	2	.0190	2	.0395	4	.0113	4
	4	.0467	4	.0110	3	.0640	2	.0171	3
Weed Seed	1	.0088	3	.0067	2	.0146	3	.0056	2
	2	.0171	1	.0052	3	.0187	1	.0025	3
	3	.0121	2	.0199	1	.0155	2	.0080	1
	4	.0058	4	.0024	4	.0088	4	.0011	4

If the samples are assumed to be drawn from normal populations, the probability of correctly separating four populations into two groups, the first group consisting of that population with the smallest variance, and the second, of the remaining three populations, is .6698 when each sample variance has 9 df and the variance in each population in the second group is twice that in the first.(1) (Thus, if these assumptions are met, the probability that all of the above rankings are correct is small indeed, $(.6698)^{12}$, or approximately 1% if all groups are mutually independent.)

The results in Table 2 do not give conclusive evidence that any one probe is better than the remaining three. (Furthermore, a comparison based on subsample totals is even more inconclusive.) The average ranks for the four probes are, respectively, 2.3, 1.6, 2.5 and 3.6. On this basis one would select probe 4.

The use of probe 4, however, causes a good deal of inconvenience to the sampler. Because of the size and structure of this probe, each bag of seed being sampled must be opened and then resewn after sampling. Thus probe 4 would be used only if it were quite clear that its use would result in a great increase in efficiency, i.e., in consistently much smaller variances.

Under these circumstances it would be desirable to use instead probe 3, the next best. The data indicate that this probe is not much different from probe 4 except on weed seeds. It is felt that the poor showing on weed seeds may be due to the particular type of weed seeds present in the lot, a type which are long, narrow seeds and known to occur frequently in clumps. This would particularly affect the results in the case of the probe 3 because of the smaller size of the slot in this probe. Since the type of weed seeds present will vary from lot to lot, it is likely that in most cases in practice probe 3 should be better than these data would indicate.

b) <u>Subsample sizes</u>. It is clear from the results given in Table 3 that the variance among 5-gram subsamples within probes within dividers is consistently much smaller than that among 2-gram subsamples. The efficiency of 5-gram subsamples relative to 2-gram subsamples is computed as $\frac{V(2\text{-gram})}{V(5\text{-gram})}$, and is theoretically expected to be 5/2 (there are 5/2 as many seeds in a 5-gram subsample). Observed relative efficiencies, denoted E, are tabulated in Table 4.

Table 4

Variable	Probe	$E = \frac{V(2\text{-gram})}{V(5\text{ gram})}$ Gamet	Boerner
Percent Pure Seed	1	4.39	4.01
	2	2.86	2.84
	3	6.38	2.68
	4	3.53	4.49
Percent Other Crop Seed	1	5.96	6.20
	2	2.71	1.06
	3	3.18	3.50
	4	4.25	3.74
Percent Weed Seed	1	1.31	2.61
	2	3.29	7.48
	3	0.61	1.94
	4	2.42	8.00

The average of the twenty-four observed relative efficiencies is 3.73, considerably larger than the expected 2.5.

A possible explanation for the greater than expected relative efficiency is that the dividers are physically incapable of randomly sorting out a subsample as small as 2 grams. This would imply that still greater gains may be expected by further increasing subsample sizes. Further experimentation along these lines may be warranted. In any case, it is apparent that the subsamples should be at least 5 grams.

c) Comparison of the subsampling devices. In Table 3, 24 variances are presented for each divider (2 subsample sizes x 4 probes x 3 variables). In 15 of these 24 cases, the Gamet yields a smaller among subsamples variance than does the Boerner. The Sign Test results in

$$\chi^2_{1\ df} = \frac{2(15-12)^2}{12}$$

$$= 1.5 ,$$

which is not significant at the 5% level. There is no evidence to reject the hypothesis that the two dividers have the same among subsamples variance.

d) The among samples variance component. In order to estimate the Among Samples component of variation, a "within and among samples" analysis of variance was computed for each probe and each variable. The break-down of degrees of freedom and mean square expectations in the analysis are as follows:

Source	df	E(MS)
Among Samples (S)	9	$\sigma^2_W+4\sigma^2_S$
Treatments	3	
Subsample Sizes (G)	1	
Dividers (D)	1	
GxD	1	
Treatments x Samples	27	
GxS	9	σ^2_W
DxS	9	$\sigma^2_W+2\sigma^2_{DS}$
GxDxS	9	$\sigma^2_W+\sigma^2_{GDS}$
Total	39	

where σ_W^2 is the among subsamples within samples component and σ_S^2 is the among samples component of variation.

