

Descriptive statistics - presenting your results I

Mehwish Hussain

Department of Biostatistics, Dow University of Health Sciences, Karachi, Pakistan.

Email: mehwish.work@gmail.com

Researchers analyse and present datasets after the collection phase of studies. This step is most crucial as it leads to the discussion on the entire report, depending on the objective of the study. Nevertheless, researchers abide great difficulties due to unawareness of statistical tools to be used as per objective.¹ Different studies report that many medical professionals are keen to learn statistical tools. However, they start repulsing from them when they encounter difficulties in analysis.^{1,2} This phobia may be due to mathematical tabulation and interpretation of the findings. Nevertheless, if researchers can learn the core concept of statistical analysis it would help them tremendously. All correct analyses are based on the objective of the study. Guidelines of International Committee of Medical Journals Editors (ICMJE), explains the importance of acceptable reporting of statistics: "Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results."³

Types of Statistical Analyses:

Mainly, there are two types of analyses involved in statistical findings. One is descriptive, another is inferential. In descriptive statistics, researcher only describes the findings of the collected data. Inferential statistics includes methods to generalize data findings to the related populations with certain level of confidence and assurance of significance of results. It is necessary to consider some sets of principles while conducting such statistical computations. Researchers should also be cautious while presenting such statistical analyses. In this article, the considerations of descriptive statistics findings will be under the pen. Inferential statistics will be mentioned in subsequent articles of the 'Learning Research' series.

Descriptive Statistical Analysis:

Descriptive analyses are performed when the study objective is involved in enhancing the reader's knowledge, comprehension and application related to the research. The phrasing of the study objective is pivotal. Objectives' exact words are related to the study design and statistics. It contains some of the related clues as defined in Bloom's taxonomy.⁴ Such as describe, identify, evaluate, examine, design, review, show, measure etc. Strictly note that the

verbs associate, relate, compare, predict, estimate etc. come under the area of inferential statistical analysis. Statistical technical terms should not be used in a nontechnical manner such as "random", "normal," "significant," "correlations," and "sample."³ Nevertheless, if used, then these terms should be elaborated in such a way that reader can understand the scenario comfortably.

Descriptive analysis involves description of data in terms of frequencies, proportions, mean, median, quartiles, standard deviation, inter-quartiles range etc. Measurement of these statistics depends on type of variables either to be qualitative or quantitative. Qualitative variables are categorical, characterized and attributable e.g. gender, socio-economic status, pain level, treatment groups etc. On the other hand quantitative variables are measurable, continuous and numerical e.g. age, height, weight, pain score etc.

Qualitative Variable Analysis:

When the data is qualitative in nature then description of data should be done in terms of frequencies and percentages. Usually, editor asks authors to share both frequencies and percentages while writing the results. Commonly used format of the same is writing the frequency with percentages in parentheses or vice versa e.g.

"Prevalence of obesity was 1450 (30.1%) {men: 1030 (35.13%), women: 420 (22.28%)}.⁵

It should also be noted that researcher must not start the sentence integers. If one has to start a sentence with a figure, words should be used. Furthermore, if author is transforming the description of proportions in fractional forms then corresponding frequency should be given in parentheses. An example of the same is:

"More than 80% respondents had known or heard about Tetanus (n= 973). Regarding predisposing factors of Tetanus infection, ... Majority of respondents had known or heard about Rabies (n= 973; 81%).⁶

The author must read the statistical analysis and data description portion mentioned in "Instruction to Authors" of the specific journal. In this section, editor provides details that how many decimal places should be reported with fractional integers.

Quantitative Variable Analysis:

Basic descriptions of quantitative variables are presented in terms of mean \pm standard deviation (SD). Note that it must not be standard error (SE) of mean. Though, when data is skewed then median with inter-quartile range (IQR) is to be reported. It must be clear to the authors the IQR is different from range. Range is a difference between largest and smallest values in a set of values; differing IQR is calculated based on quartiles of a given dataset.⁷ There are several methods to check skewness of the data. Nonetheless, histogram and box plot are used for initial stage detection of the same. If the histogram showed a "bell shaped" curve or middle line in the box of box plot is at center of the box, then variable is non-skewed (symmetric). If such mirror-image pattern is not observed then variable is to be considered as skewed (asymmetrical). Another arithmetical evidence for detecting skewness is standard deviation. If it is at-most three times of its mean, the variable is symmetric. The variable would be considered as skewed if the standard deviation is more than three times of its mean.

