

Translation from the Polish original of §9 pp.29-42 of

**Próba uzasadnienia zastosowań rachunku  
prawdopodobieństwa do doświadczeń polowych**

(On the application of probability theory to agricultural experiments.  
Essay on principles.)

by Jerzy Splawa-Neyman

**Roczniki Nauk Rolniczych Tom X (1923) 1-51**  
(Polish Agric. Forest. J.)

Translation carried out by D.M. Dabrowska, edited by T.P. Speed

Technical Report No. 235  
January 1990

Department of Statistics  
University of California  
Berkeley, California

Translation from the Polish original of §9 pp.29-42 of

**Próba uzasadnienia zastosowań rachunku  
prawdopodobieństwa do doświadczeń polowych**

(On the application of probability theory to agricultural experiments.  
Essay on principles.)

by

Jerzy Siplawa-Neyman

**Roczniki Nauk Rolniczych Tom X (1923) 1-51**  
(Polish Agric. Forest. J.)

Translation carried out by D.M. Dabrowska, edited by T.P. Speed

## INTRODUCTION REMARKS

*The translation which follows was taken from a paper constituting part of Neyman's doctoral thesis, submitted to the University of Warsaw in 1924. It is based on research he carried out at the Agricultural Institute of Bydgoszcz (formerly Bromberg) during the period 1921/22, see Reid (1982) for further details. Its primary interest lies in the clear formulation of Neyman's model for comparing the yields of a number of varieties, each sown on a number of plots in a field. A short summary of the contents of this paper can be found in Scheffé (1956, p.269), where the model is said to refer to "the completely randomized experiment with zero technical errors". Although sampling notions play a key role in the paper, the notion of randomization is not mentioned.*

*Part I of the paper reviews the basic notions of mathematical statistics: the Gaussian law, expected values, variances, independence, Tchebychev's inequality, the weak law of large numbers, the central limit theorem for averages of independent and identically distributed observations with finite variance, estimation of means and variances when sampling with and without replacement, and posterior probability intervals obtained via Bayes' theorem (using a uniform prior).*

*Perhaps the most interesting section of Part I is §5, entitled Definition of true value. Here Neyman is concerned with the relationship between scientific (including mathematical) concepts and empirical notions. The former he viewed as fictions which have value when properly defined, and he regarded it as important to distinguish between scientific and everyday uses of the same word or phrase. He says that he would like to define the true value of a crop yield, for example, to be the common expectation of a sequence of independent and identically distributed random variables corresponding to different measurements of that yield. However, he points out, this definition does not help much, as it then depends on the definition of expectation. Neyman's solution was to define true yield to be a number which possesses one property also possessed by expectation, namely that it serves as a location parameter for the normal approximation to the distribution of the mean of the abovementioned sequence of random variables. This permits approximate probability statements to be made about the true yield — by Bayes' theorem! — and he then moves on to a discussion of these notions in the context of agricultural experiments. See p.15 below for an illustration of Neyman's use of this definition.*

*Part II begins with a brief discussion of covariance, correlation and regression, in the context of a finite population. All of this was material familiar to statisticians at that time. Neyman referred to Czuber's Wahrscheinlichkeitsrechnung (1914) and Theorie der Beobachtungsfehler (1891), Markov's 1913 Calculus of Probability and a*

paper by Bernstein published in the Kharkov Mathematical Society Journal. Karl Pearson's *The Grammar of Science* (1900) was also cited.

Neyman's original ideas on the application of statistics (or probability theory) to agricultural experimentation are contained in the portion of the paper translated. Their most well-known exposition is, of course, in the paper Neyman (1935, with the cooperation of K. Iwaskiewicz and St. Kolodzieczyk), on the occasion of the bitter clash with R.A. Fisher. However the presentation in the later paper contrasts dramatically with that translated below. In 1923 Fisher's theory of designed experimentation, including his z-test and randomization, lay in the future, as did the Neyman-Pearson theory of testing hypotheses; the notion of technical error is also absent from the earlier discussion. By contrast, the 1935 paper discusses only randomized block and Latin square designs, includes technical errors, explicitly evaluates means and mean squares with respect to randomization (permutation) distributions, and pays attention to the Type I and Type II errors of the associated tests. Thus the emphasis shifts attention away from what was perhaps Neyman's greatest contribution to designed experimentation: his explicit use of hypothetical responses corresponding to what would have been observed, had the treatment allocation been different. For an insight into Neyman's own view of this work, see Reid (1982, pp.45-49). It seems possible that those present at Neyman's 1935 oral presentation of his ideas missed this basic point at first hearing or reading, hidden as it was in a broad critique of the then popular methods of Fisher, and that this explains, at least in part, the strong reaction to Neyman's criticisms. Later writers on the design and analysis of experiments Kempthorne (1952), Cox (1958), and others, e.g. Hodges and Lehmann (1970, section 9.4) have clearly found the model of pedagogical value.

