This is to certify that the

dissertation entitled

THE APPLICATION OF STATISTICAL METHODS

TO SEED TESTING

presented by

Hongyu Liu

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Crop and Soil Sciences

L.O. Copeland
Major professor

Date May 6, 1999

THE APPLICATION OF STATISTICAL METHODS TO SEED TESTING

By

HONGYU LIU

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Crop and Soil Sciences

1999

ABSTRACT

THE APPLICATION OF STATISTICAL METHODS TO SEED TESTING

By

HONGYU LIU

In order to evaluate seed quality with accuracy and precision, it is very important that, first, a representative sample be drawn and, second, that appropriate test techniques be applied with adequate precision and repeatability. These studies were designed to measure variability in results of seed quality tests and performance of different probes in seed sampling under various conditions.

Data were collected from conventional germination referee (CGR) tests and blind germination referee (BGR) tests on corn (*Zea mays* L.) and soybeans (*Glycine max* (L.) Merrill) conducted by laboratories of the Association of Official Seed Analysts (AOSA) and the Society of Commercial Seed Technologists (SCST) in the Midwest and Upper Great Lakes Region in 1994 to 1997. The results showed a positive inter-replicate bias in which a significant correlation existed among different replications within a laboratory in the CGR but not in the BGR tests. Tolerances were estimated and compared with those used by both the International Seed Testing Association (ISTA) and the AOSA.

Data were collected from cold test referees on corn and soybean conducted by up to 51 AOSA and SCST laboratories in the Midwest and Upper Great Lakes Region in 1993 to 1995. Variation in corn cold tests was lower than that of soybean. Variation in

50-seed corn cold tests was equivalent to or less than that for 100-seed warm germination test. Possibilities for standardization appear much better for corn than for soybean.

Five soybean seed lots representing different seed sizes were counted with an electronic counter and/or manually in 11-16 AOSA/SCST laboratories in 1996-98. Both manual and electronic methods produced results suitable to meet the needs of the seed industry for supplying seed count information. Sample size was more important to test performance than number of replications. Adjustment of moisture content of the seed to a constant level increased test variation significantly. The 1.5% tolerance of the National Institute of Standards and Technology was not adequate to cover the variation in seed counting.

Six seed lots representing various sampling conditions (various seed sizes, seed surface features, and mixture of different types of seed) were sampled with a total of 10 probes/triers plus hand grabbing. Performance of probes with different physical features varied among crops and sampling conditions. With certain exceptions, most probes provided representative samples from homogeneous seed lots. Seed lots containing blends of varieties or mixtures of contaminants with different seed sizes and flow characteristics produced different levels of accuracy with certain probes. Probes with smaller openings tended to provide samples that under-represented the longer, more chaffy seed types, while over-representing the shorter, more free-flowing components. The diameter of the opening is the most important feature of a probe that will enable it to provide a representative sample from such lots.

To My Parents

Whose unselfish and constant

encouragement throughout my study in the

United States has been an immeasurable value, this thesis is

affectionately dedicated.

ACKNOWLEDGEMENTS

The completion of this thesis would not have been possible if not for the support, guidance, encouragement, patience and confidence shown me by my major professor, Dr. Lawrence Copeland. I could never have thought about my graduate study without what you have done for me. I would like to express my deepest appreciation to you for all you have provided to me for my personal as well as professional development. I will forever consider myself the most fortunate of graduate students for all the wonderful opportunities you have provided.

I am indebted to my committee members: Dr. Dennis Gilliland, Dr. Donald Penner and Dr. Oliver Schabenberger for your patience, mentoring, valuable guidance, advice and the continuous encouragement throughout my years of graduate study. Whenever I needed your help and went to you I was never disappointed. I only wish you could receive this special recognition for your contribution to my graduate study.

I am also grateful to Dr. Dick Payne of National Seed Laboratory, Mr. Perry Bohn of Asgrow Seed Company, Mr. Ron Cook of Oregon State University, Mr. Jim Cramer of Society of Commercial Seed Technologist, Mrs. D. Jamieson of Illinois Seed Analysts Association, Mr. Steve McGuire of Michigan State Seed Laboratory, Mr. Joe Lamb of Great Lakes Hybrid Seed Company, Michigan Foundation Seed, and the staff of seed testing laboratories at Oregon State

University and the Agriculture Departments of the states of Washington and Idaho who provided analyses for the seed probe study.

I am also thankful to Mr. Marcelo Queijo and Mr. Marcos Morales in Dr. Copeland's project group for discussions, friendship, suggestions and encouragement during my stay here. I did not only enjoy the academic ideas but also shared diverse cultures, food and philosophy from you.

Finally, I would like to thank my wife, Xiaoyu and my daughters, Xudan and Ellen for their consistently support and patience during my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

STATISTICAL BACKGROUND FOR SEED TESTING

The foundations of science are mostly dependent on observation and description, both of which are important tools of statistics. Statistics was applied to seed testing as early as the 19th century (Rodewald, 1889). Leggatt (1942) indicated that seed analysis is essentially a statistical science, dealing as it does with quantitative data affected to a marked extent by a multiplicity of causes.

STATISTICAL DISTRIBUTIONS

Three types of statistical distributions are of principal concern to seed analysts and seed scientists, namely the Binomial, Poisson and Normal distributions.

(1) Binomial distribution. The objective for a germination test is to know if a seed will germinate or not; for a purity test, whether a particle is a particular seed or not. In general, we consider an experiment or test as having n seeds or particles, each resulting in one of two outcomes, 'success' or 'failure', in our case for example 'germinated' or 'non-germinated', 'seed interested' or 'not'.

Let p = Pr (success occurs at any given test) and assume that p remains constant from test to test. Let the variable Y denote the total number of successes in n independent tests. Then Y is said to have a binomial distribution:

$$\Pr(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad \text{where } k = 0, 1, \ldots, n.$$

The binomial distribution originated with Jacob Bernoulli in 1710 and Abraham DeMoivres in 1718 (Dodge, 1971). The mean, variance and standard deviation of Y are as follows:

$$\text{The mean} = E(Y) = \mu = np$$
$$\text{The variance} = E(Y-\mu)^2 = \sigma^2 = np(1-p)$$
$$\text{The standard deviation} = \sqrt{\sigma^2} = \sqrt{np(1-p)}$$

For different sets of values of the parameters n and p, the shape of the binomial distribution varies. For p = 0.5, the graph of the binomial distribution is symmetric. That is, the probability of obtaining 0 success and n failures will be the same as the probability of obtaining n successes and 0 failure in n tests; the probability of obtaining 1 success and (n-1) failures will be the same as the probability of obtaining (n – 1) successes and 1 failure; and so forth. If $p > 0.5$, successes are more likely than failures, and the graph of the distribution is skewed to the left. If $p < 0.5$, the argument is reversed. An interesting characteristic of the graph of the binomial distribution for p close to 0.5 is the bell shape. It can be shown that when the number of tests n becomes larger and larger, the graph of the binomial distribution always becomes bell shaped. If p is very close to 0.5, this bell shape appearance is evident even when n is fairly small. When $n \geq 50$ and $np \leq 5$, the binomial distribution can be approximated by the Poisson distribution.

(2) Poisson distribution. Suppose the average number of noxious weed seeds in a seed lot is small, say 5 in one kilogram and there are 10,000 particles per kilogram. The rate for noxious weed seed is only 0.05%. The computation of

the probability for such data in use of the binomial formula becomes laborious. Another distribution, namely the Poisson distribution can be applied to approximate the binomial distribution. In general, the Poisson distribution is approached when a population contains only a very small number of interested characteristics but a large sample has been examined. The distribution with probability function

$$\Pr(Y = k) = f(y) = \frac{\mu^y}{y!} e^{-y} \qquad \text{for } y = 0, 1, 2, \ldots$$

is called the Poisson distribution, named after Simeon Denis Poisson who introduced it in 1837 although DeMoivre may have discovered it almost a century earlier (Larsen and Marx, 1986). Instead of two parameters as in the binomial distribution, the shape of the distribution is dependent on only one, the mean μ. As μ increases, the probability shifts to the right and the distribution becomes more bell-shaped. When μ is no more than 1, the distribution is extremely skewed to the right with almost all of the probability located at Pr(0) and Pr(1). One of the important characteristics of the Poisson distribution is that the mean (μ) equals to its variance. Therefore, the sample mean, $\overline{y}$, provides an estimate of both μ and σ^2.

(3) Normal distribution. As mentioned above, when the number of tests n becomes larger, the graph of the binomial distribution always becomes bell shaped. Then, another important distribution can be used to approximate the binomial distribution, that is, the normal distribution. When the number of tests is

relatively large, calculating the binomial probability is very difficult and time consuming. Larsen and Marx indicated that "the limit proposed by Poisson was not the only, or even the first, approximation to the binomial. DeMoivre had already derived a quite different one, that is, normal distribution, in his 1718 tract, *Doctrine of Chances*. Like Poisson's work, DeMoivre's theorem did not initially attract the attention it deserved; however, it did catch the eye of Laplace, though, who generalized it and published in 1812" (Larsen and Marx, 1986). The DeMoivre-Laplace theorem states: Let X be a binomial random variable defined on n independent trials each having success probability p. For any numbers c and d,

$$\lim_{n \to \infty} \Pr\left(c < \frac{X - np}{\sqrt{npq}} < d\right) = \frac{1}{\sqrt{2\pi}} \int_c^d e^{(-x^2)/2} dx$$

A random variable X is said to have a normal distribution with parameters μ and σ if its probability density function $f_x(x)$ is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(x-\mu)/\sigma]^2} \qquad -\infty < x < \infty$$

The normal approximation is useful in seed testing. For example, Miles (1963) used the normal approximation to calculate tolerances for noxious weed seed for the rate of noxious weed seed beyond 25 for which calculation was very laborious for the Poisson distribution without the aid of modern computing technology.

TYPE I AND TYPE II ERRORS

In seed testing, for example, we must decide whether a seed lot is correctly labeled (e.g., germination, purity, counts, moisture, weed seed, etc.) or not. A seed inspector may determine that the germination of a seed lot is out of tolerance with labeled level (incorrectly labeled) when in fact it is not and the seed lot is not accepted. On the other hand, he/she may determine that the germination of a seed lot is within tolerance with labeled level (correctly labeled) when in fact it is not and therefore, the seed lot is accepted. In both cases, the seed inspector makes an incorrect decision due to errors resulting from setting up the critical regions (tolerance in this example).

For the examples illustrated above, the hypothesis that a seed lot is correctly labeled is tested. It is called the null hypothesis and designated by H_o. The violation of the H_o results in the acceptance of an alternative hypothesis H_1. In other words, if we find the hypothesis H_0 that the difference between the labeled quality and the quality of a tested population is zero, is not true, we accept the hypothesis H_1 that the difference is not zero.

Errors are an inevitable by-product of hypothesis testing. "No matter what sort of mathematical façade is laid atop the decision making process, there is no way to avoid the probability of drawing an incorrect inference" (Larsen and Marx, 1986). In general, if we reject a true H_o, a Type I error is committed while if we accept a false H_o, we make a Type II error. Usually, the probability of committing a Type I error is referred to as a test's level of significance, and is denoted as α. In other words, Type I error is the level of seed producer's or provider's risk.

Computing the probability of committing a Type I error is not a problem: there are no calculations necessary, since the probability equals whatever value the experimenter sets *a priori* for α or whatever risk the experimenter is willing to accept of rejecting a true alternative. A similar simplicity does not apply to Type II errors. First, Type II error probabilities are not specified explicitly by the experimenter; second, each hypothesis test has an entire range of Type II error probabilities, one for each value of the parameter under the alternative hypothesis (Larsen and Marx, 1986). The Type II error will be larger if alternatives are closer to the labeled quality in seed testing. The Type II error is typically denoted as β, the consumer's risk. The probability of rejecting a false H_o, $1-\beta$, is called the power of a test. Although β is not decided by the value of α only, we know that setting up a lower probability of Type I error (α) will increase the chance of making a Type II error, and vice versa. Dodge (1971) indicated that the Type II error or the consumer's risk may be very high, depending on sample size. The Type II error can be dramatically reduced by increasing the sample size without increasing the risk of a Type I error.

VARIATION, TOLERANCE AND OUTLIERS

Usually, seed testing has been developed to assess the quality of a seed lot. Seed quality is a concept made up of different attributes which are of interest to different segments of the seed industry – to the producer, the processor, the warehouse person, the merchant, the farmer, the certification authority and to the government or agency responsible for seed quality control. In all cases the

ultimate object of testing is to determine the value of a seed lot for planting or other purposes (ISTA, 1996).

It is neither possible nor necessary to test all individual seeds in a seed lot to determine its quality. When a sample from a population is tested, usually the result from the sample is not necessarily the true value of the population. The deviation of the test result from the true population value is the test "variation". In order to know whether a deviation of a test result from a proclaimed or labeled quality of seed is due to the test variation or a true difference from two seed populations, we must define the extent of the test variation from experimental data.

The test variation may come from different sources that can be classified into two basic types. The first is inherent variability in the experimental material. It is called a random sampling error. The random sampling error is unavoidable and is due to the chance element in a random selection process. It is predictable and follows well established probability patterns. Among the random sources of variation, most important is the heterogeneity of the characteristic of interest (e.g., germination percentage, purity, noxious weed seed count) throughout a seed lot. The second is the variation due to the lack of uniformity in the physical conduct of the experiment, or in other words, failure to standardize the experimental technique. It can be called a non-sampling error. Such errors are less predictable than random sampling error and may result from inaccuracies in the testing processes such as those caused by ill-calibrated instrument, systematic counting error, differences in analyst experience and qualification.

Results of purity tests and noxious weed seed examinations may vary because of differences in analyst's ability and experience. Germination test results may vary due to differences in temperature under which the tests are conducted as well as from the differences in media. Because seed is a live material it may not be as easily tested as non-live materials. For example, separating a normal seedling from an abnormal seedling can be very difficult while we can easily distinguish the tail from the head of a coin. Although theoretically such errors for seed testing are avoidable, in practice they are usually not. It makes components of variation of seed testing more complicated than simply applying well-known statistical distributions.

Variation in seed testing is accounted for through the use of tolerances which specify the extent by which repeated independent tests can at most differ and still be considered consistent with a postulated (labeled) value of a seed lot. Tolerances should cover variation from all sources, not only predictable non-avoidable random error but also other unpredictable errors.

Tolerances are not applied to cover variation from test mistakes that result in extreme values in a test. Statistically we define those extreme values resulting from test mistakes as outliers. When test variation is evaluated from experimental data those outliers may greatly influence the result and should be excluded from the data.

SAMPLE SIZE AND NUMBER OF REPLICATIONS

Increase in sample size will decrease test variation and therefore improve test performance. More importantly, a large sample size can reduce the risk of a Type II error. However, increasing sample size is restricted by such factors as test cost, time consumption, and labor supply.

Despite the fact that replication is used in experimental work, its meaning is not always clearly understood. If a treatment is applied to absolutely homogeneous material, there is no variation from experimental unit to experimental unit; in fact, only one replicate from the population of possible replicates is obtained even though several observations are made. For example in seed count tests, the fact that a sample of soybean seeds is counted several times by an electronic counter will not result in multiple replications but only one. For germination tests, four consecutively counted 100-seed replicates are not replications in the experimental sense although they are usually labeled as four replications. There is no difference between the germination percents of counting the whole 400-seed test and of counting them by dividing them into subgroups. The term 'sample size' may be more appropriate than 'replication' in seed testing. However, in this study, we still call them "replications'.

CORRELATION COEFFICIENT BETWEEN DIFFERENT READINGS WITHIN A LABORATORY

Different readings within individual laboratories for the same seed lot should be independent of each other. Otherwise, if they are correlated, the test

result will be biased. Since these readings are called "replications", the correlation coefficient is called inter-replication correlation coefficient. Estimation of an inter-replication correlation coefficient is a special case of the model 1.25b in "Applied Linear Statistical Models" by Neter et al.(1985). It was further discussed by Dr. Gilliland (Personal communication).

If $X_1, X_2, \ldots, X_n$ are binomial B(k, p) random variables, then,

$$Var(X_i) = \sigma^2 = k \times p \times (1-p)$$

where p is germination rate for the population. If X_i and X_j are correlated with the correlation coefficient ρ, then,

$$Var(X_i - X_j) = Var(X_i) + Var(X_j) - 2 \times Cov(X_i, X_j) = 2\sigma^2 - 2\rho\sigma^2 = 2\sigma^2(1-\rho),$$

assuming $Cov(X_i, X_j) = \rho\sigma_{X_i}\sigma_{X_j} = \rho\sigma^2$.

Since $E(X_i - X_j) = E(X_i) - E(X_j) = 0$ and $Var(X_i - X_j) = E(X_i - X_j)^2 = 2\sigma^2(1-\rho)$,

$$U = \frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} (X_i - X_j)^2}{\left(\frac{n(n-1)}{2}\right)}$$

is an unbiased estimator of $2\sigma^2(1-\rho)$. Therefore, $\rho = 1-(U/\sigma^2)/2$. Under the model of independent B (k, p) random variables, $E(\rho) \cong 0$ and

$$Var(\hat{\rho}) \cong \frac{2(kn)^2}{(kn-1)^2(n-1)}, \quad \text{therefore,} \quad \frac{\bar{\rho} - 0}{\sqrt{Var(\hat{\rho})/l}} \quad \text{(where } l \text{ is the number of laboratories)}$$

can be used to test: H_0: μ_ρ=0 (a hypothesis that overall the average of the correlation coefficient is zero, indicating no correlation between replications within individual laboratories).

The value of ρ can be as high as +1, the case when variation between replications equal to 0. The lower bounds, however, vary depending on p, the germination rate for the population:

$$\rho \geq \begin{cases} \dfrac{-p}{1-p} & \text{for } 0 \leq p < 0.25; \\ \dfrac{\frac{5}{6} - \frac{4}{3}p - (1-p)^2}{p \times (1-p)} & \text{for } 0.25 \leq p < 0.50; \\ \dfrac{\frac{4}{3}p - \frac{1}{2} - p^2}{p(1-p)^2} & \text{for } 0.50 \leq p < 0.75; \\ -\dfrac{1-p}{p} & \text{for } 0.75 \leq p \leq 1. \end{cases}$$

REFERENCES

1. Dodge, Y. 1971. Statistical analysis for tolerances of noxious weed seeds. Thesis of master of science in applied statistics. Utah State University. 5-30.

2. ISTA. 1996. International Rules for Seed Testing. Seed Science and Technology, Vol. 24, Supplement, Rules: 11.

3. Larsen, R. J. and M. L. Marx. 1986. An introduction to mathematics and its applications. 2nd Edition, Prentice-Hall, New Jersey: 13-303.

4. Leggatt, C. W. 1942. Statistical Methods for Seed Analysts. Edited by E. J. Doyle, Laboratory Section, Canada Department of Agriculture.

5. Miles, R. S. 1963. The handbook of tolerances and of measures of precision for seed testing. Proceeding of ISTA Vol. 28, No.3.

6. Neter, J, W. Wasserman and M. H. Kutner. 1985. Applied Linear Statistical Models. 2nd Edition, Richard D. Irwin, Inc, Illinois:6.

7. Rodewald, H. 1891. The theory of probability applied to seed testing (A translation of Landw. Versuchs-Stationen published in 1889). Agricultural Science 5(3): 74-105.

CHAPTER TWO

VARIABILITY OF GERMINATION TESTS OF CORN AND SOYBEANS

ABSTRACT

This study was conducted to measure variability in results of germination tests, to estimate tolerances needed to cover variation in test results, and to determine the minimum number of seeds required for producing germination results within the limits of established tolerances. Data were collected from conventional germination referee (CGR) tests on corn (*Zea mays* L.) and soybeans (*Glycine max* (L.) Merrill), conducted by up to 46 Association of Official Seed Analysts (AOSA) and Society of Commercial Seed Technologists (SCST) laboratories in the Midwest and Upper Great Lakes Region from 1994 to 1996. In 1997, a "blind germination referee (BGR) test" was conducted on the same crops by 23-25 laboratories in the same region in which the individual replicates of the same seed lots were unknown to the analysts performing the test. A positive "inter-replicate bias" in which a significant correlation existed among the results of replicate tests occurred in the CGR but not in the BGR tests. Tolerance limits calculated for germination above 90% from both CGR and BGR test results were close to those in the International Seed Testing Association (ISTA) Rules, but lower than those in the AOSA Rules. Tolerances calculated for germination levels below 90% from both CGR and BGR test results were generally higher than both the AOSA and ISTA tolerances. The greatest decrease of test variability was observed when sample sizes were increased from one 100-seed replicate to two 100-seed replicates.

