Statistical Science
1990, Vol. 5, No. 4, 465–480

On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9.

Jerzy Splawa-Neyman

Translated and edited by D. M. Dabrowska and T. P. Speed from the Polish original, which appeared in Roczniki Nauk Rolniczych Tom X (1923) 1–51 (Annals of Agricultural Sciences)

Abstract. In the portion of the paper translated here, Neyman introduces a model for the analysis of field experiments conducted for the purpose of comparing a number of crop varieties, which makes use of a double-indexed array of unknown potential yields, one index corresponding to varieties and the other to plots. The yield corresponding to only one variety will be observed on any given plot, but through an urn model embodying sampling without replacement from this doubly indexed array, Neyman obtains a formula for the variance of the difference between the averages of the observed yields of two varieties. This variance involves the variance over all plots of the potential yields and the correlation coefficient r between the potential yields of the two varieties on the same plot. Since it is impossible to estimate r directly, Neyman advises taking $r = 1$, observing that in practice this may lead to using too large an estimated standard deviation, when comparing two variety means.

Key words and phrases: Field experiment, varieties, unknown potential yields, urn model, sampling without replacement, correlation.

[Numbers in brackets correspond to page numbers in the original text.]

I will now discuss the design of a field experiment involving plots. I should emphasize that this is a task for an agricultural person however, because mathematics operates only with general designs. In designing this experiment, let us consider a field divided into m equal plots and let

$$U_1, U_2, \cdots, U_m$$

be the true yields of a particular variety on each of these plots. If all the members U_i are equal, each of them may be called the average yield of the field. Otherwise the average yield may be thought of as the arithmetic mean

$$a = \frac{\sum_{i=1}^{m} U_i}{m}.$$

D. M. Dabrowska is Assistant Professor, Division of Biostatistics, School of Public Health, University of California, Los Angeles, California 90024-1722. T. P. Speed is Professor and Chair, Department of Statistics, University of California, Berkeley, California 94720.

The yield from the ith plot measured with high accuracy will be considered an estimate of the number U_i.

If we could repeat the measurement of the yield on the same fixed plot under the same conditions, we could use the above definition of the true yield. [See the Introductory Remarks for a few comments on Neyman's notion of true yield.] However, since we can only repeat the measurement of a particular observed yield, and this measurement can be made with high accuracy, we have to suppose that the observed yield is essentially equal to U_i, whereas differences that occur among yields from various plots should be attributed to differences in soil conditions, especially considering that low and high yields are often clustered in a systematic manner across the field.

To compare ν varieties, we will consider that many sequences of numbers, each of them having two indices (one corresponding to the variety and one corresponding to the plot):

$$U_{i1}, U_{i2}, \cdots, U_{im} \quad (i = 1, 2, \cdots, \nu).$$

Let us take ν urns, as many as the number of varieties to be compared, so that each variety is associated with exactly one urn.

FIG. 1. *Jerzy Neyman in Poland, not long after* 1923.

In the ith urn, let us put m balls (as many balls as plots of the field), with labels indicating the unknown potential yield of the ith variety on the respective plot, along with the label of the plot. Thus on each ball we have one of the expressions

$$U_{i1}, U_{i2}, \cdots, U_{ik}, \cdots, U_{im} \tag{13}$$

[29]

where i denotes the number of the urn (variety) and k denotes the plot number, while U_{ik} is the yield of the ith variety on the kth plot.

The number

$$a_i = \frac{\sum_{k=1}^{m} U_{ik}}{m}$$

is the average of the numbers (13) and is the best estimate of the yield from the ith variety on the field.

Further suppose that our urns have the property that if one ball is taken from one of them, then balls having the same (plot) label disappear from all the other urns.

We will use this scheme many times below and will call it the scheme with ν urns.

If we dealt with an experiment with one variety, we would have a scheme with one urn. In this case, expressions denoting yields will not have a variety index.

The goal of a field experiment which consists of the comparison of ν varieties will be regarded as equivalent to the problem of comparing the numbers

$$a_1, a_2, \cdots, a_\nu$$

or their estimates by way of drawing several balls from an urn.

