SAMPLE SIZE[1]

W. T. Federer

BU-59-M

I. Introduction

The problem of determining sample size or replicate number arises whenever an experiment or a survey is to be conducted. There are many criteria for determining sample size. The suitability of a given criterion depends upon the objectives and the nature of the experiment or the investigation. The validity of a criterion has to be determined by the experimenter or by the administrator. Although some statisticians believe that it is their prerogative to determine criteria for experimenters, I do not. Hence, the various criteria will be set forth and will be discussed individually, and it will be up to the experimenter to determine which criterion fits his situation.

The ensuing discussion relates almost entirely to experiments with very little discussion of sample surveys. The results are, however, extendible to the latter type of investigation. Also, some sections are primarily concerned with determining the required number of replicates while others consider optimum allocation of samples and of replicates, and of other combinations. In general, the number of replicates is determined by the following inter-related factors:

(i) the degree of precision desired,
(ii) the amount of variability present in the experimental material,
(iii) available resources including personnel and equipment, and
(iv) size and shape of experimental unit.

The nature of the experimental material, the characters observed, and/or the expected magnitude of the treatment differences determine the degree of precision desired. The very nature of some treatments may be such that large differences are expected or the treatments may be such that low variability is expected. Some characters are more variable than others thus requiring more replication. In connection with determining the number of replicates for an experiment, the characters of interest with their respective standard deviations should be listed. Replicate number should be determined for the most variable character to be measured. This number is then sufficiently large for

---

[1]Lecture presented at PRI on 3/3/55.

all other characters. The relative importance of the characters will also affect the determination of replicate number for the experiment in that replicate number is determined for the most important character.

The amount of variability present in experimental material is determined by the experimental conditions, the characters measured, and the treatments tested. The variance of a treatment mean relative to the treatment mean or relative to the difference between treatments may be small for some characters and large for others.

Larger experimental units tend to have smaller variation than smaller ones. Technique and weighing errors do not usually have as great an effect on large plots as on small plots. An error in weight of 100 pounds on a small plot may have a relatively large effect whereas an error in weight of 100 pounds on a large plot may be of no consequence. In fact, some of the scales used for weighing the sugar cane from experimental plots are only accurate to the nearest 200 pounds.

The shape of the experimental unit usually has a relatively small effect on the variance of a treatment mean. However, long narrow field plots tend to be less variable than square ones.

II. Criteria for determining replicate number - fixed sample size.

The following list of criteria for determining replicate number is not all inclusive but it does illustrate the diverse considerations used by experimenters in determining replicate number. For these criteria it is assumed that the experiment will be conducted in a fixed period of time, - this is called a one-stage experiment as opposed to multi-stage or sequential experimentation where the results of the first stage determine whether additional stages are to be conducted.

II-1 Crystal ball procedure.

One of the most used criteria for determining replicate number is to select a number without any other considerations. The experimenter may use 7 replicates because he likes the number 7, because Joe Blow used 7 replicates, or because 7 is the number that he has been using in the past. This procedure will be denoted as the crystal ball procedure. It is not a subjective criterion for determining replicate number.

As a passing comment, the experimenter may ask the statistician to tell him how many replicates to use in a forthcoming experiment. The statistician does not question the experimenter further but merely tells him a number. The number may have no basis because the statistician is already late for lunch, or the number may be determined by the knowledge that the experimenter will not put in more than a specified number in any event. For example, the statistician might know that the experimenter only uses one replicate so he tries to get the experimenter to put in two replicates which may be entirely inadequate to detect anything but gross differences even though two replicates are twice as good as one.

Despite the reasoning, the excuses, and the prejudices there appears to be no justification for using the crystal ball procedure.

II-2 Available resources.

Some of the experimental resources that might limit the number of replicates for a given experiment are:

(i) Amount of seed or experimental material

(ii) Available personnel

(iii) Amount of experimental land, pots, equipment, available for this experiment.

The amount of experimental seed or treatment material may be limited in the early stages of an investigation. This is quite often the situation with new varieties, new chemicals, etc., and the experimenter must decide whether to run the experiment or to wait until more experimental material is available. Likewise, limited personnel or other experimental resources may give insufficient replication. The procedure usually followed here is to conduct the experiment with the available number of replicates. Then, based on the results of the first experiment, the experimenter determines whether further testing is required. This "sequential" method of testing has long been used by experimenters. The mathematical formalization of this procedure has recently been set forth by the late Prof. A. Wald in his book entitled, Sequential Sampling, and by others. One particular method for determining replication number for an experiment conducted in a sequence of two stages has been described by Stein (1945).