INTRODUCTION

A germination test is an analytical procedure to evaluate seed viability (germination) under standardized (favorable) conditions (AOSA, 1998). The percent germination reflects the planting value of a seed lot as well as its storage potential and is perhaps the single most convincing and accepted index of seed quality. Like other tests, results of repeated germination from the same seed lot can be expected to vary from test to test. The extent of variability depends on sources of error such as heterogeneity of the seed lot, sampling technique, test technique and analyst's skill, in addition to random sampling variability. It is important to compare results from different tests of the same seed lot for quality control and law enforcement purposes. This is done by application of tolerances that define the limits by which a second test may vary from the labeled quality without being considered out of tolerance.

Rodewald (1889) suggested tolerances for germination tests as well as the use of 400 seeds. However, he presented no evidence to justify either the tolerances or the sample size and it is still unclear if they were established on the basis of experimental or theoretical work. The earliest Official Rules for Testing Seeds with tolerances for germination tests were published in 1917. These tolerances were based on the paper of Rodewald with minor modifications (AOSA, 1917). The AOSA used these tolerances to cover variation among duplicate tests (differences between two 100-seed replicates within a laboratory) before 1937. Thereafter, these tolerances have been used to determine whether a subsequent germination agrees with the previous (labeled) test result.

Miles (1963) conducted several investigations using the binomial distribution as a model. The tolerances he estimated from germination referee tests were accepted by the International Seed Testing Association (ISTA, 1963). However, the original tolerances published in the AOSA Rules of 1917 based on Rodewald (1889) are still used by AOSA (AOSA, 1998).

Both the ISTA and AOSA Rules specify that four replicates of 100 seeds each be tested to estimate germinability of a seed lot. In most seed laboratories, the same analyst evaluates all four replicates in sequence. The need to test 400 seeds has never been critically examined. To test four 100-seed replicates is expensive and many analysts believe that they can obtain acceptable test results by testing fewer than 400 seeds.

The objective of this study was to evaluate the variability in routine corn and soybean germination tests to determine: (1) whether the present tolerances in both AOSA and ISTA Rules are appropriate and (2) to determine the minimum number of seeds required for reliable germination test results.

MATERIALS AND METHODS

Conventional Germination Referee (CGR) Tests

In order to measure variation in routine germination tests, data were collected and analyzed from corn and soybean germination referee tests from 1994 to 1996. AOSA and Society of Commercial Seed Technologists (SCST) laboratories in the Midwest and Upper Great Lakes Region conducted those tests. Seed lots were obtained by the Illinois Seed Analyst Association and samples were distributed to 46 participating laboratories throughout the Midwest

and Upper Great Lakes regions. Each laboratory was asked to conduct a 400-seed germination test for each seed lot following the procedures described in the AOSA Rules within a month after receiving samples. Eight seed lots for each species were tested each year. A total of 24 lots for each crop were tested (Table 1). Only data from laboratories that fully complied with the AOSA Rules were used.

Table 1. Germination ranges of seed lots used in germination referee tests.

Crop	Number of seed lots	Germination level (%)*
Conventional germination referee (CGR) test		
Corn	23	93-98
	1	81
soybean	23	88-97
	1	69
Blind germination referee (BGR) test		
corn	5	71, 82, 86, 90, and 94
soybean	5	74, 80, 85, 91 and 95

* The values of germination levels were the mean germination from the CGR and BGR tests although the germination levels of each lot for BGR tests were pre-tested

Blind Germination Referee (BGR) Tests

Since preliminary analysis showed that a positive inter-replicate bias existed in the CGR test results, a blind germination referee (BGR) was conducted among 23-25 AOSA and SCST laboratories for both corn and soybean in 1997. Five seed lots representing different germination levels (Table 1) for each crop were provided by Great Lakes Hybrids, Ovid, Michigan. Eight 100-seed samples were randomly drawn from each lot, comprising a total of 40 samples for each crop for

each laboratory. Laboratories were asked to conduct a germination test on each single 100-seed sample following the procedures described in the AOSA Rules. This was called a blind germination referee (BGR) test because the forty samples were not identifiable with respect to seed lot by the laboratory as in the conventional germination referee (CGR) test. Tests were completed within one month after samples arrived at each laboratory. Although many of the laboratories which participated in this referee also participated in the CGR, the two referees were completely different and should each provide independent estimates of variability in germination test results.

Statistical Analyses

1. Elimination of outliers from data. Any experiment or test may confront outliers which are observations that are considered too far from the population mean to be useful in describing test variability. Several approaches can be used to define outliers (Miles, 1963; Tattersfield, 1979 and Kenkel, 1989). For comparison and consistency, the method to eliminate outliers described by Miles (1963) was applied in this study. If a laboratory mean for a lot diverged more than 4 standard deviations away from the mean across laboratories, the data from that laboratory were considered as outliers and omitted from the analyses.

2. Inter-replication correlation coefficient. Estimation of inter-replication correlation coefficient is a special case of the model 1.25b in "Applied Linear Statistical Models" by Neter et al. (1985). Let p be the germination rate for a population. For a germination test of n replications and k seeds each, the

correlation coefficient (ρ) between replicates for each laboratory was calculated as

$$\text{Correlation coefficient} = \rho = 1 - \frac{U}{2\sigma^2},$$

$$\text{where } U = \frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} (X_i - X_j)^2}{\left(\frac{n(n-1)}{2}\right)} \quad \text{and} \quad \sigma^2 = k \times p \times (1-p).$$

The hypothesis that no correlation existed (H_0: ρ=0) was tested as:

$$z = \frac{\bar{\rho} - 0}{\sqrt{Var(\hat{\rho})/l}} \qquad \text{where } Var(\hat{\rho}) \cong \frac{2(kn)^2}{(kn-1)^2(n-1)}$$

and l is the number of laboratories in the test.

If a significant correlation was present, results from different replications within individual laboratories were biased. A negative correlation implies a higher variation than random sampling error while a positive correlation may show readings within a laboratory which are too similar compared to our expectation under the binomial law. We call the latter a positive inter-replication bias.

The ratio s/σ of observed vs. expected standard deviations described by Miles (1961) was used as additional evidence of the occurrence of inter-replication bias. For each sample in each laboratory, the actual variance, s^2, of n k-seed replications was computed and compared to the random sampling error, σ^2, by computing the ratio s/σ.

3. Sample sizes. The germination data for different sample sizes were generated in the following way. For single100-seed replicates, one of the four

100-seed replicate test results within an individual laboratory was randomly selected to construct a subset of data for each seed lot of the CGR tests. Likewise, for two 100-seed replicates, two of the four 100-seed replicate test results were randomly selected for the subset. Similar methods were used for all sample sizes for both CGR and BGR tests (Figure 1). Thirty subsets of data were obtained by repeating the procedure twenty nine additional times for each sample size for each seed lot. However, for sample sizes of four 100-seed replicates for CGR tests and eight 100-seed replicates for BGR tests, no subset of data could be drawn because of their combination limitation.

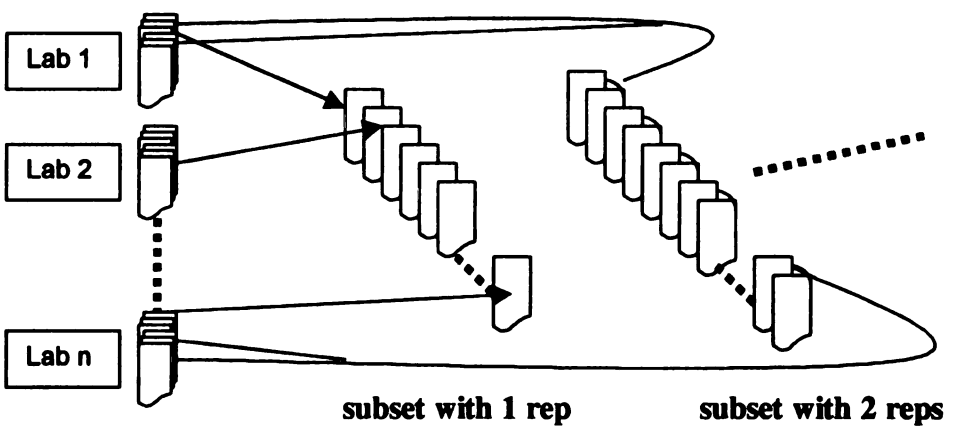

Figure 1. Construction of subsets of data with different number of replications.

The standard deviation was measured for each subset of data. The averaged standard deviation from thirty subsets of data for each sample size was used to evaluate the relationship between sample size and test variability.

4. Tolerance estimates. Tolerances were estimated with the method described by Miles (1963). The standard deviation, s, among laboratories was computed for each lot representing different levels of germination. The regression of the

ratio, f (= s/σ), on germination rate (p) is calculated for each crop. Tolerance estimates are then calculated from the equation, $T = f \times \sigma_1 \times t_\alpha$, where, $t_\alpha = 1.65$ (assuming a 5% significant one-sided test), $\sigma_1 = \{2 \times [p \times (100 - p)/n]\}^{1/2}$, and n = number of replicates.

RESULTS AND DISCUSSION

Inter-replicate Bias

A positive inter-replicate bias existed for CGR tests of both corn and soybean. Theoretically, replicates should be independent of each other and the result from one replicate should not influence the outcome of others. Table 2 shows that a very significant positive correlation coefficient existed among replications within individual laboratories for CGR tests but not for BGR tests.

Table 2. Correlation coefficient (ρ) among replications averaged across laboratories in germination tests.

Test	Corn	Soybean
CGR	0.416***	0.401***
BGR	0.060	0.069

*** At a significance level of less than 0.1%.

In the absence of experimental error, tests from homogeneous lots would show only variation due to random sampling error, compliant with the properties of the binomial distribution. By comparing observed and expected standard deviation, we can determine the precision by which tests are conducted in a laboratory. The ratio s/σ of observed vs. expected standard deviation is expected to vary around the value of one. Using this method, Miles (1961) found

that for many tests, variation within a laboratory was much less than that expected from random sampling error. He speculated that the person who counted the seedlings might remember the count from the first replicate or two and tended to make the other counts similar. He concluded that this phenomenon was probably unconscious--at least in most cases. Miles evaluated data from an international referee and confirmed this conclusion (Miles, 1961). In our study (Table 3), the ratio (s/σ) for CGR tests was asymmetrically distributed. Most (75.9% to 86.5%) of the 1200 tests exhibited ratios less than 1 and 33-48% of the tests showed ratios less than 0.5. This result implies that the test variation

Table 3. Percentage of test cases that fall in certain range of ratio s/σ (observed vs. expected standard deviation within a laboratory) in germination referee tests.

Year	Number of cases	% of test cases that fall in the ratio (s/σ) range of				
		< 0.5	0.5 - 1	1 - 1.5	1.5 - 2	>2
			Corn CGR test			
94	204	33.8	48.5	15.7	1.0	1.0
95	205	38.0	42.9	17.6	1.5	0.0
96	157	47.8	37.6	11.5	3.2	0.0
			Soybean CGR test			
94	199	32.7	43.2	18.6	5.0	0.5
95	200	40.0	46.5	13.5	0.0	0.0
96	227	37.4	43.6	15.0	2.6	0.0
			Corn BGR test			
97	115	3.5	41.7	41.7	10.4	2.6
			Soybean BGR test			
97	125	0.0	46.4	35.2	15.2	3.2

within a laboratory was small for most of the tests and very small for one third to half of the tests. We called this phenomenon a positive inter-replication bias. The results from different replicates for the same lot within a laboratory were positively correlated. For BGR tests, the ratio s/σ was symmetrically distributed around the expected value of 1. Among the 240 tests, 82.5% had ratios between 0.5 and 1.5. Only 1.7% of the 240 tests had ratios less than 0.5. It is necessary to keep in mind that heterogeneity of a seed lot would result in $s/\sigma > 1$. Thus, the positive inter-replicate bias did not occur with the BGR test by concealing the identity of the replicates.

Sample Sizes

Test variation decreased with the increase in sample sizes. Table 4 shows that standard deviation in germination percentage decreased as the number of 100-seed replicates increased from one to seven for corn BGR tests. However, the greatest drop in standard deviation among laboratories always occurred when number of replicates increased from 1 to 2. For example, for seed lot A in Table 4, the standard deviation decreased by 1.35 unit (or 39%) with an increase in the number of replicates from 1 to 2 while the additional decrease was only 0.31 unit (9%) for 3 replicates. Similar results were observed from all other tests in this study, suggesting that no more than two 100-seed replicates may be necessary for germination testing, depending on the accuracy desired (See Appendix A for the results from other seed lots). However, once the level of test variability is determined, the issue of sample size to use become somewhat arbitrary, depending on philosophical (e.g., seed law enforcement or

Table 4. Relationship between standard deviation (germination %) and number of 100-seed replicates in corn BGR testing in 1997.

Number of replications	Seed lot				
	A (92%)*	B (89%)	C (86%)	D (81%)	E (70%)
1	3.46	4.07	5.25	5.85	6.03
2	2.11	3.17	4.00	4.35	4.72
3	1.80	2.92	3.56	4.18	4.00
4	1.73	2.84	3.48	4.21	4.06
5	1.75	2.88	3.40	4.13	4.09
6	1.64	2.89	3.43	4.00	4.03
7	1.64	2.83	3.37	3.95	3.90
8	1.64	2.83	3.37	3.95	3.90

* Percentage in the parentheses was the average of germination test results for the seed lot.

quality assurance) considerations. Some of the AOSA and ISTA germination tolerances are based on a 5% significant level, implying a willingness to make a Type I error (i.e., rejecting a correctly labeled seed lot) 5% of the time. These tolerances are based on a given level of accuracy as reflected in the magnitude of tolerances required. For the most part, these tolerances are based on testing 400 seeds (four 100-seed replicates). Any decision to test fewer than 400 seeds should also recognize the necessity of using wider tolerance if the same level of accuracy is to be maintained. Suggested tolerance levels appear in the next section.

Tolerance estimates

For germination percentages between 50 and 99% for a four 100-seed replicate test, tolerances computed on the basis of variability from CGR and BGR are compared with these in the ISTA and AOSA Rules referees (Table 5). For

Table 5. Tolerance estimates for germination tests with 4 100-seed replicates at the 5% significant one-sided test, compared with the ISTA and AOSA Rules.

Germination	Soybean		Corn		AOSA	ISTA
Level	BGR test	CGR test	BGR test	CGR test	Rules	Rules
	Germination %					
99	1	1	2	1	5	2
98	2	2	3	2	5	3
97	3	3	3	3	5	3
96	3	3	4	3	5	4
95	4	4	4	4	6	4
94	4	4	5	4	6	4
93	5	5	5	5	6	5
92	5	5	5	5	6	5
91	5	6	6	6	6	5
90	6	6	6	6	6	6
89	6	7	6	7	7	6
88	6	7	6	7	7	6
87	7	8	7	7	7	6
86	7	8	7	8	7	7
85	7	8	7	8	7	7
84	8	9	7	9	7	7
83	8	9	8	9	7	7
82	8	10	8	10	7	7
81	9	10	8	10	7	8
80	9	10	8	11	7	8
79	9	11	8	11	8	8
78	9	11	8	12	8	8
77	10	12	8	12	8	8
76	10	12	9	13	8	8
75	10	12	9	13	8	9
74	10	13	9	14	8	9
73	11	13	9	14	8	9
72	11	14	9	15	8	9
71	11	14	9	15	8	9
70	11	14	9	15	8	9
69	12	14	9	16	9	10
68	12	15	9	16	9	10
67	12	15	10	17	9	10
66	12	16	10	17	9	10
65	13	16	10	18	9	10
64	13	16	10	18	9	10
63	13	16	10	18	9	10
62	13	17	10	19	9	10
61	13	17	10	19	9	10
60	14	17	10	20	9	10
59	14	18	10	20	10	11
58	14	18	10	21	10	11
57	14	18	10	21	10	11
56	14	18	10	21	10	11
55	14	19	10	22	10	11
54	15	19	10	22	10	11
53	15	19	10	22	10	11
52	15	20	10	23	10	11
51	15	20	10	23	10	11
50	15	20	10	23	10	11

germination of 90%, all estimates were identical to the AOSA and ISTA tolerances. The estimates from this study were closer to tolerances in the ISTA Rules than in the AOSA Rules for germination above 93%. For germination below 85%, estimates from CGR tests on both crops and BGR tests on soybeans were generally higher than that under both Rules. Since more rigorous (smaller) tolerances work against the seller, tolerances for germination above 93% in the AOSA Rules are favorable to the seller while those for lower germination levels work against the seller. These results indicate that tolerances in both Rules can cover the variation for germination levels above 85% but not for lower germination levels. Tolerance estimates for different number of 100-seed replicates from this study are shown in Table 6 (see next page).

SUMMARY AND CONCLUSIONS

1. Inter-replicate bias existed in routine CGR tests. The blind referee test method provided an effective approach to avoid inter-replicate bias that influence germination test results. Perhaps elements of the BGR concept could be incorporated into standard laboratory practice or other practices implemented to reduce the likelihood of inter-replicate bias. Two possibilities may be:

   (A) Conceal the identity of different replicates of the same lot.

   (B) Separate the replicates in time and/or space.

2. Variability in test results decreased with increase in the amount of seed used for the test, however, use of more than two 100-seed replicates did not reduce

Table 6. Tolerance estimates for germination tests of 1 to 4 100-seed replicates at the 5% significant one-sided test.