The simplest way of obtaining an estimate of the number a_i would be by drawing κ balls from the ith urn in such a way that after noting the expressions on the balls drawn, they would be returned to the urn. In this way we would obtain κ independent outcomes of an experiment, and their average X_i would, based on the law of large numbers, be an estimate of the mathematical expectation of the result of our trial. Let x denote a possible outcome of the experiment consisting of drawing one ball from the ith urn. [In modern terminology, lower case x, with or without subscripts, denotes a random variable, and upper case X the corresponding realized values.] We shall calculate $\mathbb{E}x$. Since the probability of drawing a ball from the ith urn is the same for all balls, and equal to $1/m$, and since all possible results of the trial are

[30]

contained in the sequence (13), so of course

$$\mathbb{E}x = \frac{1}{m} \sum_{k=1}^{m} U_{ik} = a_i$$

and the average of the results of the κ trials would be an estimate of a_i.

Unfortunately in practice, returning the balls to the urns cannot be carried out. We are obliged to sample without replacement.

Let $x_1, \cdots, x_\kappa$; $X_1, X_2, \cdots, X_\kappa$ be the possible and true outcomes, respectively, of κ trials carried out in this way. Let us assume, as is often the case in practice, that the sequence (13) contains numbers that do not differ greatly from one another and so may be considered equal. We can group the sequence in such a way that, in the first group, we put all the smallest numbers V_{i1}, there being mp_1 such numbers, in the second class the next smallest of the remaining numbers, whose common value is V_{i2} and whose number is mp_2, etc.

In this way we replace sequence (13) by

$$V_{i1}, V_{i2}, \cdots, V_{in} \tag{14}$$

representing possible outcomes of the trial where the probability that the outcome of the first trial is V_{ik} is p_k.

Let us assume that on the first ball drawn we have the number V_{ik}. What is the probability of the outcome of the next trial?

First of all, the urn contains one fewer balls. Further, the number of elements in the kth class of (14) is reduced by one. Therefore the probability p_r^1 that the outcome of the second trial is equal to V_{ir}, where $r \neq k$ turns out to be

$$p_r^1 = \frac{mp_r}{m-1} = p_r + \frac{p_r}{m-1}$$

whereas the probability of the result V_{ik} in the same trial

[31]

is

$$p_r^1 = \frac{mp_k - 1}{m-1} = p_k - \frac{1-p_k}{m-1}.$$

In the end, after $\kappa - 1$ trials being carried out in the same way, we will find the probability $p_{k,0}^{\kappa-1}$ that the outcome of the κth trial is V_{ik}, where V_{ik} has not been drawn so far, is equal to

$$p_{k,0}^{\kappa-1} = \frac{mp_k}{m-\kappa+1} = p_k + \frac{(\kappa-1)p_k}{m-\kappa+1}$$

and the probability $p_{sl}^{\kappa-1}$ that a number V_{is} which has been drawn l times previously, is equal to

$$p_{sl}^{\kappa-1} = \frac{mp_s - l}{m-\kappa+l} = p_s + \frac{(\kappa-1)p_s - l}{m-\kappa+1}.$$

We see that knowledge of the outcomes of preceding trials has an effect on the probability of outcomes of

subsequent trials, *so that trials conducted in this way are not independent.* If we assume that the number m is very large in relation to κ, so that $\kappa/(m - \kappa)$ is negligible in comparison with the probabilities p_s, then it follows from the above formulas that information about previous trials will not affect probabilities of subsequent trials. Therefore the trials will turn out to be independent, and we will be able to apply the law of large numbers, and our definition of a true yield, and along with it known formulas from probability theory. If each of the ν varieties is sown on κ plots, then $m = \nu\kappa$ and the condition for the independence of experiments will be that the ratio $1/(\nu - 1)$ is small, in other words, the number of varieties ν to be compared is large.

Should we draw from this the conclusion that in the case where the number of varieties is small, probability theory cannot be applied?
[32]

Of course not. It follows from previous considerations, however, that for small ν or m the application of the common formulas should be justified in a manner different from that which we have just described, or that these formulas should be modified.

I will derive new formulas below. I will mention here a certain misunderstanding which is frequently repeated in the agricultural literature, whose explanation is connected to the above argument.