In view of the fact that Scheffé (ibid) described the model in this paper as corresponding to the "completely randomized experiment", the reader may wonder if there is in this paper any explicit reference to randomization. We could find none. In our view, there is an implicit assumption of randomization, in that the plots which are assigned to particular treatments are supposed — for the purposes of calculating means and variances — to be simple random samples satisfying the prescribed constraints. But implicit is not explicit: randomization as a physical act, and later as a basis for analysis, was yet to be introduced by R.A. Fisher. Fisher's priority is explicitly recognized in the 1935 paper, where (p.131) Neyman, discussing the conditions under which an observed difference between two treatment means can be regarded as an unbiased estimate of their true difference, writes

*'The difficulty has been overcome by the device proposed by R.A. Fisher, which consists in making the  $\eta$ 's random variables with mean equal to zero. For this purpose the plots within each block are randomly distributed among*

*the different objects [treatments]''.*

*Why publish a translation of Neyman's introduction of this model just now? In our view, it is the model for designed experiments and observational studies which permits most (all?) of the important issues associated with their analysis and interpretation to be clearly defined, discussed, and elucidated. Apart from the earlier writers cited, we feel that this has been convincingly demonstrated most recently in a series of papers by D. Rubin, beginning with Rubin (1974, 1977, 1978), see especially Holland and Rubin (1983, 1988) and papers by Rosenbaum, alone or jointly with Rubin, cited in Holland (1986). The desirability of drawing attention to this fine body of work, and placing it more clearly in the tradition from which it arose, would seem to justify carrying out the translation. We hope that the paper will also have some appeal to those interested in the evolution of the statistical notions associated with experiments and other studies.*

#### **ACKNOWLEDGEMENTS**

*We would like to thank Professor J. Baksalary for supplying a photocopy of the paper and Professors D. Freedman, E. Lehmann and S. Stigler for their encouragement and helpful comments in relation to this project.*

#### **REFERENCES**

*COX, D.R. (1958). Planning of experiments. New York: John Wiley.*

*CZUBER, E. (1891). Theorie der Beobachtungsfehler. Leipzig: B.G. Teubner.*

*CZUBER, E. (1914). Wahrscheinlichkeitsrechnung und ihre Anwendung auf Fehlerausgleichung Statistik und Lebensversicherung. (3rd ed.) Leipzig: B.G. Teubner.*

*HODGES, J.L. and LEHMANN, E. (1970). Basic concepts of probability and statistics, 2nd ed. San Francisco: Holden-Day.*

*HOLLAND, P.W. (1986). Statistics and causal inference (with discussion). J. Amer. Statist. Assoc. **81**, 945-970.*

*HOLLAND, P.W. and RUBIN, D.B. (1983). On Lord's paradox. Principals of Modern Psychological Measurement ed. H. Wainer and S. Messick. Hillsdale, N.J.: Lawrence Erlbaum Assoc. 3-25.*

*HOLLAND, P.W. and RUBIN, D.B. (1988). Causal inference in retrospective studies.*

Evaluation Review **12**, 203-231.

KEMPTHORNE, O. (1952). The design and analysis of experiments. New York: John Wiley.

MARKOV, A.A. (1913). Calculus of Probability (In Russian) St. Petersburg.

NEYMAN, J. with the cooperation of K. Iwaskiewicz and St. Kolodziejczyk (1935). Statistical problems in agricultural experimentation (with discussion). Suppl. J.R. Stat. Soc. **2**, 266-278.

PEARSON, K. (1900). The Grammar of Science. 2nd ed. rev. and enlarged. London: Adam and Charles Black.

REID, C. (1982). Neyman - from life. New York: Springer - Verlag.

RUBIN, D.B. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. J. Educ. Psychology **66**, 688-701.

RUBIN, D.B. (1977). Assignment of treatment group on the basis of a covariate. J. Educ. Statistics **2**, 1-26.

RUBIN, D.B. (1978). Bayesian inference for causal effects: the role of randomization. Ann. Statist. **6**, 34-58.

SCHEFFÉ, H. (1956). Alternative models for the analysis of variance. Ann. Math. Statist. **27**, 251-271.