If the experiment is to be conducted only with the number of replicates dictated by available resources with no further testing, serious consideration

should be given to the idea of not running the experiment at all but to put the limited resources on other experiments.

II-3 Number of degrees of freedom in the error mean square.

A criterion that has some usefulness, especially for experiments with few treatments, is to use sufficient replication so that the error mean square is associated with at least 12 to 16, preferably 20, degrees of freedom. The reasoning here is that the variance of the error mean square will not be so large, relatively, for 20 or more degrees of freedom. The variance of an error mean square, E, is 2 (error m.s.)^2/(degrees of freedom) = $2E^2/df$. From the following table we see that relatively small decreases in the variance of an error mean square are made with greater than 20 degrees of freedom:

Degrees of freedom = df	2/df
2	1.00
4	.50
8	.25
10	.20
16	.125
20	.10
50	.04
100	.02

The error degrees of freedom may be increased by increasing the number of replicates or the number of treatments. Thus, with 20 or more treatments 2 replicates would be sufficient to satisfy this criterion.

II-4 Standard error equal to a specified percentage of the mean.

Some experimenters use sufficient replication to obtain a standard error of a mean which is not over a specified percentage of the mean. For this criterion it is required that a relatively good estimate of the coefficient of variation be available. Given this estimate a replicate number is selected which yields the desired percentage. For example, suppose that the coefficient of variation is 10 per cent and that it is desired to have a standard error of a mean which is less than 2.5 per cent of the mean. Thus $10/\sqrt{16} = 10/4 = 2.5$, or 16 replicates would be required to obtain a standard error which is 2.5% of the treatment mean.

The requirements for this criterion are:

(i) the estimated standard deviation, s, and

(ii) the estimated mean, $\bar{x}$.

With these two statistics the coefficient of variation is computed as $s/\bar{x}$ = c.v.

II-5 Error of accepting the wrong hypothesis set equal to specified percentage.

For this criterion we need the following information:

(i) estimated difference among treatments, d, that is important to detect,

(ii) estimated error variance for treatments s^2,

(iii) size of type I error (usually 5%),

(iv) size of type II error, and

(v) the possible hypotheses specified.

The estimated error variance should be fairly close to the true error variance. If no error variance is readily available Harris, Horvitz, and Mood (1948) have a procedure for obtaining an estimate of the error variance and the associated degrees of freedom. Use may be made of this procedure for some of the methods described below.

II-5.1 "Usual Method"

The "usual method" for determining numbers of replicates is to use the following formula:

$$r = 2(s\, t_{05}/d)^2,$$

when t_{05} is the tabulated value for Student's t at the 5 per cent level. For example, suppose that it is desired to detect a difference of 5 units and that the estimated error standard deviation is 5 units. Since we assume that the coefficient of variation is relatively well known, we use t = 2. (The late R. J. Borden has tabulated the number of replicates for the various numbers of degrees of freedom and the corresponding t values.) Then $4 = 2[5(2)/5]^2 = 8$ replicates. Also, we could use estimates of these quantities in per cent of the mean. Suppose that the coefficient of variation is 10% and the smallest difference that it is desired to detect is 10%, then $r = 2[.10(2)/.10] = 8$. In both examples, a difference of one standard deviation unit was considered to be the smallest difference that was of practical importance. This is a fairly large difference.

The type I error in the above example is 5% and the type II error approximately 50%. This means that with 8 replicates we would detect a difference as small as one standard deviation unit in only 50% of the cases. For most types of experimental work we would like to detect the specified difference d in a higher proportion of the cases.

II-5.2 Methods for increasing the confidence that the computed interval will be less than or equal to d.

II-5.2.1 Harris-Horvitz-Mood

In order to raise the proportion of cases in which we would be able to detect a value of d or larger, Harris, Horvitz, and Mood have multiplied the formula in section II-5.1 by an F value; thus,

$$r = 2(s\, t_{05,\ df2}/d)\ F_{1-c}\ (df_2,\ df_1),$$

when df_1 represents the degrees of freedom associated with the estimated error variance, df_2 represents the degrees of freedom for error in the proposed experiment, c = confidence with which it is desired to detect the prescribed difference, d, between treatments. The value of r obtained is of sufficient size to give a confidence interval less than or equal to 2d with an assurance of c per cent. For example, suppose that the completely randomized design is to be used, that c = 90%, that v = 6 treatments are to be tested, that d = 20, that $s = \sqrt{141.6}$ with 40 degrees of freedom, and that the first estimate of r = 6 and consequently $df_2 = 30$; then,

$$r = 2(141.6)(2^2)\ [F_{1-90}\ (30,40) = 1.54]/20^2 = 4.4$$

Since df_2 was over-estimated try r = 5 and $df_2 = 6(5-1) = 24$. Then

$$r = 2(141.6)(2.06)^2(1.61)/400 = 5.$$

With 5 replicates we would have a 90% assurance that the lsd would be less than or equal to d.

