Letter to Editor
Open Access
Iranian Journal of Diabetes and Obesity (IJDO)

Statistical Rules in Scientific Reports (The basics)

Alireza Pakgohar[1]*, Hossein Mehrannia[2]

[1]Departmment of Statistics, Payame Noor University (PNU). Tehran, Iran.
[2]Department of Statistics, Behbahan Branch, Islamic Azad University, Behbahan, Iran.

Abstract

Objective: Scientific papers usually contain information and data that we call them statistics. We expect statistics to provide us with a suitable description of the data by summarizing.
Scientific journals have specific frameworks for this work in mention to writers and readers can understand statistical concepts with a common terminology.
In this paper we guide reader to write a scientific issue without any confusion and crowding and we propose some notes to report a scientific descriptive statistics, and make a table properly and draw a visual graph.

Keywords: Descriptive statistics, Scientific report, Methodology of research, Hypothesis test, Sample size

QR Code:

Citation: Pakgohar A, Mehrannia H. Statistical Rules in Scientific Reports (The basics). IJDO 2023; 15 (1) :1-6

URL: http://ijdo.ssu.ac.ir/article-1-772-en.html

10.18502/ijdo.v15i1.12205

Article info:
Received: 10 October 2022
Accepted: 21 December 2022
Published in March 2023

Corresponding Author:
Alireza Pakgohar, [1]Departmment of Statistics, Payame Noor University (PNU). Tehran, Iran.
Tel: (98) 913 318 0781
Email: a_pakgohar@pnu.ac.ir
Orcid ID: 0000-0003-2512-1581

[DOI: 10.18502/ijdo.v15i1.12205]

Introduction

In writing the research paper, there is frequently a massive quantity of data that must be incorporated to meet the research's purpose. Instead of stuffing your research paper with all this information, you can employ visual assets to make it simpler to read and use to your advantage to make it more appealing to readers.

Every research that is done includes keywords that are highlighted in two places. First, it is in the abstract section that introduces the reader to the fundamental words of the research, then in the introduction section of the article, where writers explains them.

The main words of the research frame consider the importance of the content. Sometimes, specialized terms such as the name of medicine or disease are used, which are of secondary importance, but the reader must be guided by providing brief information about them.

Similarly, any research should clearly define what the research variables are. State them clearly and specify the role of each in the article. Finally, in the research method, three main elements of research source, research sample and sampling method should be specified**.**

Additional tips:

- All variables should be defined clearly and precisely.
- Complex variables must be described obviously in the text.
- It is very useful to provide a phrase in the text from which the variable is derived.
- Different definitions can give very different results.

Parametric Hypothesis tests

Parametric and non-parametric statistical tests are very important. Parametric statistical tests have stronger power than their non-parametric counterparts. In other words, the parametric test with a stronger probability accepts the alternative hypothesis. Therefore, it should be specified in the articles whether the statistical tests are parametric or non-parametric. In parametric tests, from the scale of measurement to the distribution of the dependent variable data, it must have certain conditions. Obviously, research variables should be quantitative in parametric test, that is, their minimum measurement scale should be interval. But regarding the data distribution, we should go back to the central limit theorem, which considers the normality distribution of the sample mean condition. So, it is clear that if the data distribution is normal, one of the conditions for the implementation of the parametric test has been considered. Which test should be used to check the goodness of fit of the normality of the data distribution?

Many researchers use the Shapiro-Wilk (S-W), Anderson-Darling and Kolmogorov-Smirnov (K-S) tests. These are very popular and have proven their benefits but K-S test in compare to S-W test is less convenient tests for samples with n<50. However, they are not a suitable test for a large sample. Because they easily rejects the normality of data distribution and prohibits us from applying a parametric test (1). So it is better to use the kurtosis and skewness indices of the data distribution (2). In this method, it is sufficient that the absolute value of these two indices is less than 2 to accept the normality of data distribution. As a thumb rule, statistical inferences are made about mean and variance with skewness and kurtosis, respectively.

Finally there are many various goodness of fit tests with different sample size. You can referee to (3).

Sample size

The sample size in a study indicates the validity of the results. In qualitative sampling methods, the methods of determining the sample size are obtained with rules of thumb and considering the theoretical saturation (4), and in probabilistic methods, Type I and II errors (α, β), power of test (π or θ) and effect

[DOI: 10.18502/ijdo.v15i1.12205]

size measure determine the sample size. percentage that we consider for ourselves before conducting a study (e or d) are part of the sample size computation. The G*Power is a famous and useful calculator software for the size of sample (5).

Many studies report the sample size, but it is not clear on what basis and with what validity this sample size was determined. This problem is one of the common mistakes of the authors of the article and it is taken into account in the statistical evaluation of the article by reviewers. Sometimes, the number of samples greater than or equal to 30 is selected from the rule of the central limit, which, although it has some flaws, but at least it is necessary to pay attention to the central limit theorem (6). Many studies are conducted on laboratory samples and animal studies, in which case it is clear that a sample size of $n \geq 30$ considered an unethical method in the research profession.

