World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(REVIEW ARTICLE)

Requirements of the experiment

João Gilberto Corrêa da Silva *

Department of Mathematics and Statistics, Federal University of Pelotas, Pelotas, RS, Brazil.

World Journal of Advanced Research and Reviews, 2024, 22(02), 405–423

Publication history: Received on 20 March 2024; revised on 03 May 2024; accepted on 06 May 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.22.2.1311

Abstract

The purpose of the experiment is to derive inferences about the causal effects of a subset of the characteristics of the units of a target population (explanatory characteristics - treatment factors) on characteristics that express the performance of these units (response characteristics). To achieve this objective, the effects of the treatment factors in the sample must manifest themselves as in the target population. This attempt requires the construction of a valid and representative sample of the target population. It is achieved if the experiment plan and its execution satisfy the experiment requirements. These requirements are considered in this article.

Keywords: Experiment plan; Estimation of errors affecting effects of treatment factors; Precision; Internal validity; External validity; Simplicity; Manifestation of the effects of treatments; Prediction of statistical inference procedures; Inference uncertainty measures.

1. Introduction

The modern methodology of experimental research originated with the innovative ideas of Fisher [1,2]. One of his relevant contributions was the recognition and importance attributed to the planning of the experiment and, particularly, to the efficiency gain that can be provided by a well-planned experiment. The basic experiment designs proposed by Fisher began to be applied in various situations, particularly in agricultural research. The wide dissemination of these designs and the more complex designs added by Yates [3,4,5], at a time when resources for data analysis were precarious, led to their widespread use.

With the growing availability of computing resources, it became possible to formulate the most appropriate experiment design for each situation and, consequently, to design more efficient experiments. However, the adoption of this approach requires an understanding of the scientific approach to experimental research and particularly the conceptual and methodological foundations of the experiment, the principles of experiment design, and the requirements of the experiment.

The scientific approach requires the establishment of a plan prior to the execution of the experiment, that is, the establishment in advance, in written form, of the complete set of decisions and actions that must be taken and proceeded for the execution of the experiment. Such a plan must be consistent with the objectives of the experiment, determined by the problem and the scientific hypothesis, and be formulated in such a way as to ensure the inferences to achieve these objectives.

To fulfill this purpose, it is necessary that the plan of the experiment satisfies some important requirements, particularly estimation of errors that affect the effects of experimental factors, precision and validity. These requirements demand the foundation of the experiment on a rational, precise, complete and coherent conceptual basis, and the correct use of the methodology built throughout the evolution of experimental research. This approach forms a sequence of steps

* Corresponding author: João Gilberto Corrêa da Silva

logically arranged and defined in line with the context of the research Silva [6]. Silva [7,8] proposed a conceptual and methodological basis consistent with the real meanings and the logical sequence of experimental research. This rationale leads to the correct identification of the components of experimental error that affect the effects of experimental factors and provides precise and unbiased inferences about the effects of treatment factors. These requirements also require the appropriate adoption of principles of experiment design (Silva [9]).

This article considers the required properties of the experiment, which must be foreseen in its planning and guaranteed in its execution. Compliance with these requirements allows the elaboration of the plan and the execution of the experiment that provides the ability to detect the real effects of treatment factors.

This article is based on the contributions of Fisher [1,2], Yates [3,4,5], Silva [6,7,8,9], Bailey [10], Cochran & Cox [11], Fisher & Yates [12] and Kuehl [13], which are explicitly referred in the text, and Cox [14], Christensen [15], Federer [16], Giesbrecht [17], Gould [18], Hinkelmann & Kempthorne [19], Kempthorne [20,21,22], Little & Hills [23], Mead [24], Mitchell &, Jolley [25], Pearce [26,27], Petersen [28], Selwyn [29], Shadish, Cook & Campbell [30].

2. Requirements of the experiment

The experiment must provide the relevant inferences to its objectives at minimal cost. For this purpose, the experiment must satisfy the following essential properties or requirements: 1-establishment of the plan before the start of the execution of the experiment, 2-estimation of errors affecting the effects of treatment factors, 3-precision, 4-validity, 5-simplicity, resource saving and feasibility, 6-manifestation of the real effects of the treatments, and 7-prediction of statistical inference procedures and provision of uncertainty measurement.

These requirements may seem obvious, but it is surprising how often they are underestimated or ignored. The consequences are failures of experiments that entail wasted resources and time. These requirements are considered below.

2.1. Establishment of the plan before the start of the execution of the experiment

Planning requires the prior detailed and clear specification of the objectives of the experiment. This is necessary so that the definitions about the stages of the sequential process of the experiment can be defined in an objective and coherent way for the achievement of these objectives.

The plan of the experiment must be formulated in detail and described in an appropriate protocol clearly so that the decisions and actions necessary until the derivation of the inferences are predicted (Silva [9]). In particular, the plan must provide for the resources needed and the times when they must be available. If these resources are not available or if they are insufficient for the experiment to satisfy the requirements for generating the inferences that constitute its objectives, these objectives must be revised to make them feasible, or the experiment must not be carried out.

2.2. Estimation of errors affecting effects of treatment factors

For the derivation of inferences regarding a treatment factor effect, the experiment must provide two sources of variation of the observed values of the response variable: a) a source that originates two components of this variation: this treatment factor effect and the error confounded with it, and b) a source that originates only the error that affects that treatment factor effect. Therefore, inferences about the effects of treatment factors require estimation of the errors that affect these effects. These errors comprise experimental error and may also include effects of random experimental factors, if present in the condition structure.

Thus, an essential requirement of the experiment is to provide for the estimation of the errors, particularly the experimental error affecting relevant effects of treatment factors. To this end, the experiment plan must ensure that levels of treatment factors are present in more than one experimental unit, i.e., it must ensure repetitions of these levels. In some situations of several treatment factors, estimates of errors affecting important effects of treatment factors can be provided by factor interactions supposedly non-existent. This is the case with experiments in uniform environments where it is known that certain high-order interactions are non-existent or irrelevant. This situation can occur, for example, in research of industrial processes. Based on this argument, components of variation that express such interactions are used as error estimates for inferences regarding effects of relevant treatment factors. Of course, the validity of these inferences depends on the correction of this argument. If it is not correct, the error will be overestimated and, consequently, the inferences will be biased.

2.3. Precision

The precision of the experiment is related to the sensitivity to detect effects of treatment factors manifested in the sample. It depends on the magnitude of the experimental error that affects the effects of treatment factors. It is higher the smaller this fraction of the experimental error.

In experiment design with one treatment factor in a single stratum, the magnitude of the experimental error that affects treatment factor effects is measured by the variance of the experimental error affecting these effects expressed per experimental unit, which is denoted by σ^2. The precision of the experiment is expressed by the inverse of this variance: $1/\sigma^2$. Usually, the population variance σ^2 is not known and in its place is used an estimate of it, that is, an appropriate value to approximate it determined from the sample. This estimate of the variance of the experimental error per unit is the variation of the observed values of the response variable that is due to extraneous characteristics, divided by the corresponding number of independent information units, which is commonly called degrees of error freedom. This estimate is denoted by s^2. Then, the precision of an experiment with this experiment design is expressed by the inverse of the variance s^2 of the experimental error, that is, $1/s^2$.

This expression of precision is not easily interpretable, since its unit of measure is the square of the unit of measure of the response variable. It may be more convenient to express the precision of the experiment by the inverse of the standard deviation, which is the square root of the variance σ^2, i.e., σ that is estimated by s. Precision is even more conveniently interpreted by the inverse of the coefficient of variation, which is the standard deviation expressed as a percentage of the mean: CV = 100 s/m, where m is the mean of the observed values of the response variable.

In comparative experiments, the most important inferences refer to comparisons of treatment means. The magnitude of the experimental error that affects a difference in treatment means is expressed by the corresponding variance or its square root, which is called the standard error. Standard error is a measure of the imprecision of inferences about a difference in treatment means. In experiments with a single experimental factor, the same number of repetitions for all treatments and treatment effects affected by experimental error of a single stratum, the standard error of the estimate of the difference of two treatment means is expressed by $s\sqrt{2/r}$, where s is the standard deviation and r is the number of repetitions common to both treatments.

