Journal of Pharmacognosy and Phytochemistry

Available online at www.phytojournal.com

E-ISSN: 2278-4136
P-ISSN: 2349-8234
JPP 2018; SP3: 290-296

Dr. Alok Shrivastava
College of agriculture, Navsari agricultural University, Campus Bharuch, Gujarat, India

Dr. V Singh
College of agriculture, Navsari agricultural University, Campus Bharuch, Gujarat, India

Dr. KG Patel
College of agriculture, Navsari agricultural University, Campus Bharuch, Gujarat, India

Dr. Deepa Hiremath
College of agriculture, Navsari agricultural University, Campus Bharuch, Gujarat, India

Dr. Dinisha Abhishek
College of agriculture, Navsari agricultural University, Campus Bharuch, Gujarat, India

Correspondence
Dr. Alok Shrivastava
College of agriculture, Navsari agricultural University, Campus Bharuch, Gujarat, India

National conference on "Conservation, Cultivation and Utilization of medicinal and Aromatic plants"
(College of Horticulture, Mudigere Karnataka, 2018)

Common errors while applying design of experimentation in special reference to aromatic and medicinal plant

Dr. Alok Shrivastava, Dr. V Singh, Dr. KG Patel, Dr. Deepa Hiremath and Dr. Dinisha Abhishek

Abstract
Statistical science is concerned with the aspect of theory of design of experiments and sample investigation and drawing valid inferences from using various statistical methods. The statisticians design the experiments, trials and analyze the data and interpret the facts. The statisticians design the experiments, trials and analyze the data and interpret the facts. Statistical design and technique helps to describe the involvement of complex phenomena and behaviour of agricultural growth. The impact of associated factor can be analyzed with the help of simple statistical design, sampling techniques with inferential statistics. The techniques of drawing valid interpretation depend on how the data has been gathered and also depending upon the research objective. This paper describes the basic concept of statistical research design of design of experiment and techniques used for analysis and interpretation of investigations and also discuss the experimental.

Keywords: aromatic and medicinal plant, Statistical science, phenomena

Introduction
Agricultural research and interpretation is based on a collection of statistical tools used to elucidate the associations with the outcome. A deeper understanding of this science is that of discovering causal relationships. Statistical technique of a research study means, planning the study in scientific manner so that the objectives of the study are fulfilled to facilitate meaningful interpretations of the data collected during the research (Singh and Masaku, 2012) [12].

Area and types of research: Generally, there are two types of research investigations, experimental and non-experimental/observational study. The experimental study, involves a planned interference with the natural course of events so that it can be observed. In the observational study, the investigator is more passive observer interfering as little as possible with the phenomena and wishes to record. There are different types of investigation in the agricultural research. Agricultural research investigations can be broadly classified in to four sections.
1. Planning of research proposals.
2. Execution of the investigation.
3. Appropriate analysis of data or outputs.
4. Meaningful interpretations of the results

Basic steps in the proposals of research
1. Definition of research problems,
2. Formulation of objectives and hypothesis.
3. Methodology of research for the particular problems.
4. Selection of variables for the research study.
5. Coverage of all possible subject matters associated with research objectives.
6. Well defined tools and techniques for data analysis.
7. Well defined study population, sample, control,
8. sample size and time coverage.

9. Formulation of analytical methods for the data and
10. Planning for resources.
11. Anticipation and estimation of possible errors and
12. Evolving appropriate actions to rectify the errors. (Singh and Masaku, 2012) [12].

Different stages at which errors/mistakes creeps in the data or methodology

The sources of error in applying statistical procedures are legion and include all of the following: (Phillip and Hardin, 2012) [11].

1. a) Replying on erroneous reports to help formulate hypotheses
   b) Failing to express qualitative hypotheses in quantitative form
   c) Using the same set of data both to formulate hypotheses and to test them
2. a) Taking samples from the wrong population or failing to specify in advance the population(s) about which inferences are to be made
   b) Failing to draw samples that are random and representative
3. Measuring the wrong variables or failing to measure what you intended to measure Using inappropriate or inefficient statistical methods. Examples include using a two - tailed test when a one - tailed test is appropriate and using an omnibus test against a specific alternative
4. a) Failing to understand that p - values are functions of the observations and will vary in magnitude from sample to sample
   b) Using statistical software without verifying that its current defaults are appropriate for your application failing to adequately communicate your findings
5. a) Extrapolating models outside the range of the observations
   b) Failure to correct for confounding variables
   c) Use the same data to select variables for inclusion in a model and to assess their significance failing to validate models:

Medicinal and aromatic plants in India

Medicinal and aromatic plants have played an important role in the socio-cultural, spiritual and health-care needs of rural and tribal people of the emerging and developing countries. In many developing countries, a large section of the population still relies on traditional systems of medicine to meet their health-care needs. Also, more and more people in the developed countries have turned to alternative therapies and herbal medicines resulting in many fold increase in the demand of medicinal plants and their products in these parts of the world.