This misunderstanding consists in the *unjustified assertion* that probability theory can be applied to solve problems similar to the one discussed *only if the yields from the different plots follow the Gaussian law.*

This assertion arose because, consciously or unconsciously, a different framework was used from the one mentioned above when applying probability theory.

More precisely, the yields from different plots were considered as independent measurements of one and the same number—the true yield of the variety on the field—and the measurement was assumed to be subject to errors in the sense of Laplace. To justify this framework, experiments were carried out consisting of sowing a large number of identical plots with a single variety, and it was investigated whether the yields followed the Gaussian law, as would be true if the framework above reflected experimental practice. (I will not discuss in detail here the meaning of agreement with the Gaussian law; the reader should refer to publications devoted to this topic.) Such experiments had both positive and negative results, and in those cases where positive results were questionable, the discrepancies were justified as being an unusual event. Even among the greatest optimists, I found words suggesting doubts. [Here Neyman refers to Gorskiego and Stefaniowa in the 1917 volume of the same journal.]

We have to say that in many cases the yields do not follow the Gaussian law. This is highly likely
[33]

a priori. Further, the consistency with the law of random errors should not justify a framework which is based on an assumption of independence of the measurements. In discussing this matter we will quickly get to a discussion of the assumption and constraints on the number of plots on the field or on the number of varieties compared.

In this way we conclude that *consistency with the Gaussian law is not sufficient to justify the application of known formulas, and even this (consistency) is open to doubt.*

The proposed framework even makes it superfluous, since it is enough to assume that our measurements are independent, and for that we need a large number of plots on the field.

I will now discuss the case where the ratio

$$\frac{\kappa}{m - \kappa}$$

is not so small as to be negligible, and so the experiments cannot be considered independent. Consider the design with one urn. First of all we have to say that the arithmetic mean from κ measurements may be considered an estimate of the mean

$$a = \frac{\sum_{i=1}^{m} U_i}{m}.$$

[The notation here is slightly confusing. There is no connection between the subscript i on U_i and that on the random variable x_i. Indeed the latter subscript is superfluous at this point, although the author undoubtedly has the ith urn in mind; cf. (16) and (17) below.] For that, as follows from Tchebychev's theorem, it is enough that

$$\mu^2 = \mathbb{E}(x_i - a)^2$$

tends to zero as $\kappa \to \infty$. [μ^2 is a generic expression for variance (cf. the modern use of σ^2), here of the random variable x_i which is the average of κ trials.]

We calculate μ^2:

$$\mu^2 = \mathbb{E}x_i^2 - a^2 = \frac{1}{\kappa^2}\left[\sum_{k=1}^{\kappa} x_{ik}^2 + 2\sum_{k,r} x_{ik}x_{ir}\right] - a^2,$$

where the sum $\sum x_{ik}x_{ir}$ runs over all nonidentical expressions of the type $x_{ik}x_{ir}$ with $k \neq r$. [Here x_{ik} is the random variable corresponding to the kth

of the κ trials and $x_i = (1/\kappa)\sum_{k=1}^{\kappa} x_{ik}$.] Of course

$$\mu^2 = \frac{1}{\kappa}\left(\frac{\sum_{k=1}^{m} U_k^2}{m} + \frac{2(\kappa-1)}{m(m-1)}\sum_{k,r} U_k U_r\right) - a^2$$

[34]

$$= \frac{m-\kappa}{\kappa(m-1)}\,\frac{\sum_{k=1}^{m}(U_k-a)^2}{m} = \frac{m-\kappa}{\kappa(m-1)}\,\sigma_U^2.$$

Dividing this expression in the numerator and denominator by m, and remembering that $m > \kappa$, we conclude that

$$\lim_{\kappa\to\infty}\mu^2 = \lim\frac{1-(\kappa/m)}{\kappa(1-(1/m))} = 0.$$

Thus in this case the arithmetic mean of several outcomes of the trial may be regarded as an estimate to the expected value a.

Let us make another comment. It is possible that, apart from the arithmetic mean just discussed, there exists an different function $F_{(x,\kappa)}$ of the results of the κ experiments for which $\mathbb{E}F_{(x,\kappa)} = a$, which could also be regarded as an estimate of the number a. It is also possible that the standard deviation of the function F is smaller than μ. In this case, as it follows from the law of large numbers, $F_{(x,\kappa)}$ may be associated with a better estimate of a than the arithmetic mean. Therefore we can look for the function $F_{(x,\kappa)}$ which will give the best estimate.