Fisher [1,2] proposed an alternative expression for precision or amount of information supplied by an experiment design as an approximation to $1/\sigma^2$: $\frac{\nu+1}{(\nu+3)s^2}$, where s^2 is an estimate of σ^2 and ν is the degrees of freedom of this estimate. He argues that the estimated variance s^2 is itself subject to sampling error, and an allowance for such error should be made by using the t-distribution, instead of the normal distribution, which considers not only the estimate s^2, but also the degrees of freedom ν upon which this estimate was based.

The above precision expressions are appropriate for experiments with orthogonal designs with one experimental factor, in which effects of this treatment factor are restrict to one stratum. It also applies to precision in each stratum of experiments with orthogonal design with more than one treatment factor and more than one formation of experimental units when effects of each treatment factor are restricted to one stratum, with s^2 symbolizing the estimate of the variance of the experimental error in the corresponding stratum.

In experiments with more than one treatment factor in one or more strata, precision may vary with the factor and with factor level, if levels have different numbers of repetitions. An appropriate definition of precision for these circumstances applies to treatment means, with a similar expression: $\frac{r_i}{s^2}$ or $\frac{(\nu+1)}{(\nu+3)}\frac{r_i}{s^2}$ suggested by Fisher, where the estimated experimental error variance s^2 is replaced by the standard error of a treatment mean: $s^2_{\bar{y}_i} = \frac{s^2}{r_i}$, ν is the degrees of freedom of the estimate s^2 and r_i is the number of repetitions of treatment i.

In general, the expression of precision depends on design. In experiments with more complex designs, as balanced incomplete block designs and factorial design with confounding, effects of a treatment factor can be situated in more than one stratum. Consequently, the estimate of the variance of the error that affects these effects is a composition of variance estimates in these strata. Considerations on this subject can be found in Bailey [10], Cochran & Cox [11], Fisher & Yates [12] and Kuhel [13].

2.3.1. Origins of imprecision

Precision depends on the magnitude of errors that affect the effects of treatment factors. Thus, sources of inflation of these errors are damage to precision. Generally, they have the following origins: response variables, experimental factors and their levels, number and dimensions of experimental units, extraneous characteristics of the experimental material, that is, characteristics related to: initial sample, treatment vehicles, application of treatments, environment, actions for operation of the units, and measurement and recording of data.

When planning an experiment, the researcher must identify potential sources of imprecision based on experience gathered from similar experiments, information available in the literature and other sources. In some situations, particularly in experiments in new areas, it may be recommended to carry out preliminary experiments to evaluate the impact of extraneous characteristics on precision, and to define and improve experimental techniques to control these characteristics.

2.3.2. Actions to increase precision

Once the important sources of imprecision are identified, the next step is to define the actions to control them to achieve the appropriate precision. This precision should be sufficient to provide a high probability of detecting important effects of treatment factors. The appropriate resources for this purpose depend on the circumstances of each experiment. The sources of imprecision and actions that may contribute to increased precision are considered below.

Appropriate choice of response variables - The discrimination of the levels of a response characteristic depends on the scale of measurement of the variable that is defined to express it. Response variables must express the characteristics they represent with the precision appropriate to their relevance. For example, the incidence of a fungal disease in wheat plants can be expressed by a variable of only two levels: without and with incidence, by a variable of four levels: without infection and with low, medium and high infection, by a discrete variable with a higher number of levels, or by a continuous variable. The measurements provided by these alternative variables distinguish the degrees of infection manifested in the observation units with different precision.

Appropriate choice of treatment factors and their levels - The choice of a treatment factor with levels that relate quantitatively may have relevant implications for the precision of inferences. This is the case, for example, when the levels are doses of a fertilizer or an antibiotic. For illustration, consider the simple situation in which the response to increasing levels of the factor is linear, that is, the ratio between the increment of the response y and the increment of the level of the factor x is constant, so that inference about the slope b of the line $E(y) = a + bx$ that represents the postulated relationship for the target population is of interest. The levels x_1, x_2,... of x that provide the maximum precision of the inferences about the parameter b are the values of x that make minimal the variance of the estimator of b, i.e., $\mathrm{Var}(\hat{b}) = \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}$, where $\bar{x}$ is the mean of these levels and σ^2 the variance of the error that affects effects of factor x. The precision of these inferences is maximum when the levels are the values of x that maximize $SSX = \sum_i (x_i - \bar{x})^2$. Assuming that the total number of observations is fixed, SSX can be increased by increasing the amplitude of the range of values of x, and for a fixed value of this amplitude, SSX is maximized when half of the values of x lie at each end of the interval. This is the optimal treatment structure for inferences regarding the ratio of increments of a linear relationship. This procedure extends to determining the optimal treatment structure for inferences about parameters or parametric functions of more complex response functions.

Precision gains can also be achieved by considering two or more treatment factors in the same experiment, rather than performing separate experiments for each of these factors. For example, research of the effects of plowing depth and soil fertilization with nitrogen on corn yield can be carried out through two experiments, one to compare deep ploughing (A_1) and shallow ploughing (A_2), and another to compare fertilization with a specific amount of N (N_1) with the absence of fertilization (N_0). Alternatively, these two comparisons can be made in a single experiment with two factors: plowing and fertilization with N, with four treatments: 1 - A_1N_1, 2 - A_1N_0, 3 - A_2N_1 and 4 - A_2N_0. Then, the effect of soil fertilization with nitrogen is evaluated by comparing treatments 1 and 3 with treatments 2 and 4. On the assumption of experimental material with the same composition of extraneous characteristics, this experiment provides a more precise estimate of the effect of nitrogen fertilization than an experiment with the same number of repetitions devoted to research only this effect. In fact, the precision of this estimate provided by the two-factor experiment is $\sqrt{2}$ times the precision provided by the experiment with the N-fertilization factor alone. In general, experiments with two or more treatment factors provide more precise estimates of the effects of these factors than separate experiments each devoted to one of the factors.

Choice of homogeneous initial sample - The adoption of a sample of homogeneous units at the beginning of the experiment is highly effective to achieve high precision. For example, experiments with uniform animals and plants are more precise than experiments with animals and plants with varying individual characteristics. This procedure is often appropriate for basic experiments; however, it can be inconvenient for technological experiments, such as experiments that aim to generate technologies for adoption recommendation. In these experiments the sample must represent the heterogeneity present in the target population, which is considerably accentuated in many situations.

Appropriate experiment size - The number of experimental units determines the number of degrees of freedom for estimating the variance of the experimental error. If this increase in the degrees of freedom outweighs the increment of variation due to the extraneous characteristics, the variance of the experimental error is reduced.

The increase in the size of the experiment can be obtained by increasing the number of repetitions and by considering supplementary experimental factors. The standard error of the difference of the means of two treatments based on the same number of repetitions and affected by experimental error of a single stratum is inversely proportional to the square root of the number of repetitions. Thus, in this situation, the standard error is halved by quadrupling the number of repetitions, while to reduce the standard error to one tenth, it is necessary to multiply the number of repetitions by one hundred. In wide-range experiments, the size of the experiment is increased by conducting the experiment in various sections of space and time. In these experiments, the increase in the number of repetitions for important treatment factors is obtained by including the intrinsic factors local and time. Additional treatment factors may also be used for this purpose. For example, in an experiment to investigate the efficiency of fungicides to control vine downy mildew, greater precision can be achieved by increasing the number of fungicide repetitions in a location and year, by conducting the experiment in several locations and years, or by including an additional treatment factor, such as cultivar. Although theoretically the standard error can be made arbitrarily small by increasing the number of repetitions, this way of increasing precision has limited use as it is costly and often impractical.

The preceding discussion stresses the importance of striving to reduce estimates of standard deviations and errors. It should be noted, however, that while the standard deviations and errors must be small enough to permit the derivation of convincing inferences, they must not be too small. In fact, unnecessarily small standard deviations and errors may imply waste of experimental material.

Very often, inferences of interest are estimation and hypothesis tests regarding differences in treatment means. The use of the expression of the standard error of the estimate of the difference of two means of treatments, given previously, or its extension to the situation of different numbers of repetitions, allows to predict, when the experiment is being planned, the number of repetitions necessary to achieve a given precision for these inferences, or the precision that will be obtained with a particular number of experimental units. For this, it is necessary to know the variability of the units that allows to evaluate the magnitude of the standard deviation. Approximate information of this variability can be obtained from similar experiments already carried out.