While the possibility that there may be a treatment by samples interaction cannot be ignored, it seems reasonable to assume that there is no sample by subsample size interaction, so that

$$E[MS(GxS)]=\sigma_W^2+0 .$$

In this case an estimate of the among samples variance component is in each instance

$$\hat{\sigma}_S^2= \frac{MS(S)-MS(SxG)}{4} .$$

The estimated among samples variance components in each case and the average of the four estimates for each variable are given in Table 5.

Table 5

Variable	Probe	$\hat{\sigma}_W^2$	$\hat{\sigma}_S^2$
Percent Pure Seed	1	.0434	.00153
	2	.0487	.0288
	3	.0167	.00498
	4	.0357	.00808
	Average	.0361	.0108
Percent Other Crop Seed	1	.0472	-.00357
	2	.0416	.0123
	3	.0128	.00710
	4	.0397	.00653
	Average	.0353	.00559
Percent Weed Seed	1	.00480	.00348
	2	.00506	.00257
	3	.00325	.00506
	4	.00277	.00075
	Average	.00397	.00297

For pure seed and other crop seed the average among samples component of variation is considerably smaller than the average within samples component. In the case of weed seeds the average among samples component, while again smaller than the average $\hat{\sigma}^2_W$, is relatively larger. This may again be due to the particular type of weed seeds in the lot. In general since the σ^2_S are all small, it appears that for lots of seed not less homogeneous than the lot used in this study, one-half pound samples are sufficiently large to assure a "representative" sample.

4. Summary

An experiment was conducted to study four probes used to sample bags of Birdsfoot Trefoil seed, two laboratory instruments used for subsampling, variation among samples, and subsample sizes. Probes, subsampling devices, and subsample sizes are selected on the basis of the respective subsample variances. The experimental results and conclusions drawn were as follows:

(1) On the basis of average ranks of the subsample variances probe 4 should be selected. It is noted, however, that probe 4 has certain undesirable operating characteristics and that probe 3 is not worse than probe 4 except on weed seeds. Since the type of weed seed present will vary from lot to lot, it is concluded that one may very likely do equally as well using probe 3 as using probe 4.

(2) The average efficiency of 5 gram relative to 2 gram subsamples was found to be 3.73, considerably larger than the expected relative efficiency of 2.5. This may be due to the fact that the subsampling instruments are physically incapable of randomly sorting out a subsample as small as 2 grams. We conclude that subsamples should be of at least 5 grams.

(3) No difference between the Gamet divider and the Boerner mixer was dectected.

(4) The among samples variance components were found to be relatively small in most cases. It is concluded that for lots of seed such as this, one-half pound samples are sufficiently large.

5. Some Suggestions for Future Research

We present the following remarks and observations concerning a few of the problems left unanswered or only partially solved:

(1) In selecting a probe, much larger sample sizes are desirable in order to increase to probability of a correct ranking. A larger experiment, possibly excluding probes 1 and/or 2, may be desirable.

(2) The reason that the relative efficiency of the 5 gram to the 2 gram subsample was much greater than expected is unknown. Further work is needed to determine if still greater gains would be achieved using 6, 8, or even 10 gram subsamples.

(3) Larger sample sizes could enable the experimenter to detect a difference between the Gamet and Boerner mixers, if a difference exists.

(4) The conclusions above may or may not be applicable in general to Birdsfoot Trefoil seed. Other lots should be sampled.

(5) The among samples component of variation should be estimated for a very large lot of seed. One-half pound samples may not be large enough for larger lots.

(6) A study of the uniformity of lots in general is desirable. If the lots are always homogeneous, a one-half pound sample is more than large enough, regardless of the size of the lot.

An experiment including lots of seed of known compostion and different methods of mixing seed (carried out jointly with a company such as Culver and Sons) would help to answer some of the questions suggested in (5) and (6).

References

1. Bechhofer, R. E., and M. Sobel, "A single-sample multiple decision procedure for ranking variances of normal populations." Ann. Math. Stat. 25:273-89. (1954)

2. Seed Certification Handbook, New York Certified Seed Growers Coop., Inc., Ithaca, N. Y.

3. Snedecor, G. W., Statistical Methods, 5th ed., The Iowa State College Press, Ames, Iowa. (1956)