Sometimes, the sample size may be less than 10. In this case, it is possible to use non-parametric tests although that It appears unlikely that K-S have enough power to accept the alternative hypothesis for sample size $n<10$. The idea that non-parametric analyzes apply on non-normal distributions is wrong, but they do not have any assumption for the distribution. Furthermore, their power and reliability can be approximately less or equivalent to parametric tests. So you probably haven't got much to lose by choosing a nonparametric test and ignoring the parametric assumptions such as the normal distribution.

Some rules for characterizing sample size and statistical power

The following rules outline the key relationships between the specify of sample size and power of test

- The statistical power increases by larger sample
- When the effect size is small, we have to increase power by larger sample.
- Larger sample has a contributory role to evaluation of a program with low take-up.
- High variation in outcomes, needs a larger sample.
- Power is maximized when the sample is equally split between groups for example control and experimental groups.
- For a given sample size, the statistical power reduces when randomizing at the cluster level as versus to the individual level. The more similar the outcomes of individuals within clusters are, the larger the sample needs to be.

Descriptive statistics

Statistical descriptive is a term that refers to data analysis that aides illustrate, describe or brief summary evaluated of data in a meaningful way such that, for instance, criteria's that may come out from data. However, descriptive statistics do not permit us to make conclusions over the data we have evaluated or to reach results about research hypotheses. These are just a way of describing mentioned data. To summarize the data, average ($\bar{X}$) , median (Md), mode (Mo), variance, etc. are measured and graphs are also drawn if needed.

- To display variables that have multiple categories or values we can get following: For categorical variables, we can comprise classes into broad groups or/and combine the rare values into the "other" category to avoid presenting highly detailed graphs and tables.
- For continuous variables, first, the range of any numerical variable with large values can be divided into a small-scale number of interval measures and the data can be obtained in each of the processes described that was explained, for cases where a variables have only a small number of classes or categories is showing amount labels or values. The 2nd, and probable option most common, option for representing a numerical variable that takes many values is to provide just merely statistics of cumulative distribution, including its Md, $\overline{X}$, and standard deviation (SD) (7). This is possible only by

[DOI: 10.18502/ijdo.v15i1.12205]

presenting these statistics in a table. It is not permitted to display summary statistics for a numerical variable in a chart. Beside, to Md, $\overline{X}$ and SD, it would be relevant to provide Min and Max values, upper and lower quartile values, and perhaps a measure of skewness.

- The simplest way to show the relationship between variables is that where both variables have a small number of values or categories. In a simple crosstab table, the values or categories of one variable may act as rows, while the values or categories of another variable may act just as columns.
- The $\overline{X}$ and SD should not be used for qualitative variables. For nominal variables, can be used from Mo and percentage as a central tendency, and for ordinal variables, Md and Mo. The suitable index for calculating dispersion for the nominal variable is entropy and standard error of the sample proportion, and for the ordinal variable, quartile and decimal coefficient of dispersion besides entropy measure.
- Besides of $\overline{X}$, the SD coefficient and the standard error coefficient of the mean (S.E) can be used. It is suggested to use the SD with applying the notation ($\overline{X} \pm$ SD).
- The percentage distribution of discrete variables must be clearly defined whether they are a percentage of the population or a group (8).

Scientific Tables

Tables and graphs can obtain some descriptive statistics of main information of regarding variables. These statistics consist, but are not restricted to, central tendency as $\bar{X}$, Md, percentage distribution, and the dispersion measure SD.

Graphs and tables are the main material in any statistical report. They must reflect all the main types of information collected in the study. Prefund analysis of narrower topics must be left to more focus certain reports. A table should contains a small amount of text limited to some information of relevant the table.

This section teaches you all you need to know on how to make a scientific table to include in your research paper. The proper table format is extremely basic and straightforward to accomplish, here's a simple guideline to help you (9,10):

Number: If you have more than one table, number them sequentially (Table 1, Table 2).

Referencing: Each table must be referred to in the text with a capital T: "as seen in Table 1".

Title: Make sure the title corresponds to the topic of the table. Tables should have a precise, informative title that serves as an explanation for the table. Titles can be short or long depending on their subject.

Column headings: Column headings should be helpful and clear when representing the type of data provided. The reader's attention is drawn progressively from the headline to the column title. A clear title helps the reader immediately understand what the table is about.

Table body: This is the major section of the table that contains numerical or textual data. Make your table such that the elements read from top to bottom, not across.

Needed information: Make sure to include units, error values, and number of samples, as well as explain any abbreviations or symbols used in the tables. In all statistical tables, it is necessary to report the statistical value and the probability value. Reports should not have many tables and arrays of numbers in each table.

The following must be used in the symbols of statistical tables:

Lines: Limit the use of lines, only use what's necessary. Avoid drawing column lines between table cells.

Additional tips:

- If your study comprises tables and graphs,

Index	Frequency	Percent	Mean	Standard deviation	Standard error of mean	P-value
Notation	f	%	$\bar{X}$	SD	S.E	*P*

[DOI: 10.18502/ijdo.v15i1.12205]

refrain from putting the same information in both.