Control of experimental techniques - This experimental control procedure may be appropriate for the reduction of variation due to extraneous characteristics of various origins. The actions that may be effective for this purpose are considered below.

- Homogenization of the extraneous characteristics of the initial sample - For example, use of basic fertilization, that is, uniform fertilization before planting to reduce the variation of soil fertility in an experiment with plants; adoption of a pre-experimental period in which the animals are kept under the same management to standardize the individual characteristics of the animals.
- Appropriate dimensions of experimental units - In experiments where the experimental material exhibits considerable spatial variability, such as field agricultural experiments, precision can be significantly affected by the size and shape of the experimental units. In these circumstances, the appropriate choice of the dimensions of the experimental units can contribute to increase precision. For example, in an experiment to be conducted on terrain with marked heterogeneity of soil characteristics in one direction, higher precision can be achieved with the use of rectangular plots with larger dimension in the perpendicular direction.
- Homogenization of the extraneous characteristics conveyed with the treatments - In general, treatments are associated with extraneous characteristics along with which they are transmitted. For example, in a cultivar comparison experiment the cultivar factor levels are cultivars each of which is defined as a set of specific characteristics that are inherent to it, particularly genetic characteristics; in an experiment on the control of an infection of animals with antibiotics, antibiotics are made up of sets of characteristics inherent to their respective active principles. However, the cultivars are conveyed by seeds that comprise, in addition to the characteristics related to the genetic entity cultivar, characteristics related to health, purity, germination, vigor, etc.; the active principles of antibiotics are conveyed together with characteristics of the substances to which they are attached. The variation arising from these extraneous characteristics constitutes a treatment error, which contributes to the experimental error. This variation should be controlled by control of experimental techniques; for example, in the experiment of comparison of cultivars, the seeds must be uniform as to vigor,

sanity, purity and other characteristics not inherent to cultivar; in the animal infection control experiment, antibiotics must be homogeneous in quality, particularly regarding the substances with which they are conveyed and the expiry date.

- Control of the application of treatments - Differences regarding the application of treatments to experimental units also constitute treatment error and contribute to experimental error. For example, in the cultivar comparison experiment, sowing should be carried out uniformly in the experimental units with regard to quantity and distribution of seeds, date of planting, etc.; in an experiment of nutrition of lambs, the diets should be applied in the experimental units according to the respective specifications regarding the quantity and mode of supply; in an experiment on the effect of humidity on seed quality after a period of storage, the seeds with the specified moistures should be evenly distributed in the corresponding containers to constitute the experimental units; in an insect control experiment with insecticides, special apparatus may be required to ensure that all experimental units receive the insecticides assigned to them uniformly, particularly with regard to quantity.
- Homogenization of the manifestation of environmental characteristics - For example, uniformization of air temperature and humidity in experiments conducted in controlled environments, such as greenhouse, laboratory and protected facilities; application of insecticides, fungicides and herbicides to control the incidence of insects, diseases and invaders, and protection against predators in field agricultural experiments; application of prophylactic medications, such as vaccines and dewormers, in animal experiments.
- Control of the implementation of techniques necessary for the operation of the units - Faulty experimental techniques can make a substantial contribution to the experimental error. This potential component of experimental error can be reduced with the controlled execution of these techniques. In particular, plant cultivation techniques (planting, application of insecticides, fungicides and herbicides, harvesting, etc.) and animal management techniques (provision of food and water, application of medicines, etc.) should be employed uniformly. For example, the number of seeds should be homogeneous for all plots, and the stocking and supply of feed and water should be uniform for all compartments (boxes, paddocks, cages) that constitute the experimental units.
- Use of intentional experimental techniques to control the operation of units - The use of special techniques that are not usual in the target population can contribute to increased precision. For example, use of border on the plot or protection to avoid confounding of effects of treatments in neighboring plots and reduction of plants that emerge to the same number per plot, in field agricultural experiments; replacement of affected animals by disturbing extraneous characteristics (predators, for example) and adoption of unusual drugs in the target population to avoid diseases and parasites, in animal experiments.
- Use of appropriate measuring procedures and instruments - The precision of inferences can be considerably affected by the precision of measurements. For example, in an experiment with adult sheep, the measurement of body weight with a kilogram precision scale does not allow to distinguish animal weights that differ by up to 0.999... kg, as is the case of the weight of two animals with 49.6 kg and 50.4 kg; The weights of these two animals will both be recorded with rounding to 50 kg. In general, measurement procedures for continuous characteristics should be appropriate so that data are recorded with adequate numbers of significant digits. As a rule, it is recommended to record data with at least three significant digits. Thus, in the experiment with chickens, body weight should be expressed as a variable with a unit of measurement of decagrams and measured on a scale that provides precision of this order. Evaluations performed by evaluators are subject to variability due to the evaluator's ability. Variations of this origin can be reduced by training evaluators, using reference standards, and adopting more than one evaluator or more than one assessment per observation unit. The precision of the measurement process can also be increased using more than one observation unit per experimental unit, or by evaluations performed by more than one evaluator in each experimental unit, and the adoption of the mean as the response in the experimental unit.

Local control and statistical control - These experimental control procedures are the resources that can commonly be used most conveniently and most profitably for increased precision. Very often, one or a few characteristics of the experimental material are the predominant sources of extraneous variation that can considerably inflate the experimental error. For example, in a cultivar comparison experiment, these sources of variation may be soil fertility and moisture, and stand (number of plants); in an animal nutrition experiment, such sources may be the origin, age and body weight of the animal. In these situations, local control may be appropriate to control variations attributable to extraneous characteristics related to soil and to origin and age of the animal, while statistical control may be convenient to control extraneous variations due to plant stand and animal weight. The appropriate use of local control and statistical control in these circumstances allows to separate from the variation of the experimental error that affects effects of treatment factors the variation due to relevant extraneous characteristics, and to eliminate this variation from the estimates of these effects. Thus, these experimental control procedures can provide increased precision in experiments with experimental material representative of the heterogeneity present in the target population.

In general, the adoption of an experiment design that exerts effective local control may allow the control of relevant sources of extraneous variation. For example:

- If the experimental material is heterogeneous, but blocks of sufficiently homogeneous units can be formed, each of which comprising the complete collection of the treatments, this design with complete blocks allows to eliminate the relevant source of the extraneous variation of the estimates of the effects of the treatments and of the estimate of the variance of the experimental error that affects these effects.
- If the experimental material is very heterogeneous or the number of treatments is very high, it may be appropriate to use a design with incomplete blocks, with the constitution of blocks of smaller size than the number of treatments. In this case, two situations should be considered: b_1) If there is a single treatment factor and the relevant inferences refer to the comparisons of each treatment with each of the other treatments, it is convenient to adopt a balanced design that guarantees equal precision for all these comparisons. b_2) If the experiment comprises several treatment factors and high-order interactions are irrelevant, a confounding design that sacrifices inferences regarding these interactions may be appropriate to provide high precision for inferences about important effects.

The choice of experimental design can also be profitably made to achieve convenient precision in many other circumstances. For example:

- In the case of two experimental factors with different relative importance, the adoption of two formations of experimental units in which one of them constitutes subdivisions of the other can be used to assign higher precision to the inferences about the most important factor by allocating the levels of this factor to the units constituted by these subdivisions; the levels of the least important factor are assigned to the larger experimental units.
- In experiments in which treatments have different importance for the inferences of interest, higher precision for these inferences can be achieved by allocating appropriate numbers of repetitions for the treatments. This is the case, for example, when one of the treatments is the reference for the other treatments and the inferences of interest are the comparisons of the reference treatment (usually called control treatment) with each of the other treatments; the precision of these comparisons can be considerably increased by assigning a greater number of repetitions to the control treatment.

2.4. Validity

The experiment should have ability to reveal the actual effects of treatment factors and, particularly, differences between the effects of treatments in the target population. This ability is higher the greater the accuracy of the inferences derived from the sample for the target population, that is, the greater the precision and the validity of these inferences.

- \- An experiment is valid if the inferences derived from the sample to the target population are valid, i.e., unbiased.