Medicinal and aromatic plants are important factors in sustainable development, environmental protection and public health. Traditional medicine, based on the use of medicinal plants has developed and turned into the investigation of the active principles and pharmaco-dynamic action of medicinal plants. Thus, to achieve optimal results at cultivation, harvesting and processing levels, the collaboration between farmers and researchers became necessary. (ICS-UNIDO, 2003) [8].

Error associate with experiment

Errors are normally classified in three categories: systematic errors, random errors, and blunders.

Systematic errors

Systematic errors are due to identified causes and can, in principle, be eliminated. Errors of this type result in measured values that are consistently too high or consistently too low. Systematic errors may be of four kinds:

1. Instrumental. For example, a poorly calibrated instrument such as a thermometer that reads 102 ºC when immersed in boiling water and 2 ºC when immersed in ice water at atmospheric pressure. Such a thermometer would result in measured values that are consistently too high.
2. Observational. For example, parallax in reading a meter scale.
3. Environmental. For example, an electrical power brown out that causes measured currents to be consistently too low.
4. Theoretical. Due to simplification of the model system or approximations in the equations describing it. For example, if your theory says that the temperature of the surrounding will not affect the readings taken when it actually does, then this factor will introduce a source of error. (*http://www.physics.nmsu.edu/research/lab*)

Random errors

Random errors are positive and negative fluctuations that cause about one-half of the measurements to be too high and one-half to be too low. Sources of random errors cannot always be identified. Possible sources of random errors are as follows:

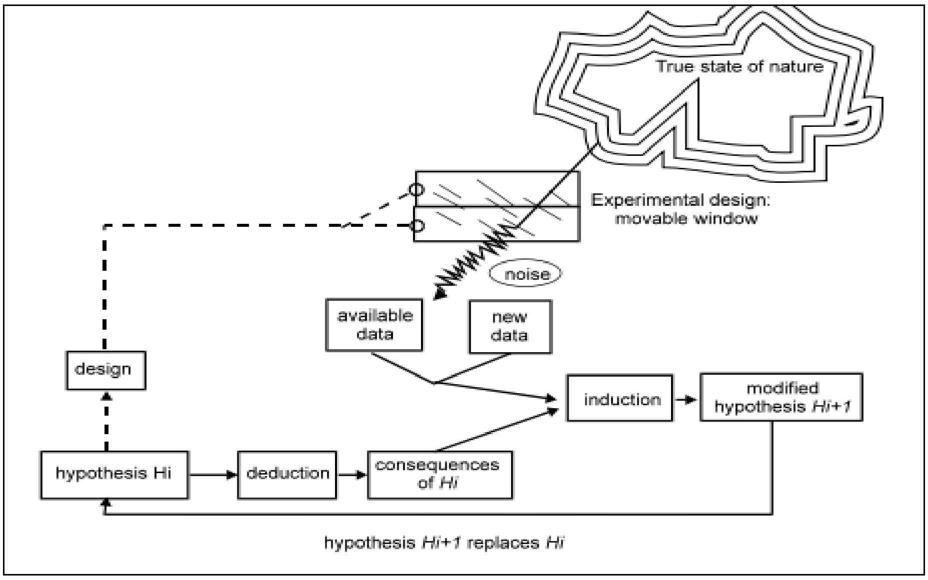
For example, errors in judgment of an observer when reading the scale of a measuring device to the smallest division.

Design of experimental techniques commonly used in agricultural research

The role of experimental design

Experimental design concerns the validity and efficiency of the experiment. The experimental design in the following diagram is represented by a movable window through which certain aspects of the true state of nature, more or less distorted by noise, may be observed. The position and size of the window depend on the questions being asked by and the quality of the experiment. A poorly used design may not generate any useful information for meaningful analysis. A wisely designed experiment can provide factual evidence which can easily be analyzed and understood by the researcher.