Using a somewhat related technique Harris, Horvitz, and Mood (1948) prepared tables for determining sample size. After specifying d, c, the size of the type I error (usually .05), and the estimated error variance, the number of replicates may be computed with the aid of their tables.

II-5.2.2 Tukey

The preceding methods make use of Student's t in computing sample size. Since the t distribution refers to experiments involving only 2 treatments, it is necessary to consider the range of more than two sample means. The statistic, q, for this has been tabulated by Pearson and Hartley (1954). These values of the various values for ranges, q, have been used by Tukey (1953) to determine sample size. Thus,

$$r = q^2_{05,\ df_2}\ s^2 F_{1-c}(df_2 df_1)/d^2\ .$$

The value $q^2_{05,\ df_2}$ replaces the value $2t^2_{05,\ df_2}$ in the preceding section. To illustrate the method, consider the example of section II-5.2.1. Trying $r = 6$ and $df_2 = 30$ we obtain

$$r = (4.30)^2(141.6)(1.54)/400 = 10$$

Since the value for df_2 was underestimated try $r = 9$ and $df_2 = 6(9-1) = 48$. Then

$$r = (4.2)^2(141.6)(1.48)/400 = 9.2.$$

Since r is greater than 9 we next try r = 10 and find that 10 replicates are sufficiently large for our purposes. With 10 replicates we would have the assurance that the 95% confidence half interval, which = $q_{05,\ df_2}$ x $\sqrt{\text{error variance}/r = 10 \text{ replicates}}$, will be less than or equal to d = 20 in 90% of all such experiments.

II-5.2.3 Tang

Tang (1938) has prepared tables for determining sample size. For Tang's method it is required that the sum of squares among the v treatments be specified. This means that the deviations of the treatment means from the overall mean must be prescribed. Since this is rather difficult to do without first running the experiment the method does not appear to be readily usable by the experimenter.

II-5.2.4 Cochran-Cox

Cochran and Cox (1950) present a readily usable table for determining replicate number (table 2.1). It is required that the following information is available:

(i) the estimated coefficient of variation,

(ii) the value of d in % of the mean that it is important to detect,

(iii) the size of the type I error, and

(iv) the size of the type II error.

If it is desired to have an lsd less than or equal to d (see section II-5.2.1), the reader is referred to section 2.22 of their book. However, to illustrate the use of their table (table 2.1) suppose that we do not know beforehand which treatment will be better than others and that we use s = 4%, 8%, and 12%, and d = 5%, 10%, and 15% for various values of c = 50%, 80%, 90%. The following numbers of replicates were obtained from Cochran and Cox's table 2.1 for c = 80% and 90% while the values for r for c = 50% were computed from the formula in section II-5.1 using t = 2. The type I error = 5%.

		d = 5%	d = 10%	d = 15%
c.v. = 4%	c = 50%	5	2	1
	c = 80%	11	4	3
	c = 90%	15	5	3
c.v. = 8%	c = 50%	21	5	3
	c = 80%	41	11	6
	c = 90%	x	15	7
c.v. = 12%	c = 50%	46	12	5
	c = 80%	x	24	11
	c = 90%	x	31	15

x = greater than 50.

The above table illustrates the importance of having a small coefficient of variation in detecting relatively small differences. To detect a difference of 15% of the mean in experiments with a c.v. = 8% with a confidence of 80 per cent (i.e., in 4 out of 5 experiments) requires at least 6 replicates. However, to detect a difference of 5% in the same type of experiment requires 41 replicates.

II-5.3 Minimax procedure.

II-5.3.1 Selection of hypothetical genetic ratios.