Absolute validity requires that the initial experimental material, i.e., the experimental material on which the experiment is to be conducted, is a random sample of the target population and the experimental units are randomly associated with the levels of the treatment factors, so that the experimental error affecting the effects of these factors comprises exclusively effectively randomized extraneous characteristics. Real experiments usually do not satisfy these two properties: the experimental material does not consist of random sampling, and the experimental error also comprises potentially disturbing extraneous characteristics. For this reason, the evaluation of validity always involves subjective judgment. Thus, the property of validity of the experiment, or of inferences derived from the experiment, can only be achieved approximately.

The bias of the experiment has two origins or components: the **intrinsic bias** that comes from the experimental error, that is, from the biased confounding of effects of extraneous characteristics with effects of treatment factors, and the **extrinsic bias** that arises from the sampling error, that is, from the deviations of the sampled population from the target population. The bias of the experiment is smaller the smaller the bias originated from the experimental error and the sampling error.

Considering these two sources of bias, that is, intrinsic bias and extrinsic bias, one can distinguish, respectively, the internal validity and the external validity of the experiment. Internal validity means validity of inferences from the sample to the sampled population. The validity of the inferences of the sample to the target population also requires the external validity, that is, the validity of the extension of the inferences from the sampled population to the target population. Thus, the process of inference from the sample to the target population comprises two steps:

- Generalization of the sample to the sampled population, and
- Extension of this generalization from the sampled population to the target population. Therefore, internal validity requires the absence of bias in the first step; the external validity, the absence of bias in the second step.

2.4.1. internal validity

Internal validity refers to the unbiasedness of inferences regarding the causal effects of treatment factors for the population represented by the sample, that is, for the sampled population:

- The experiment has internal validity if the inferences about effects of treatment factors from the sample to the sampled population are unbiased.

Thus, internal validity requires the unbiasedness of estimates of the effects of treatment factors and estimates of experimental error affecting these effects.

Repetitions is necessary for the experiment to provide estimates of the experimental error. However, it is not sufficient to guarantee the unbiasedness of these estimates and of the estimates of the effects of treatment factors. For these estimates to be unbiased, the design of the experiment must ensure that the variation of the observed values of the response variable is random, except for effects of experimental factors and effects of extraneous characteristics controlled by local control and statistical control. This means that internal validity requires that the experimental material does not comprise disturbing extraneous characteristics. In these circumstances, inferences about effects of treatment factors, particularly about differences in these effects, derived for the sampled population are unbiased. If this assumption is ensured, these inferences can be derived by statistical methods.

2.4.2. Origins of intrinsic bias

Impairments to internal validity come from biased confounding of effects of extraneous characteristics with effects of treatment factors and biased estimates of the errors that affect these effects. In general, these losses have the following origins: experimental units; extraneous characteristics of the experimental material referring to: treatment vehicles, application of the treatments, environment, techniques for the operation of the units, researcher and measurement and recording of data.

The plan of the experiment should consider the potential sources of intrinsic bias and establish the appropriate actions for its control. The possible origins of this bias can be indicated by previous experiments, information provided by the literature and other sources.

2.4.3. Actions to increase internal validity

Actions to control potential sources of internal bias should be predicted and described in detail in the experiment protocol. These actions fundamentally comprise appropriate definition of experimental units, control of experimental techniques and randomization, and, in some circumstances, statistical control.

Appropriate definition of experimental units - Experimental units are the units of information about experimental error that affects effects of experimental factors. The correct definition of experimental units is essential for the unbiased estimation of experimental errors that affect the effects of treatment factors. For illustration, consider an experiment to comparing four loads of grazing animals with a paddock for each load and ten animals per paddock. Sometimes, the variation between animals within a paddock is used to estimate the experimental error for stocking comparisons. This variation, however, underestimates the experimental error to compare stockings, since the variation between animals within the same paddock is usually lower than the variation between animals in different paddocks, which also includes variation between paddocks.

Control of experimental techniques - The control of experimental techniques to increase precision can also favor internal validity. Thus, the control of experimental techniques for increasing precision should be planned to also contribute to avoiding intrinsic bias. Actions for this purpose are considered below.

- Appropriate constitution of experimental units - Internal validity requires the absence of confounding of treatment effects, which should be controlled by experimental techniques that prevent treatment effects from being contaminated by treatment effects in neighboring experimental units. For example, in experiments with plants in which the treatments are cultivars of different sizes, fertilizers, irrigation methods, insecticides, fungicides and herbicides, may be required the use of borders or spacing between the experimental units to avoid that effects of treatments in one unit are affected by effects of treatments in neighboring experimental units. In experiments to control animal diseases that can be transmitted between animals, individuals who constitute different experimental units should be kept in separate compartments (paddocks, boxes, cages, etc.) to avoid confounding of treatment effects. In certain experiments, it may be necessary to adopt protection to avoid contamination of the effect of a treatment in an experimental unit by the effects of treatments in neighboring units; for example, use of screens in pest control experiments to prevent insect migration between experimental units with different treatments.

- Control of the application of treatments - Failures in the application of treatments can lead to biased confounding of effects of treatments with effects of extraneous characteristics. Intrinsic bias of this origin can be avoided by controlling experimental techniques. For example:
  - In a lamb nutrition experiment in which diets differ only in qualitative composition, the different diets should be applied to lambs in equal quantities and in a similar way;
  - In an experiment to compare nitrogen sources, nitrogen fertilizers from different sources should have the same nitrogen content and should be applied to the experimental units uniformly with regard to the amount and form of application;
  - In an insect control experiment with insecticides, different insecticides should only differ in active principles and be uniformly applied in terms of quantity and mode of application.
- Control of environmental characteristics - Effects of environmental characteristics may become relevant and be tendentiously confounded with effects of treatment factors. For example, variations in wind intensities, temperature and precipitation, incidences of insects, fungal diseases, invasives and predators in agricultural field experiments; diseases and worms in animal experiments. Intrinsic bias of this origin should be avoided or reduced by control of experimental techniques. This control can be exercised efficiently in experiments conducted in artificial environments, such as greenhouses, laboratories, and protected facilities. In experiments conducted in natural environments, such as field experiments with plants and animals, control is usually limited to preventive measures so that characteristics of the environment that are controllable manifest uniformly; for example: application of insecticides, fungicides and herbicides to control the incidence of insects, diseases and invaders, and protection against predators in experiments with plants; application of prophylactic medications, such as vaccines and dewormers, in animal experiments.
- Control of the execution of techniques necessary for the operation of the units - Variations or failures of these techniques can have two implications for experimental error: to introduce variations of a random nature or of a systematic nature. In the first case, they contribute only to the increase of the estimate of the experimental error, which impairs the precision; in the second, for the bias of this estimate. For example, if in an experiment with plants cultivation techniques, such as planting, weeding and harvesting, extends over a considerable time interval, their execution treatment by treatment may imply biased confounding of effects of environmental characteristics with effects of experimental factors.
- Use of experimental techniques for the purpose of controlling the functioning of the units - In some circumstances, it may be appropriate to employ unusual techniques in the target population in the sample for the purpose of preventing potentially disturbing extraneous characteristics from becoming disturbing. For example: use of border on the plot to avoid contamination of effects of different treatments in neighboring plots and reduction of plants that emerge to the same number per plot, in field agricultural experiments; replacement of animals affected by predators and diseases to control stocking in compartments with different treatments, in animal experiments.
- Control of effects of the units and the researcher - Units may be affected by effects arising from preference or rejection to treatments that are not inherent effects of treatments. This can occur in human experiments. Bias arising from the effects of the units can be avoided by omitting to the units information about the treatments they receive. The researcher effect is also a potential source of intrinsic bias. This bias can arise when there is some conscious or unconscious interference that causes benefit or harm to some treatments. In general, special care allows to avoid failures of this origin that imply intrinsic bias. In extreme situations, it may be convenient to omit information to the researcher about the treatments.
- Use of appropriate measurement procedures and instruments - Measurement procedures and instruments and data recording should be free of bias. The response characteristics should be measured according to the respective definitions, which should be established in the protocol of the experiment. Thus, for example, in an experiment with fruit plants in which plants with different treatments may have maturations on different dates, measurements of the weight of production and other characteristics of the fruit should be carried out treatment by treatment, or unit of observation per unit of observation, to the extent that the fruits reach the defined state of maturation. Characteristics that must be measured on the same date may be sensitive to the instant of

measurement or to the state of the unit of observation at that instant; for example, grain characteristics such as production weight and humidity can be considerably affected by the humidity of the environment, particularly when rain occurs; milk characteristics may be affected by the time of milk collection and also by changes in the animal due to stress. In these cases, it is recommended that the measurement of all units be carried out in a sufficiently short time interval so that effects of the measurement instant or the state of the unit do not result in biased confounding with treatment effects. Characteristics should be measured under comparable conditions; for example, if seed characteristics are determined from samples taken from the observation units, the sampling procedure should be the same for all units; if fruit characteristics are measured in samples of the fruits produced individually by the plants, these samples must include fruit harvested from the same positions.