The diagram emphasizes that, although the conjectured state of nature may be false or at least inexact, the data themselves are generated by the true state of nature. This is the reason why the process of continually updating the hypothesis and comparing the deduced states of nature with actual data can lead to convergence on the right answers.



Design of experiments deals with the study of methods for comparing the treatment, varieties, factors etc. under different experimental situations faced by agricultural research worker. The main objective of any experimental design is to provide the maximum amount of information relevant to the problem under the investigation. Experimental design provides maximum amount of information at minimum cost. There are three basic principles of experimental design. Replication Randomization Local Control Replication means repetition of the basic experiment. It is useful for more precise estimate of the mean effect of any factor and it is also useful for estimation of experimental error and determination of confidence interval. Randomization is the technique or device for eliminating the bias.
Replication means repetition of the basic experiment. *Haphazard is not randomized*. It is useful for more precise estimate of the mean effect of any factor and it is also useful for estimation of experimental error and determination of confidence interval.
Randomization is the technique or device for eliminating the bias.

Local Control: The purpose of the local control is to make the experimental design more efficient and reduce the experimental error. Experimental Design and Design of Experiment: We are concerned with the analysis of data generated from an experiment. It is wise to take time and effort to organize the experiment properly to ensure that the right type of data and enough of it is available to answer the questions of interest as clearly and efficiently as possible. This process is called experimental design. We should also attempt to identify known or expected sources of variability in the experimental units since one of the main aims of a designed experiment is to reduce the effect of these sources of variability on the answers to questions of interest. That is, we design the experiment in order to improve the precision of our answers.
(*http://www.plantsciences.ucdavis.edu/agr205/lectures/2011*)

Treatment: In experiments, a treatment is something that researchers administer to experimental units. For example, a rice field is divided into four, each part is 'treated' with different fertilizer to see which produces the most rice; a teacher practices different teaching methods on different groups in the class to see which methods yield the best results; a doctor treats a patient with different treatments to see which is most effective.

Factor: A factor of an experiment is a controlled independent variable; a variable whose levels are set by the experimenter or researcher. A factor is a general type or category of treatments. Different treatments constitute different levels of a factor. For example, three different groups of farmers are subjected to different training methods. The farmers are the experimental units, the training methods, treatments; where the three types of training methods constitute three levels of the factor 'type of training'.

Main Effect: This is the simple effect of a factor on a dependent variable. It is the effect of the factor alone averaged across the levels of other factors.

Interaction: An interaction is the variation among the differences between means for different levels of one factor over different levels of the other factors.

Some important characteristics of a well-planned experiment are (Cox, 1958)

1. **Degree of precision**. The probability should be high that the experiment will be able to measure differences with the degree of precision the experimenter desires. This implies an appropriate design and sufficient replication.
2. **Simplicity**. The design should be as simple as possible, consistent with the objectives of the experiment.
3. **Absence of systematic error**. Experimental units receiving one treatment should not differ in any systematic way from those receiving another treatment so that an unbiased estimate of each treatment effect can be obtained.
4. **Range of validity of conclusions**. *C*onclusions should have as wide a range of validity as possible. An experiment replicated in time and space would increase the range of validity of the conclusions that could be drawn from it. A factorial set of treatments is another way of increasing the range of validity of an experiment.
5. **Calculation of degree of uncertainty**. The experiment should be designed so that it is possible to calculate the possibility of obtaining the observed result by chance alone.

Steps in experimentation (Little and Hills 1978) **[10]**

- Define the problem

- Determine the objectives
- Select the treatments
- Select the experimental material
- Select the experimental design
- Select the experimental unit and number of replications
- Ensure proper randomization and layout
- Ensure proper means of data collection
- Outline the statistical analysis before doing the experiment
- Conduct the experiment
- Analyze the data and interpret the results
- Prepare complete and readable reports

Specific issues of experimental design

Once the objectives, interesting questions, and the hypothesis are defined, the scope, type, and requirements of an experiment are also more or less determined. Thus, the experiment should be designed to meet those requirements.