Translation from the Polish original of §9 pp.29-42 of  
**Próba uzasadnienia zastosowań rachunku  
prawdopodobieństwa do doświadczeń polowych**  
(On the application of probability theory to agricultural experiments.  
Essay on principles.)  
by Jerzy Splawa-Neyman

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

[29]

9. 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

$$U_1, U_2, \dots, U_m$$

be the true yields of a particular variety on each of these plots. If all the numbers  $U_i$  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

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

The yield from the  $i$ -th 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<sup>1)</sup> of the 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  $v$  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):

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

---

<sup>1)</sup> See the introduction for a few comments on Neyman's notion of true yield.

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

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

$$U_{i1}, U_{i2}, \dots, U_{ik}, \dots, U_{im} \quad (13)$$

[30]

where  $i$  denotes the number of the urn (variety), and  $k$  denotes the plot number, whilst  $U_{ik}$  is the yield of the  $i$ -th variety on the  $k$ -th 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  $i$ -th 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  $v$  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  $v$  varieties will be regarded as equivalent to the problem of comparing the numbers

$$a_1, a_2, \dots, a_v$$

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

The simplest way of obtaining an estimate to the number  $a_i$  would be by drawing  $\kappa$  balls from the  $i$ -th 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  $\kappa$  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<sup>2)</sup> of the experiment consisting of drawing one ball from the  $i$ -th urn. We shall calculate  $\mathbf{E} x$ . Since the probability of drawing a ball from the  $i$ -th urn is the same for all balls, and equal to  $\frac{1}{m}$ , and since all possible results of the

---

<sup>2)</sup> In modern terminology, lower case  $x$ , with or without subscripts, denotes a random variable, and upper case  $X$  the corresponding realized values.



trial are

[31]

contained in the sequence (13), so of course

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

and the average of the results of the  $\kappa$  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  $x_1, \dots, x_\kappa; X_1, X_2, \dots, X_\kappa$  be the possible and the true outcomes, respectively, of  $\kappa$  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

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

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 contain one fewer balls. Further, the number of elements in the  $k$ -th 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

[32]

is

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

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  $\kappa$ -th 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

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

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  $\kappa$ , so that  $\frac{\kappa}{m - \kappa}$  is negligible in comparison with the probabilities  $p_s$ , then it follows from the above formulae that information about previous trials will not affect probabilities of subsequent trials, and so the trials will turn out to be independent, whence we will be able to apply the law of large numbers, and our definition of a true yield, and along with it known formulae from probability theory. If each of the  $v$  varieties are sown on  $\kappa$  plots, then  $m = v\kappa$  and the condition for the independence of experiments will be that the ratio  $\frac{1}{v - 1}$  is small, in other words, the number of varieties  $v$  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?

[33]

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

I will derive new formulae 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 in 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,<sup>3)</sup> I found words suggesting doubts.

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

[34]

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 formulae, 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  $\kappa$  measurements may be considered an estimate of the mean<sup>4)</sup>

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

For that, as follows from Tchebychev's theorem, it is enough that<sup>5)</sup>

---

<sup>3)</sup> Here Neyman refers to Gorskigo and Stefaniowa in the 1917 volume of the same journal.

<sup>4)</sup> 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  $i$ -th urn in mind, cf. (16) and (17) below.

<sup>5)</sup>  $\mu^2$  is a generic expression for variance (cf. the modern use of  $\sigma^2$ ), here of the random variable  $x_i$  which is the average of  $\kappa$  trials.

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

tends to zero as  $\kappa \rightarrow \infty$ .

We calculate  $\mu^2$ :

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

where the sum  $\sum x_{ik} x_{ir}$  runs over all non-identical expressions<sup>6)</sup> of the type  $x_{ik} x_{ir}$  with  $k \neq r$ . Of course

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

[35]

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

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

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

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 a different function  $F_{(x, \kappa)}$  of the results of the  $\kappa$  experiments for which  $\mathbf{E} F_{(x, \kappa)} = a$ , 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  $\mu$ . 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<sup>7)</sup>

<sup>6)</sup> Here  $x_{ik}$  is the random variable corresponding to the  $k$ -th of the  $\kappa$  trials and  $x_i = \frac{1}{\kappa} \sum_{k=1}^{\kappa} x_{ik}$ .

<sup>7)</sup> In what follows the random variable corresponding to the  $i$ -th and  $k$ -th of the  $\kappa$  trials are now denoted by  $x_i$  and  $x_k$  respectively. Neyman refers to Markov (1913) at this point.

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

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.