- Measurement procedures should be as objective as possible, measurement instruments should be calibrated periodically, data should be recorded carefully, and transcription of data should be avoided. Measurements made by evaluators should be carried out with the necessary care to avoid bias, particularly bias that may result from subjectivity. Resources for this purpose are, for example, training of evaluators, use of reference standards, use of more than one evaluator per unit and omission of information to the evaluator regarding the treatment being evaluated in each unit. On the other hand, the potential bias arising from data recording can be avoided with special care and, particularly, with procedures that avoid data transcription.

Statistical control - Environmental characteristics can become relevant so that its effects tendentiously affect the effects of treatment factors. If the units affected by disturbing extraneous characteristics are few and identifiable, the resulting bias can be controlled by statistical control. This procedure consists of the omission of observations in these units and the use of statistical analysis procedures that make the appropriate adjustment of the observed values of the response variable. Thus, the recording of relevant occurrences during the execution of the experiment makes it possible to take them into account in the inferences derived from the experiment.

Randomization - Randomization is an efficient resource for controlling the bias that may result from the confounding of effects of extraneous characteristics with effects of treatment factors. Randomization should be used in the assignment of experimental units to treatments and in the execution of experimental techniques that may imply biased confounding with treatment effects. In this second situation, randomization is used to determine the order of execution of experimental techniques in experimental units, when the order may imply bias. For example, randomization can avoid bias arising from the order of execution of plant cultivation techniques, such as planting, weeding, and harvesting, and from the order of animal management techniques, such as shear, weighing, when these operations extend over a considerable time interval.

2.4.4. External validity

- The experiment has **external validity** if the extension of the inferences from the sampled population to the target population is unbiased.

The inferences derived from the experiment are valid for the sampled population, that is, the population from which the experimental material is considered a representative sample. In general, restrictions on the choice of experimental material imply that this population differs from the target population. Thus, the extension of the inferences to the target population involves additional uncertainty to that resulting from the experimental error.

External validity is crucial for technological experiments. It may not be as relevant to basic experiments and early experiments of a research program. In general, these experiments have as their purpose the research of basic questions to be further considered in experiments conducted in natural environments, before recommending results for practical conditions.

It should be noted that in comparative experiments the representativeness of the sample is essential regarding its implications on differences in treatment effects; it is usually not essential with respect to the individual effects of treatments.

2.4.5. Origins of extrinsic bias

Impairment to external validity stems from bias resulting from failures in the representation of the target population by the experimental material. External bias can originate in the planning and execution of the experiment. Sources of internal bias can also imply external bias. Losses to external validity have the following origins: response variables, experimental factors and their levels, experimental units, extraneous characteristics of the experimental material relating to initial sample, treatment vehicles, application of treatments, environment, techniques for operation of the units and measurement and recording of data.

As well as the sources of imprecision and intrinsic bias, the possible sources of extrinsic bias must be identified and considered in the planning of the experiment so that the appropriate actions for its control are predicted and specified in the protocol of the experiment. Indications of potential sources of extrinsic bias can be obtained from similar experiments and information in the literature.

2.4.6. Actions to increase external validity

Actions to increase external validity can favor or impair internal validity and precision. Thus, when planning and executing these actions, the researcher must take the necessary care to achieve the appropriate balance regarding the impacts that actions proper to increase external validity may have for precision and internal validity.

Several resources can be used to increase the external validity of the experiment, namely:

Appropriate definition of response variables - The representation of a property of units by a variable can be misleading. To avoid bias of this origin, the response variables must express the relevant properties of the response characteristics they represent. For example, in a peach experiment, the weight of the fruits of the lower branches of the plant may not be an appropriate variable to express the weight of fruit production; in an experiment with corn, the average height of the five most vigorous plants may not be a valid variable to express the height of the plants in the plot.

The scales of measure must be appropriate for the correct representation of the properties of the characteristic that the variables express. In some situations, it is more appropriate to represent the levels of a characteristic by ratios rather than absolute numbers. For example, the amount of sugar in grape is more appropriately expressed in proportion or percentage than by weight, since this depends on the amount of grape produced; For similar reason, when the number of animals in a compartment (cage, for example) is variable, the amount of food consumed should be expressed relative to an animal. The levels of the variables must appropriately represent the levels of the characteristics they represent and particularly the relationships between these levels. The researcher should take precautions to avoid biases that may arise from the choices of the levels of response variables, particularly when variables are measured subjectively, such as variables that express organoleptic properties of beverages and foods, and when categorical variables express continuous characteristics, such as degree of infection of a disease with four levels: no infection, weak infection, medium and high infection.

Appropriate choice of experimental factors and their levels. The definitions of the experimental factors for the sample must be the same as those established for the target population and each factor level chosen for the sample must be the same as that which corresponds to it in the target population. These requirements call for clear and precise definitions of the experimental factors and their respective levels under consideration in the target population. For example, in an experiment on the effect of diet supplementation on body development and meat production of sheep, supplementation of a fixed daily amount or supplementation at will may be considered. These two forms of supplementation are distinct and their effects on the animal's response may be quite different. The decision between these two alternative forms of supplementation must be made in line with the objectives of the experiment.

In some situations, the representation of the target population may require the extension of the sample with the consideration of a relevant characteristic as an experimental factor not directly related to the objectives of the experiment. This is the case when there is an expectation that the effects of relevant treatment factors may depend on some characteristic that varies in the units of the target population. For example, in an experiment on the control of a wheat fungal disease in which fungicide is the relevant treatment factor, the effects of fungicides may depend on the cultivar; in this case, it may be convenient to adopt two or more cultivars that represent the variation of susceptibility of the cultivars in the target population to this disease and consider cultivar as an additional treatment factor. In an experiment about feed supplementation of lambs in which the effects of supplementation may depend on sex and breed it may be convenient to constitute the sample by male and female animals of the breeds present in the target population; then, sex and race become additional experimental factors. In agricultural experiments for research of cultivation techniques, such as soil fertilization and control of diseases and pests, it is generally convenient to repeat in several places in the region of interest and in several years; in these experiments of wide scope, local and year should be considered as additional experimental factors.

Choice of an initial sample representative of the target population - This is usually a most critical condition for the external validity of the experiment. In experimental research, very commonly, the sample does not consist of units chosen from the target population, but of units built to simulate these units. In these circumstances, the external validity is highly dependent on the degree of similarity achieved between these constructed units and units of the target population, especially regarding their initial compositions. Thus, in field experiments with plants, the plots should be constructed in such a way as to represent real crops; particularly, plots should be constructed with appropriate size and shape to allow similar cultivation conditions as the real crops. Similarly, if an experiment on the effect of storage time on soybean seed quality is carried out in a facility specially built for the research, the environmental conditions of these facilities must be like those in the target population. Particularly, it is highly important that the sample represents the variability present in the target population. Thus, if the target population has a broad spatial and temporal scope, as is the case of experiments on plant cultivation techniques or animal management, it is very often necessary that the sample comprises several locations and time intervals of similar coverage to those in the target population. In some situations, it may be convenient for the experiment to be carried out in the target population's own units, as areas and facilities of properties in these two examples. Restrictions on the choice of locations and time intervals can have serious implications for the external validity of technological experiments.

It should be noted that in comparative experiments the representativeness of sample is essential regarding its implications for differences in treatment effects; it may not be as essential with respect to the individual effects of treatments. It is also relevant to observe that in many experiments the target population does not consists of units with characteristics like those of the units present at the time the experiment is carried out, but units that will exists in the future or modified and evolved units. In these cases, the initial sample should represent this ideal target population.

Control of experimental techniques - Control of experimental techniques to increase internal validity can also contribute to external validity. However, control for the increase of precision may impair the external validity. Therefore, it is important that the control of experimental techniques that is planned for the benefit of one of these requirements also favor the other requirements, or at least do not harm them. The actions for increasing external validity are considered below.