Germination level (%)	Soybean BGR				Soybean CGR				Corn BGR				Corn CGR			
	1 rep	2 reps	3 reps	4 reps	1 rep	2 reps	3 reps	4 reps	1 rep	2 reps	3 reps	4 reps	1 rep	2 reps	3 reps	4 reps
99	3	2	2	1	4	3	2	2	4	3	2	2	3	2	2	1
98	5	3	3	2	5	4	3	3	6	4	3	3	4	3	3	2
97	6	4	4	3	6	5	4	3	7	5	4	4	6	4	3	3
96	7	5	4	4	8	5	4	4	8	6	5	4	7	5	4	3
95	8	6	5	4	9	6	5	4	9	6	5	4	8	5	4	4
94	9	6	5	4	10	7	6	5	10	7	6	5	9	6	5	4
93	10	7	6	5	10	7	6	5	11	7	6	5	10	7	6	5
92	10	7	6	5	11	8	7	6	11	8	6	6	11	8	6	5
91	11	8	6	5	12	9	7	6	12	8	7	6	12	8	7	6
90	12	8	7	6	13	9	8	7	12	9	7	6	13	9	7	6
89	13	9	7	6	14	10	8	7	13	9	7	6	14	10	8	7
88	13	9	8	6	15	11	9	7	13	9	8	7	14	10	8	7
87	14	10	8	7	16	11	9	8	14	10	8	7	15	11	9	7
86	15	10	8	7	17	12	10	8	14	10	8	7	16	12	9	8
85	15	11	9	7	17	12	10	9	15	10	9	7	17	12	10	8
84	16	11	9	8	18	13	10	9	15	11	9	8	18	13	11	9
83	16	12	9	8	19	13	11	9	16	11	9	8	19	14	11	9
82	17	12	10	8	20	14	11	10	16	11	9	8	20	14	12	10
81	18	12	10	9	21	14	12	10	16	11	9	8	21	15	12	10
80	18	13	10	9	21	15	12	11	17	12	10	8	22	16	13	11
79	19	13	11	9	22	16	13	11	17	12	10	8	23	16	13	11
78	19	14	11	9	23	16	13	11	17	12	10	9	24	17	14	12
77	20	14	11	10	24	17	14	12	17	12	10	9	25	18	14	12
76	20	14	12	10	24	17	14	12	18	12	10	9	26	18	15	13
75	21	15	12	10	25	18	14	13	18	13	10	9	27	19	15	13
74	21	15	12	10	26	18	15	13	18	13	10	9	28	20	16	14
73	22	15	13	11	27	19	15	13	18	13	11	9	29	20	17	14
72	22	16	13	11	27	19	16	14	19	13	11	9	30	21	17	15
71	23	16	13	11	28	20	16	14	19	13	11	9	30	22	18	15
70	23	16	13	11	29	20	17	14	19	13	11	9	31	22	18	15
69	24	17	14	12	29	21	17	15	19	14	11	10	32	23	19	16
68	24	17	14	12	30	21	17	15	19	14	11	10	33	23	19	16
67	25	17	14	12	31	22	18	15	19	14	11	10	34	24	20	17
66	25	18	15	12	31	22	18	16	20	14	11	10	35	25	20	17
65	26	18	15	13	32	23	18	16	20	14	11	10	36	25	21	18
64	26	18	15	13	33	23	19	16	20	14	11	10	37	26	21	18
63	26	19	15	13	33	24	19	17	20	14	12	10	37	27	22	18
62	27	19	16	13	34	24	20	17	20	14	12	10	38	27	22	19
61	27	19	16	13	35	24	20	17	20	14	12	10	39	28	23	19
60	28	20	16	14	35	25	20	18	20	14	12	10	40	28	23	20
59	28	20	16	14	36	25	21	18	20	14	12	10	41	29	24	20
58	28	20	16	14	36	26	21	18	20	14	12	10	42	29	24	21
57	29	20	17	14	37	26	21	18	20	14	12	10	42	30	24	21
56	29	21	17	14	37	26	22	19	21	15	12	10	43	31	25	21
55	29	21	17	14	38	27	22	19	21	15	12	10	44	31	25	22
54	30	21	17	15	39	27	22	19	21	15	12	10	45	32	26	22
53	30	21	17	15	39	28	23	20	21	15	12	10	45	32	26	22
52	30	22	18	15	40	28	23	20	21	15	12	10	46	33	27	23
51	31	22	18	15	40	28	23	20	21	15	12	10	47	33	27	23
50	31	22	18	15	41	29	23	20	21	15	12	10	47	34	27	23

test variability substantially. Although we did not test the significance of the increase in precision (decrease in variability) with decreased number of 100-seed replicates, we feel that the comparison in standard deviation values alone will be meaningful to seed analysts. In this context, a 200-seed test appears adequate for providing reliable germination results for corn and soybean. However, the use of a 200-seed test will necessitate the use of different tolerances.

3. Tolerance estimates for germination levels above 90% from both CGR and BGR test results were close to tolerances under ISTA Rules, but lower than those under the AOSA Rules. These studies indicate the need for further research to determine whether the AOSA germination tolerances should be reconsidered.

4. Tolerance estimates for lower germination levels from both CGR and BGR test results were generally higher than the present tolerances in both the ISTA and AOSA Rules. Our results indicate that consideration should be given to increasing the tolerances of both the AOSA and ISTA for germination below 85%.

5. When different sample size are tested, proper tolerances for that size should be applied.

REFERENCES

1. AOSA. 1917. Rules for Testing Seed. Proceeding of AOSA 1917: 15.

2. AOSA. 1998. Rules for Testing Seed. Journal of Seed Technology 19.

3. ISTA. 1963. International Rules for Seed Testing. Proceeding Of ISTA 28(2).

4. Kenkel, J. L. 1989. Introductory Statistics for Management and Economics. 3rd Edition. PWS-KENT Publishing Company, Boston: 88:151.

5. Miles, R. S. 1961. Germination variation and tolerances. 51st Annual Meeting, Proceeding of AOSA: 86-91.

6. Miles, R. S. 1963. The handbook of tolerances and of measures of precision for seed testing. Proceeding of ISTA Vol. 28, No.3.

7. Neter, J., W. Wasserman and M. H. Kutner. 1985. Applied linear statistical models. 2nd Edition, Richard D. Irwin, Inc.: 6.

8. Rodewald, H. 1891. The theory of probability applied to seed testing (A translation of Landw. Versuchs-Stationen published in 1889). Agricultural Science 5(3): 74-105.

9. Tattersfield, J. G. 1979. Assessment of results of referee tests on germination. Seed Science and Technology, 7: 247-257.

CHAPTER THREE

VARIATION IN COLD TESTS OF SOYBEAN AND CORN SEED VIGOR

ABSTRACT

Seed vigor is widely recognized as an important attribute of seed quality. However, in practice, vigor test results are not labeled on containers of seed offered for sale. This is because the lack of standardized methods for vigor testing is thought to produce excessive variability among repeated test results, especially among tests performed in different laboratories. However, one vigor test, the cold test is routinely conducted on almost all corn and much of the soybean seed sold in the United States because it is known to be a valuable in-house method of assessing seed vigor where testing methods can be standardized. This study was designed to determine the extent of variation in cold test results conducted in routine tests among different laboratories throughout the Midwest corn and soybean area. Data were collected from cold referee tests on corn (*Zea mays* L.) and soybean (*Glycine max* (L.) Merrill) conducted by up to 51 member laboratories of Association of the Official Seed Analysts (AOSA) and Society of Commercial Seed Technologists (SCST) in the Midwest and Upper Great Lakes Region from 1993 to 1995. The results show that a significant positive correlation between replications within a laboratory existed in 100-seed tests but not in 50-seed tests for both crops. Variability in cold tests of corn was much lower than that of soybean. Variation among results of four 50-seed corn cold tests was equivalent to

or less than that among results of warm germination tests. Based on this study, four 50-seed replicates are suggested for corn cold tests. Tolerances for cold test levels of 90% or above were calculated.

INTRODUCTION

Germination tests are successful in predicting seed quality in at least two aspects. First, they have a high level of repeatability and second, they provide information about the potential germination of a seed lot under optimum conditions. The major problem with the germination test is its inability to detect potential differences in performance among higher germinating seed lots (Hampton and TeKrony, 1995), especially under adverse conditions. Interest in vigor testing began when seed scientists realized that some aspects of seed quality could not be detected by the standard germination test (Byrum and Copeland, 1995). Therefore, various vigor tests are used to detect these quality differences. The cold test is the most widely used vigor test for corn, soybean and sorghum in North America and Europe (TeKrony, 1983; Ferguson, 1990; Hampton, 1992) and is widely accepted by the seed industry in many other parts of the world (Hampton and TeKrony, 1995). Today, the cold test is performed on almost all corn seed sold in the United States. It is also commonly used for soybean seed lots planted throughout the Midwest (Byrum and Copeland, 1995).

However, the cold test method is still not a standardized test throughout the seed industry or from laboratory to laboratory. Thus, there are many reports of inconsistencies in test results among different laboratories conducted on the

same seed lot (Bradnock, 1975; Ader and Fuchs, 1978; Burris and Navratil, 1979; Fiala, 1987). Factors such as soil type, pH, substrate moisture content, crop rotation, substrate soil : sand ratio, temperature, oxygen supply in the substrate, seedborne diseases, fungicide seed treatment and duration of the cold period have been identified that may contribute to the reported variability in cold test results among different seed testing laboratories (Nijënstein, 1995).

Because of the extent of variability that is thought to exist among cold test results, cold test levels are never labeled on seed offered for sale. For labeling purposes, a test method should be standardized and the variability in test results should be accounted for by tolerances. One of the few actual studies in which measurements were made of cold test variability by conducting a blind referee test on corn, showed that cold test variability did not vary substantially more than that for standard germination test results (Byrum and Copeland, 1995). Thus, it was concluded that cold test results might be covered by the same tolerances used for standard germination tests (warm germination test). Additional studies were suggested to further determine the variability in routine cold test results and to explore the possibility of establishing tolerances suitable for cold test results.

The objectives of this study were to determine the extent of the variability among cold test results of corn and soybean seeds in routine tests to compare the variation in warm germination test results and to determine the potential factors that might contribute to test variation.

MATERIALS AND METHODS

Data sources and vigor test methods

Data on corn and soybean cold test results were collected from member laboratories of the Association of Official Seed Analysts (AOSA) and Society of Commercial Seed Technologists (SCST). These data were comprised of results from a cold test referee in which samples from the same seed lot were delivered to and tested by participating laboratories. From 1993 to 1995, 14 corn and 13 soybean seed lots were tested in up to 27 laboratories for corn and 51 laboratories for soybeans throughout the Midwest and Upper Great Lakes region. Although officially standardized methods still do not exist in either AOSA or ISTA Rules, a handbook has been developed in which preliminary suggested cold test procedures are described (Hampton and TeKrony, 1995). The principle of the cold test is to expose seeds to cold temperatures (10°C, 7 days) in non-sterile field soil at approximately 60-70% of water-holding capacity followed by a 4-7 day grow-out period under ideal conditions (25°C). Two basic cold test methods are suggested: the rolled towel and the tray methods. For the rolled towel method, paper towels of the same weight and thickness as used for warm germination test are used. For the tray method, a 45 x 66 x 2.75 cm tray made of fibreglass, plastic or metal with sheets of creped cellulose paper is utilized. A minimum of four replicates of 50 seeds each for the rolled towel or four 100-seed replicates for the tray method are suggested in the handbook. The handbook also indicate that the soil component is a very important aspect of the test and the selection of proper soil source is critical to the reproducibility. It should be

from a field site that has supported some vegetation, preferably the crop to be tested. The data we collected showed that almost all (98% and 100% for soybean and corn, respectively) of the participating laboratories used either 50-seed or 100-seed tests. However, only about 75% of the laboratories applied 4-replication tests. The number of replications used by the laboratories varied from 1 to 16. Although different number of replicates and seeds per replicate were used by different laboratories, only those data from tests representing 4 replications of 50 or 100 seeds each were analyzed. Although some of the laboratories used sterile soil, we were unable to omit their data to have a minimum of ten laboratories in each test.

Statistical analysis

1. Inter-replication correlation coefficient. Estimation of inter-replication correlation coefficient is a special case of the model 1.25b in "Applied Linear Statistical Models" by Neter et al. (1985). Let p be the vigor germination rate for a population. For a vigor test of n replications of k seeds each, the correlation coefficient between replicates within a laboratory was estimated as:

$$Correlation\ coefficient = \rho = 1 - \frac{U}{2\sigma^2},$$

$$\text{where } U = \frac{\sum_{i=1}^{n} \sum_{j=i+1}^{n} (X_i - X_j)^2}{\left(\frac{n(n-1)}{2}\right)} \quad \text{and} \quad \sigma^2 = k \times p \times (1-p).$$

The hypothesis that no correlation existed (H_0: $\rho=0$), was tested as:

$$z = \frac{\bar{\rho} - 0}{\sqrt{Var(\hat{\rho})/l}} \quad \text{where} \quad Var(\hat{\rho}) \cong \frac{2(kn)^2}{(kn-1)^2(n-1)} \quad \text{and } l \text{ is the number of laboratories in the test.}$$

If a significant correlation coefficient existed, results from different replications within individual laboratories were biased. A negative correlation implies a higher variation than random sampling error while a positive correlation may show biased readings within a laboratory. We call the latter a positive inter-replication bias.

2. Data distribution

The expected random sampling standard deviation $\sigma = (pq/n)^{1/2}$, was computed for each sample using the mean of cold test results across laboratories for p and q = 1 - p. For each sample, the deviation, d, of each laboratory vigor test results from the all-laboratory mean was divided by σ for that sample. Percentages of data distribution in terms of σ were determined.

3. Variability and tolerance estimation

To determine variability for estimating tolerances, outliers which are observations that are considered too far from the population mean to be useful in describing test variability were defined by the method described by Miles (1963). If a laboratory mean for a lot differed more than 4 standard deviations from the mean across laboratories, the data from that laboratory were considered as outliers and were omitted from the analyses.

To study the impact of number of replicates on variation, data from the cold test results were generated as follows. One of the four 50- or 100-seed replicate test results within an individual laboratory was randomly selected to

construct a subset of data for a single 50- or 100-seed replicate for each seed lot. Likewise, for two 50- or 100-seed replicates, two of the four 50- or 100-seed replicate test results were randomly selected for the subset. Similar methods were used for other replicate sizes (Figure 1, page 19). A total of 30 subsets of data were obtained by repeating the procedure twenty-nine additional times for each replicate size for each seed lot.

Standard deviation for each subset of data was obtained by the UNIVARIATE Procedure of SAS (SAS Institute, 1995). The mean standard deviation from thirty subsets of data for each replicate size was used to evaluate the relationship between number of replicates and variability.

Tolerances were estimated with the method described by Miles (1963). The standard deviation, s, among laboratories was computed for each lot representing different levels of germination. The regression of the ratio, f (= s/σ), on germination percent (p) is calculated. Tolerance estimates are then calculated from the equation, $T = f \times \sigma_1 \times t_\alpha$, where, $t_\alpha = 1.65$ (assuming a 5% significant one-sided test), $\sigma_1 = \{2 \times [p \times (100 - p)/n]\}^{1/2}$, and n = number of replicates.

RESULTS AND DISCUSSION

Inter-replication correlation coefficient

Table 7 shows that a significant positive correlation existed in 100-seed vigor test results but not in 50-seed vigor test results for both crops. It may imply that personal bias more likely occurred when sample sizes were larger.

Table 7. Correlation coefficient (ρ) among replications averaged across laboratories in cold test results.

Crop	Sample size	ρ	P_r (H_0: ρ=0)
Corn	50 seeds	-0.03932	0.6424
	100 seeds	0.37819	0.0000
Soybean	50 seeds	-0.05143	0.7658
	100 seeds	0.35633	0.0000

Variability of the cold tests results

Percentage of data distribution of cold test results in Table 8 shows that for soybean less than 50% of 100-seed test results and 65.4% of 50-seed test results were within four standard deviations. This compares with 87.4% of the warm germination test results that were within the same standard deviation range. However, for corn, 96.4% of the results of four 50-seed replicate cold tests were within four standard deviations. Compared to 92.2% in this range for warm germination results, the results of four 50-seed replicate cold test had less variability than that of four 100-seed replicate warm germination tests for corn. The variation in four 100-seed replicate tests was higher than that in warm germination tests. These results seem to indicate that the cold test may be able to perform as well or even better than standardized warm germination tests on corn if four 50-seed replicate tests are used.

Table 8. Percentage of data distribution of test results in cold referee and warm germination (WG) referee tests on corn and soybeans based on 4-replicate tests.

Data range	Corn			Soybean		
	Cold Test		WG*	Cold Test		WG
	50-seed	100-seed		50-seed	100-seed	
± 1σ**	49.1	20.8	35.4	15.7	9.0	30.5
± 2σ	76.4	45.4	61.1	29.9	25.3	56.8
± 3σ	87.3	60.8	82.4	48.8	34.3	75.7
± 4σ	96.4	76.2	92.2	65.4	48.8	87.4
± 5σ	98.2	84.6	95.8	74.8	60.2	94.8
± 6σ	98.2	90.0	97.8	84.3	70.5	96.3
± 7σ	98.2	93.8	98.5	90.6	79.5	97.4
± 10σ	100.0	96.9	100.0	100.0	90.4	100.0

* WG = Warm germination referee test.
** σ = standard deviation.

Sample size

Generally, test variation decreased with increase in number of replicates (Figure 2). The substantial decrease in test variation with an increase in number of replicates indicates the necessity of four replicates for 50-seed corn cold tests.

Tolerance estimate

Tolerances for four 50-seed replicate corn cold tests are given in Table 9. The Table shows that these calculated tolerances for corn cold tests are similar to tolerances estimated from warm germination test results and therefore, close to those in the ISTA Rules. Since only one seed lot tested had a vigor level below 90%, it may not be reasonable to estimate tolerances for vigor levels below 90% from these data. It also appears meaningless to estimate a tolerance table for soybean cold tests and for 100-seed corn cold tests from these data, since greater variability existed for those tests.

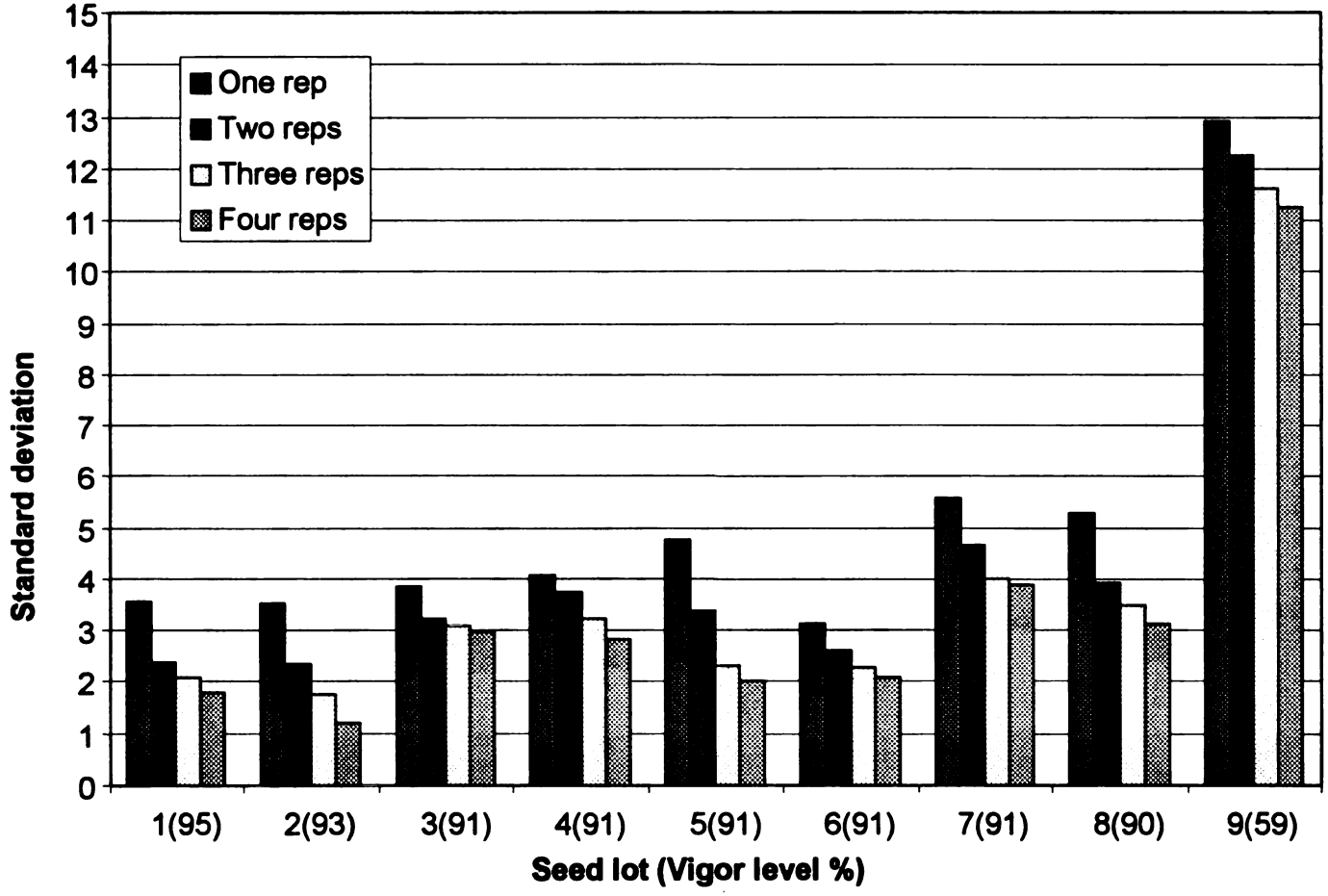

Figure 2. Relationship between number of replicates and standard deviation for 50-seed cold tests on corn in 1993-95 [outliers (3.6% of the original data) were eliminated].

Table 9. Tolerances estimated for corn cold tests and their comparison with tolerances estimated from conventional germination referee tests and those in the Rules.

Germination or vigor level	Tolerance for			
	Corn cold test (four 50-seed replicate test)	Warm germination test (four 100-seed replicate test) from or in		
(%)		CGR*	AOSA Rules	ISTA Rules
99	1	1	5	2
98	2	2	5	3
97	2	3	5	3
96	3	3	5	4
95	4	4	6	4
94	4	4	6	4
93	5	5	6	5
92	6	5	6	5
91	6	6	6	5
90	7	6	6	6

* CGR = The conventional germination referee tests conducted in 1994-95 among AOSA/SCST member laboratories. See previous chapter for more details.