[36]

Of course

$$\begin{aligned} M^2 &= \mathbb{E} \left[ \sum_{i=1}^{\kappa} \lambda_i (x_i - a) \right]^2 \\ &= \sum_{i=1}^{\kappa} \lambda_i^2 \mathbb{E} (x_i - a)^2 + 2 \sum_{i,k} \lambda_i \lambda_k \mathbb{E} (x_i - a) (x_k - a) \\ &= \sigma_U^2 \left[ \sum_{i=1}^{\kappa} \lambda_i^2 - \frac{2}{m-1} \sum_{i,k} \lambda_i \lambda_k \right], \end{aligned}$$

since

$$\begin{aligned} \mathbb{E} (x_i - a)^2 &= \frac{\sum_{k=1}^m (U_k - a)^2}{m} = \sigma_U^2; \\ \mathbb{E} (x_i - a) (x_k - a) &= \frac{2 \sum_{i=1}^{m-1} \sum_{k=i+1}^m (U_i - a) (U_k - a)}{m(m-1)} \\ &= \frac{-\sum_{i=1}^m (U_i - a)^2}{m(m-1)} = -\frac{\sigma_U^2}{m-1}. \end{aligned}$$

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

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

[37]

$$\text{i.e. } \lambda_i = \frac{1}{\kappa}, \quad (i = 1, 2, \dots, \kappa)$$

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

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

An estimate of the standard deviation  $\mu$  will be found by calculating<sup>8)</sup>

$$\begin{aligned} \mathbb{E} (x_{ik} - x_i)^2 &= \mathbb{E} \left( \frac{\kappa - 1}{\kappa} x_{ik} - \frac{1}{\kappa} \sum_k' x_{ir} \right)^2 \\ &= \frac{(\kappa - 1)m}{\kappa(m - 1)} \sigma_U^2 = \frac{(\kappa - 1)m}{m - \kappa} \mu^2, \end{aligned}$$

where

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

Therefore the estimate of the standard deviation of the arithmetic mean can be denoted by  $\mu'''$  whose square is equal to<sup>9)</sup>

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

This formula should be used instead of formula (6), when  $\kappa$  is not negligible compared with  $m$ , as is most common.<sup>10)</sup>

<sup>8)</sup> The text now reverts to the notation described in Footnote 4.

<sup>9)</sup> Here  $X_{ik}$  is the  $k$ -th observed outcome,  $k = 1, \dots, \kappa$  and  $X_i = \frac{1}{\kappa} \sum_{k=1}^{\kappa} X_{ik}$ .

<sup>10)</sup> 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) remains unchanged in this case since<sup>11),12)</sup>

$$\begin{aligned} E(x_i - x_0)^2 &= E\left(\frac{v-1}{v}x_i - \frac{1}{v}\sum_i x_k\right)^2 \\ &= \frac{v-1}{v}(E x_i^2 - E x_i x_k). \end{aligned}$$

[38]

Since the numbers  $x_i$  and  $x_k$  are independent, therefore

$$E x_i x_k = (E x_k)^2 = a^2.$$

Thus

$$\begin{aligned} E(x_i - x_0)^2 &= \frac{v-1}{v}(E x_i^2 - a^2) = \frac{v-1}{v}\mu^2 \\ &= \frac{v-1}{v} \frac{m-\kappa}{\kappa(m-1)} \sigma_U^2 \end{aligned}$$

and as an estimate of  $\mu^2$  we may use

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

In the case when the  $X_i$  follow the Gaussian law, multiplying  $\mu''$  or  $\mu'''$  by 0.67449, we get  $E$  - an estimate of the probable average error.<sup>13)</sup>

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  $v$  urns. It is easy to see that<sup>14)</sup>

$$E(x_i - x_j) = a_i - a_j$$

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:

<sup>11)</sup> Formula (8) gives the usual unbiased estimate of the variance based upon a sequence of independent random variables with common mean and variance.

<sup>12)</sup> For the next few formulae,  $x_i, x_k$  etc. are members of a set of  $v$  independent random variables with expectation  $a$ , variance  $\mu^2$ , and  $x_0 = \frac{1}{v} \sum_{k=1}^v x_k$ .

<sup>13)</sup>  $E$  is thus an estimate of the inter-quartile range. The expression  $\sigma^{12}$  just above was defined earlier in the paper, and is the usual biased estimate of a population variance, whereas  $\mu''^2$  is the corresponding unbiased estimate.