- Appropriate dimensions and constitution of the experimental units - The components of the units should manifest in the sample behavior like that of the target population. In experiments with animals and plants, for example, the dimensions and composition of the units must be appropriate so that the animals and plants are in conditions related to the density of individuals, environment and management that allow them the usual behavior in the facilities and crops considered in the target population.
- Control of the extraneous characteristics conveyed with treatments - Treatments should manifest in the sample according to their definitions. Compliance with this requirement is usually unfeasible, since treatments are commonly conveyed with extraneous characteristics. However, in comparative experiments it is usually essential only the homogeneity of these extraneous characteristics between treatments. The researcher must exercise appropriate control to achieve this homogeneity. For example, if the treatments are cultivars, the seeds must be homogeneous in vigor, sanity and purity; if the treatments are diets, the diets must be formulated according to their definitions and be homogeneous with respect to extraneous characteristics regarding to quality, quantity and form of supply.
- Control of the application of the treatments - Even under the assumption of homogeneity of the extraneous characteristics conveyed with the treatments, in some circumstances it may be difficult or unfeasible to apply the treatments according to their definitions. For example, in an experiment of the effect of soybean seed moisture on its physiological quality after a storage period with moisture levels of 10%, 12%, 14% and 16%, the levels achieved in the sample are usually not exactly these; the nominal level of 10% can be implemented as 9.8 or 10.2%, for example. If the researcher is aware that this is the treatment applied and takes it appropriately into account in statistical analysis procedures, this bias is corrected and has no implications for inferences.
- Control of the environmental characteristics - Extraneous characteristics of the environment may manifest in the sample differently from the way they manifest in the target population. In these circumstances, these characteristics are sources of extrinsic bias. For example, occurrences of hail and wind, temperature and precipitation outside the limits considered in the target population, incidences of insects, fungal diseases, invasive and predatory in field agricultural experiments, and of diseases and worms in animal experiments. In some circumstances, extrinsic bias of this origin can be avoided or reduced by control of experimental techniques adopted as a preventive measure; for example, use of insecticides, fungicides, herbicides, and protection from predators in the first example, and drugs and dewormers in the second. However, in experiments in natural environments, this control is limited; for example, climate-related characteristics are generally not controllable.
- Control of the execution of techniques necessary for the operation of the units - These techniques must be carried out on the sample in the same way as on the target population. Failed techniques can lead to extrinsic bias. Particularly, plant cultivation techniques (planting, application of insecticides, fungicides and herbicides, and harvesting, for example) and animal management techniques (provision of food and water, application of

medicines, etc.) must be carried out in the way they are employed in the crops and breeding units considered in the target population.

- Use of purposeful experimental techniques to control the functioning of the units - The use of unusual techniques in the target population can provide a better representation of the functioning of the sampling units. For example: use of border on the plot in experiments with plants to provide that the plants manifest in the experimental units behavior like that of the crops; production of artificial infection of animals in an experiment on the control of an infection to simulate the behavior of infected animals in the target population; insect infestation in an experiment on the resistance of cultivars to a pest for a similar purpose.

2.4.7. Relationship between internal validity, external validity and precision

The validity of the inferences derived from the sample to the target population requires internal validity and external validity. The previous discussion pointed out that meeting these two requirements is sometimes conflicting.

Internal validity is necessary for inferences about causal relationships that constitute the objective of the experiment. For this reason, it is an essential minimum requirement of the experiment. It has implications for the answer to the question: do treatment factors have effects on the response variables in the specific conditions represented by the sample? External validity is related to the answer to this question for the units object of the inferences: do these effects extend to the target population? While internal validity is the essential criterion and the issue of extension to the target population, like the issue of inductive inference, is never completely resolved, it is highly desirable that the experiment meets these two requirements.

The relative importance of internal validity and external validity depends on the objectives of the experiment and the state of knowledge. They are distinct for basic experiment and technological experiment, and for initial phase of research programs and when knowledge has been accumulated. In the basic experiment the internal validity assumes high importance, because the objective is the verification of the existence or not of the causal relationship formulated by the research hypothesis; external validity is often of less interest. In the technological experiment, external validity tends to be more relevant because of the importance attributed to the application of the knowledge originated from the research to real contexts. For example, a preliminary trial of new drugs to control a disease in humans can be done in an experiment with guinea pigs. However, at a more advanced stage of the research, this test should be performed with humans representative of the population of individuals affected by this disease. Similarly, in the initial phase of a research program the identification of causal effects of treatment differences is of high importance. In this phase, the search for possibly effective treatments tolerates experimental plans and designs that do not meet some principles and requirements and allows many false statements of treatment efficiencies so as not to neglect the identification of possibly effective treatments. As new knowledge accumulates, external validity assumes greater importance and, in the last phase of the research, external validity is a priority, mainly to characterize the performance of treatments in the real situations.

In fact, internal validity and external validity are not so incompatible. This incompatibility, more apparent than real, can be resolved by the balanced use of the control of experimental techniques and by randomization. In particular, the researcher must plan each of the actions of control of experimental techniques to achieve the appropriate balance between the consequences for internal and external validity.

Some actions that benefit internal validity or external validity can adversely affect precision. Usually, however, this negative impact can be avoided with the balanced use of control of experimental techniques and using local control and statistical control. In certain technological experiments in which the representation of the variability of the units of the target population in space and time requires that the sample be constituted by units disposed in different places and periods of a time interval, the variability of the units of that origin can be controlled by proper consideration of the structure of the units.

2.5. Simplicity, economy of resources and feasibility

The objectives of the experiment must be achieved with maximum simplicity and minimum cost. First, the experiment plan must be feasible with the financial, human and material resources available. This is a necessary condition to ensure that the plan can be executed and completed. Efforts and resources have been wasted by alteration, suspension or non-obedience of experiment plans due to lack or unavailability of resources at times when they are needed. Modifications of plans very often imply the non-fulfillment of important requirements with consequent losses for the achievement of the objectives of the research.

In some situations, there are practical constraints that must be considered in the planning of the experiment. For example, if the execution of the experiment is going to be done by people with little skill and experience it may be difficult to achieve the understanding of complex tasks, even if they are described in detail in the protocol and in instructions that are provided; if the experiment is conducted on private properties, such as an agricultural enterprise or an industry, it may be important to interfere as little as possible in the production process. These restrictions may lead to the choice of alternatives that do not meet important requirements. The adoption of these alternatives should be well considered, considering the relative importance of practical conveniences and desirable properties of the experiment. It can only be admitted if the minimum properties relevant to the inferences are satisfied.

Feasibility is not a justification for flaws that prevent or invalidate the inferences that constitute the object of the experiment. For illustration, consider an experiment to compare four stockings of grazing animals and suppose that only four paddocks are available so that each stocking should be assigned to only one paddock. With this restriction, the experiment will not provide the estimation of the appropriate experimental error for inferences about stocking effects. Therefore, it will not provide valid inferences regarding these effects.

In long-term experiments, it is convenient the periodic evaluations of the progress of the research for decision regarding its continuation, conclusion or suspension. In some situations, before the deadline established for completion, experiments reveal results that meet the objectives, or that highlight the inconvenience of continuing the research effort, possibly because of some event that caused irreversible damage, or that recommend changes regarding experimental factors or their levels. Thus, periodic evaluations can avoid the waste of resources and time resulting from the continuation of experiments that will not lead to useful new information.

Simplicity is a property related to feasibility, and economy of resources. The simplicity of the research methods and procedures has implications for the optimization of the use of resources and the practical facilities for conducting the experiment. In particular, the structure of the experimental conditions, the structure of the units and the structure of the experiment should be as simple as possible, if they are appropriate for the achievement of the objectives of the experiment. The importance of this simplicity is most salient in experiments of broad scope, which must be conducted in several locations and for several years with the collaboration of teams with heterogeneous training. In these experiments, the simplicity of methods and procedures facilitates their understanding, can contribute to the better quality of the execution and avoid bias resulting from errors in the implementation of actions defined in the plan of the experiment. Greater contribution to avoid failures of this origin can be supplied by training of collaborators and written, detailed and clear instructions.