SUMMARY

A positive correlation among replications within individual laboratories existed in 100-seed vigor tests but not in 50-seed tests for both crops. It implies that a positive inter-replication bias more likely occur with larger sample sizes than with smaller sample sizes. Variation in cold test results was different for corn and soybean seed. The accuracy for corn was higher than that for soybean. Variation for 50-seed corn tests was equivalent to or less than that for the warm germination test. Based on this study, a four 50-seed replicate cold test is suggested for corn. Tolerances for vigor levels of 90% or above are suggested

in Table 9. Finally, the cold test for soybean appears far from being standardized on the basis of the variation in test results. However, possibilities for standardization appear much better for corn.

REFERENCES

1. Ader, F. and Fuchs, H. 1978. Einige Hinweise zur Bedeutung der Erde als Substrat für die Kaltprüfung von Mais. Seed Science and Technology, 6, 877-893.

2. Bradnock, W. T. 1975. Report of the vigor committee, 1971-1974. Seed Science and Technology, 3(1): 124-127.

3. Burris, J. S. and Navratil, R. J. 1979. Relationship between laboratory cold-test methods and field emergence in maize inbreds. Agronomy Journal, 17: 985-988.

4. Byrum, J. R. and L. O. Copeland. 1995. Variability in vigor testing of maize (Zea maize L.) seed. Seed Science and Technology, 23, 543-549.

5. TeKrony, D. M. 1983. Seed vigor testing. Journal of Seed Technology, 8, 55-60.

6. Ferguson, J. 1990. Report of seed vigor subcommittee. Journal of Seed Technology, 14, 182-184.

7. Fiala, F. 1987. Report of the vigor test committee 1983-1986. Seed Science and Technology, 15(2): 507-522.

8. Hampton, J. G. 1992. Vigor testing within laboratories of the International Seed Testing Association: a survey. Seed Science and Technology, 20, 199-203.

9. Hampton, J. G. and D. M. TeKrony. 1995. Handbook of Vigor Test Methods. The International Seed Testing Association, P.O. Box 412, 8046 Zurich, Switzerland: 51-65.

10. Miles, R. S. 1963. The handbook of tolerances and of measures of precision for seed testing. Proceeding of ISTA Vol. 28, No.3.

11. Nijënstein, J. H. 1995. Soil cold test - European perspective: Seed Vigor Testing Seminar (Edited by H. A. van de Venter). Copenhagen, Denmark, 7 June, 1995. International Seed Testing Association, Zurich, Switzerland. 34-52.

12. SAS Institute. 1995. SAS Procedures Guide. Version 6, 3rd Edition. SAS Institute Inc., Cary, NC, USA: 617-634.

CHAPTER FOUR

VARIABILITY IN SEED COUNTS OF SOYBEANS AS DETERMINED BY MANUAL VS. ELECTRONIC METHODS

ABSTRACT

Soybean (*Glycine max* (L.) Merrill) seed counts are routinely conducted by seed suppliers throughout the seed industry for the purpose of providing their customers with information that can help them determine planting rates needed to achieve desired plant populations. These studies were conducted to determine the level of variability among both manual and electronic seed count methods for help in establishing meaningful tolerances that can be used for truth-in-labeling purposes. Tests were also conducted to determine the best combination of sample and seed sizes needed for conducting the tests. The results showed that both manual and electronic methods can be used to meet the needs of the seed industry for supplying seed count information. Sample size was more important to test performance than number of replications. Adjusting moisture content of the seed to a constant level increased test variation significantly. The 1.5% tolerance of the National Institute of Standards and Technology was not adequate to cover the variation encountered in seed counting. No correlation between seed size and test variation was observed.

INTRODUCTION

Seed count tests are offered by more and more seed laboratories for their clientele because of the interest in precision planting for a wide range of vegetable and agronomic crops. A survey of 34 laboratories by the joint Association of Official Seed Analysts (AOSA) Referee/ Society of Commercial Seed Technologists (SCST) Research Committees (McGuire, personal communication) in the Midwest Region of the United State showed that only 7 did not offer seed count services. Many laboratories conduct hundreds and even thousands of seed counts, especially for corn and soybeans. Small grains were the third most frequent kind tested. Different methods are used to obtain seed counts for soybeans. About half of the laboratories use electronic counters. Less than half test moisture at the time of the count and report the result. Twenty-four of the 34 respondents clearly thought the AOSA should have official procedures or recommendations for seed count tests.

Establishment of truth-in-labeling for seed count information requires the use of tolerances. It is well known that repeated tests on a seed sample for one or more quality factors do not necessarily give exactly the same results each time because of the variation which results from both random and non-random sources. However, it was not until 1929 that statistical methods for handling such variability were first introduced into official seed testing rules (Collins, 1929), although the use of tolerances was suggested much earlier (Rodewald, 1891). Since that time, a body of literature has developed to help seed analysts

understand how seed testing results vary and to establish tolerances that are used to determine when differences are statistically significant, i.e., whether apparent differences represent a real difference between two values or can be explained by chance alone (Banyai et al., 1988; Dodge,1971; Leggatt, 1935; Miles, 1963). If a second test is outside of tolerance from the first (or labeled quality), a real difference is considered to exist between the two tests, while within tolerances, the variation most likely is due to chance alone. Tolerances for such factors as germination, purity, and noxious weed seed, have been established by either the International Seed Testing Association (ISTA) or the AOSA, or both (ISTA, 1996; AOSA, 1998).

Since no established AOSA/SCST tolerances are available for seed counts, the 1.5% tolerance of the National Institute of Standards and Technology (NIST) has been applied in some states. In 1994, Illinois (Lair, personal communication) reported 263 potential seed count violations with the use of the NIST tolerance. Another survey study (Payne, personal communication) showed that at least 10% and up to 36% seed count test results were out of the NIST tolerance during the period of 1994 to 1997. A survey from Asgrow Seed Company (Bohn, personal communication) showed that only 32% of the 256 seed lots tested had differences between labeled and tested seed counts within ±1%, 56% within ±2%, 71% within 3% and 90% within ±4%. Many questions have arisen about count methods and the tolerances applied. Many people in the seed industry believe that the NIST tolerances are too strict and not

statistically valid for biological materials. Consequently, research was needed to answer such questions.

The objectives of this study were to: (1) evaluate comparative results obtained by manual vs electronic methods, (2) determine proper sample size and number of replications which should be used for the test, and (3) provide information for the establishment of tolerances for seed counts.

MATERIALS AND METHODS

Samples from a soybean seed lot were counted by 11 AOSA/SCST laboratories in 1995-96 and samples from another soybean seed lot with larger seeds were counted by 15 laboratories in 1996-97. All of the laboratories involved offer routine seed count services. In 1998, 16 laboratories conducted an additional seed count referee test on soybeans. Each participating laboratory used the electronic counter that was used for its routine seed count services.

1. **1995-96 tests**

A randomly selected 500-g sample from one soybean lot, with a seed size of approximately 7500 seeds/kg (3400 seeds/lb) was mailed in a sealed polyethylene bag to each laboratory from the referee coordinator. A purity test according to the AOSA Rules on the entire 500-g sample was performed, then each of the subsequent tests was conducted on the pure seed portion. A moisture test was made at the time of each count. For the manual count, eight

100-seed replicates were randomly counted out without replacement by hand from the 500-g sample as prescribed in the ISTA Rules (ISTA, 1996). The weight of each sample was recorded to three decimal places in grams. This was repeated five additional times to obtain a total of six replicate tests. For the tests with electronic counters, counts were made on six randomly selected successive samples with replacement on sample sizes of 375, 250 and 125 g with the electronic counter. Results for six replicates of tests for each sample size were recorded.

2. **1996-97 tests**

Based on the preliminary results from the previous year, a 1000-g random sample of a soybean seed lot with a larger seed size of approximately 5700 seeds/kg (2600 seeds/lb) was mailed in a sealed polyethylene bag to each participating laboratory from the referee coordinator. Fifteen AOSA/SCST laboratories completed the referee. While most of the procedures used were the same as those in 1995-96, modifications included: (1) for the manual count, the number of replicates was decreased from eight to four and four different sample sizes were used instead of one in previous year in order to evaluate impact of the sample size on the results. Therefore, four random samples for each sample size of 100, 200, 300 and 400 seeds were counted by hand and weighed; (2) for electronic counts, a sample size of 500 g was added, while the number of replicates for each sample size was maintained the same.

3. 1998 tests

Three different soybean seed lots representing seed sizes of 5199, 6240, and 7610 seeds/kg (2360, 2830 and 3450 seeds/lb, respectively) were counted. Three 1000-g samples, one for each size, were provided to each participating laboratory. The samples remained in their polyethylene bags until tested. Purity and moisture content were determined before the following procedures were conducted on the pure seed portion.

A. Counts on four replications of 500-g each. The 1000-g sample was divided though a Boerner (or Gamet) divider into two 500-g samples. Each sample was counted separately. Then, the two samples were recombined and the process repeated three additional times to obtain four replications.

B. Counts on four 375-g samples. The 1000-g sample was divided though the divider twice, achieving four 250-g samples; then, one of the 250-g samples was divided one additional time, giving two 125-g samples. One of the 125-g samples was combined with one of the three 250-g samples obtained previously, giving a sample of 375 g. Any deviation from 375-g was corrected manually. After performing a seed count on this 375-g sample, all of samples were recombined, mixed well, and the process repeated three additional times to give a total of four counts on samples of 375 g each.

C. Counts on four 250-g samples. The 1000-g sample was divided twice into four 250-g samples. Exactly 250 g were achieved by manually moving seed to or from the container. After performing a seed count on this sample, the entire

sample was recombined, mixed, and the process repeated three additional times to give four replications.

D. Finally, the process was repeated as described above to obtain a 125-g sample and seed counts performed before recombining with the entire sample and repeating the process three additional times to give counts on four 125-g samples.

Seed count for each sample was calculated as:

$$\text{Seed count (seeds/kg)} = \frac{1000 \times \text{number of seed counted in the sample}}{\text{sample weight (g)}} \times (\text{pure seed \%})$$

Both seed control officials and seed company representatives have expressed concern about the effect that variation in seed moisture content could have on the accuracy of seed counts. It was thought that changes in the percent moisture of a sample due to environmental conditions, would cause corresponding increases or decreases in the number of seeds per unit. In order to evaluate the impact of moisture content of the seed, the following formula was used to convert the data into a comparative seed count on the basis of 10 % moisture content:

$$\text{Seed count (seeds/kg)} = \frac{1000 \times (1-0.1) \times \text{number of seed counted in the sample}}{\text{sample weight (g)} \times (1-\text{mc\% tested}/100)} \times (\text{pure seed \%})$$

Observations that were disproportionately small or large were defined as outliers. Many approaches can be used to detect outliers (Miles, 1963; Tattersfield, 1979 and Kenkel, 1989). For this study the box plot method is

applied (Kenkel, 1989; SAS, 1990). The bottom and top edges of the box are located at the sample 25th and 75th percentiles with the sample media as the center of the box. The distance between the 25th and the 75th percentiles is called an interquartile range. Any data that are more than 2 interquartiles away from the media will be defined as outliers. Such outliers were excluded when calculating test results to avoid their influence on population means. In order to detect the impact of outliers in this study, both data with or without outliers were analyzed separately. The ANOVA model II $y_{ij} = \mu_i + e_{ij}$ for j^{th} reading of the i^{th} laboratory and $\sigma^2_{\bullet} = \sigma^2_{a} + \sigma^2_{w}$ was applied (Neter et al., 1996). $\sigma^2_{\bullet}$ was the total variance, σ^2_{a} and σ^2_{w} were the variance among and within laboratories, respectively. Suggested tolerances (two-sided test at the 5% significant level) for comparison of results obtained by different laboratories were calculated with the formula:

$$\text{Tolerance (\%)} = (1.96 \times \sqrt{2 \times \sigma^2_{\bar{e}}}) / (\text{seeds/kg.})$$

Suggested tolerances among replications within a laboratory (two-sided test at the 5% significant level) were calculated with the formula:

$$\text{Tolerance (\%)} = 100 \times \left(1.96 \times \sqrt{\text{number of replicates} \times \sigma^2_{\text{w}} / \text{n}} \right) / (\text{seeds/kg})$$

where n is the number of readings (replications in each laboratory).

RESULTS AND DISCUSSION

Impact of the outliers on seed count results

As defined above, outliers are the observations that are considered too far from the population mean to be useful in describing test variability. Thus, they are extreme numbers, either much higher or much lower than the population mean. The standard deviation among and within laboratories decreased an average of 35.2% and 12.1%, respectively, after outliers were eliminated (Table 10, next page). In some cases, the reduction in standard deviation was as high as 99.2% (variation within a laboratory from 500-g electronic counts in 1997) after outliers were eliminated.

The existence of outliers might also diminish the efficiency of increasing sample size or number of replications. Table 11 shows that if outliers were included, the standard deviation among and within laboratories for the sample sizes of 500 g was 1.60 and 2.31, respectively, which was much higher than that for smaller sample sizes in the same test. However, after outliers were eliminated, the standard deviation decreased with the increase in sample size. If all outliers were used, the resulting tolerances would have been as high as 14 to 15% (data not shown). Thus, we believe that variation from outliers should not be considered for establishing tolerances; instead, the test technique of laboratories producing outliers should be improved with the aid of a regular program for calibrating electronic seed counters, along with routine participation

Table 10. Changes of test standard deviation after outliers were eliminated.

Seed lot	Sample size	Standard deviation changes (±%)	
		Among laboratories	Within a laboratory
		1998 (Electronic)	
1	125 g/test	-53.9	-9.5
	250 g/test	-24.9	-11.5
	375 g/test	0.0	0.0
	500 g/test	0.0	0.0
2	125 g/test	-89.1	-56.4
	250 g/test	-14.6	-33.6
	375 g/test	0.0	0.0
	500 g/test	0.0	0.0
3	125 g/test	-93.2	7.8
	250 g/test	-43.3	5.6
	375 g/test	-49.4	2.7
	500 g/test	-79.2	12.5
		1997 (electronic)	
4	125 g/test	0.0	0.0
	250 g/test	-0.4	-0.2
	375 g/test	-4.9	-0.5
	500 g/test	-83.3	-99.2
		1997 (manual)	
4	100 seeds/test	-63.9	-60.9
	200 seeds/test	-83.0	-42.8
	300 seeds/test	-75.1	-0.7
	400 seeds/test	-42.3	8.2
		1996 (electronic)	
5	125 g/test	0.0	0.0
	250 g/test	0.0	0.0
	375 g/test	-8.2	0.1
		1996 (manual)	
5	100 seeds/test	0.0	0.0
Average		**-35.2**	**-12.1**

Table 11. Standard deviations before and after outliers were eliminated with different sample sizes for electronic counts on soybean in 1997.

Sample size (g/test)	Standard deviation (%)			
	Before outliers eliminated		After outliers eliminated	
	Among†	Within†	Among	Within
125	0.70	0.77	0.70	0.77
250	0.75	0.40	0.75	0.40
375	0.78	0.24	0.76	0.24
500	1.60	2.31	0.65	0.21

† Among = among laboratories; Within = within a laboratory

in referee and seed count educational programs. A wide range of electronic counters is used throughout the industry that collectively, along with improper seed counting techniques (including improper calibration of equipment), may contribute to the magnitude of variability observed in this study. Although laboratories were asked to calibrate their electronic counters, inconsistencies in doing so may have also contributed to the number of outliers found in this study.

The impact of seed moisture content on seed count variability

The adjustment of the seed moisture content increased the standard deviation in test results among laboratories by an average of 41.6% (Table 12). The results of moisture content among different laboratories for individual seed lots varied from 7% to over 10%, a wider range than expected, since samples were kept in sealed bags that were not opened until when they were tested. This wide range could be due to different types of moisture meters used by the

Table 12. Changes of the standard deviation among and within laboratories for seed counts when adjusting moisture content (MC) to a level of 10% for all tests of soybean in 1996-98.

Seed size (seeds/kg)	Sample size	No MC adjusted		MC adjusted		Change (±%) if MC adjusted	
		Among	Within	Among	Within	Among	Within
			1995-96 (electronic)				
7560	125g	1.04	0.80	1.60	0.81	53.4	1.0
	250g	0.62	0.67	1.40	0.68	125.3	1.3
	375g	0.80	0.37	1.37	0.37	71.3	1.3
			1995-96(manual)				
7606	100 seeds	0.94	0.95	0.96	0.95	2.7	1.0
			1996-97 (electronic)				
5764	125g	0.70	0.77	1.24	0.77	78.5	-0.2
	250g	0.75	0.40	1.11	0.40	48.7	0.1
	375g	0.76	0.24	0.92	0.24	20.9	-0.1
	500g	0.65	0.21	0.90	0.21	39.4	0.0
			1996-97(manual)				
5747	100 seeds	0.95	1.16	1.29	1.20	35.8	3.9
	200 seeds	0.50	0.64	0.83	0.69	67.4	7.8
	300 seeds	0.47	0.74	0.57	0.69	20.8	-6.8
	400 seeds	0.70	0.67	0.76	0.64	8.7	-3.9
			1998(electronic)				
6240	125g	1.12	0.55	1.43	0.54	26.7	-0.3
	250g	0.84	0.50	1.16	0.50	37.3	0.3
	375g	0.83	0.30	1.16	0.29	39.3	-0.2
	500g	0.70	0.30	1.06	0.30	50.4	0.2
5199	125g	1.85	0.72	2.07	0.72	11.5	0.0
	250g	1.38	0.62	1.41	0.62	2.1	0.2
	375g	1.07	0.63	1.07	0.63	-0.4	0.1
	500g	1.02	0.52	0.94	0.52	-8.2	0.0
7610	125g	1.41	0.69	1.70	0.69	21.1	-0.1
	250g	1.05	0.46	1.68	0.47	59.2	0.1
	375g	0.88	0.42	1.55	0.41	76.1	-0.1
	500g	0.75	0.49	1.58	0.49	110.8	-0.3
Average						**41.6**	**0.2**

laboratories and the lack of a standardized procedure for conducting moisture tests.

Comparison of manual vs. electronic counts

The manual and electronic methods in this study were comparable in test performance although electronic counts may be more cost effective (e.g., time, labor) than manual counts. Both manual and electronic counts provided reliable results if sample size was adequate (Table 12). In most cases, test variation within a laboratory was higher for manual counts than for electronic counts while variation among laboratories was smaller for manual counts than for electronic counts. This shows that improvement in manual counting should focus on the test method itself, while improvements in electronic counts should focus more on attempts to attain consistency among laboratories. The use of standardized equipment and routine equipment calibration should be an effective approach in reducing test variability among laboratories.

Sample size vs. number of replicates

In general, increase in sample size and number of replicates decreased test variation. However, the results of this 3-year study for both manual and electronic counts showed that increase in sample size was more important than number of replicates. For example, when the number of replicates for electronic counts (Figure 3, see appendix A for figures of other seed lots tested) increased

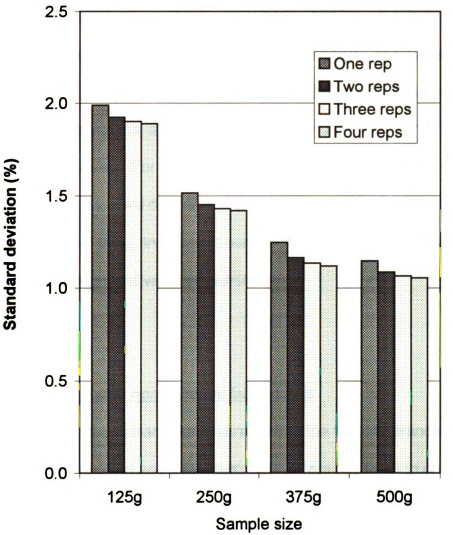

Figure 3. Standard deviation for electronic counts as a function of sample sizes and number of replicates for seed lot 2 (5199 seeds/kg) of soybean tested in 1998.

from 1 to 4, the decrease in standard deviation was 0.1 (from 5 to 8.7%) for all four sample sizes tested in 1998. However, the standard deviation decreased 0.47 (24.5%) with increase in sample size from 125 to 250 g/test, 0.29 (19.8%) from 250 to 375 g/test and 0.08 (6.7%) from 375 to 500 g/test. The same trend was observed for results from all other seed counts in this study. Based on limited data for manual counts, a sample size of at least 200 seeds and two replicates should be tested. For electronic counts, a sample size of 500 g/test resulted in a smallest standard deviation in all tests (the 500-g sample was not included in the 1995-96 test). Increase in number of replicates beyond one 500-g test did not reduce standard variation substantially.