Graphical Representation of Data:

Tabular or graphical presentation of results is a great asset that an author can use to present complicated and large volumes of findings. The pie chart and bar chart are used to present proportions and frequencies obtained for qualitative data. Though, pie chart is not much preferable as it represents one variable only. Also, with large number of categories, the pie chart representation becomes quite vague. The colours of the slice/bar of the chart must be light. Since, most of the journals published graphs in black and white colours. Thus, to distinguish diverse categories of the variables, different pattern can be induced while drawing them (Figure).⁸ Data labels must be given in terms of either frequencies or percentages. Both X (horizontal) and Y (vertical) axes must be labeled. The title of the figure should be given beneath the graph. It should be detailed but concise.

Tabular Presentation of Data:

There are many formats for the tabular presentation. Tables containing description of less than three variables are usually not considered good for data presentation. Also, tables having many cells with zero frequencies or having a lot of categories with small number of counts should not be displayed in articles. Too small and too large tables should be avoided. The display of variable's names and categories can be presented in two ways. Examples of presenting such datasets are shown in Tables 1 and 2. Variable names should be given in the first column. In second column, variable's categories should be mentioned (Table-1).⁹ Number of rows of the first and second columns must be merged within corresponding columns and should be equal to the number

Table-1: Example of tabulated data presentation.

		Total (n)	MCC n (%)	Area AKU n (%)
None		67	63(64.5)	-4
Class I-V		24	20(21.8)	-4
Education* Class VI-VII		5	5(5.3)	0
Class IX-X		28	6(6.4)	-22
Class XI-XII		8	0(0.0)	-8
Graduate		64	2(2.2)	-62
Employed*	Yes	21	5(5)	16(16)
	No	175	91(95)	84(84)
Number of children	0-1	64	19(19)	45(45)
	1-2	44	22(23)	22(22)
	2-3	35	16(17)	19(19)
Breast feed^o	3-4	20	15(16)	5(5)
	> 4	33	24(25)	9(9)
	0-1	43	24(25)	19(19)
	1-2	38	14(14.6)	24(24)
Age of the Infants 3-4 (months)	2-3	38	19(19.8)	19(19)
	4-5	28	13(13.5)	15(15)
	5-6	19	9(9.3)	10(10)
		30	17(17.7)	13(13)
Sex of the Infant	Male	119	61(63.5)	58(58)
	Female	77	35(36.5)	42(42)
Place of birth of the Infant*	Home	45	41(42.7)	4(4)
	Hospital	151	55(57.3)	96(96)
Delivery of the infant assisted by*	Dai	40	38(38.5)	2(2)
	Doctor	141	43(44.8)	98(98)
		15	15(15.6)	0(0)

of variable's categories. Another way of presentation is that variable's name and categories are given together in the first column. It should be done in a manner that in one row the name of the variable is given in bold format. Successive rows below mention the categories of that variable in normal text format (Table-2).¹⁰ In later columns statistics for each category should be displayed in both types of presentations. Each cell should contain both frequencies with percentages in parentheses when describing categorical variables. The mean \pm SD or median (IQR)

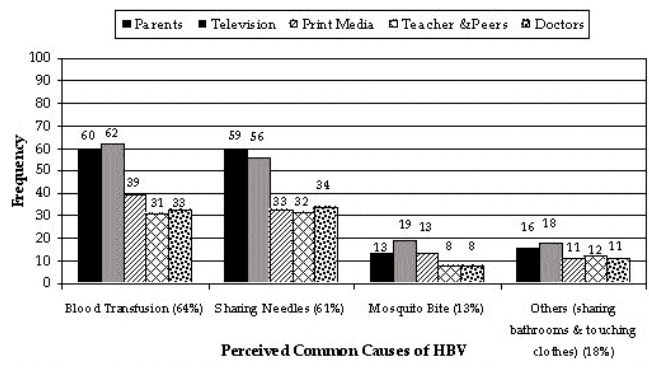


Figure: Bar chart showing different patterns for diverse subcategories.

Table-2: Example of tabulated data presentation using 4 columns.