These considerations also apply to the methods of analysis of the results. Fortunately, the requirements of efficiency of experiment design and simplicity of analysis methods are generally highly correlated. In general, for the usually more efficient experimental designs, appropriate statistical analysis procedures are available, provided that certain assumptions are satisfied. The electronic computing resources that are provided by computers and statistical analysis packages enables the use of the most appropriate analysis methods for each experiment.

2.6. Manifestation of the real effects of treatments

An important property of the experiment is the ability for real effects of treatments to manifest in the sample. This property includes requirements related to the structure of the experimental conditions and compliance with the requirements of precision, internal validity and external validity. The relevance of these requirements justifies the reiteration of some previous considerations.

- Treatments should be distinguished by simple and identifiable attributes. Failure to meet this condition may imply that inferences reveal differences in the effects of treatments that have two or more interpretations. For example, consider an experiment on the effect of nitrogen fertilization on top dressing in irrigated rice crops whose treatments are four doses of nitrogen, preceded by basic fertilization to soil preparation common to all experimental units, which includes nitrogen. In these circumstances, if the inferences do not reveal differences in treatment effects, the researcher will have no way of knowing whether this result originated from the lack of effect of nitrogen applied in top dressing or from the nitrogen needed to have been supplied by basic fertilization. Generally, ambiguities of this origin can be avoided by including additional treatments or establishing a factorial structure for the treatments. Thus, in the example, the dubiety of interpretation can be avoided by considering a treatment structure consisting of combinations of the levels of two factors: factor 1 - fertilization in soil preparation with two levels: without and with nitrogen fertilization, and factor 2 - fertilization in cover with the levels constituted by the doses originally defined.

- b) Treatments should be implemented in the sample according to the definitions established in the plan of the experiment. In comparative experiments, the fundamental condition for external validity is the agreement between the differences of treatments implemented in the sample and the differences between the treatments defined in the plan of the experiment. This is a necessary condition for the external validity of inferences regarding treatment comparisons. Its fulfillment requires homogeneity of the extraneous characteristics conveyed with the treatments, which can be achieved with the use of careful control of experimental techniques.
- c) Effects of treatments, if any, should manifest in the sample. The most elaborate experiment may prove to be ineffective if the sample does not provide the manifestation of the real effects of the treatments. For example, in an experiment comparing the effectiveness of insecticides in controlling a pest, the differences in the effects of insecticides can only be manifested if there is the presence of insects in the experimental area. Similarly, in an experiment on the effectiveness of drugs for the control of a disease in animals, treatments cannot reveal differences if they are applied to animals that do not manifest this disease. In these circumstances it is necessary to use additional control treatments (absence of insecticide and antibiotic, in the examples) that inform about the presence in the experimental material of the conditions necessary for the treatments to manifest their effects.
- Effects of treatments require time for their manifestation and are usually not constant in this time. In addition, the variation of these effects over time may differ between treatments. The amplitude of the experimental period should be determined so that treatments have time to manifest their relevant effects and reveal the variation of these effects. In some experiments the variation of the effects of treatments over time is inherent to the objectives of the research. For example, in a lamb nutrition experiment, the effects of diets on the weight development of the animals in a time interval may be important; in an experiment on the effect of the seed storage temperature of a soybean cultivar, inferences about differences in temperature effects and the variation of these differences over time may be relevant.
- d) The effects of treatments must be free from confounding with effects of other treatments. Effect of treatment in one unit must not be affected by the effects of treatments in other units. This means that there can be no interference from the effects of treatments. In some situations, treatment effects have a spatial or temporal amplitude that may imply that the effect of a treatment in one unit is affected by treatments assigned to other units. For example, confounding of treatment effects may occur a) in experiments with plants related to pest, disease and invasive control, and irrigation and fertilization; and b) in animal experiments on disease and parasite control. In these cases, interference between the units may result from the contamination of treatment effects, that is, from the passage of insecticides, fungicides, herbicides, water, fertilizers and anthelmintics, or from the displacement of insects, fungi, invasive plants, parasites, water and fertilizer between the units. Confounding of effects of treatments of this origin must be controlled by experimental techniques, with the appropriate isolation of experimental units. For example, by means of bordering, spacing or protection, in field experiments with plants and animals. In some situations, it may be convenient to check for transmission of diseases, pests and parasites. Intercalation of blank plots, i.e., without insecticide, fungicide, herbicide or anthelmintic, for example, can be useful for this purpose.
- In some experiments the same individuals are used as units at successive intervals of the experimental period, receiving different treatments. This structure of units performs local control of the individual characteristics of the animals. However, it may imply confounding of effects of treatments applied to the same individual. This confounding can be avoided by spacing between the intervals of application of the treatments and by appropriate procedures for measuring response characteristics. Suppose, for example, a dairy cow nutrition experiment with four different diets in which each of the animals receives the four diets in the corresponding orders defined at four successive two-week intervals and milk production is measured by the average production in the last two or three days of these intervals. A basic assumption for the validity of the inferences of this experiment is that milk production at the end of each two-week interval is not affected by the effect of the diet that the animal received in the previous interval. If this assumption is false, inferences about differences in treatment effects will be biased. The possibility of interference of treatment effects between intervals can be diminished by control of appropriate experimental techniques, such as intercalation of intervals of sufficient amplitude in which all animals are submitted to a common procedure instead of treatments.
- This interference of treatment effects can also be considered by the formulating an experiment design that takes in account the residual effects of treatments. For example, by an experiment design in which each of the treatments succeeds each of the other treatments the same number of times. This design provides the balance of the residual effects of the treatments so that these effects affect all treatments equally.
- e) Treatment effects must manifest in the sample without confounding with effects of relevant extraneous characteristics. This is a requirement for internal validity and external validity. The experiment plan and execution must ensure that all relevant extraneous characteristics are controlled by experimental techniques, local control or statistical control, or are randomized. In particular, the presence or introduction of any source

of extraneous variation during the execution of the experiment that may constitute systematic variation must be avoided. Thus, any experimental technique or interference that may constitute a source of relevant extraneous variation must be avoided or subject to experimental control. Sources of variation beyond the control of the experimenter must be recorded for appropriate consideration in the analysis and interpretation phase of the results.

2.7. Prediction of statistical inference procedures and provision of measures of the uncertainties of these inferences

The experiment plan must provide for the statistical inference procedures to be adopted in the data analysis phase. Statistical inferences are inductive inferences; therefore, they are uncertain. An important property of statistical methods is that the uncertainty of the inferences they provide can be known or established by the researcher.

Degrees of uncertainty of inferences are expressed by estimates of standard errors that affect estimates of effects of treatment factors, particularly estimates of treatment means and differences of these means, or by limits of decision errors regarding these effects for probabilities fixed by the researcher.

Effects of treatment factors on values of a response variable in the units of the sampled population can be constant for all units or realizations of a random variable with a certain probability distribution, respectively if the effect originates only from fixed factors or from fixed and random factors. In any of these situations the effects on the population are unknown. In the first case, inferences of interest refer to the magnitude of population effects; in the second, the variability of these effects.

The reason for conducting the experiment is to infer about the actual existence and about the magnitude or variability of each relevant effect. As has been pointed out, objective inferences derived by statistical methods validly apply to the sampled population, not necessarily to the target population. The extension of these inferences from the sampled population to the target population requires subjective judgment.

The processes of statistical inference are basically of two types: estimation and hypothesis testing. The estimation process consists of determining an approximation or estimate of the magnitude or variability of the effect, based on the observed values of the response variable. Because an experimental factor effect is invariably confounded with effects of extraneous characteristics, an estimate usually differs from the actual value of the effect. Standard error estimates are measures of the uncertainty of inferences by the estimation process. An alternative estimation process consists of determining an interval, called a confidence interval, that contains the effect with a conveniently fixed probability. In this case, measures of uncertainties of inferences are the probabilities fixed a priori by the researcher.

Hypothesis testing is a decision process between two alternative hypotheses about the effect: the effect does not exist, or the effect exists. A statistical hypothesis comprises these two hypotheses, which are usually called, respectively, null hypothesis and alternative hypothesis and denoted by H_0 and H_A:

$$\begin{cases} H_{0:} \text{ the effect does not exist} \\ H_A\text{: the effect exists} \end{cases}$$

As the effect manifests confounded with effects of extraneous characteristics, the decision may be concordant or discordant with the actual situation (which is unknown), which means that it may be correct or incorrect. This decision-making process leads to one of four possibilities, two of which correspond to right and the other two to error. This decision process is illustrated in Table 1.