Tolerances

A. *Tolerance for test results from different laboratories*.

Results of this study clearly demonstrated that the 1.5% tolerances of the National Institute of Standard and Technology (NIST) are too small to cover seed count variation for both electronic and manual methods. Projected tolerances at the 95% level of certainty based on the variability found in this study appear in Table 13. None of the projected tolerances are equal to or less than 1.5%. Perhaps this explains why 10% to 36% of test results of seed counts in the 1994-97 survey were in violation to the NIST tolerance and may also explain results of the 1998 Asgrow survey. It is important to note that this study

Table 13. Estimation of seed count tolerances among laboratories and among replicates within a laboratory without adjusting moisture content of the seed.

Seed lot	Sample size (per test)	Tolerance (%)						
		Among replicates within a laboratory			Among laboratories			
		2-rep	3-rep	4-rep	1-rep	2-rep	3-rep	4-rep
1998 (electronic)								
1	125-g	1.5	1.9	2.1	3.5	3.3	3.2	3.2
	250-g	1.4	1.7	2.0	2.7	2.5	2.5	2.4
	375-g	0.8	1.0	1.2	2.4	2.4	2.4	2.3
	500-g	0.8	1.0	1.2	2.1	2.0	2.0	2.0
2	125-g	2.0	2.4	2.8	5.4	5.3	5.2	5.2
	250-g	1.7	2.1	2.4	4.2	4.0	3.9	3.9
	375-g	1.7	2.1	2.5	3.4	3.2	3.1	3.1
	500-g	1.4	1.8	2.0	3.1	3.0	2.9	2.9
3	125-g	1.9	2.3	2.7	4.3	4.1	4.0	4.0
	250-g	1.3	1.6	1.8	3.1	3.0	3.0	2.9
	375-g	1.2	1.4	1.6	2.7	2.5	2.5	2.5
	500-g	1.4	1.7	1.9	2.4	2.3	2.2	2.1
1996-97 (electronic)								
4	125-g	2.1	2.6	3.0	2.9	2.5	2.3	2.2
	250-g	1.1	1.3	1.6	2.4	2.2	2.2	2.2
	375-g	0.7	0.8	0.9	2.2	2.2	2.2	2.2
	500-g	0.6	0.7	0.8	1.9	1.9	1.8	1.8
1996-97 (manual)								
4	100 seeds	3.2	3.9	4.5	4.2	3.5	3.2	3.1
	200 seeds	1.8	2.2	2.5	2.3	1.9	1.7	1.7
	300 seeds	2.1	2.5	2.9	2.5	2.0	1.8	1.7
	400 seeds	1.9	2.3	2.6	2.7	2.4	2.2	2.2
1995-96 (electronic)								
5	125-g	2.2	2.7	3.1	4.8	4.3	4.1	4.1
	250-g	1.9	2.3	2.6	3.3	2.8	2.7	2.6
	375-g	1.0	1.3	1.4	3.2	3.1	3.0	3.0
1995-96 (manual)								
5	100 seeds	2.6	3.2	3.7	4.9	4.2	4.0	3.8

only measured the variability in seed counts from samples from a single bag.

This assumes acceptable uniformity (homogeneity) within the seed lot from bag

to bag. It must be recognized that count variability will be larger if heterogeneity exists among bags in the lot. However, it should be noted that although a larger tolerance will give more certainty of avoiding the rejection of a correct label (statistically a Type I error), it will also increase the likelihood of accepting an incorrect label (Type II error). Thus, further study is needed to focus on the heterogeneity among bags within commercial seed lots.

B. *Tolerances for results from different replicates within a laboratory*.

Projected tolerances between replicates within a laboratory are given in columns 3-5 of Table 12 for the 95% level of certainty. The maximum differences allowed between any two replicates for the 250-g sample size for electronic counts varied from 1.1 to 1.9% and, for a sample size of 200 seeds/test for manual counts, was 1.8%. Thus, projected tolerances for inter-replicate variability are smaller than those for inter-laboratory test results because of the smaller levels of variability involved.

Seed size and test variation

No correlation between seed size and test variation was observed from this study.

CONCLUSIONS

1. Routine seed count referee tests should be maintained to continuously monitor and improve test performances of laboratories because some laboratories are more likely to give outliers than others. Tolerances

should not be expected to cover the variation from outliers.

2. Moisture content of the seed should not be adjusted to a certain level since more variability could result from the moisture test itself. Such adjustments greatly increased test variation in this study.

3. Both manual and electronic counts can provide reliable results, provided the sample size is adequate. However, increase in sample size is more important than the number of replicates for both methods.

4. The 1.5% tolerance of the National Institute of Standard and Technology (NIST) is too restrictive to cover seed count variation for both electronic and manual methods. Tolerances for different sample sizes and methods based on individual tests were estimated and listed on Table 13. For example, projected tolerances for a one-replicate test of 500 g ranged from 1.9% to 3.1%. Thus, the widest tolerance, 3.1% is suggested for a full coverage of test variation.

5. No correlation between seed size and test variation was observed.

REFERENCES

1. AOSA. (1998). Rules for Testing Seeds. Journal of Seed Technology: 19.

2. Banyai, J., J. Fischer and Zs. Lang. (1988). Tolerances based on Poisson Distribution. Seed Science and Technology 16:321-329.

3. Collins, G. N. (1929). The application of statistical methods to seed testing. USDA Circulars 29:1-17.

4. Dodge, Y. (1971). Statistical analysis for tolerances of noxious weed seeds.

Master thesis, Utah State University.

6. ISTA. (1996). International Rules for Seed Testing. Seed science and technology. Vol. 24, Supplement, Rules.

7. Kenkel, J. L. (1989). Introductory statistics for management and economics. PWS-KENT 3rd Edition: 131-133.

8. Leggatt, C. W. (1935). Contributions to the study of statistics of seed testing. Proceeding of ISTA. 1935:38-48.

9. Miles, R. S. (1963). The handbook of tolerances and of measures precision for seed testing. Proceeding of ISTA. Vol. 28, No. 3.

10. Neter, J., M. H. Kutner, C. J. Nachtsheim and W. Wasserman. (1996). Applied linear statistical models. 4th Edition, IRWIN, 958-976

11. Rodewald. (1891). The theory of probability applied to seed testing (a translation of Landw. Versuchs-Stationen published in 1889). Agricultural Science 3(5): 74-105.

12. SAS. 1990. SAS Procedures Guide. Version 6, Third Edition. SAS Institute Inc., Cary, NC.

13. Tattersfield, J. G. 1979. Assessment of results of referee tests on germination. Seed Science and Technology, 7: 247-257.

CHAPTER FIVE
STUDY OF RELATIVE EFFICIENCY OF DIFFERENT PROBES FOR SEED SAMPLING

ABSTRACT

This study was designed to compare the relative effectiveness of different seed sampling probes/triers in common use among seed control officials and the seed industry. A secondary objective was to identify the most effective sampling instruments for representative kinds of seed to help establish criteria for evaluating sampling instruments for different sampling conditions. Six seed lots representing different sampling conditions (various seed size, seed surface features, and mixture of different types of seed) were sampled with a total of 10 probes/triers plus hand grabbing. Results showed that performance of probes with different physical features varied among crops and sampling situations. With certain exceptions, most probes provided representative seed samples from homogeneous seed lots when properly used. However, representative samples are unlikely to be obtained by any probe from heterogeneous seed containers. Furthermore, seed lots containing blends of varieties or mixtures of contaminants with different seed size and flow characteristics, can not be sampled accurately with certain probes. Probes with smaller openings tended to provide samples that under-represented the longer, more chaffy seed types, while over-representing the shorter, more free-flowing components. We believe that the diameter of the opening is the most important feature of a probe that will enable it to provide a representative sample from such lots. Finally, all probes should be long enough to reach across the entire width or length of the container.

INTRODUCTION

The first requirement in generating any seed test is to carefully obtain a representative sample. No matter how accurately a seed analysis is made, it can only show the quality of the sample that is submitted for analysis (ISTA, 1938). Hence, it is of fundamental importance that the sample properly represents the quality of the seed lot from which it is drawn. While the correct result would be obtained by analyzing the entire lot, practicality dictates that a small sample be drawn in such a way that truly represents the quality of the entire lot.

A survey of 34 states by the American Association of Seed Control Officials (AASCO) (Nees, 1990) showed that a wide range in equipment is used for sampling seeds of various crop types existed even among seed control officials. An even wider range of sampling instruments is used for quality control in the seed industry. These include a variety of probes and triers, ranging from small 15.2-cm probes to double-sleeve triers up to 122 cm in length, depending on the type of seed and container.

Several studies have been conducted on sampling methods. In the 1930's, M.T. Munn observed the movement of seeds in bags when sampled with various instruments and found that the most practical and satisfactory type of probe or trier for closed bags was the German type (Munn, 1935). Later, a type of probe called the "sticker" was designed to obtain more representative samples (Leggatt, 1938). Debney (1960) introduced a dynamic sampling technique which was shown to be more accurate than any other device for

sampling from bags at that time. Although a simple sampling trial carried out at Dublin illustrated that the triers could produce biased samples even from homogeneous lots (Mullin, 1965), a study in Australia showed that certain triers did not differ significantly with regard to different seed components (Bean, 1970). At almost the same time Grisez and Hardin (1972) found that different sampling accuracy existed among sampling tools and methods.

This research was conducted to compare the relative effectiveness of various seed sampling probes/triers in common use among seed control officials and the seed industry and to identify the most effective sampling instruments for different kinds of seed to help establish criteria for evaluating sampling instruments for different sampling conditions.

MATERIALS AND METHODS

1. Probes: Ten probes were tested for their effectiveness and precision in sampling from different kinds of seeds. These are illustrated in Figure 4, and their characteristics are described in Table 14.

2. Seed lot preparation:

(A). Ryegrass: One thousand eight hundred and fourteen kilograms of perennial ryegrass seed were thoroughly mixed with 22.7 kg of stained (red) perennial ryegrass by a commercially accepted blending facility. Then, the seed was bagged into eighty 22.7-kg bags.

(B). A second ryegrass seed lot weighing 1,792 kilograms was thoroughly mixed with 22.7 kg of stained (red) perennial ryegrass by a commercially accepted

blending facility. Then, the seed was bagged into eighty 22.7-kg bags. As the bags were being filled, a 227-g sample of stained (green) ryegrass seed was placed in the center core of each with a "stove-piping" method using a small steel cylinder 61 cm long and 2 cm in diameter.

Table 14. Probes used for sampling probe study.

Probe No.	Length (mm)	Diameter (mm)	Opening		
			Number	Length (mm)	Width (mm)
1	1017	21	6	65	12
2	813	15	3	220	12
3	461	13	1	48	12
4	203	15	1	84	7-11
5	305	23	1	112	7-21
6	750	12	9	44	6
7	483	12	5	44	6
8	458	12.7	1	50	12.5
9	966	12.7	1	788	12.5
10	864	23	2	330	21

Note: See Figure 4 (page 65) for more details.

(B) Mixture: One thousand eight hundred fourteen kilograms of perennial ryegrass, red fescue, and Kentucky bluegrass seed (1:1:1) were thoroughly mixed with 7.71 kg each of stained (red) perennial ryegrass, red fescue, and bluegrass seeds (1:1:1) by a commercially accepted blending method. Then the seed was bagged into eighty 22.7-kg bags.

(C) Wheat: One thousand eight hundred and fourteen kilograms of wheat were thoroughly mixed with 22.7 kg of stained (red) wheat by a commercially accepted blending methods. Then the seed was bagged into eighty 22.7-kg bags.

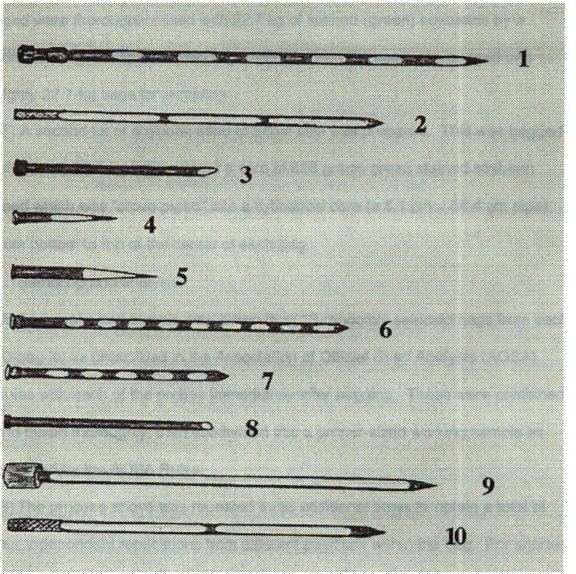

Figure 4. Probes tested in the study (Scale ≈ 1:8).

(D) Soybeans: One thousand eight hundred and fourteen kilograms of soybean seed were thoroughly mixed with 22.7 kg of stained (green) soybeans by a commercially acceptable blending method. Then the seed was bagged into eighty 22.7-kg bags for sampling.

(E) A second lot of soybean seed of equal size was prepared. This was bagged into eighty 22.7-kg bags around a core of 626 grams green stained soybean seed which was "stove-piped" into a cylindrical core (a 5.1 cm × 58.4 cm pipe) from bottom to top of the center of each bag.

1. Sampling and analysis:

(A) An equal-sized sample was drawn from 13 randomly selected bags from each 80-bag lot as prescribed in the Association of Official Seed Analysts (AOSA) Rules with each of the probes immediately after bagging. These were combined and mixed thoroughly, then subdivided into a proper-sized working sample as specified by the AOSA Rules.

(B) The process above was repeated three additional times to obtain a total of four independent replications from different positions within the bag. For shorter probes, samples of the four replicates were drawn from two positions along the side of the bag at equal distances from each respective corner while for others, the sample was taken diagonally through the bag from each of the four corners (Figure 5).

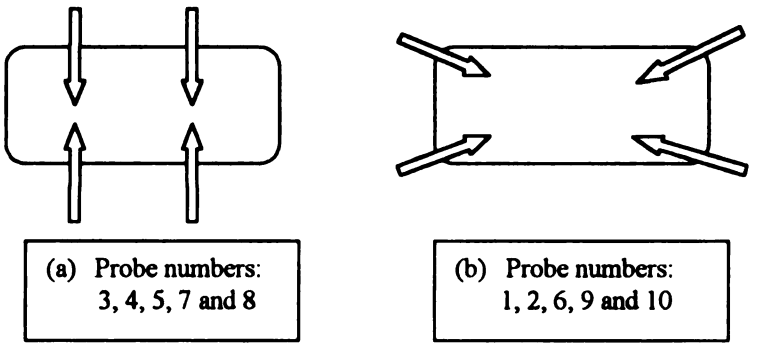

Figure 5. Inserted positions for shorter types (a) and longer types (b) of probes on a seed bag.

(C) A hand sample was taken by the "grab" method before the bag was sealed by randomly choosing another set of 13 bags and grasping a handful of seed from the bag for seed lots A (perennial ryegrass), B (perennial ryegrass, stove-piped) and C (mixture of Kentucky bluegrass, perennial ryegrass and red fescue).

(D) From the sampling procedures above we obtained 204 samples which were further divided into smaller working samples for the following analysis:

- Thirty-two 50-g samples of the Lot A (perennial ryegrass) analyzed for the number of red colored ryegrass contaminants.
- Thirty-two 50-g samples of the Lot B (perennial ryegrass, stove-piped) analyzed for the number of both red and green colored ryegrass contaminants.
- Two sets of samples from seed lot C (mixture of Kentucky bluegrass, perennial ryegrass and red fescue): A) thirty-six 30-g samples analyzed for the number of the red colored seeds of each species; and B) thirty-six 3-g samples analyzed for the percent composition of each of the three components.

- Twenty-eight 500-g samples of wheat from seed lot D analyzed for the number of red colored wheat caryopses.
- Forty 500-g samples of soybean from seed lots E and F analyzed for the number of green colored soybeans.

(E) Samples from seed lot A were analyzed at the Washington State Department of Agriculture Seed Laboratory, samples from seed lot B were analyzed at the Idaho State Department of Agriculture Seed Laboratory. Samples from seed lot C were analyzed by the Oregon State University Seed Laboratory and the remainder were analyzed in the Seed Research Laboratory at Michigan State University.

2. Statistical Analysis:

The number of colored seed found was compared with the expected number for each sample from each lot. The average difference between the number of colored seed found and expected was squared to obtain a common comparable index of accuracy. The variance was calculated from the average of the four replicates as an index of precision. The overall performance of the different sampling probes and hand grab method was evaluated by combining the mean square error (MSE) of accuracy and precision together and the ratios of the MSEs of two probes were compared as a measure of their relative efficiency (RE), i.e., RE (%) = MSE_i/MSE_j, where i, j = probe numbers or hand sampling, and $i \neq j$.

Finally, the Chi-square test was used to evaluate the agreement between test results and expected values.

RESULTS AND DISCUSSION

Perennial ryegrass contaminated with red-colored ryegrass

In general, all probes except probe number 2 provided samples in which the incidence of red-colored seeds was over-represented (Table 15). Probe number 2 was the most efficient, 38% more efficient than probe number 6, the second ranked probe, and 82% more than probe number 7 (Table 16). The rank of the probes in order of decreasing accuracy was 2, 7, 6, hand grabbing, 4, 3, 1 and 5. Probe number 6 was most precise, followed by probe numbers 1, 2, 5, 7, 3, 4 and hand grabbing.

Table 15. Performance of various probes and hand sampling on colored perennial ryegrass seed.

Probe	Expected (Number of seeds)	Found	Bias (%)	Precision	MSE	χ^2
Hand	331	351	6.0	656	1056	11*
1	331	359	8.5	179	963	11.3*
2	330	327	-0.8	275	284	2.6
3	331	356	7.6	431	1056	11.6**
4	330	353	7.0	623	1152	12.3**
5	330	367	11.2	322	1691	19.1**
6	330	349	5.8	38	399	4.6
7	330	340	3.0	419	519	4.8

Expected = sample weight (g) x seeds/g x colored seeds (%);
Found = number of colored seeds in the sample;
Bias % = (Found - Expected) x 100%/Expected;
Precision = Variance within individual laboratories;
MSE = Mean Square Error = Precision + (Found - Expected)^2;
*,**Significantly different from expected colored seeds at 5% and 1% levels, respectively.

Chi-square tests showed that the colored seed detected with probe numbers 2, 6, and 7 was not significantly different from the expected values, while those from other probes and hand grabbing were significantly different at

either the 5% or 1% levels of confidence. Although probe number 1 was relatively precise, it was not very accurate.

Table 16. Relative efficiency (R.E.%) of probes for detecting colored seeds of perennial ryegrass.

Probe	6	7	1	Hand	3	4	5
2	38	82	243	270	275	308	490
6		32	148	168	171	195	326
7			88	103	106	124	224
1				8	9	19	72
Hand					1	10	59
3						9	57
4							45

Note: R.E % of a probe in a row over a probe in a column = (the MSE of the probe in the column - the MSE of the probe in the row) x 100% / the MSE of the probe in the column.

Purity components of Kentucky bluegrass/perennial ryegrass/red fescue mixture

The Chi-square test showed that proportions of the purity components detected with all probes were not significantly different from those expected (Table 17). However, it was important to note that all probes (probes 1-8) resulted in consistent over-estimation of Kentucky bluegrass by 4.47 to 12.68% and perennial ryegrass by 2.74 to 10.66% except probe number 5 (a 305-mm probe) which resulted in an under-estimation for perennial ryegrass of 0.6%. All probes resulted in a consistent under-estimation of red fescue by 6.20 to 18.04%. None of the probes tested provided an unbiased sample for the

Table 17. Performance of probes sampling purity components of Ky bluegrass (K), P. ryegrass (R) and red fescue (F) mixture.