<sup>14)</sup> From now on,  $x_i$  and  $x_j$  are the averages of  $\kappa$  trials corresponding to varieties  $i$  and  $j$ , sampled as in the scheme with  $v$  urns.

$$\begin{aligned}\mu_{x_i-x_j}^2 &= \mathbb{E} [x_i - x_j - (a_i - a_j)]^2 = \mathbb{E} (x_i - a_i)^2 + \\ &\quad \mathbb{E} (x_j - a_j)^2 - 2 \mathbb{E} (x_i - a_i) (x_j - a_j) \\ &= \mu_{x_i}^2 + \mu_{x_j}^2 - 2 [\mathbb{E} x_i x_j - a_i a_j].\end{aligned}$$

The expression in the brackets will be calculated separately:

$$\begin{aligned}\mathbb{E} x_i x_j &= \frac{1}{\kappa^2} \mathbb{E} \left( \sum_{k=1}^{\kappa} x_{ik} \sum_{l=1}^{\kappa} x_{jl} \right) = \mathbb{E} x_{ik} x_{jl} \\ &= \frac{\sum_{k=1}^{m-1} \sum_{l=k+1}^m (U_{ik} U_{jl} + U_{il} U_{jk})}{m(m-1)}\end{aligned}$$

[39]

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

Taking into account

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

we get

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

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

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

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

$$\mu_{x_i-x_j}^2 = \mu_{x_i}^2 + \mu_{x_j}^2 + \frac{2}{m-1} r \sigma_{U_i} \sigma_{U_j} = \mu_{x_i}^2 + \mu_{x_j}^2 + \frac{2\kappa r}{m-\kappa} \mu_{x_i} \mu_{x_j}. \quad (17)$$

It is easy to see that  $\mu_{x_i-x_j}^2$  tends to zero with  $\mu_{x_i}, \mu_{x_j}$ . It is of interest to see the relation between the standard deviations of the differences of the partial averages computed using formulae (6), (12)<sup>15)</sup> and the above ones. Let us denote

<sup>15)</sup> Formula (12) states that the variance of the difference of two independent



$$R_{x_i}^2 = \frac{\sigma_i^2}{\kappa - 1}, \quad R_{x_i - x_j}^2 = R_{x_i}^2 + R_{x_j}^2.$$

[40]

Of course<sup>16)</sup>

$$\begin{aligned} \mu_{x_i - x_j}^2 &= \frac{v - 1}{v} [R_{x_i}^2 + R_{x_j}^2 + \frac{2r}{v - 1} R_{x_i} R_{x_j}] \\ &= R_{x_i - x_j}^2 - \frac{1}{v} [R_{x_i}^2 + R_{x_j}^2 - 2r R_{x_i} R_{x_j}]. \end{aligned}$$

It is easy to see that

$$R_{x_i}^2 + R_{x_j}^2 - 2r R_{x_i} R_{x_j} > 0,$$

since

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

Therefore we conclude that

$$\mu_{x_i - x_j}^2 < R_{x_i - x_j}^2.$$

We can further determine that with given  $R_{x_i}$ ,  $R_{x_j}$  the variance  $\mu_{x_i - x_j}^2$  increases as  $v$  and  $r$  increase.

We achieve the largest value of the ratio

$$q = \frac{\mu_{x_i - x_j}^2}{R_{x_i - x_j}^2} = 1 - \frac{1}{v R_{x_i - x_j}^2} [R_{x_i}^2 + R_{x_j}^2 - 2r R_{x_i} R_{x_j}]$$

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

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

In this case

$$q = 0, \quad \mu_{x_i - x_j}^2 = 0, \quad R_{x_i - x_j}^2 = 2R_{x_i}^2.$$

random variables is the sum of their variances.

<sup>16)</sup> The right-hand side is really an estimate of the left-hand side.

[41]

We see that the standard deviation of the difference of partial averages computed using the standard formulae 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  $r = 1$ .<sup>17)</sup> 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,<sup>18)</sup> which, although has been generalized<sup>19)</sup> 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  $\kappa$  selected plots, we mean a

[42]

---

<sup>17)</sup> This corresponds to what is frequently termed unit-treatment additivity, see e.g. Kempthorne (1952), Cox (1958) and Holland (1986).

<sup>18)</sup> A standard form of the central limit theorem.

<sup>19)</sup> Here Neyman refers to Markov (1913) for the exposition of an unpublished result of S. Bernstein.

number  $\Delta$  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  $\delta$  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

$$\frac{1}{\sigma_{x_i - x_j} \sqrt{2\pi}} \int_{T_1}^{T_2} e^{-\frac{t^2}{2\sigma_{x_i - x_j}^2}} dt,$$

where<sup>20)</sup>

$$\sigma_{x_i - x_j}^2 = \frac{m - \kappa}{m(\kappa - 1)} [\sigma_i^2 + \sigma_j^2 + \frac{2\kappa}{m - \kappa} \sigma_i \sigma_j]$$

and  $T_1 < T_2$  are arbitrary numbers.

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.

---

<sup>20)</sup> A misprint (or inconsistency) in the following has been eliminated, cf. formulae (16) and (17) above.