Table 1 Alternative situations that can occur in the process of testing a hypothesis.

Real situation	**Decision**	
	H_0: Effect does not exist	**H_A: Effect exists**
Effect does not exist	Correct	Incorrect Type 1 Error
Effect exists	Incorrect Type 2 Error	Correct

The error that corresponds to the incorrect decision of declaring that the effect exists when it does not exist is called a type 1 error. The other decision error, corresponding to declaring that the effect does not exist when it exists is called a type 2 error.

In this decision process the researcher must establish the probabilities of these two types of error, according to their relative importance. Naturally, he would like to assign very small values to the two errors. It happens, however, that when the probability of one of these errors is reduced, the other is increased. Very often, the researcher has the expectation of the presence of real effect of a treatment factor and runs the experiment to prove this expectation objectively. In these circumstances, the researcher wants to assign a high probability of declaring the existence of the effect if that effect exists. This probability is called test power. On the other hand, it is usually more difficult to control type 2 error than type 1 error. So, very often, the researcher fixes only the probability of the type 1 error at a conveniently small value, but not too small so that the probability of the type 2 error does not result inconveniently high. This probability fixed for the type 1 error is called the significance level of the test.

The significance level is a measure of the uncertainty of the inference generated by the hypothesis testing process, which is fixed by the researcher.

The validity of these uncertainty measures requires the random origin and unbiasedness of the estimates of the standard errors that affect the effects of treatment factors and, consequently, a valid estimate of the experimental error.

In comparative experiments commonly important hypotheses refer to differences in effects or averages of treatments. For example, in the case of two treatments with population means μ_1 and μ_2, if there is no a priori reason that if $\mu_1 \neq \mu_2$ then $\mu_1 > \mu_2$ (or $\mu_1 < \mu_2$), the statistical hypothesis is expressed by:

$$\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_A: \mu_1 - \mu_2 \neq 0 \end{cases} \; or \; \begin{cases} H_{0:}\ \mu_1 = \mu_2 \\ H_A: \mu_1 \neq \mu_2 \end{cases}$$

Hypothesis H_0 will be rejected if the difference between the estimates of the means μ_1 and μ_2 provided by the sample is sufficiently large, whatever its signal, to be attributed only to the experimental error. This hypothesis is called **bilateral hypothesis**. In some situations, a priori reason states that if $\mu_1 \neq \mu_2$ then $\mu_1 > \mu_2$ (or $\mu_1 < \mu_2$). In this circumstance, the hypothesis is called **unilateral hypothesis,** and the alternative hypothesis has one of the following two expression:

H_A: $\mu_1 - \mu_2 > 0$ or H_A: $\mu_1 > \mu_2$, if μ_1 cannot be less than μ_2,

or

H_A: $\mu_1 - \mu_2 < 0$ or H_A: $\mu_1 < \mu_2$, if μ_1 cannot be greater than μ_2.

Then, the null hypothesis is rejected only if the sign of the difference between the estimates of the means of the two treatments is the same established by the alternative hypothesis for the difference of the corresponding population means.

For example, hypotheses regarding population means of the weight of grain production of cultivars are usually bilateral, since, in general, there is no reason to establish a priori that, if the means of two cultivars are different, one of them (known a priori) is higher than the other. On the other hand, a hypothesis referring to the difference between the mean of an insecticide, fungicide or antibiotic and the mean of a control treatment corresponding to the absence of pesticide is usually unilateral, since very often there is a priori reason to establish that the pesticide can have no other effect than the control of the insect, fungus or disease agent. Bilateral hypotheses are more common than unilateral hypotheses.

3. Conclusion

The experiment achieves its objective if the effects of treatment factors manifest themselves in the sample as they manifest in the target population and has the capacity to detect these effects efficiently. This attempt is achieved if the experiment plan and its execution satisfy the experiment requirements: estimation of errors affecting effects of treatment factors, precision, validity, simplicity, manifestation of the effects of treatments and prediction of statistical inference procedures.

References

[1] Fisher RA. The arrangement of field experiments. Journal of the Ministry of Agriculture of Great Britain, v.33, p.503-513, 1926.

[2] Fisher RA. The design of experiments. Edinburgh: Oliver and Boyd, 1935. 248p.

[3] Yates F. The principles of orthogonality and confounding in replicated experiments. Journal of Agricultural Science XXIII, part I, 108-145, 1933.

[4] Yates F. Complex experiments (with discussion). Supplement, Journal of the Royal Statistical Society 2, 181-247, 1936.

[5] Yates F. The design and analysis of factorial experiments. Harpenden, England: Imperial Bureau of Soil Sciences. Technical Communication No. TI, 1937.

[6] Silva JGC. Experimental research. World Journal of Advanced Research and Reviews. 16(03), 239–256, 2022. https://wjarr.com/sites/default/files/WJARR-2022-1152.pdf.

[7] Silva JGC. Experiment: Conceptual basis. Delhi/London: Journal of Experimental Agriculture International. 42(6): 7-22, 2020. https://doi.org/10.9734/jeai/2020/v42i630530.

[8] Silva JGC. Experiment: Methodological basis. Raipur, India: International Journal of Science and Research. 11(3): 1085-1097, 2022. https://www.ijsr.net/get_abstract.php?paper_ id=SR22316090857.

[9] Silva JGC. Principles of the experimental design. World Journal of Advanced Research and Reviews. 22(01), 470–486, 2024. https://wjarr.com/sites/default/files/WJARR-2024-1094.pdf.

[10] Bailey RA. Design of comparative experiments. Cambridge, UK: Cambridge University Press; 2008.

[11] Cochran WG, Cox GM. Experiment design. 4.ed. New York: John Wiley, 1957. 481p.

[12] Fisher RA, Yates F. Statistical tables for biological, agricultural and medical research, 6th ed. Edinburgh: Oliver and Boyd, 1963. 155p.

[13] Kuehl RO. Design of experiments: Statistical principles of research design and analysis. 4.ed. Pacific Grove, California: Druxbury, 1994. 666p.

[14] Cox D R. Planning of experiments. New York: John Wiley, 1958. 308p.

[15] Christensen LB Experimental methodology. 7.ed. Boston: Allyn and Bacon, 1997. 590p.

[16] Federer WT. Statistics and society, data collection and interpretation. New York: Marcel Dekker, 1973. 399p.

[17] Giesbrecht FG, Gumpertz ML Planning, construction, and statistical analysis of comparative experiments. Hoboken, New : John Wiley, 2004. 693p.

[18] Gould JE. Concise handbook of experimental methods for the behavioral and biological sciences. Boca Raton, Florida: CRC. 2002. 430p.

[19] Hinkelmann K, Kempthorne, O. Design and analysis of experiments. New York: John Wiley, 1994. v.1, 495p.

[20] Kempthorne O. Why randomize? Journal of Statistical Planning and Inference, v.1, p.1-25, 1977.

[21] Kempthorne O. The design and analysis of experiments. Huntington, NY: Robert E. Krieger Publishing Company, 1979. 631 p.

[22] Kempthorne O. Sampling inference, experimental inference and observation inference. Sankhya, 40, Serie B, Parts 3 and 4, p.115 145, 1979.

[23] Little TM, Hills FJ. Agricultural experimentation: Design and analysis. New York: John Wiley, 1978. 350p.

[24] Mead R. The design of experiments: Statistical principles for practical application. Cambridge: Cambridge University, 1988. 620p.

[25] Mitchell M, Jolley J. Research design explained. 3. ed. Fort Worth, Texas: Harcourt Brace, 1996. 624p.

[26] Pearce, SC. The agricultural field experiment: A statistical examination of theory and practice. Chichester: John Wiley, 1983. 335p.

[27] Pearce SC Field experimentation with fruit trees and other perennial plants. Commonwealth Agricultural Bureau, Farnham Royal, Bucks, England. Technical Communication, 23. 1953. 131p.

[28] Petersen RG . Design and analysis of experiments. New York: Marcel Dekker, 1985. 429p.

[29] Selwyn MR Principles of experimental design for the life sciences. Boca Raton, Florida: CRC, 1996. 160p.

[30] Shadish WR, Cook TD, Campbell DT. Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin, 2002. 623p.