	Expected	Found			Bias			Precision			MSE			Overall		χ^2	
Probe	(%)*	K(%)	R(%)	F(%)	K(%)	R(%)	F(%)	K	R	F	K	R	F	MSE	Precision	Test	Prob.
1	33.0	36.1	34.2	28.6	9.4	3.6	-13.3	4.3	3.5	0.4	13.9	4.9	19.8	38.6	8.2	0.93	0.82594
2	33.0	37.1	33.9	28.0	12.4	2.7	-15.2	0.4	0.8	1.6	17.2	1.6	26.6	45.4	2.8	1.30	0.72913
3	32.9	35.1	34.3	29.3	6.7	4.3	-10.9	3.4	0.7	1.2	8.2	2.7	14.2	25.1	5.3	0.61	0.89411
4	33.0	34.5	36.5	28.0	4.5	10.6	-15.2	4.3	6.7	1.5	6.6	19.0	26.5	52.0	12.5	1.21	0.75061
5	33.0	35.3	32.8	31.0	7.0	-0.6	-6.1	1.7	0.5	0.5	7.0	0.5	4.5	12.0	2.7	0.29	0.96179
6	33.0	36.4	34.3	28.2	10.3	3.9	-14.5	1.4	0.7	0.9	13.0	2.4	23.9	39.3	3.0	1.12	0.77226
7	33.0	37.2	34.8	27.0	12.7	5.5	-18.2	1.0	0.5	1.9	18.6	3.7	37.9	60.3	3.4	1.72	0.63253
8	33.0	36.0	35.4	27.6	9.1	7.3	-16.4	2.5	3.3	1.3	11.5	9.1	30.5	51.0	7.1	1.34	0.71969
Hand	32.8	28.8	32.4	37.2	-12.2	-1.2	13.4	0.8	3.8	4.7	16.8	4.0	24.1	44.8	9.3	1.10	0.77707

Expected (%) = (100- (weed seed % + other crop % + inert matter %))/3;
Found (%) = percentage of colored seeds in the sample;
Bias % = (Found - Expected) × 100%/Expected;
Precision = Variance within individual laboratories;
MSE = Mean Square Error = Precision + (Found - Expected)^2;

mixture of components of different flowability characteristics because the shorter, more free-flowing components were consistently more likely to be sampled than longer, more chaffy seed types.

In contrast to the various probes, the manual grab method consistently provided samples in which longer more chaffy seeds were over-represented compared to shorter, more free-flowing seeds which tended to be under-represented.

Although none of these deviations from expected levels were significant, there was a consistent bias toward certain kinds of seed depending on their flowability characteristics for all sampling methods. Overall, this bias was less pronounced for the 305-mm probe (probe number 5) with the sharpest end and widest opening.

Red colored contaminants in the mixture of Kentucky bluegrass/Perennial ryegrass/red fescue

Probe number 3 provided the best overall results, with 21% more efficiency than probe number 2 and 31% more than probe number 1 (Table 18). All of the probes were more efficient than hand sampling. However, Chi-square tests showed that the number of colored seeds detected was significantly different from the expected values at the 1% level (Table 19, page 74). Since all contaminants were so greatly under-estimated, we believe that both the accuracy and efficiency, especially for perennial ryegrass, were generally reduced by the analyst's inability to detect slightly-stained seed and that the

results of this test did not truly reflect the absolute efficiency of the probes tested. However, the relative efficiency and precision of the various probes (Table 18) should have been unaffected. This also suggests that further research should be done to more definitively test this conclusion.

Table 18. Relative efficiency (R.E.%) of probes for detecting colored seeds from the mixture lot.

Probe	2	1	4	6	5	7	8	Hand
3	21	31	41	94	167	172	193	294
2		8	17	60	120	125	142	225
1			8	49	104	109	124	202
4				37	88	93	107	179
6					37	40	51	103
5						2	10	48
7							8	45
8								35

Note: R.E % of a probe in a row over a probe in a column
= (the MSE of the probe in the column - the MSE of the probe in the row) x 100% / the MSE of the probe in the column.

Wheat with contaminants of red colored wheat

All probes provided samples in which the incidence of red-colored wheat seed was over-represented (Table 20, page 75). Some gave reasonable accuracy but were not very precise (e.g., probe number 3), whereas others (e.g., probe number 9) were precise but lacked accuracy.

The number of colored seeds obtained with probe numbers 1, 10, 3 and 9 was significantly below that expected at both the 5% or 1% levels of confidence, while the results with other probes were not (Table 20). Probe number 2 was

Table 19. Performance of probes for detecting colored seeds of Ky bluegrass (K), P. ryegrass (R) and R. fescue (F) mixture.

Probe	Expected			Found			Bias			Precision			MSE			Overall		χ^2
	K	R	F	K	R	F	K	R	F	K	R	F	K	R	F	Precision	MSE	Test
	(number of seeds)						(%)											
1	386	67	113	306	4	81	-20.7	-94.0	-28.3	314	2	55	6714	3971	1079	11764	371	85**
2	386	67	113	292	3	74	-24.4	-95.5	-34.5	81	10	96	8917	4106	1617	14641	188	97**
3	386	67	113	308	5	80	-20.2	-92.5	-29.2	777	10	256	6861	3854	1345	12061	1044	83**
4	386	67	113	300	4	81	-22.3	-94.0	-28.3	144	2	288	7540	3971	1312	12823	434	87**
5	386	67	113	271	5	77	-29.8	-92.5	-31.9	5457	6	160	18682	3850	1456	23987	5622	104**
6	386	67	113	333	4	69	-13.7	-94.0	-38.9	8745	0	78	11554	3969	2014	17537	8823	84**
7	386	67	113	289	3	66	-25.1	-95.5	-41.6	8645	11	226	18054	4107	2435	24596	8882	105**
8	386	67	113	258	5	71	-33.2	-92.5	-37.2	4512	6	41	20896	3850	1805	26551	4559	115**
Hand	386	67	113	223	4	94	-42.2	-94.0	-16.8	4101	2	642	30670	3971	1003	35644	4745	131**

*Expected = sample weight (g) × seeds/g × colored seeds (%);
Found = number of colored seeds in the sample;
Bias % = (Found - Expected) × 100%/Expected;
Precision = Variance within individual laboratories;
MSE = Mean Square Error = Precision + (Found - Expected)^2;
*,**Significantly different from expected colored seeds at 5% and 1% levels, respectively.

most efficient, with 16% and 28% greater efficiency than probe numbers 4 and 5, respectively (Table 21). Probe number 3 was the least efficient among all probes tested.

Table 20. Performance of various probes for detecting colored wheat seeds.

Probe	Expected (number of seeds)	Found	Bias (%)	MSE	Precision	χ^2
1	137.1	120.9	-11.8	391.8	129.4	10.46*
2	137.1	124.0	-9.6	252.6	81.0	6.79
3	137.1	134.6	-1.8	802.1	795.9	17.60**
4	137.1	123.5	-9.9	293.4	108.4	7.78
5	137.1	130.4	-4.9	324.2	279.3	7.43
9	137.1	114.2	-16.7	570.3	45.9	16.36**
10	137.1	125.2	-8.7	350.1	208.5	8.73*

Expected = sample weight (g) x seeds/g x colored seeds (%);
Found = number of colored seeds in the sample;
Bias % = (Found - Expected) x 100%/Expected;
Precision = Variance within individual laboratories;
MSE = Mean Square Error = Precision + (Found - Expected)^2;
*,**Significantly different from expected colored seeds at 5% and 1% levels, respectively.

Table 21. Relative efficiency (R.E.%) of probes for detecting colored wheat seeds.

Probe	4	5	10	1	9	3
2	16	28	39	55	126	217
4		10	20	33	95	173
5			8	20	76	147
10				11	63	128
1					46	105
9						40

Note: R.E % of a probe in a row over a probe in a column
= (the MSE of the probe in the column - the MSE of the probe in the row) x 100% / the MSE of the probe in the column.

Soybean with contaminants of green-colored soybeans

All probes except probe number 6 were very effective in detecting the level of green-colored soybean seed (Table 22). Only the colored seed detected by using probe number 6 was significantly different from the expected value at the 5% level. Others were not significantly different. Probe number 1 was the best, with 30% more efficiency than probe number 10, followed by probe numbers 5, 2 and 6 (Table 23). Although probe number 6 gave the closest to the expected seed number (Table 22), results among different replications varied greatly, indicating poor repeatability (precision). Among all 5 probes tested, probe number 6 had the smallest opening diameter (Table 14).

Table 22. Performance of various probes for detecting colored soybean seeds.

Probe	Expected (number of seeds)	Found	Bias (%)	MSE	Precision	χ^2
1	36.2	34.0	-6.1	10.7	5.9	1.1
2	36.2	33.9	-6.4	67.4	62.1	7.74
5	36.2	34.7	-4.1	17.8	15.5	1.72
6	36.2	37.3	3.0	108.6	107.4	8.45*
10	36.2	38.6	6.6	13.6	7.8	1.13

Expected = sample weight (g) x seeds/g x colored seeds (%);
Found = number of colored seeds in the sample;
Bias % = (Found - Expected) x 100%/Expected;
Precision = Variance within individual laboratories;
MSE = Mean Square Error = Precision + (Found - Expected)^2;
*Significantly different from expected colored seeds at 5% level.

Table 23. Relative efficiency (R.E.%) of probes for detecting colored soybean seeds.

Probe	10	5	2	6
1	30	69	540	935
10		29	391	693
5			280	514
2				62

Note: R.E % of a probe in a row over a probe in a column
= (the MSE of the probe in the column - the MSE of the probe in the row) x 100% / the MSE of the probe in the column.

Stove-piped soybean and ryegrass seed lots

The fact that variation was very high in both seed lots indicates that no probe gave a representative sample from extremely heterogeneous seed bags. The colored seeds detected with all probes were significantly different from expected values, and were either over- or under-estimated (Tables 24 and 25).

Table 24. Performance of probes on soybean sampling with stove-piped colored seeds.

Probe	Expected (number of seeds)	Found	Bias (%)	MSE	Precision	χ^2
1	80.9	182.4	125.5	10630	328	521.2**
2	80.9	74.5	-7.9	940.9	900	35.4**
5	80.9	100.5	24.2	5462.2	5078	207.1**
6	80.9	120.9	49.4	3764	2164	159.2**
10	80.9	48.4	-40.2	1763.3	707	78.7**

Expected = sample weight (g) x seeds/g x colored seeds (%);
Found = number of colored seeds in the sample;
Bias % = (Found - Expected) x 100%/Expected;
Precision = Variance within individual laboratories;
MSE = Mean Square Error = Precision + (Found - Expected)^2;
*Significantly different from expected colored seeds at 5% level.

Table 25. Performance of various probes and hand sampling on perennial ryegrass with red colored seed and stove-piped green seed.

Probe	Expected (number of seeds)	Found	Bias (%)	MSE	Precision	χ^2
			Red-colored seed			
Hand	334.2	270.3	-19.1	4305	222	50.9**
1	334.5	309.3	-7.5	1241	606	13.2**
2	333.4	410.5	23.1	24119	18175	235.8**
3	334.7	323.3	-3.4	967	837	9.0*
4	335.0	262.5	-21.6	10265	5009	108.2**
5	335.8	318.8	-5.1	2182	1893	20.9**
6	333.6	306.3	-8.2	907	162	10.8*
7	334.8	303.0	-9.5	2652	1641	18.8**
			Stove-piped green seed			
Hand	330.1	3048.5	823.5	8581996	1192297	100347**
1	330.3	114.8	-65.2	46882	442	567**
2	329.2	449.3	36.5	41675	27251	417.9**
3	330.6	659.5	99.5	266094	157919	2751.6**
4	330.8	203.8	-38.4	42320	26191	432.7**
5	331.6	257.0	-22.5	75385	69820	698.3**
6	329.5	342.3	3.9	31444	31280	288**
7	330.7	275.7	-16.6	61923	58876	382.5**

Expected = sample weight (g) x seeds/g x colored seeds (%);
Found = number of colored seeds in the sample;
Bias % = (Found - Expected) x 100%/Expected;
Precision = Variance within individual laboratories;
MSE = Mean Square Error = Precision + (Found - Expected)^2;
*Significantly different from expected colored seeds at 5% level.

CONCLUSIONS

Based on the result of this study, the following conclusions are drawn:

1. The performance of probes with different physical features varied among crops and sampling situations.

2. For perennial ryegrass seed, only probe numbers 2, 6 and 7 performed well. They provided samples from which the number of the stained components was not significantly different than expected. Use of the other probes resulted in samples that varied from expected levels by varying amounts.

3. For the mixture of Kentucky bluegrass, perennial ryegrass and red fescue, all probes as well as the hand grab method, provided representative samples. However, only probe number 5 provided consistently unbiased samples. All other probes provided **consistently biased** samples **although the bias was not statistically significant for any probe**. Smaller and more free-flowing components (Kentucky bluegrass and perennial ryegrass) were more likely to be sampled than longer and narrower, more chaffy seed types (e.g., red fescue). On the other hand, the hand grab method provided samples in which the longer, chaffier seed (red fescue) was consistently over-represented and the shorter, more free-flowing components were consistently under-represented. The best probe for sampling from this mixture was probe number 5 a 305-mm (12-inch) probe with the sharpest end and a wide (8-21 mm) opening. It is likely (though not proven) that this probe would also be best for other seed mixtures containing both shorter, more free-flowing seeds and longer less free-flowing seed.

4. For wheat (middle-sized seed) probe numbers 2, 4 and 5 provided

representative samples. However, probe number 2 was the most efficient.

5. For soybean seed, all probes except probe number 6 were very efficient in providing representative samples. Probe number 6 was the only one with openings less than 10 mm in diameter. This indicates that all probes with large openings (diameter) should provide representative samples, although probe number 1 was the most precise in sampling soybeans. Since soybean has a rather large seed, probes with relatively small openings (less than 15 mm diameter) should be avoided.

6. For the extremely heterogeneous stove-piped soybean and perennial ryegrass seed lots, all probes as well as the hand grab method failed to provide a representative sample. The numbers of colored seeds detected varied greatly above or below expected values, showing the difficulty in obtaining a representative sample from extremely heterogeneous containers with any sampling method.

7. It should be expected that similar results would be obtained in sampling from other species and/or mixtures of species with similar characteristics.

8. Finally, the overall conclusion of this research is that all of the probes tested will provide a statistically representative sample if properly used for appropriate seed types and sampling situations.

SUMMARY

The performance of probes with different physical features varied among crops and sampling situations. Representative samples are unlikely to be obtained by any probe from heterogeneous seed containers. On the other hand, with certain exceptions, most probes and sampling methods should provide a representative seed sample from a homogeneous seed lot if properly used. However, some seed lots, e.g., those containing blends of varieties or mixtures of contaminants with different seed size and flow characteristics, can not be sampled accurately with certain probes. Probes with smaller openings tended to provide samples that under-represented the longer, more chaffy seed types, while over-representing the shorter, more free-flowing components. We believe that the diameter of the opening is the most important feature of a probe that will enable it to provide a representative sample from such lots. Furthermore, all probes should be long enough to reach across the entire width or length of the container. Finally, regardless of the sampling instrument used, it is very important to follow proper sampling procedures.

REFERENCES

1. Bean, J. E. (1970). Seed trier efficiency trial. Proc. of ISTA 35(3): 673-681.

2. Debney, E. W. (1960). Dynamic seed sampling spears in the United Kingdom. Proc. of ISTA 25: 174-181.

3. Grisez, J. P. and E. E. Hardin. (1972). A comparison of four methods of sampling small seeds. Proc. of ISTA 37(3): 661-667.

4. Leggatt, C. W. (1938). A new sampler for sampling seed in the sack. Proc. of AOSA 1938: 192-195.

5. Mullin, J. F. (1965). Investigation into some common methods of lot sampling. Proc. of ISTA 30(2): 207-213.

6. Munn, M. T. (1935). Observations upon the movement of seeds in bags when sampled with instruments. Proc. of ISTA 7: 15-18.

7. Nees, L. W. (1990). Seed inspector's qualification and training committee report. AASCO, 1990.

8. ISTA. (1938). International Rules for Seed Testing. Proc. of ISTA 10: 408-410.

APPENDIX A

FIGURES FOR GERMINATION, VIGOR AND SEED COUNT TESTS

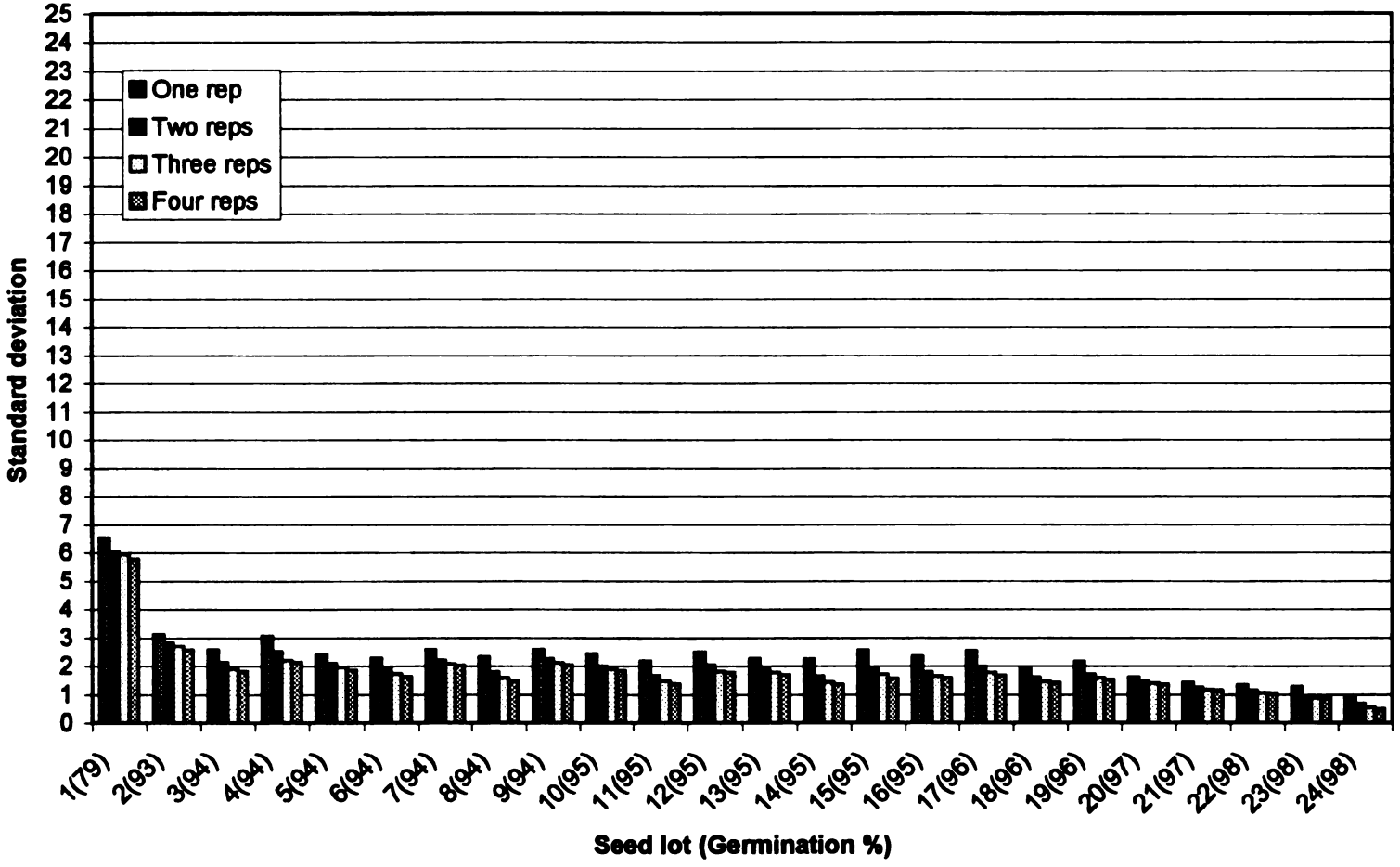

Figure A-1. Relationship between number of replicates and standard deviation for 100-seed conventional germination referee (CGR) tests on corn in 1994-96 [outliers (7.8% of the original data) were eliminated].