We shall consider a linear function

$$F_{(x,\kappa)} = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_\kappa x_\kappa.$$

[In this equation and what follows the random variable corresponding to the ith and kth of the κ trials are now denoted by x_i and x_k respectively. Neyman refers to Markov (1913) at this point.] In order that a number $F_{(x,\kappa)}$ could be considered as an estimate to a, it is sufficient that

$$\mathbb{E}F_{(x,\kappa)} = \mathbb{E}\sum_{k=1}^{\kappa}\lambda_k x_k = a$$

i.e.,

$$\sum_{i=1}^{\kappa}\lambda_i = 1.$$

In order for this estimate to be the best it is necessary that

$$M^2 = \mathbb{E}(F_{(x,\kappa)} - a)^2$$

be a minimum.

[35]

Of course

$$M^2 = \mathbb{E}\left[\sum_{i=1}^{\kappa}\lambda_i(x_i-a)\right]^2$$

$$= \sum_{i=1}^{\kappa}\lambda_i^2\mathbb{E}(x_i-a)^2 + 2\sum_{i,k}\lambda_i\lambda_k\mathbb{E}(x_i-a)(x_k-a)$$

$$= \sigma_U^2\left[\sum_{i=1}^{\kappa}\lambda_i^2 - \frac{2}{m-1}\sum_{i,k}\lambda_i\lambda_k\right],$$

since

$$\mathbb{E}(x_i-a)^2 = \frac{\sum_{k=1}^{m}(U_i-a)^2}{m} = \sigma_U^2;$$

$$\mathbb{E}(x_i-a)(x_k-a) = \frac{2\sum_{i=1}^{m-1}\sum_{k=i+1}^{m}(U_i-a)(U_k-a)}{m(m-1)}$$

$$= \frac{-\sum_{i=1}^{m}(U_i-a)^2}{m(m-1)} = -\frac{\sigma_U^2}{m-1}.$$

From the identity

$$\sum_{i=1}^{\kappa}\lambda_i = 1,$$

it follows that

$$2\sum_{i,k}\lambda_i\lambda_k = 1 - \sum_{i=1}^{\kappa}\lambda_i^2,$$

so

$$M^2 = \frac{\sigma_U^2(m\sum_{i=1}^{\kappa}\lambda_i^2 - 1)}{m-1}$$

$$= \frac{\sigma_U^2[m-\kappa+m\sum_{i,k}(\lambda_i-\lambda_k)^2]}{\kappa(m-1)}$$

is smallest when

$$\lambda_i = \lambda_1, \quad (i = 2, 3, \cdots, \kappa)$$

[36]

i.e.,

$$\lambda_i = \frac{1}{\kappa}, \quad (i = 1, 2, \cdots, \kappa)$$

$$F_{(x,\kappa)} = \frac{\sum_{i=1}^{\kappa}x_i}{\kappa}, \qquad M^2 = \frac{m-\kappa}{\kappa(m-1)}\,\sigma_U^2.$$

We see that for the case considered, the arithmetic mean of κ experiments is the best estimate of the number a.

An estimate of the standard deviation μ will be found by calculating [The text now reverts to the notation described on Neyman's page 34.]

$$\mathbb{E}(x_{ik} - x_i)^2 = \mathbb{E}\left(\frac{\kappa - 1}{\kappa} x_{ik} - \frac{1}{\kappa} \sum_k' x_{ir}\right)^2$$

$$= \frac{(\kappa - 1)m}{\kappa(m-1)} \sigma_U^2 = \frac{(\kappa - 1)m}{m - \kappa} \mu^2,$$

where

$$\sum_k' x_{ir} = \sum_{r=1} x_{ir} - x_{ik}.$$

Therefore the estimate of the standard deviation of the arithmetic mean can be denoted by μ''' whose square is equal to

$$\mu'''^2 = \frac{m - \kappa}{m(\kappa - 1)} \frac{\sum_{k=1}^{\kappa} (X_{ik} - X_i)^2}{\kappa} \tag{16}$$

$$= \frac{m - \kappa}{m(\kappa - 1)} \sigma_i^2$$

[Here X_{ik} is the kth observed outcome, $k = 1, \cdots, \kappa$ and $X_i = (1/\kappa) \sum_{k=1}^{\kappa} X_{ik}$.]