	Low GI	High GI	P value
SOL (min)	16.3±16.6	21.4±17.1	NS
REM latency (min)	167.3±53.8	153.4±62.5	NS
TST (min)	269.0±28.4	265.0±25.8	NS
TWT (min)	6.1±2.6	7.8±5.1	NS
First half			
Sleep efficiency (%)	92.3±5.7	90.5±5.8	NS
Arousal index (no./h)			
REM	11.5±4.7	11.4±11.6	NS
NREM	9.6±2.3	12.7±4.8	0.049
Total	9.9±2.2	12.9±4.6	0.04
Sleep stage 1 (min)	7.7±5.3	7.4±2.5	NS
Sleep stage 2 (min)	126.9±21.2	138.4±13.8	NS
Sleep stage 3-4 (min)	18.9±10.0	14.2±7.4	NS
REM sleep (min)	25.9±10.0	20.5±10.4	NS
Second half			
Sleep efficiency (%)	95.1±1.6	91.7±7.8	NS
Arousal index (no./h)			
REM	17.5±7.4	14.1±6.7	NS
NREM	11.3±3.6	12.6±4.3	NS
Total	13.2±4.2	13.7±4.1	NS
Sleep stage 1 (min)	12.1±7.9	14.3±11.4	NS
Sleep stage 2 (min)	160.3±21.9	153.5±20.0	NS
Sleep stage 3-4 (min)	11.8±14.0	7.4±6.9	NS
REM sleep (min)	85.4±10.0	86.8±17.8	NS
Changes between first and second halves			
Sleep efficiency (%)	2.8±5.4	1.3±8.6	NS
Arousal index (no./h)			
REM	6.0±5.8	4.6±9.9	NS
NREM	1.7±3.6	0.4±4.8	NS
Total	3.4±4.2	0.8±4.7	NS
Sleep stage 1 (min)	4.4±4.4	6.9±11.6	NS
Sleep stage 2 (min)	33.4±34.8	15.1±27.5	NS
Sleep stage 3-4 (min)	-7.1±19.7	-6.8±7.1	NS
REM sleep (min)	59.5±16.5	66.3±13.8	NS

GI: Glycaemic index; SOL: Sleep Onset Latency; ROL: REM onset latency; SE: Sleep Efficiency (the percentage ratio between total sleep time and time in bed); NREM: Non-rapid eyes movement; REM: Rapid Eyes Movement, NS: Not Significant ($p > 0.05$).

should be presented corresponding to quantitative variable. Long phrases and abbreviations should be referred in footnote of the table with the marking of symbols such as *, †, ‡ etc.¹⁰ The title should be given above the table.

Few words of caution must be shared with the authors here. First, the authors must never repeat the study findings in table or figures if already adequately mentioned in the text of the article. A good researcher first reviews the dataset and checks what can be best presented in figures or tables. Later, the text is written accordingly. Authors need to mention the tables and graphs in the main text of the article.

References

- Altman DG. Statistics in medical journals: some recent trends. *JPMA* 2000; 19: 3275-89.
- Curran-Everett D, Benos DJ. Guidelines for reporting statistics in journals published by the American Physiological Society. *Am J Physiology-Heart and Circulatory Physiology* 2004; 287: H447-9.
- International Committee of Medical Journal Editors. Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Manuscript Preparation and Submission: Preparing a Manuscript for Submission to a Biomedical Journal. (Online) 2009 (Cited 2012 May 6). Available from URL: http://www.icmje.org/manuscript_1prepare.html.
- Bloom BS. Bloom's Taxonomy. In: Victoria Uo, ed volume 2012. Boston: Allyn and Bacon, 2004.
- Ustu Y, Ugurlu M, Aslan O, Aksoy YM, Kasim I, Egici MT, Sanisoglu SY. High prevalence of obesity in Tokat, a northern province of Turkey. *JPMA* 2012; 62: 435-40.
- Wasay M, Malik A, Fahim A, Yousuf A, Chawla R, Daniel H, Rafay M, Azam I, Razzak J. Knowledge and attitudes about Tetanus and Rabies: A population-based survey from Karachi, Pakistan. *JPMA* 2012; 62: 378-82.
- Stat Trek. How to Measure Variability in Statistics. (Online) 2012 (Cited 2012 May 6). Available from URL: <http://stattrek.com/descriptive-statistics/variability.aspx>.
- Thaver AM, Kamal A. Impact of information sources on the knowledge of adolescents about hepatitis B. *JPMA* 2010; 60: 1072-5.
- Ali S, Ali SF, Imam AM, Ayub S, Billoo AG. Perception and practices of breastfeeding of infants 0-6 months in an urban and a semi-urban community in Pakistan: a cross-sectional study. *JPMA* 2011; 61: 99-104.
- Jalilolghadr S, Afaghi A, O'Connor H, Chow CM. Effect of low and high glycaemic index drink on sleep pattern in children. *JPMA* 2011; 61: 533-6.