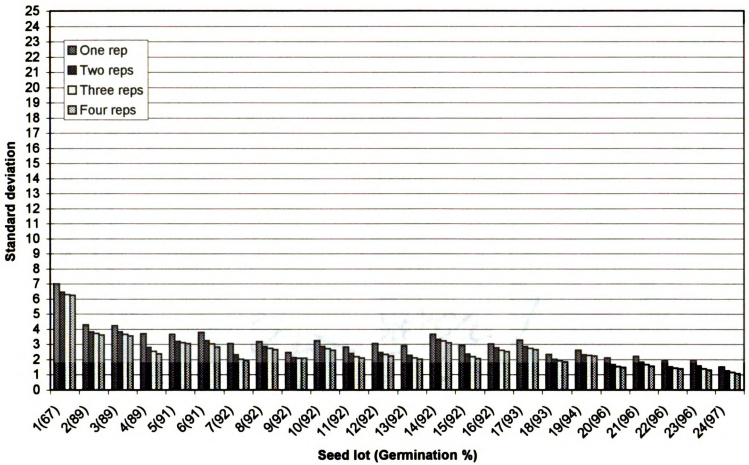

Figure A-2. Relationship between number of replicates and standard deviation for 100-seed conventional germination referee (CGR) tests on soybean in 1994-96 [outliers (12.6% of the original data) were eliminated].

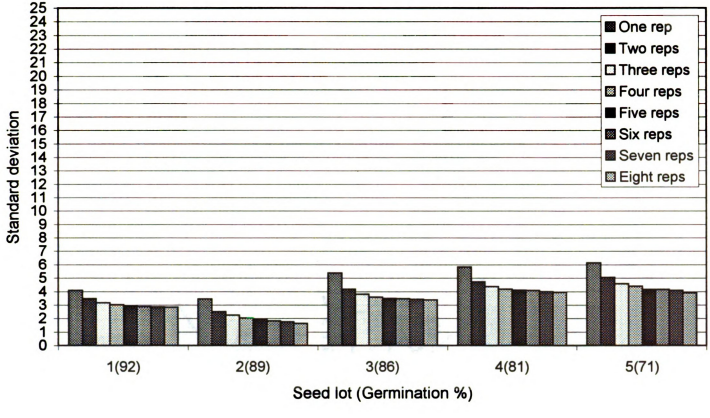

Figure A-3. Relationship between number of replicates and standard deviation for 100-seed blind germination referee (BGR) tests on corn in 1997.

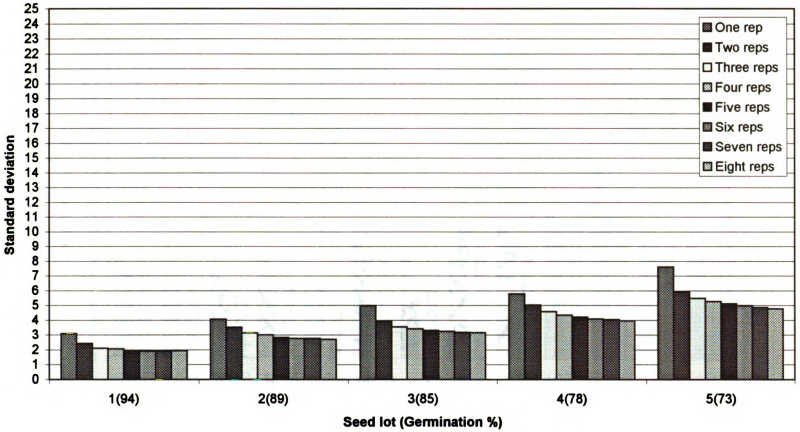

Figure A-4. Relationship between number of replicates and standard deviation for 100-seed blind germination referee (BGR) tests on soybean in 1997.

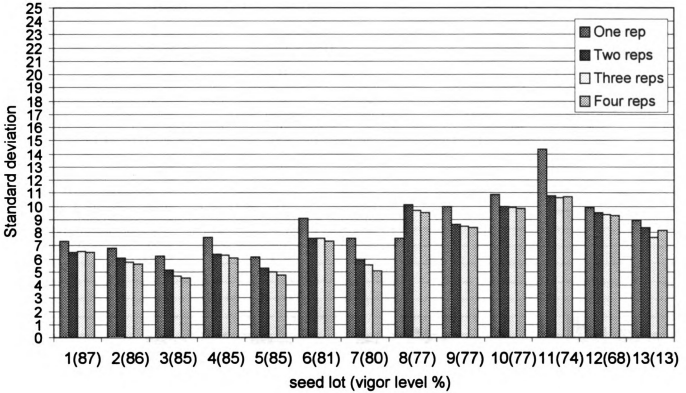

Figure A-5. Relationship between number of replicates and standard deviation for 50-seed cold tests on soybean in 1993-95 [outliers (35.6% of the original data) were eliminated].

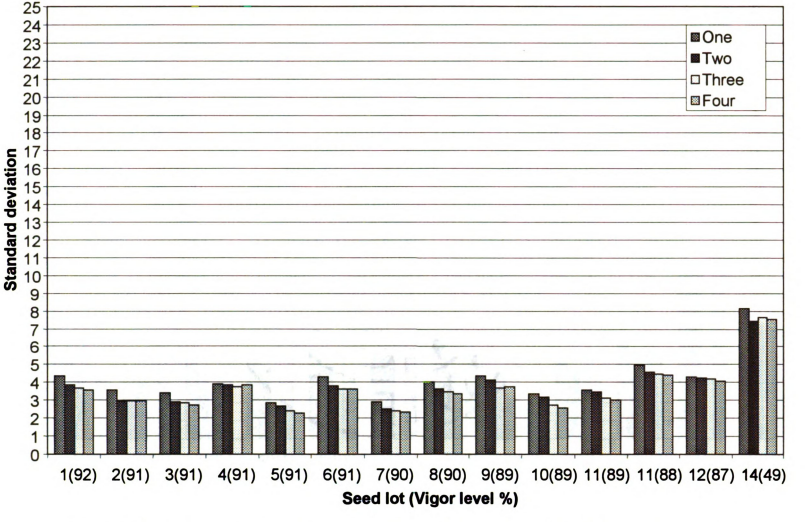

Figure A-6. Relationship between number of replicates and standard deviation for 100-seed cold tests on corn in 1993-95 [outliers (23.8% of the original data) were eliminated].

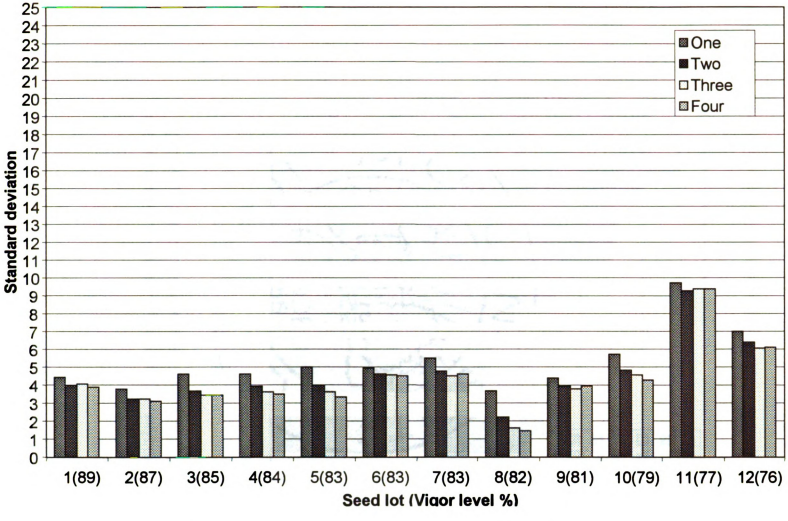

Figure A-7. Relationship between number of replicates and standard deviation for 100-seed cold tests on soybean in 1993-95 [outliers (51.2% of the original data) were eliminated].

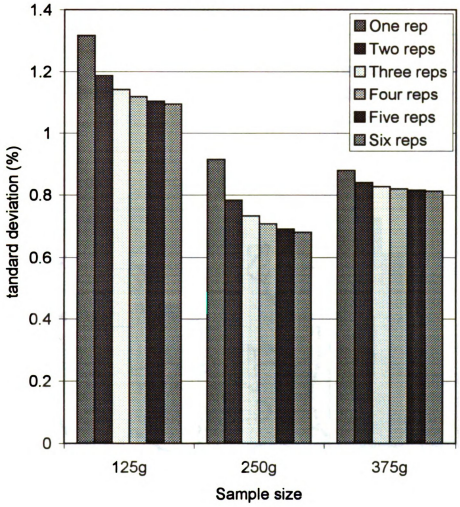

Figure A-8. Standard deviation for electronic counts as a function of sample sizes and number of replicates on soybean in 1995-96.

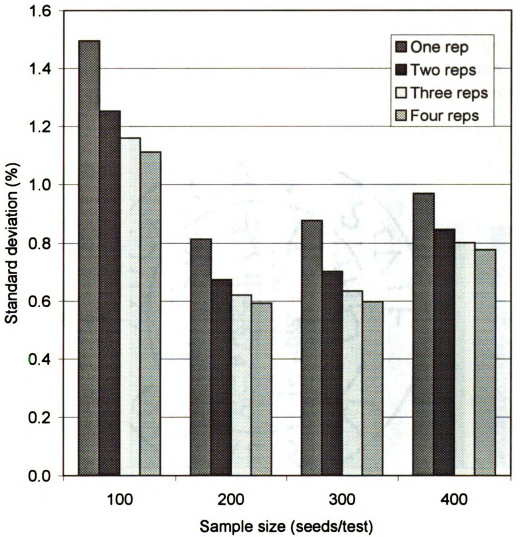

Figure A-9. Standard deviation for manual counts as a function of sample sizes and number of replicates for soybean tested in 1996-97.

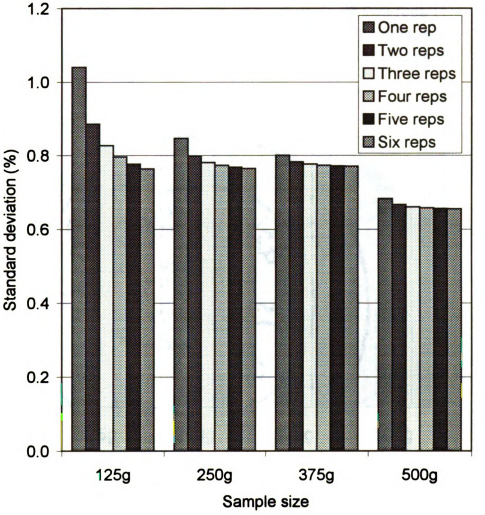

Figure A-10. Standard deviation for electronic counts as a function of sample sizes and number of replicates for soybean tested in 1996-97.

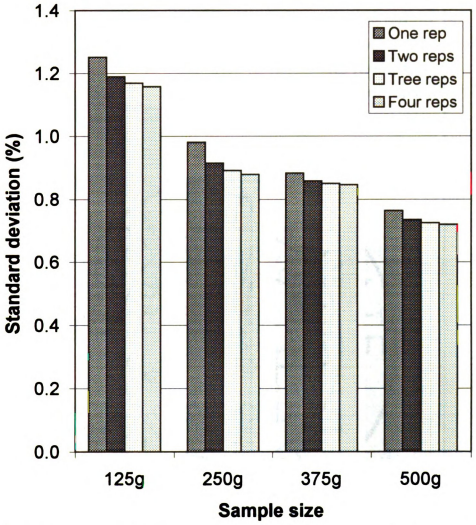

Figure A-11. Standard deviation for electronic counts as a function of sample sizes and number of replicates for seed lot 1 (6240 seeds/kg) of soybean tested in 1998.

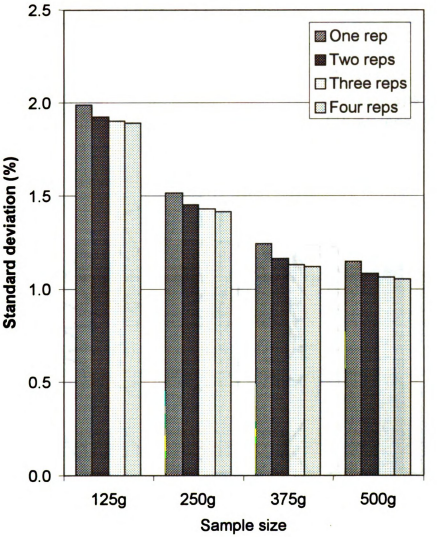

Figure A-12. Standard deviation for electronic counts as a function of sample sizes and number of replicates for seed lot 2 (5199 seeds/kg) of soybean tested in 1998.

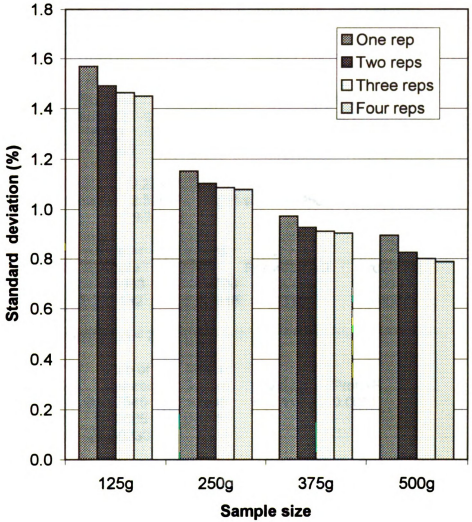

Figure A-13. Standard deviation for electronic counts as a function of sample sizes and number of replicates for seed lot 3 (7610 seeds/kg) of soybean tested in 1998.

APPENDIX B STATISTICAL ANALYSIS TABLES

The relationship between germination (or vigor) levels and standard deviation

Model: model I
Dependent variable = Standard deviation

Table A-1. Regression analysis for corn CGR tests.

(1) Number of replication = 1

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	23.55322	23.55322	438.114	0.0001
Error	22	1.18273	0.05376		
C Total	23	24.73595			

Root MSE	0.23186	R-square	0.9522
Dep Mean	2.37607	Adj R-sq	0.9500
C.V.	9.75826		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	29.433683	1.29356096	22.754	0.0001
MEAN	1	-0.286132	0.01367016	-20.931	0.0001

(2) Number of replication = 2

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	21.57545	21.57545	512.119	0.0001
Error	22	0.92685	0.04213		
C Total	23	22.50230			

Root MSE	0.20526	R-square	0.9588
Dep Mean	1.97456	Adj R-sq	0.9569
C.V.	10.39498		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	27.868511	1.14499499	24.339	0.0001
MEAN	1	-0.273831	0.01210034	-22.630	0.0001

(3) Number of replications = 3

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	21.20799	21.20799	459.967	0.0001
Error	22	1.01437	0.04611		
C Total	23	22.22236			

Root MSE	0.21473	R-square	0.9544	
Dep Mean	1.82175	Adj R-sq	0.9523	
C.V.	11.78684			

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	27.417933	1.19427507	22.958	0.0001
MEAN	1	-0.270675	0.01262073	-21.447	0.0001

(4) Number of replications = 4

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	20.14405	20.14405	391.786	0.0001
Error	22	1.13115	0.05142		
C Total	23	21.27520			

Root MSE	0.22675	R-square	0.9468
Dep Mean	1.74278	Adj R-sq	0.9444
C.V.	13.01087		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	26.660040	1.25970624	21.164	0.0001
MEAN	1	-0.263493	0.01331206	-19.794	0.0001

Table A-2. Regression analysis for soybean CGR tests.

(1) Number of replications = 1

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	24.29208	24.29208	151.851	0.0001
Error	22	3.51941	0.15997		
C Total	23	27.81149			

Root MSE	0.39997	R-square	0.8735
Dep Mean	3.12150	Adj R-sq	0.8677
C.V.	12.81327		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	19.758398	1.35255851	14.608	0.0001
MEAN	1	-0.181765	0.01475032	-12.323	0.0001

(2) Number of replications = 2

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	22.06183	22.06183	135.390	0.0001
Error	22	3.58490	0.16295		
C Total	23	25.64673			

Root MSE	0.40367	R-square	0.8602	
Dep Mean	2.67461	Adj R-sq	0.8539	
C.V.	15.09269			

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	18.553511	1.36715228	13.571	0.0001
MEAN	1	-0.173487	0.01490987	-11.636	0.0001

(3) Number of replications = 3

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	21.80926	21.80926	127.321	0.0001
Error	22	3.76847	0.17129		
C Total	23	25.57773			

Root MSE	0.41388	R-square	0.8527
Dep Mean	2.52712	Adj R-sq	0.8460
C.V.	16.37744		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	18.321119	1.40227284	13.065	0.0001
MEAN	1	-0.172572	0.01529398	-11.284	0.0001

(4) Number of replications = 4

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	21.86639	21.86639	133.564	0.0001
Error	22	3.60172	0.16371		
C Total	23	25.46811			

Root MSE	0.40462	R-square	0.8586
Dep Mean	2.43801	Adj R-sq	0.8522
C.V.	16.59619		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	18.275546	1.37287152	13.312	0.0001
MEAN	1	-0.173035	0.01497231	-11.557	0.0001

Table A-3. Regression analysis for corn BGR tests.

(1) Number of replicates = 1

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	3.58116	3.58116	5.890	0.0936
Error	3	1.82404	0.60801		
C Total	4	5.40520			

Root MSE	0.77975	R-square	0.6625
Dep Mean	4.97803	Adj R-sq	0.5501
C.V.	15.66384		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	14.318206	3.86433312	3.705	0.0342
MEAN	1	-0.111360	0.04588513	-2.427	0.0936

(2) Number of replicates = 2

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2.63021	2.63021	5.224	0.1064
Error	3	1.51058	0.50353		
C Total	4	4.14079			

Root MSE	0.70960	R-square	0.6352
Dep Mean	3.98775	Adj R-sq	0.5136
C.V.	17.79437		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	12.106574	3.56643590	3.395	0.0426
MEAN	1	-0.096811	0.04235847	-2.286	0.1064

(2) Number of replicates = 3

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2.21012	2.21012	4.739	0.1177
Error	3	1.39902	0.46634		
C Total	4	3.60914			

Root MSE	0.68289	R-square	0.6124
Dep Mean	3.63045	Adj R-sq	0.4832
C.V.	18.81011		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	11.006631	3.40198713	3.235	0.0480
MEAN	1	-0.088010	0.04042720	-2.177	0.1177

(4)Number of replicates = 4

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2.14372	2.14372	4.444	0.1256
Error	3	1.44710	0.48237		
C Total	4	3.59082			

Root MSE	0.69453	R-square	0.5970
Dep Mean	3.44191	Adj R-sq	0.4627
C.V.	20.17850		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	10.693882	3.45400937	3.096	0.0535
MEAN	1	-0.086554	0.04105740	-2.108	0.1256

(5) Number of replicates = 5

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	1.87104	1.87104	3.574	0.1551
Error	3	1.57049	0.52350		
C Total	4	3.44153			

Root MSE	0.72353	R-square	0.5437
Dep Mean	3.32410	Adj R-sq	0.3916
C.V.	21.76621		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	10.093916	3.59548686	2.807	0.0674
MEAN	1	-0.080794	0.04273600	-1.891	0.1551

(6) Number of replicates = 6

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2.03458	2.03458	3.598	0.1541
Error	3	1.69625	0.56542		
C Total	4	3.73083			

Root MSE	0.75194	R-square	0.5453
Dep Mean	3.27472	Adj R-sq	0.3938
C.V.	22.96206		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	10.350957	3.74547492	2.764	0.0699
MEAN	1	-0.084437	0.04451232	-1.897	0.1541

(7) Number of replicates = 7

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	1.95608	1.95608	3.349	0.1647
Error	3	1.75220	0.58407		
C Total	4	3.70828			

Root MSE	0.76424	R-square	0.5275
Dep Mean	3.22047	Adj R-sq	0.3700
C.V.	23.73078		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	10.156553	3.80549887	2.669	0.0758
MEAN	1	-0.082762	0.04522414	-1.830	0.1647

(8) Number of replicates = 8

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	1.67501	1.67501	2.560	0.2079
Error	3	1.96285	0.65428		
C Total	4	3.63786			

Root MSE	0.80888	R-square	0.4604
Dep Mean	3.13672	Adj R-sq	0.2806
C.V.	25.78742		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	9.569640	4.03676430	2.371	0.0984
MEAN	1	-0.076749	0.04796728	-1.600	0.2079

Table A-4. Regression analysis for soybean BGR tests.