This formula should be used instead of formula (6), when κ is not negligible compared with m, as is most common. [Formula (6) is analogous to (16), but was derived under the assumption of independence of the observations, and so is without the factor $(m - \kappa)/m$.]

On the other hand, if the experiments are conducted with replacement the formula (8) [Formula (8) gives the usual unbiased estimate of the variance based upon a sequence of independent random variables with common mean and variance. For the next few formulas, x_i, x_k etc. are members of a set of ν independent random variables with expectation a, variance μ^2, and $x_0 = (1/\nu) \sum_{k=1}^{\nu} x_k$.] remains unchanged in this case since [35]

$$E(x_i - x_0)^2 = \mathbb{E}\left(\frac{\nu - 1}{\nu} x_i - \frac{1}{\nu} \sum_i' x_k\right)^2$$

$$= \frac{\nu - 1}{\nu} (\mathbb{E}x_i^2 - \mathbb{E}x_i x_k).$$

[37]

Since the numbers x_i and x_k are independent, therefore

$$\mathbb{E}x_i x_k = (\mathbb{E}x_k)^2 = a^2.$$

Thus

$$\mathbb{E}(x_i - x_0)^2 = \frac{\nu - 1}{\nu} (\mathbb{E}x_i^2 - a^2) = \frac{\nu - 1}{\nu} \mu^2$$

$$= \frac{\nu - 1}{\nu} \frac{m - \kappa}{\kappa(m-1)} \sigma_U^2,$$

and as an estimate of μ^2 we may use

$$\mu''^2 = \frac{\nu}{\nu - 1} \sigma'^2.$$

In the case when the X_i follow the Gaussian law, multiplying μ'' or μ''' by 0.67449, we get E, an estimate of the probable average error. [E is thus an estimate of the inter-quartile range. The expression σ'^2 just above was defined earlier in the paper, and is the usual biased estimate of a population variance, whereas μ''^2 is the corresponding unbiased estimate.]

It should be emphasized that the problem of determining the difference between the yields of two varieties becomes more complicated in this case. Let us consider the scheme with ν urns. [From now on, x_i and x_j are the averages of κ trials corresponding to varieties i and j, sampled as in the scheme with ν urns.] It is easy to see that

$$\mathbb{E}(x_i - x_j) = a_i - a_j,$$

so that the expected value of the difference of the partial averages of yields from two different varieties is equal to the difference of their expectations. It can also be determined that this difference is an estimate of $a_i - a_j$, but the expression for the standard deviation becomes more complicated:

$$\mu_{x_i - x_j}^2 = \mathbb{E}[x_i - x_j - (a_i - a_j)]^2$$

$$= \mathbb{E}(x_i - a_i)^2 + \mathbb{E}(x_j - a_j)^2$$

$$- 2\mathbb{E}(x_i - a_i)(x_j - a_j)$$

$$= \mu_{x_i}^2 + \mu_{x_j}^2 - 2[\mathbb{E}x_i x_j - a_i a_j].$$

The expression in the brackets will be calculated separately:

$$\mathbb{E}x_i x_j = \frac{1}{\kappa^2} \mathbb{E}\left(\sum_{k=1}^{\kappa} x_{ik} \sum_{l=1}^{\kappa} x_{jl}\right) = \mathbb{E}x_{ik} x_{jl}$$

$$= \frac{\sum_{k=1}^{m-1} \sum_{l=k+1}^{m} (U_{ik} U_{jl} + U_{il} U_{jk})}{m(m-1)}$$

[38]

$$= \frac{\sum_{k=1}^{m} U_{ik} \sum_{l=1}^{m} U_{jl} - \sum_{k=1}^{m} U_{ik} U_{jk}}{m(m-1)}.$$

Taking into account

$$\sum_{k=1}^{m} U_{ik} = m a_i, \qquad \sum_{k=1}^{m} U_{jk} = m a_j,$$

we get

$$\mathbb{E}(x_i x_j) - a_i a_j = \frac{a_i a_j - (1/m) \sum_{k=1}^{m} U_{ik} U_{jk}}{m - 1}.$$