Prasert Na Nagara (1953) has prepared tables giving the sample size necessary to select the genetic ratio describing the sample data from among a number of genetic ratios for various error rates. The procedure is quite simple. First, decide on the allowable error of selecting the wrong ratio. Then, from the tables determine sample size large enough to have an error of selecting the wrong hypotheses less than the allowable error. After the sample results are available they are compared with the division points in Na Nagara's tables to determine which hypotheses fit the data. For example, suppose that we specify that our sample data fit one of the 3 test cross ratios -- 1:1, 3:1, or 7:1 and suppose that we wish to make our error of selecting the wrong hypotheses equal to 5 per cent or less. From the tables we find that we need a sample size of at least 110 individuals. Suppose that we observe 110 or more individuals, and if the ratio of the two phenotypes is less than 63% we select the 1:1 ratio as being the correct ratio, if the ratio falls between 63% and 82% we select the 3:1 genetic ratio, and if the ratio is greater than 82% we select the 7:1 genetic ratio.

II-5.3.? Selection of correct hypothesis equal to specified per cent.

Bechhofer (1954), Sobel (1954), Dunnett, Paulson (1949, 1952), and Somerville (1953) have all contributed to the problem of selecting a sample size large enough so that the error of selecting the wrong hypothesis is not over a fixed percentage. For the method described by Bechhofer (1954) the following information is required to determine the necessary sample size:

(i) true error variance
(ii) difference, d, which it is important to detect,
(iii) the error rate for selecting the wrong hypothesis.

Tables for determining samples have been prepared for the case where the best mean, the best 2 means, the best 3 means, etc. are selected from samples of sizes 2 to 15. A table for correctly ranking three means with given probabilities has also been prepared.

Somerville (1953) considers the financial losses incurred from selecting the incorrect hypothesis. In addition to the above information it is assumed that the financial losses associated with selecting the wrong hypothesis are also known.

The limitations of these methods at present are that the true error variance and the loss considerations are unknown for many situations. However, these methods may prove to be quite useful for a number of experimental situations.

III. Criteria for allocating relative number of varieties and replicates, of replicates, locations, and/or years, of samples and replicates, and of replicates and fields -- fixed sample size.

Several criteria are possible for allocating various experimental resources. The two criteria considered here are

(i) maximum genetic advance and
(ii) minimum variance for fixed cost.

Many others are possible but these two will illustrate the procedure for allocating experimental material.

III-1 Maximum genetic advance.

III-1.1 Replicate and variety numbers.

The plant breeder must choose between using more replicates and fewer strains and using fewer replicates and more strains since the number of experimental plots is usually limited by one or more factors. Faced with this

dilemma, he might wish to select a procedure that will allow him to make maximum genetic progress during his lifetime or during the lifetime of the project. In other words, he would use a procedure that will allow a reasonably good chance for selecting varieties which are higher yielding than the presently used varieties.

If the plant breeder selects one variety (or k varieties) at random from a very large population he makes zero progress, and if he selects the highest yielding variety (or k varieties) from tests of v varieties each in r replicates, then his average progress compared to the former procedure (the random selection of a variety) may be determined from the formula

$$\frac{s_v^2 \quad \bar{x}_v}{\sqrt{s_v^2+s_e^2/r}} = G,$$

where s_v^2 is a measure of the heritable genetic variance, s_e^2 is a measure of the environmental variance, and $\bar{x}_v$ is obtained from a table of means of ranges (Table XX, Fisher and Yates, 1938). As v increases so does $\bar{x}_v$ and as r increases the fraction s_e^2/r decreases. Also, the value of $\bar{x}_v$ depends upon the number of varieties selected. Given that a fixed number of plots are to be used there is a value for r which gives the highest value for G.

To illustrate suppose that the ratio $s_v^2/s_e^2 = 1.1440$, that the 2 highest yielding varieties are always going to be selected in every experiment, and that 200 plots are to be grown for each test. Then the values of G for various numbers of r and v are (from Federer, 1951, table 9):

No. of varieties	No. of replicates	G x s_e
25	8	1.77
50	4	1.99
100	2	2.08
200	1	2.02

Thus, of the 4 schemes 100 entries in each of 2 replicates would yield the maximum genetic advance.

III-1.2 Replicate, location, year, and variety numbers.

The necessary formulae for determining the optimum number of replicates, locations, years and varieties is given by Sprague and Federer (1951). The results are illustrated with examples. In addition, these authors present

formulae for determining the optimum number of replicates, locations, and years for a fixed number of varieties and a fixed budget. The criterion used here is maximum genetic advance per dollar spent.

III-2 Minimum variance for a fixed cost.

The solution for number of replicates and fields is given below since this is one problem that pertains directly to the research program of the Pineapple Research Institute.