(1) Number of replicates = 1

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	11.53389	11.53389	137.021	0.0013
Error	3	0.25253	0.08418		
C Total	4	11.78642			

Root MSE	0.29013	R-square	0.9786
Dep Mean	5.10477	Adj R-sq	0.9714
C.V.	5.68353		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	21.904900	1.44107673	15.200	0.0006
MEAN	1	-0.200228	0.01710538	-11.706	0.0013

(2) Number of replicates = 2

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	7.29431	7.29431	354.535	0.0003
Error	3	0.06172	0.02057		
C Total	4	7.35604			

Root MSE	0.14344	R-square	0.9916
Dep Mean	4.16870	Adj R-sq	0.9888
C.V.	3.44082		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	17.804472	0.72702103	24.490	0.0001
MEAN	1	-0.162574	0.00863420	-18.829	0.0003

(3) Number of replicates = 3

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	6.72516	6.72516	384.960	0.0003
Error	3	0.05241	0.01747		
C Total	4	6.77757			

Root MSE	0.13217	R-square	0.9923
Dep Mean	3.77308	Adj R-sq	0.9897
C.V.	3.50306		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	16.825381	0.66786205	25.193	0.0001
MEAN	1	-0.155646	0.00793288	-19.620	0.0003

(4) Number of replicates = 4

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	6.01871	6.01871	374.221	0.0003
Error	3	0.04825	0.01608		
C Total	4	6.06696			

Root MSE	0.12682	R-square	0.9920
Dep Mean	3.61822	Adj R-sq	0.9894
C.V.	3.50504		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	15.982001	0.64163826	24.908	0.0001
MEAN	1	-0.147413	0.00762028	-19.345	0.0003

(5) Number of replicates = 5

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	5.91554	5.91554	662.957	0.0001
Error	3	0.02677	0.00892		
C Total	4	5.94231			

Root MSE	0.09446	R-square	0.9955
Dep Mean	3.48279	Adj R-sq	0.9940
C.V.	2.71223		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	15.748692	0.47825294	32.930	0.0001
MEAN	1	-0.146190	0.00567775	-25.748	0.0001

(6) Number of replicates = 6

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	5.56865	5.56865	536.417	0.0002
Error	3	0.03114	0.01038		
C Total	4	5.59979			

Root MSE	0.10189	R-square	0.9944
Dep Mean	3.40285	Adj R-sq	0.9926
C.V.	2.99420		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	15.330835	0.51702211	29.652	0.0001
MEAN	1	-0.142152	0.00613763	-23.161	0.0002

(7) Number of replicates = 7

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	5.15431	5.15431	528.848	0.0002
Error	3	0.02924	0.00975		
C Total	4	5.18355			

Root MSE	0.09872	R-square	0.9944
Dep Mean	3.35319	Adj R-sq	0.9925
C.V.	2.94416		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	14.875355	0.50297678	29.575	0.0001
MEAN	1	-0.137310	0.00597087	-22.997	0.0002

(8) Number of replicates = 8

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	4.81536	4.81536	753.145	0.0001
Error	3	0.01918	0.00639		
C Total	4	4.83454			

Root MSE	0.07996	R-square	0.9960
Dep Mean	3.31041	Adj R-sq	0.9947
C.V.	2.41543		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	14.465820	0.40805661	35.451	0.0001
MEAN	1	-0.132953	0.00484460	-27.443	0.0001

Table A-5. Regression analysis for corn vigor tests with 50 seeds per replicate.

(1) Number of replicates = 1

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	69.01359	69.01359	125.276	0.0001
Error	7	3.85625	0.55089		
C Total	8	72.86984			

Root MSE	0.74222	R-square	0.9471
Dep Mean	5.18249	Adj R-sq	0.9395
C.V.	14.32172		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	29.110613	2.15210624	13.527	0.0001
MEAN	1	-0.271870	0.02429001	-11.193	0.0001

(2) Number of replicates = 2

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	73.79827	73.79827	192.801	0.0001
Error	7	2.67938	0.38277		
C Total	8	76.47765			

Root MSE	0.61868	R-square	0.9650
Dep Mean	4.28103	Adj R-sq	0.9600
C.V.	14.45172		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	28.386096	1.74822027	16.237	0.0001
MEAN	1	-0.274000	0.01973310	-13.885	0.0001

(3) Number of replicates = 3

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	71.20864	71.20864	188.097	0.0001
Error	7	2.65001	0.37857		
C Total	8	73.85865			

Root MSE	0.61528	R-square	0.9641
Dep Mean	3.75437	Adj R-sq	0.9590
C.V.	16.38847		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	27.675578	1.75619879	15.759	0.0001
MEAN	1	-0.271684	0.01980948	-13.715	0.0001

(4) Number of replicates = 4

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	70.30120	70.30120	152.355	0.0001
Error	7	3.23001	0.46143		
C Total	8	73.53122			

Root MSE	0.67929	R-square	0.9561
Dep Mean	3.46292	Adj R-sq	0.9498
C.V.	19.61602		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	27.383161	1.95111028	14.035	0.0001
MEAN	1	-0.271494	0.02199539	-12.343	0.0001

Table A-6. Regression analysis for corn vigor tests with 100 seeds per replicate.

(1) Number of replicates = 1

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	17.87761	17.87761	53.233	0.0001
Error	12	4.03004	0.33584		
C Total	13	21.90765			

Root MSE	0.57951	R-square	0.8160
Dep Mean	4.14165	Adj R-sq	0.8007
C.V.	13.99236		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	13.465483	1.28727235	10.460	0.0001
MEAN	1	-0.107215	0.01469485	-7.296	0.0001

(2) Number of replicates = 2

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	14.89389	14.89389	44.165	0.0001
Error	12	4.04677	0.33723		
C Total	13	18.94066			

Root MSE	0.58072	R-square	0.7863
Dep Mean	3.79510	Adj R-sq	0.7685
C.V.	15.30175		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	12.273223	1.28513879	9.550	0.0001
MEAN	1	-0.097476	0.01466750	-6.646	0.0001

(3) Number of replicates = 3

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	18.16749	18.16749	49.029	0.0001
Error	12	4.44660	0.37055		
C Total	13	22.61409			

Root MSE	0.60873	R-square	0.8034
Dep Mean	3.64113	Adj R-sq	0.7870
C.V.	16.71809		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	12.957105	1.34037601	9.667	0.0001
MEAN	1	-0.107094	0.01529472	-7.002	0.0001

(4) Number of replicates = 4

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	17.44496	17.44496	41.670	0.0001
Error	12	5.02377	0.41865		
C Total	13	22.46873			

Root MSE	0.64703	R-square	0.7764
Dep Mean	3.57527	Adj R-sq	0.7578
C.V.	18.09735		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	12.768073	1.43454946	8.900	0.0001
MEAN	1	-0.105663	0.01636860	-6.455	0.0001

Table A-7. Regression analysis for soybean vigor tests with 50 seeds per replicate.

(1) Number of replicates = 1

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	4.10031	4.10031	0.778	0.3965
Error	11	57.95594	5.26872		
C Total	12	62.05625			

Root MSE	2.29537	R-square	0.0661
Dep Mean	8.63806	Adj R-sq	-0.0188
C.V.	26.57275		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	10.894987	2.63638249	4.133	0.0017
MEAN	1	-0.030122	0.03414554	-0.882	0.3965

(2) Number of replicates = 2

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	5.36635	5.36635	1.436	0.2560
Error	11	41.10408	3.73673		
C Total	12	46.47043			

Root MSE	1.93306	R-square	0.1155	
Dep Mean	7.72139	Adj R-sq	0.0351	
C.V.	25.03517			

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	10.313233	2.22825376	4.628	0.0007
MEAN	1	-0.034560	0.02883868	-1.198	0.2560

(3) Number of replicates = 3

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2.98246	2.98246	0.719	0.4147
Error	11	45.65890	4.15081		
C Total	12	48.64136			

Root MSE	2.03735	R-square	0.0613
Dep Mean	7.45758	Adj R-sq	-0.0240
C.V.	27.31924		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	9.379082	2.33620272	4.015	0.0020
MEAN	1	-0.025591	0.03018994	-0.848	0.4147

(4) Number of replicates = 4

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	5.72736	5.72736	1.336	0.2722
Error	11	47.16021	4.28729		
C Total	12	52.88757			

Root MSE	2.07058	R-square	0.1083
Dep Mean	7.36010	Adj R-sq	0.0272
C.V.	28.13245		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	10.037556	2.38664227	4.206	0.0015
MEAN	1	-0.035619	0.03081700	-1.156	0.2722

Table A-8. Regression analysis for soybean vigor tests with 100 seeds per replicate.

(1) Number of replicates = 1

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	17.40599	17.40599	13.307	0.0045
Error	10	13.08045	1.30804		
C Total	11	30.48644			

Root MSE	1.14370	R-square	0.5709
Dep Mean	5.28506	Adj R-sq	0.5280
C.V.	21.64019		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	32.850054	7.56369316	4.343	0.0015
MEAN	1	-0.334168	0.09160667	-3.648	0.0045

(2) Number of replicates = 2

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	16.85334	16.85334	9.226	0.0125
Error	10	18.26751	1.82675		
C Total	11	35.12085			

Root MSE	1.35157	R-square	0.4799
Dep Mean	4.56606	Adj R-sq	0.4279
C.V.	29.60042		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	31.235613	8.78902686	3.554	0.0052
MEAN	1	-0.323516	0.10651046	-3.037	0.0125

(3) Number of replicates = 3

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	16.18786	16.18786	6.968	0.0247
Error	10	23.23094	2.32309		
C Total	11	39.41879			

Root MSE	1.52417	R-square	0.4107
Dep Mean	4.36607	Adj R-sq	0.3517
C.V.	34.90946		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	30.783044	10.01708030	3.073	0.0118
MEAN	1	-0.320411	0.12137984	-2.640	0.0247

(4) Number of replicates = 4

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	17.21996	17.21996	7.156	0.0233
Error	10	24.06415	2.40641		
C Total	11	41.28411			

Root MSE	1.55126	R-square	0.4171
Dep Mean	4.29162	Adj R-sq	0.3588
C.V.	36.14633		

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	31.337402	10.12032127	3.096	0.0113
MEAN	1	-0.328178	0.12268125	-2.675	0.0233

APPENDIX C
SAS CODE FOR MAJOR STATISTICAL PROCEDURES

1. To generate data from original data sheet provided by AOSA/SCST.

```
data new;
 set raw ;
  array a{4} abn1-abn4;
   array d{4} ded1-ded4;
    do rep = 1 to reps;
     abnorm = a{rep};
      dead  = d{rep};

    /* Use "0" to replace "." values. */

      if a{rep}=. then a{rep}=0;
     if d{rep}=. then d{rep}=0;
    germten = 100-a{rep}-d{rep};
   output;
  end;
 run;

 data selected;
  set new;
   if germten < 100;
    run;

proc sort data=selected;
 by lotid labid;
  run;

data selec;
 set selected;
  by lotid labid;
   if first.labid then nn=1;
    else nn+1;
     if last.labid and (nn=4) then output;
      run;
```

2. Use Miles (1963) method to eliminate outliers which are four standard deviations away from the seed lot mean

```
/* Following procedures are created for statistical analysis of both germination and vigor test
results */

proc sort data=import1;
 by lotid labid ;
  run;
```

```
proc means data=import1;
  var germten;
    by lotid labid ;
      output out=one mean=germ;
       run;

proc univariate data=one;
   var germ;
     output out=med mean=mean;
       by lotid  ;
         run;

data selec;
   merge one med;
     d=germ-mean;
       sigma=sqrt(mean*(100-mean)/400);
        /* for 50-seed four replicate test, 400 is replaced with 200 */
         ratio=d/sigma;
          by lotid;
           run;

proc sort data=selec;
   by lotid;
     run;

data three;
   set selec;
      if ratio= <4 and ratio= >-4;
        by lotid ; run;

data outliers;
   set selec;
    if ratio > 4 or ratio <-4;
      by lotid;
       title 'Outliers from 100-seed corn vigor tests in 1993-96';run;

data two ;
  merge three import1;
   if ratio ne . then output;
     by lotid  labid;run;

proc sort data=two;
  by lotid ;run;
```

3. Generate random data to construct subset data of different sizes of replicate

```
   /* To generate standard deviation for each subset data */
/*-------------For more than one replicate-------------*/

%macro samplesi(num,r);
```

```
data simu;
 set two;
  ran=ranuni(&num);
   by lotid ;
    run;

proc sort data=simu;
 by lotid labid ran;run;

data simu;
 set simu;
  by lotid  labid ;
   if first.labid then cnt=1;
  else cnt+1;
   run;

%macro new1(repl,co);
 data subset;
  set simu(where=(cnt < &repl));
   by lotid ;
    run;

proc means data=subset;
 var germten;
  by lotid  labid ;
   output out=one mean=germ;
    run;

proc univariate data=one;
 var germ;
  output out=med&repl&co mean=mean std=std;
   by lotid ;
    run;

data med&repl&co;
 set med&repl&co;
  n=&repl-1;
   t=&co;
    run;

   %mend new1;

%new1(5,&r)
%new1(4,&r)
%new1(3,&r)
  data medd&r;
    set med5&r med4&r med3&r;
     run;

        %mend samplesi;
```

```
      %samplesi(1234,01) %samplesi(2234,02) %samplesi(3234,03) %samplesi(4234,04)
      %samplesi(5234,05) %samplesi(6234,06) %samplesi(7234,07) %samplesi(8234,08)
      %samplesi(9234,09) %samplesi(2345,10) %samplesi(1234,11) %samplesi(223,12)
      %samplesi(323,13) %samplesi(423,14) %samplesi(523,15) %samplesi(623,16)
      %samplesi(723,17) %samplesi(823,18) %samplesi(923,19) %samplesi(2340,20)
      %samplesi(14,21) %samplesi(24,22) %samplesi(34,23) %samplesi(44,24)
      %samplesi(54,25) %samplesi(64,26) %samplesi(74,27) %samplesi(84,28)
      %samplesi(94,29)
      %samplesi(104,30)

data final;
 set medd01 medd02 medd03 medd04 medd05 medd06 medd07 medd08 medd09 medd10
     medd11 medd12 medd13 medd14 medd25 medd16 medd17 medd18 medd19 medd20
     medd21 medd22 medd23 medd24 medd25 medd26 medd27 medd28 medd29 medd30;
      run;

/* ---------- for one rep ----------*/

%macro onerep(num,r);
    data simu;
 set two;
  ran=ranuni(&num);
   by lotid;
    run;

proc sort data=simu;
 by lotid labid ran;
  run;

data simu;
 set simu;
  by lotid labid;
   if first.labid then cnt=1;
    else cnt+1;
     run;

data subset;
 set simu(where=(cnt <2));
  by lotid;
   run;

proc means data=subset;
 var germten;
  by lotid;
   output out=one&r mean=germ std=std;
    run;

%mend onerep;
   %onerep(12,1) %onerep(13,2) %onerep(14,3) %onerep(15,4) %onerep(16,5)
   %onerep(17,6) %onerep(18,7) %onerep(19,8) %onerep(20,9) %onerep(234,10)
```

```
   %onerep(235,11)  %onerep(236,12)  %onerep(237,13)  %onerep(238,14)
   %onerep(239,15)  %onerep(232,16)  %onerep(231,17)  %onerep(233,18)
   %onerep(240,19)  %onerep(241,20)  %onerep(512,21)  %onerep(513,22)
   %onerep(514,23)  %onerep(5134,24)  %onerep(5121,25)  %onerep(5123,26)
   %onerep(2234,27)  %onerep(5424,28)  %onerep(5554,29)  %onerep(5234,30)
data final2;
        set one1 one2 one3 one4 one5 one6 one7 one8 one9 one10 one11 one12 one13 one14
        one15 one16 one17 one18 one19 one20 one21 one22 one23 one24 one25 one26 one27
        one28 one29 one30;
      run;
data out;
 set final final2;
  run;

    /*========== for the averages of STD and MEANS ==========*/

data neww;
 set import;
  proc sort;
   by lotid n;
    run;

proc means data=neww;
 var mean;
  by lotid n;
   output out=mean mean=mean;
    run;

proc means data=neww;
 var std;
  by lotid n;
   output out=stand mean=stand;
    run;

data newww;
 merge  mean stand;
  by lotid n;
   proc print; run;


data newww;
 set newww;
  proc sort;
   by n;
    proc print; run;

proc reg;
  model stand=mean;
  by n;
   title1 'The relationship between vigor levels and standard deviation';
```

```
title2 '50-seed (30 samples) vigor test on soybean without outliers in 1994-96';
 run;
```

4. To calculate the variation among and within laboratories of seed counts. (The part to construct subset data of different replicate sizes was similar to that of germination and vigor tests, thus not given below)

```
data new;
 set import;
  proc sort data=new;
   by lotid size labid ;
    run;

proc univariate plot data=new ;
 var counts;
  by  lotid size;
   output out=one mean=means std=std;
    title 'Seed count data from 1998 referee without mc adj.';
     run;

proc mixed data=new;
 class rep labid;
  model counts=;
   random labid;
    make out=v1 covparms;
     by lotid size;
      run;

 data four;
  set one v1;
   proc print data=four;
    title 'Test variation from soybean seed count referee in 1998';
     run;
```

5. Calculation of correlation coefficient among replications within individual laboratories

```
data raw;
 set soy50;
  gt=(g1+g2+g3+g4)/400;run;

data soye50;
 set raw;
  vb=50*(gt)*(1-gt); /* For 100-seed test replace 50 with 100*/
   vo=((g1-g2)**2+(g1-g3)**2+(g1-g4)**2+
    (g2-g3)**2+(g2-g4)**2+(g3-g4)**2)/6;
     rho=1-(vo/(2*vb));
      run;

data firsquar;
 set soye50 (where=(gt<0.25 or gt=0.25 ));
  if  rho < (-gt)/(1-gt)
   then rho=(-gt)/(1-gt);  run;

data secoquar;
 set soye50 (where=(gt<0.5 and gt>0.25 or gt=0.5));
  if  rho < (5/6-(4/3)*gt-(1-gt)**2)/(gt*(1-gt))
   then rho=(5/6-(4/3)*gt-(1-gt)**2)/(gt*(1-gt)); run;

data thirquar;
 set soye50 (where=(gt<0.75 and gt>0.5 or gt=0.75));
  if  rho <((3/4)*gt-0.5-gt**2)/gt*(1-gt)
   then rho=((3/4)*gt-0.5-gt**2)/gt*(1-gt); run;

data fortquar;
 set soye50 (where=(gt>0.75 or gt=1));
  if  rho <-(1-gt)/gt
   then rho=-(1-gt)/gt; run;

data soyee50;
 set firsquar secoquar  thirquar fortquar;run;

/* Prepare a graph of the correlation coefficients against labid
   separately for each seedlot */

proc plot data=soyee50;
 plot rho*labid/vref=0;
  run;

/* Test the hypothesis that the average coefficients are not
    significantly different from 0. */

proc univariate data=soyee50;
 var rho;
  output out=mean mean=mean n=n signrank=signrank;
```

```
  title 'Soybean vigor 50-seed test ';
   title2 'Use the lower bound equations by Dr. Gilliland';
    run;

/*By simulation of random binomial sampling, if z>=1.65, accept H1,otherwise accept Ho */

 data simul;
  set mean;
  z=(mean-0)/(sqrt((2(4*100)**2)/((4*100-1)**2)(4-1))/(sqrt(n)));
   if z>=1.65 then sign='*';
    else sign='n';
     proc print;
       run;

/* Test that coefficients are as likely to be positive
   as they are to be negative*/

data si;
 set soyee50;
  if rho>0 then rhh='p';else rhh='n';
   proc freq data=si;
    tables rhh/testp=(0.5,0.5);
     run;
```