Specifically, experimental design is concerned with the following issues:

1. The size of the study: number of replications, and the size and shape of experimental units.
2. Type and number of measurements: availability of a measuring device, precision and accuracy of the measurement, and the timing of making measurements.
3. Treatments: the type of treatments, the levels of treatment, and the number of treatments.
4. Assignment of treatments to experimental units: completely random, restricted randomization, etc.
5. Error control: the error control can be accomplished by blocking techniques, the use of concomitant observations, the choice of size and shape of the experimental units, and the control of the environment using a growth chamber or greenhouse
6. Relative precision of designs involving few treatments: to compare two experimental designs one compares amounts of information.

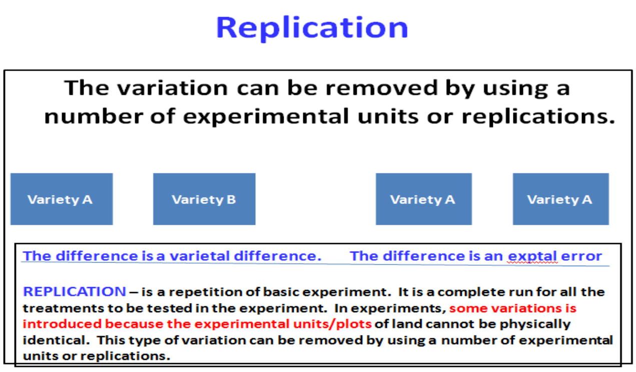
Among the above considerations, replication and randomization are the most important basic principles in designing experiments.

Essential components of design of an experiment

- **Estimate of error**
  1. Replication
  2. Randomization
- **Control of error**
  1. Blocking
  2. Proper plot technique

Data analysis

- Proper interpretation of results



Functions of replication

1. To provide an estimate of the experimental error. The experimental error is the variation which exists among observations on experimental units treated alike. When there is no method of estimating experimental error, there is no way to determine whether observed differences indicate real differences or are due to inherent variation.
2. To improve the precision of an experiment by reducing the standard deviation of a treatment mean. Increased replication usually improves precision, decreasing the lengths of confidence intervals and increasing the power of statistical tests.
3. To increase the scope of inference of the experiment by selection and appropriate use of more variable experimental units. Example: replication in time and space in yield trials.
4. To effect control of the error variance. The aim is to assign the total variation among experimental units so that it is maximized among groups and, simultaneously, minimized within. Experimental error must not be inflated by differences among groups

Distinguishing between replications, subsamples, and repetitions

Replication refers to the number of experimental units that are treated alike (experimental unit or experimental plot is the unit of material to which one application of a treatment is applied). Misconception of number of replications has often occurred in experiments where subsamples or repeated observations on some unit were mistaken as experimental units. A 1982 survey of articles in plant pathology journals showed no true replications in 19% of papers on phytopathology and 37% of the papers on plant disease. Without proper replication, no valid scientific conclusion can be drawn from such a study.

There are two important points to remember in determining what constitutes a replication.

1. Each replication must be independent of every other.
2. Each replication must be part of a randomized trial; that is, any one plot must have the same chance of getting each treatment.

Example: Field trial. We want to compare 3 fertilizer treatments, denoted A, B, and C. Except as discussed in Case 4, the field is an ordinary one, and assumed uniform. The crop is an annual.

1	2	3
4	5	6
7	8	9
10	11	12

N

Case 1: 4 plots are selected at random for each fertilizer. This is done by moving through a table of random numbers. The first 4 between 1 and 12 selected are assigned to treatment 1, the next 4 to treatment 2, and the last 4 to treatment 3. This is a completely randomized design. There are 4 replications for each treatment. If the experiment is run in succeeding years then each year provides a different set of replications.

Case 2: Same as case 1 except the crop is a perennial. In this case the succeeding years are not replications. The fact that the crop is a perennial violates the independence requirement above. This is a repeated measures experiment.

Case 3: Same as case 1 except each of the 12 plots are further divided into three subplots. These subplots are not replications, they are subsamples. The randomization requirement above is violated.

Case 4: The numbers 1, 2, and 3 are arranged in a random order. Each treatment is applied to all plots in a north-south column containing the randomly selected number (e.g. treatment 1 is applied to plots 2, 5, 8, and 11, etc.). In this case there is no replication for the treatment. The separate plots are subsamples.