Given that there are r replicates in each of several fields on the plantation. The following analysis of variance holds:

Source of variation	d.f.	Expectation of mean square
Fields	f-1	$s_e^2 + ts_{rf}^2 + rs_{ft}^2 + rts_f^2$
Replicates within fields	f(r-1)	$s_e^2 + ts_{rf}^2$
Treatments	t-1	$s_e^2 + rs_{ft}^2 + rfs_t^2$
Treatments x fields	(f-1)(t-1)	$s_e^2 + rs_{ft}^2$
Treatments x replicates within fields	f(t-1)(r-1)	s_e^2
Total	2ft-1	

The error variance for a given treatment mean is

$$\frac{s_e^2}{rf} + \frac{s_{ft}^2}{f} = V(\bar{x}). \qquad (1)$$

Assume that it costs C_r units to plant t plots (all plots in one replicate) and that it costs C_f units to travel from field to field on a per field basis. The total cost for planting r replicates in each of f fields is

$$rfC_r + fC_f = C. \qquad (2)$$

Assuming that cost is fixed and that we wish to minimize equation (1) for a given amount of expenditures we proceed as follows: The function to be minimized is

$$\theta = \frac{s_e^2}{rf} + \frac{s_{ft}^2}{f} - \lambda(C - rfC_r - fC_f) \qquad (3)$$

The partial derivatives of (3) with respect to r,f, and λ are:

$$\frac{d\theta}{dr} = \frac{-s_e^2}{fr^2} + \lambda f C_r \tag{4}$$

$$\frac{d\theta}{df} = \frac{-1}{f^2}\left(\frac{s_e^2}{r} + s_{ft}^2\right) + \lambda(rC_r + C_f) \tag{5}$$

$$\frac{d\theta}{d\lambda} = -C + rfC_r + fC_f \tag{6}$$

The above 3 equations are set equal to zero, and we solve for r and f. From equations (4) and (5) we obtain 2 equations for λ: thus,

$$\lambda = \frac{s_e^2}{r^2 f^2 C_r} \tag{7}$$

$$\lambda = \frac{1}{f^2(rC_r + C_f)}\left(\frac{s_e^2}{r} + s_{ft}^2\right) \tag{8}$$

Subtracting (7) from (8) we obtain an equation in terms of r and f. Thus,

$$\frac{1}{(rC_r + C_f)}\left(\frac{s_e^2}{r} + s_{ft}^2\right) - \frac{s_e^2}{r^2 C_r} = 0 \tag{9}$$

or

$$r^2 C_r\left(\frac{s_e^2}{r} + s_{ft}^2\right) - s_e^2(rC_r + C_f) = 0 \tag{10}$$

Solving for r we obtain

$$r = \sqrt{C_f s_e^2 / C_r s_{ft}^2} \tag{11}$$

Then, from equations (2) or (6), the value for $f = \frac{C}{rC_r + C_f}$ (12)

Substitution of experimental values in equations (11) and (12) results in optimum values for r and f subject to the criterion set forth above.

If the cost function is defined by equation (2) then we could use the criterion of "minimum cost for a fixed (or specified) variance". Also, the solution for optimum number of samples from a plot may be obtained from

formula (11) if the cost function is of the form given by formula (2). Formula (11) is a familiar equation to sample surveyors.

Yates and Zacopanay (1935) have presented methods to determine the optimum percentage of a plot to be harvested for a specified cost function. Theirs is a graphical procedure which is treated by the above method in Chapter III of Experimental Design by Federer (1955).

IV. Optimum plot size and shape.

Smith (1938) has determined optimum plot size subject to the criterion that the amount of information per unit of cost be a minimum. The cost equation,

$$rxC_s + rC_p = \text{total cost} = C_t,$$

where C_s is the cost of sampling one unit and C_p is the cost of sampling the plot independent of the number of units sampled, is of the same form given in the previous section [formula (2)]. The amount of information is the reciprocal of the variance. The value for optimum plot size is

$$x = bC_p/(1-b)C_s,$$

where b is the linear regression coefficient of log of plot sizes on log of variances of plot sizes.

Use of the criterion "maximum amount of information per unit of cost" leads to the solution obtained in the previous section for sample number per plot. This is another illustration of the fact that several criteria may lead to the same solution for sample size.

Taylor (1951) has presented a rather thorough discussion on plot shape. He also presents a good review of the literature on this subject. In general, long narrow plots tend to be more efficient than square plots (Federer, 1955, Ch. III).