Thus if we denote by r the correlation coefficient between the yield of two varieties on the same plot

$$r = \frac{(1/m)\sum_{k=1}^{m} U_{ik}U_{jk} - a_i a_j}{\sigma_{U_i}\sigma_{U_j}},$$

we get

$$\mathbb{E}x_i x_j - a_i a_j = -\frac{1}{m-1} r\sigma_{U_i}\sigma_{U_j},$$

and for the standard deviation of the difference of the two averages we get

$$\mu^2_{x_i - x_j} = \mu^2_{x_i} + \mu^2_{x_j} + \frac{1}{m-1} r\sigma_{U_i}\sigma_{U_j} \tag{17}$$

$$= \mu^2_{x_i} + \mu^2_{x_j} + \frac{2\kappa r}{m-\kappa}\mu_{x_i}\mu_{x_j}.$$

It is easy to see that $\mu^2_{x_i - x_j}$ tends to zero with μ_{x_i}, μ_{x_j}. It is of interest to see the relation between the standard deviations of the differences of the partial averages computed using formulas (6), (12) and the above ones. [Formula (12) states that the variance of the difference of two independent random variables is the sum of their variances.] Let us denote

$$R^2_{x_i} = \frac{\sigma_i^2}{\kappa - 1}, \qquad R^2_{x_i - x_j} = R^2_{x_i} + R^2_{x_j}.$$

[39]

Of course

$$\mu^2_{x_i - x_j} = \frac{\nu - 1}{\nu}\left[R^2_{x_i} + R^2_{x_j} + \frac{2r}{\nu - 1} R_{x_i}R_{x_j}\right]$$

$$= R^2_{x_i - x_j} - \frac{1}{\nu}[R^2_{x_i} + R^2_{x_j} - 2rR_{x_i}R_{x_j}].$$

[The right-hand side is really an estimate of the left-hand side.] It is easy to see that

$$R^2_{x_i} + R^2_{x_j} - 2rR_{x_i}R_{x_j} > 0,$$

since

$$R^2_{x_i} + R^2_{x_j} \pm 2R_{x_i}R_{x_j} = (R_{x_i} \pm R_{x_j})^2 \geq 0, \quad r < 1.$$

Therefore we conclude that

$$\mu^2_{x_i - x_j} < R^2_{x_i - x_j}.$$

We can further determine that for given R_{x_i}, R_{x_j} the variance $\mu^2_{x_i - x_j}$ increases as ν and r increase.

We achieve the largest value of the ratio

$$q = \frac{\mu^2_{x_i - x_j}}{R^2_{x_i - x_j}}$$

$$= 1 - \frac{1}{\nu R^2_{x_i - x_j}}[R^2_{x_i} + R^2_{x_j} - 2rR_{x_i}R_{x_j}]$$

only if

$$R_{x_i} = R_{x_j}, \quad r = 1,$$

when

$$q = 1, \quad \mu_{x_i - x_j} = R_{x_i - x_j}.$$

The smallest value of this ratio q is equal to zero, which can be achieved when

$$R_{x_i} = R_{x_j}, \quad r = -1, \qquad \nu = 2.$$

In this case,

$$q = 0, \qquad \mu^2_{x_i - x_j} = 0, \qquad R^2_{x_i - x_j} = 2R^2_{x_i}.$$

[40]

We see that the standard deviation of the difference of partial averages computed using the standard formulas is usually too large. It can be conjectured that in many cases this has led to the observed difference

$$X_i - X_j$$

being thought a random fluctuation, when in fact it exceeded many times the value of the standard deviation computed using the correct formula, i.e., in cases when a real difference between the yields of the two varieties being compared may be regarded as existing.

When applying (17) there is a difficulty, since we do not have a direct way of calculating r. In cases where it can be assumed that the two varieties being compared react in the same way to the soil conditions, we should take $r = 1$. [This corresponds to what is frequently termed unit-treatment additivity; see, e.g., Kempthorne, 1952, Cox, 1958, and Holland, 1986.] If we want to use the value of r computed through experiment, we will face the problem of introducing some assumptions about the nature of the variation of soil conditions over the field and the distribution of plots which are sown with comparable varieties. I hope to return to these questions in one of my future papers. They lead to a different design which ensures greater precision.