The definition of replication can have an even broader meaning. For instance, in a breeding program of yield trials, varieties are frequently compared in a number of locations and years. In this case, the varieties are treatments and the experimental units are plots in location and year. Over the years, these plots may not be the same spots in locations, or not even in the same locations. The point here is that sometimes the locations and years may also be considered as replications so as to enable us to examine the consistency of the varietal yield performance. However, it should be noted that sometimes the locations may be of special interest and these are considered as treatments rather than replications. (*http://home.iitk.ac.in/~shalab/anova/chapter4-anova-experimental-design-analysis.pdf*)

Selection of experimental site

The following points should be kept in mind while deciding experiment al site

- Slope/gradient
- History of previous crop and description of experimental site
- Graded areas
- Presence of large trees, poles and structure
- Unproductive site

Soil heterogenity

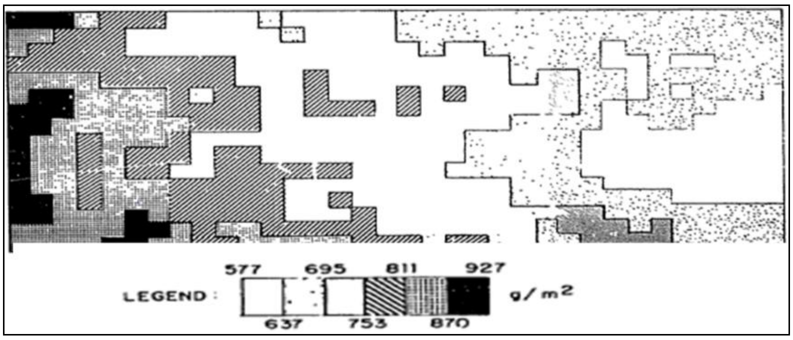
- Uniformity trail

1. Soil productivity contour map
2. Serial correlation
3. Mean square between strips
4. Smith Index soil heterogeneity

II Data from field experiments

Uniformity trial

- Indicators of uniformity
  - Topography
  - Visual appearance of the crop
  - Soil and sub soil examination
  - Field history
- Evaluating pattern of soil heterogeneity
  - Soil productivity contour maps
  - Serial correlation
  - Mean square between strips
  - Smiths index of soil heterogeneity



Serial correlation: Serial correlation procedure is generally used to test the randomness of a data set. Higher value is better as it shows fertility gradient while lower values indicate fertile areas occur in spots.

Vertical			Horizontal		
Pair (i)	X_i	X_{i+1}	Pair (i)	X_i	X_{i+1}
1	842	803	1	842	844
2	803	773	2	844	808
3	773	912	3	808	822
⋮	⋮	⋮	⋮	⋮	⋮
35	589	614	17	968	917
36	614	633	18	917	965
37	633	681	19	965	947
⋮	⋮	⋮	⋮	⋮	⋮
646	861	965	646	619	633
647	965	917	647	633	614
648	917	842	648	614	842

Mean square between strips

Units are first combined into horizontal and vertical strips and Variability between the strips in each direction is measured by the mean square. The relative size of the two MS, one for horizontal strips and another for vertical strips, indicates the possible direction of the fertility gradient and the suitable orientation for both plots and blocks.

Smith index soil heterogeneity (Zero to 1)

Smith Index used to derive optimum plot size. Larger the value of the index the lower is the correlation between the adjacent plots indicating the fertile spots are distributed randomly. (Gómez and Gómez, 1984) [2].

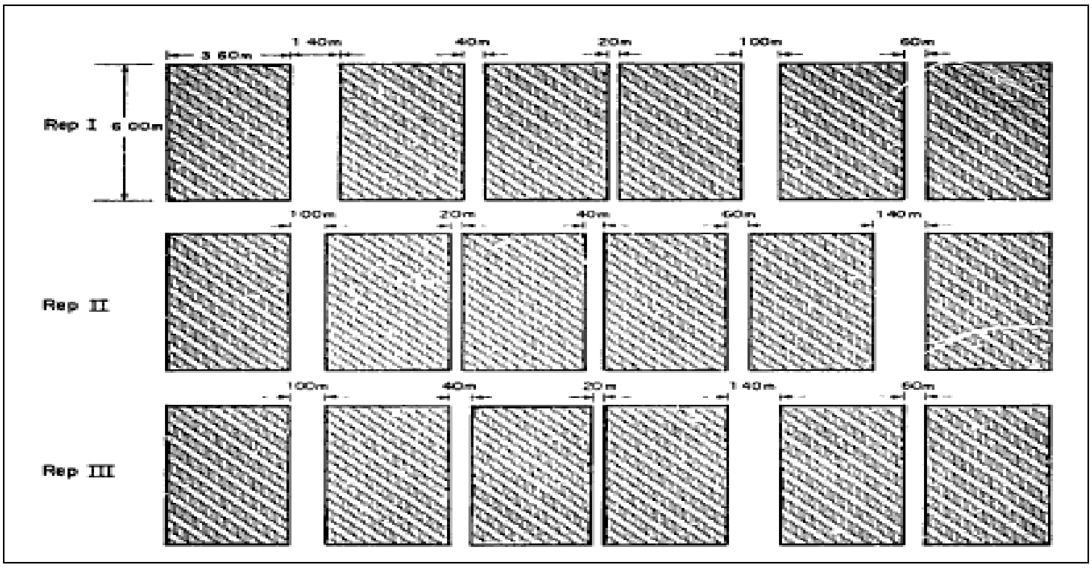
Adjusting /coping with Soil Heterogeneity