V. Sequential sampling and sequential experimentation.

The foregoing discussions relate to the determination of sample size for a one stage experiment or a one-stage sample. I believe that we shall be doing more and more two-stage, three-stage, etc. sample or replicate number determination in the future, perhaps by 1960. I do not mean to imply that experimenters have not been doing multi-stage experimentation both with sample or replicate numbers and with treatments, but rather that they have not followed

some formalized statistical procedure for this. The reason for not using such a method is obvious; there are few or no methods available to them. Papers on this important subject are just now beginning to attack some of the simpler problems. The more complex problems haven't been touched. Perhaps the mathematical statistician will soon have procedures for sequential experimentation.

The mathematical formalization of sequential sampling was started by Prof. A. Wald who describes a number of the essential principles in his book entitled Sequential Analysis. One technique that does appear to hold considerable promise for the experimenter is the two-stage experimentation scheme described by Stein (1945). A simple illustration of Stein's method is given by Cochran in his Sampling Techniques book. Also, the work of Yates, Grundy and Healy (Biometrics, Sept. 1954) may prove to be of considerable value in sequential experimentation.

Comment

This manuscript is a rough sketch of the various criteria used to determine sample size or replicate number. It is realized that a thorough, polished discussion of the various criteria would be invaluable to experimenters. The present paper must be considered as a first, incomplete attempt.

VI. Literature cited.

Bechhofer, R. E.
A single-sample multiple decision procedure for ranking means of normal populations with known variances.
Ann. Math. Stat. 25:16-39, 1954.

Bechhofer, R. E. and Sobel, M.
A single-sample multiple decision procedure for ranking variances of normal populations.
Ann. Math. Stat. 25:273-289, 1954.

Cochran, W. G.
Size of experiments.
USDA lecture series, mimeo, 1950.

Cochran, W. G.
Sampling techniques.
Wiley, N. Y., 1953.

Cochran, W. G. and Cox, G. M.
Experimental designs.
Wiley, N. Y., 1950.

Federer, W. T.
Experimental design -- theory and application.
Macmillan, N. Y., 1955.

Federer, W. T.
Evaluation of variance components from a group of experiments with multiple classifications.
Iowa Agr. Exp. Sta. Res. Bul. 380, 1951.

Fisher, R. A. and Yates, F.
Statistical tables for biological, agricultural, and medical research.
Oliver and Boyd, London, 1938.

Harris, M., Horvitz, D. G., and Mood, A. M.
On the determination of sample sizes in designing experiments.
J. Am. Stat. Assoc. 43:391-402, 1948.

Na Nagara, Prasert
Testing Mendelian ratios.
M.S. thesis, Cornell Univ. Library, June, 1953.

Paulson, E.
A multiple decision procedure for certain problems in the analysis of variance.
Ann. Math. Stat. 20:95-98, 1949.

Paulson, E.
On the comparison of several experimental categories with a control.
Ann. Math. Stat. 23:239-246, 1952.

Paulson, E.
An optimum solution to the k-sample slippage problem for the normal distribution.
Ann. Math. Stat. 23:610-616, 1952.

Pearson, E. S. and Hartley, H. O.
Statistical tables.
Biometrika office, 1954.

Smith, H. F.
An empirical law describing heterogeneity in yields of agricultural crops.
J. Agr. Sci. 28:1-23, 1938.

Snedecor, G. W.
Statistical methods.
Iowa State College Press, Ames, 1946.

Somerville, P. N.
Ph.D. thesis, North Carolina Univ. Library, Chapel Hill, 1953.

Sprague, G. F. and Federer, W. T.
A comparison of variance components in corn yield trials. II. Error, year x variety, location x variety, and variety components.
Agron. J. 43:535-541, 1951.

Stein, C.
A two sample test for a linear hypothesis whose power is independent of the variance.
Ann. Math. Stat. 16:253-258, 1945.

Tang, P. C.
The power function of the analysis of variance tests with tables and illustrations of their use.
Stat. Res. Mem. 2:126-157, 1938.

Tukey, J. W.
The problem of multiple comparisons.
Ditto, Princeton Univ. 396 pp., 1953.

Yates, F.
Modern experimental design and its function in plant selection.
Empire J. Exp. Agr. 8:223-230, 1940.

Yates, F. and Zacopanay, I.
The estimation of efficiency of sampling, with special reference to sampling for yield in cereal experiments.
J. Agr. Sci. 25:545-577, 1935.