- Do not duplicate your information from a table in a text. The interpretation of the information in the table should come in the text, and if it is not interpretable, therefore, the existence of the table is not necessary unless it contains a lot of data and the textual report of which is out of the reader's interest.
- Confirm table aesthetically appealing and easy by leaving enough space between rows and columns and using a basic yet effective structure.
- If your table has a lot of data and information, categorize and divide them into columns.
- Consider merging tables with repeated information or deleting items that may not be necessary. Our preference is to have fewer tables in an article.
- Employ the worth footnotes to highlight important information for each cell. Use an alphabetical footnote marker when your table contains numerical data.
- Indicate the *, ** and *** symbols as the significant level at 0.05 of 0.01 and 0.001, respectively below of the table as a footnote.
- If the table it shows contains previous research, cite the reference to avoid plagiarism.
- The report of the result of a statistical hypothesis test should include value of statistic and p-value. This general rule is mandatory both in the text and in the table.
- Avoid using additional information that is usually reported by statistical software. In hypothesis tests and statistical models, value of statistic and p-value are vital. For other indicators, if it is possible to reach other indicators by reporting one index, there is no need to report them, and if there is an index that is more important, provide it. For example, in a linear regression model, among the information about correlation (r), as well relative measures: coefficient of determination (r^2) and adjusted r^2, it is sufficient to report the adjusted r^2.
- The information in the tables must be sufficient in which the researcher can present them correctly without referring to the text of the report. It is vital because many researchers use photocopy the tables without referring them in paper.
- In addition to notes at the bottom of a table, a series of definitions and explanations might appear in the text accompanying the tables.
- It may be appears a series of definitions and explanations in addition to notes at the bottom of the table, in the text accompanying the tables.
- Sometimes, graphs can made more interesting and more intuitive data in comparison to table. A common way to display data graphically for a single variable with a small number of values or categories, is a column histogram or chart, where the partial frequency of each category or value is presented by the column’s height.

Conclusion

This paper referees the definition of descriptive statistics and the related concepts. We discussed sample size and its relationship to type I and II errors and statistical power. Researchers considering to assumptions about the parameters of the population distribution can applying the parametric and non-parametric test.

Not only does the correct choice of hypothesis test improve the probability of detecting a difference or association, but researchers are increasingly asked to provide information about sample size and other statistical test assumptions. In addition, descriptive statistics are an important part of primary data analysis and provide the basis for comparing variables with inferential statistical tests.

We have provided recommendations for an accurate and standardized statistical description. Also, we have provided the

[DOI: 10.18502/ijdo.v15i1.12205]

guidelines for preparing a table containing useful information for a scientific report. We discussed about we should put in a table and what we should not use. Finally, everything that needs to be said in relation to the scientific text and scientific table has been written.

Acknowledgments

We are grateful to all of those with whom I have had the pleasure to work during this and other related projects. Specially, Dr Mohadeseh Khalili has provided us extensive professional guidance and taught us a great deal about sample size techniques**.**

Funding

The authors did not disclosed receipt of the following financial support for the research, authorship, and/or publication of this article**.**

Conflict of Interest

All authors have participated in (a) conception and design, or analysis and interpretation of the data; (b) drafting the article or revising it critically for important intellectual content; and (c) approval of the final version.

The authors have no affiliation with any organization with a direct or indirect financial interest in the subject matter discussed in the manuscript.

References

1. Mishra P, Pandey CM, Singh U, Gupta A, Sahu C, Keshri A. Descriptive statistics and normality tests for statistical data. Annals of cardiac anaesthesia. 2019;22(1):67.
2. Keller G. Statistics for management and economics. Cengage Learning; 2022.
3. Pakgohar AL, Habibirad A, Yousefzadeh F. Goodness of fit test using Lin-Wong divergence based on Type-I censored data. Communications in Statistics-Simulation and Computation. 2020 Sep 1;49(9):2485-504.
4. Pakgohar A, Khalili M. Investigation of sample size in qualitative sampling methods. Popularization of Science. 2021;12(1):270-97
5. Kang H. Sample size determination and power analysis using the G* Power software. Journal of educational evaluation for health professions. 2021;18.
6. Kar SS, Ramalingam A. Is 30 the magic number? Issues in sample size estimation. National Journal of Community Medicine. 2013;4(01):175-9.
7. Glewwe P, Levin M. Presenting simple descriptive statistics from household survey data. Household Sample Surveys in Developing and Transition Countries. 2005:335-54.
8. Andrade C. Understanding the difference between standard deviation and standard error of the mean, and knowing when to use which. Indian Journal of Psychological Medicine. 2020;42(4):409-10.
9. Abbadia, Jessica. (2022). How to make a scientific table. Step-by-step and Formatting. Mind the graph. Available at https://mindthegraph.com/blog/how-to-make-a-scientific-table/
10. Kumar R. Research methodology: A step-by-step guide for beginners. Sage;2018.

[DOI: 10.18502/ijdo.v15i1.12205]