For the time being, we will conclude that since it is impossible to calculate directly an estimate of r, it is necessary to take $r = 1$; the method of comparing varieties or fertilizers by way of comparing average yields from several parallel plots has to be considered inaccurate.

Returning to the problem of determining the value of the true yield, we conclude that we are interested primarily in the true value of the difference between the yields of two varieties. Rejecting the assumption of independence of experiments, we cannot use theorem 2 [a standard form of the central limit theorem], which, although it has been generalized [here Neyman refers to Markov, 1913, for the exposition of an unpublished result of S. Bernstein] to some cases of dependent experiments,

does not apply to the case we are considering here. From these explanations it follows that it would be safe to adopt the following definitions. By the term "true value" of the difference of the yields of two varieties, sown on κ selected plots, we mean a
[41]

number Δ associated with the difference of the observed partial averages $X_i - X_j$ in such a way that the probability P_t of preserving the inequality

$$|X_i - X_j - \Delta| < t\sigma_{x_i - x_j}$$

is greater than

$$1 - \frac{1}{t^2}$$

for all $t > 0$.

We can determine empirically that the difference of partial averages of the plots sampled shows a fair agreement with the Gaussian law distribution. This encourages us to name the true difference in yields of two varieties a number δ associated with the difference of the corresponding partial averages, under the condition that the probability of preserving the inequality

$$T_1 < X_i - X_j - \delta < T_2$$

equals

$$\frac{1}{\sigma'_{x_i - x_j}\sqrt{2\pi}} \int_{T_1}^{T_2} \exp\left(-\frac{t^2}{2\sigma'^2_{x_i - x_j}}\right) dt,$$

where,

$$\sigma'^2_{x_i - x_j} = \frac{m - \kappa}{m(\kappa - 1)}\left[\sigma_i^2 + \sigma_j^2 + \frac{2\kappa r}{m - \kappa}\sigma_i\sigma_j\right]$$

and $T_1 < T_2$ are arbitrary numbers. [A misprint (or inconsistency) in the preceding equation has been eliminated; cf. formulas (16) and (17).]

We should remember, however, that this definition is not properly justified.

Of course everything that has been said about the comparison of varieties applies to the comparison of fertilizers.
[42]

Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies

Donald B. Rubin

Dorota Dabrowska and Terry Speed are to be most warmly thanked for bringing this fundamentally important but previously recondite early work of Jerzy Neyman to the attention of the statistical community. It is an honor to be asked to discuss this document, which reinforces Neyman's place as one of our greatest statistical thinkers and clarifies the debt all modern statisticians interested in causal inference owe to Jerzy Neyman. There are several specific contributions in this article (hereafter referred to as Neyman, 1923) that I feel are particularly noteworthy. To delineate these for my discussion, I first provide a brief summary using a mix of Neyman's notation and more standard current notation. I then discuss Neyman's original definition of causal effects in randomized experiments, extensions of it to experiments with interference between units and versions of treatments, and further extensions to observational studies. Three other contributions in Neyman (1923) are also analyzed: his proposal for the completely randomized experiment, his proposal for using repeated-sampling evaluations over randomization distributions, and his specific results on variance estimation in the completely randomized experiment. Throughout, I attempt to relate these contributions of Neyman's to proceeding and contemporary work of R. A. Fisher and others, and to subsequent work, including my own cited in the Dabrowska and Speed introduction. My conclusions regarding the relationship of Neyman (1923) to other work are briefly summarized in the final section.

Donald B. Rubin is Professor and Chairman, Harvard University, Department of Statistics, Science Center, 1 Oxford Street, Cambridge, Massachusetts 02138.

1. AN OVERVIEW OF NEYMAN (1923)

Neyman begins with a description of a field experiment with m plots on which v varieties might be applied: "$\cdots$ U_{ik} is the yield of the ith variety on the kth plot"; U_{ik} is a "potential yield" (Neyman's term) not an observed yield because i indexes all varieties and k indexes all plots, and each plot is exposed to only one variety. Throughout, the collection of poten-