- Proper choice of plot size and shape
- Factors influencing plot size
- Soil heterogeneity
- Variability becomes smaller as plot size becomes larger
- Gain in precision decreases as plot size increases
- Practical consideration
- Kind of crop
- Number of treatments
- Type of machinery
- Land area
- Factors influencing plot shape (Length to width ratio)
- Sufficiently wide to remove the border effect
- Higher precision is obtained in rectangular or square plots
- Spotty fertility pattern requires square plots
- Block size
  - Plot size and number of treatments
  - Experimental design adopted
- Shape of block
  - Block should be oriented perpendicular to the fertility gradient
  - Spotty fertility pattern- square block
- Number of replication
  - Inherent variability of the experimental material
  - Experimental design
  - Number of treatments
  - Degree of precision desired

Competition effects, measuring and control

- Non planted borders
- Varietal competition
- Fertilizer competition
- Missing hill Competition effects

Non planted borders: Area between plots around the experimental areas without crop serve as markers



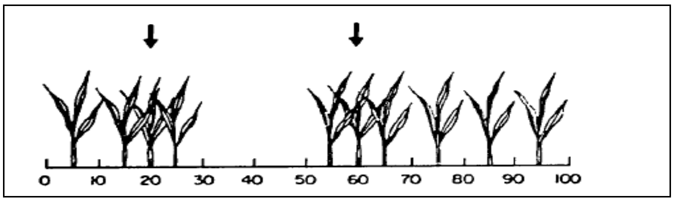
Defects in statistical methods

- Statistical analysis is applicable to random sample
- Inverse relation between real error and estimates of error
- Limitations of F test followed by LSD test

Mechanical errors

Furrowing of Row Spacing- alters the plant population and resultant yield

- Selection of seedlings- uniformly grown seedlings to be used
- Thinning- selecting plants to be retained, Correct stage of thinning, spacing of crop



- **Transplanting- number of seedlings per hill and uniformity of the seedlings**

Border Effect: Leads to wrong data and interpretation

Problem of plant population aging lead to misleading result of the experiment (Khadi, 2016) [9].

Other common error associated with experiment

Common statistical errors

Before we proceed to common statistical errors, let us understand why it is important that we carry out data analysis meticulously. Data analysis is a process that seeks to identify relationships, associations, differences, variance or trends that may exist within the data. The purpose is to see if the results can be generalized to the population or in other words “how true” or real these findings are.

It is useful remember the following before data analysis-

1. Organize all collection forms, and material used to record the data in one place
2. Check the data for completeness and accuracy
3. Note missing data if any and decide whether or not to remove from the analysis and document this
4. Assign unique identifiers to the data

Errors in Publication: Researcher should ensure publication of work done as papers that do not see the light of the day have wasted precious time and resources of all stakeholders and will have failed to advance Evidence Based Practice. Getting the statistics right in the publication is crucial.

Errors in Data analysis: The choice of Parametric versus non-parametric methods and the importance of assumptions Statistics as a discipline uses models and assumptions. Prior to applying any parametric test, the researcher needs to check if the assumption of normality is met as these tests are to be used only when the data is normally distributed.

Errors in data analysis

The choice of Parametric versus non-parametric methods and the importance of assumptions Statistics as a discipline uses models and assumptions. Prior to applying any parametric test, the researcher needs to check if the assumption of normality is met as these tests are to be used only when the data is normally distributed.

Using the wrong statistical test: Statistical tests are not only numerous, but also have similar sounding names. Each test is to be used only if certain assumptions are satisfied. For example, the student's t-test is a widely used parametric test that is of two types- the unpaired [also called the two-sample t test] and the paired t test and data needs to be normally distributed for its use.

Presentation of categorical data: Categorical Variables may be dichotomous or binary [for example -male and female] or non-binary [mild, moderate and severe pain]. These are described as proportions of the total number of participants [along with 95% Confidence Intervals]. They can also be expressed in the form of a bar or pie chart. Often times, binary categorical data is best presented as a 2 x 2 table.

Outliers and their reporting: An outlier is essentially an abnormal value that lies far away from the rest of the values in the sample. Outliers are important are they can have a significant impact and alter results of the analysis dramatically. They could be true outliers, a typographical error that resulted during data entry [which needs to be corrected], or a wrong measurement. All outliers need to be carefully considered. Given that here is little consensus on how outliers are to be analyzed, it is important that are outliers are reported with honesty and where appropriate an analysis with and without the outlier be performed and reported.
(*https://www.researchgate.net/post/Parametric_and_non-parametric_tests*).
Reporting only P values, not reporting the exact p value and confusing it with the effect size and not reporting Confidence Intervals.

Interpreting data correctly
Correlation and Causation

Common mistakes to is to just assume that just because we find a correlation between two variables, one causes the other. This is often described in statistical parlance as "Correlation does not imply causation". An often quoted example in this regard is the "Correlation" of Sun Signs in Astrology with outcomes by the researchers of the Second International Study of Infarct Survival Trials Collaborative Group [ISIS-2]. 18 Overall, the study showed a significant benefit of aspirin over placebo [$p < 0.00001$]. However, based on the date of birth entered in the case record forms, when the researchers classified all patients as per the Sun Sign they were born under, two Sun Signs- Gemini and Libra showed no apparent benefit with aspirin while another Sun sign Capricorn, showed a nearly 50% reduction in mortality! It would be inappropriate to say that one Sun Sign appears to benefit more with aspirin than the other. Thus, these relationships should be viewed merely as associations and not cause [Sun Signs] and effect [mortality].

The challenge of small sample sizes: All sample size calculations should be done before starting the study regardless of the number of groups being studied. A small sample size does not necessarily make the study a weak or a poor one. Rather, the ability to generalize and draw inference about the population of interest simply becomes more difficult. Also, with small sample sizes, one must be careful not to overstate the strength of evidence or go beyond what you have observed to draw overarching conclusions. (*https://wikivividly.com/wiki/Correlation_does_not_imply_causation*)

"The purpose of statistical science is to provide an objective basis for the analysis of problems in which the data depart from the laws of exact causality"

D. J. Finney,

An Introduction to Statistical Science in Agriculture

References

1. Cox DR. Planning Experiments. New York: John Wiley & Sons, 1958.
2. Gómez KA, Gómez AA. Statistical procedures for agricultural research. 2nd Edition. International Rice Research Institute. Phillipines. 1984, 479-503.
3. http://home.iitk.ac.in/~shalab/anova/chapter4-anova-experimental-design-analysis.pdf
4. http://www.physics.nmsu.edu/research/lab
5. http://www.plantsciences.ucdavis.edu/agr205/lectures/2011
6. https://wikivividly.com/wiki/Correlation_does_not_imply_causation
7. https://www.researchgate.net/post/Parametric_and_non-parametric_tests
8. ICS-UNIDO. The International Centre for Science and High Technology and United Nations Industrial Development Organization. Medicinal plants and their Utilization. Chapter 1, 2003.
9. Lecture notes of Dr. B.M. Khadi Winter School 'Application of Advanced Statistical Tools in Agricultural Research, 2016.
10. Little TM, Hills FJ. Agricultural Experimentation: Design and Analysis. John Wiley and Sons, New York, NY. 1978, 350.
11. Phillip IG, Hardin JW. Common Errors in Statistics (and How to Avoid Them). John Willey & Sons. Fourth Edition, 2012.
12. Singh AS, Masaku MB. An Insight in Statistical Techniques and Design in Agricultural and Applied Research. World Journal of Agricultural Sciences. 2012; 8(6):568-584.