APPLICATION OF BASIC STATISTICAL TESTS IN RESEARCH AND DATA ANALYSIS IN THE COMMERCE DISCIPLINE

Dr. Alpa Vinodray Joshi
ATMIYA University, Rajkot

Dr. Amisha Ghelani
ATMIYA University, Rajkot

ABSTRACT

Statistical methods in conducting research include planning, design, data collection, analysis, meaningful interpretation, and reporting of research findings. Statistical analysis gives meaning to insignificant numbers and thus breathes life into lifeless data. Results and conclusions are accurate only when appropriate statistical tests are used. The purpose of this article is to introduce the reader to the basic research tools that are used to conduct various studies. The article briefly discusses variables, understanding quantitative and qualitative variables, and measures of central tendency. It provides an overview of sample size estimation, power analysis and statistical errors. Finally, the parametric and non-parametric tests used in the data analysis are summarized.

INTRODUCTION

Statistics is a discipline that deals with collecting, organizing, analysing and drawing conclusions from samples of the entire population. This requires an appropriate research design, appropriate selection of the research sample and selection of an appropriate statistical test. Adequate statistical information is necessary to properly design an epidemiological study or clinical trial. Incorrect statistical methods can lead to wrong conclusions, which can lead to unethical practice.

VARIABLES

Variable is a characteristic that varies from one individual member of population to another individual. Variables such as height and weight are measured by some type of scale, convey quantitative information and are called as quantitative variables. Sex and eye colour give qualitative information and are called as qualitative variables [Figure 1].

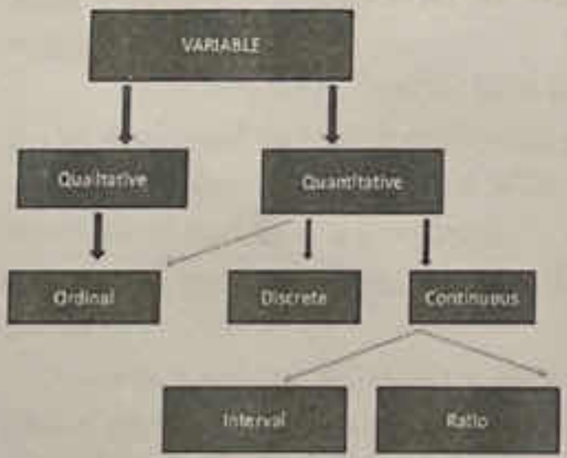

Figure 1
Classification of variables
Quantitative variables

Quantitative or numerical data are subdivided into discrete and continuous measurements. Discrete numerical data are recorded as a whole number such as 0, 1, 2, 3,... (integer), whereas continuous data can assume any value. Observations that can be counted constitute the discrete data and observations that can be measured constitute the continuous data. Examples of discrete data are number of episodes of respiratory arrests or the number of re-intubations in an intensive care unit. Similarly, examples of continuous data are the serial serum glucose levels, partial pressure of oxygen in arterial blood and the oesophageal temperature.

A hierarchical scale of increasing precision can be used for observing and recording the data which is based on categorical, ordinal, interval and ratio scales [Figure 1].

Categorical or nominal variables are unordered. The data are merely classified into categories and cannot be arranged in any particular order. If only two categories exist (as in gender male and female), it is called as a dichotomous (or binary) data. The various causes of re-intubation in an intensive care unit due to upper airway obstruction, impaired clearance of secretions, hypoxemia, hypercapnia, pulmonary oedema and neurological impairment are examples of categorical variables.

Ordinal variables have a clear ordering between the variables. However, the ordered data may not have equal intervals.

Interval variables are similar to an ordinal variable, except that the intervals between the values of the interval variable are equally spaced. A good example of an interval scale is the Fahrenheit degree scale used to measure temperature. With the Fahrenheit scale, the difference between 70° and 75° is equal to the difference between 80° and 85°: The units of measurement are equal throughout the full range of the scale.

Ratio scales are similar to interval scales, in that equal differences between scale values have equal quantitative meaning. However, ratio scales also have a true zero point, which gives them an additional property. For example, the system of centimetres is an example of a ratio scale. There is a true zero point and the value of 0 cm means a complete absence of length. The thyromental distance of 6 cm in an adult may be twice that of a child in whom it may be 3 cm.

STATISTICS: DESCRIPTIVE AND INFERENTIAL STATISTICS

Descriptive statistics are used to describe the relationship between variables in a sample or population. Descriptive statistics summarize data in the form of mean, median, and mode. Inferential statistics use data taken from a random sample of the population to describe the entire population and draw conclusions. This is valuable when it is not possible to look at every member of the entire population. The examples if descriptive and inferential statistics are illustrated in Table 1.

Table I

Example of descriptive and inferential statistics

Descriptive statistics	
	The intracranial pressures (mmHg) of 10 patients admitted with severe head injury in Intensive Care Unit are 13, 32, 35, 42, 30, 19, 32, 27, 36 and 31. These data can be summarised to best represent the observations. We can rank the observations from lowest to highest 13,19,27,30,31,32,32, 35, 36 and 42. We get now a clearer idea of the intracranial pressures in severe head injury. The idea about the commonly observed values 9 (the smaller and larger values less represent our sample)
	The sample mode (most commonly observed value) is 32
	Mean value is 29.7 mmHg
	The median is the middle value. If there is an even number of observations, then the median is calculated as the average of the two middle values. The median is 31+32/2=31.5 mm Hg
Inferential statistics	
	If one plans to study the association of learning disabilities after exposure to anaesthesia before the age of 4 years, it will be feasible to compare the learning disabilities between children who have received anaesthesia and those who have not received anaesthesia
	It is impossible to measure the learning disability in all children of an entire population. However, it is possible to measure the learning disability in a representative random sample in different schools and draw inferences that could be applicable to the whole population

Descriptive statistics

The extent to which observations are clustered around a central location is described by the central tendency, and the spread to the extremes is described by the degree of dispersion. Measures of central tendency Measures of central tendency are mean, median and mode. The mean (or arithmetic mean) is the sum of all points divided by the number of points. Extreme variables can profoundly affect the average. For example, a single patient remaining in the ICU for approximately 5 months due to septicemia may affect the average length of stay in the ICU for patients with organophosphorus poisoning. Extreme values are called outliers.The formula for the mean is

Mean, $$\overline{x} = \frac{\sum x}{n}$$

where x = each observation and n = number of observations. Median is defined as the middle of a distribution in a ranked data (with half of the variables in the sample above and half below the median value) while mode is the most frequently occurring variable in a distribution. Range defines the spread, or variability, of a sample. It is described by the minimum and maximum values of the variables. If we rank the data and after ranking, group the observations into percentiles, we can get better information of the pattern of spread of the variables. In percentiles, we rank the observations into 100 equal parts. We can then describe 25%, 50%, 75% or any other percentile amount. The median is the 50^{th} percentile. The interquartile range will be the observations in the middle 50% of the observations about the median (25^{th} - 75^{th} percentile). Variance is a measure of how spread out is the distribution. It gives an indication of how close an individual observation clusters about the mean value. The variance of a population is defined by the following formula:

$$\sigma^2 = \frac{\sum (X_i - X)^2}{N}$$

where σ^2 is the population variance; X is the population mean, X_i is the i^{th} element from the population and N is the number of elements in the population. The variance of a sample is defined by slightly different formula:

$$s^2 = \frac{\sum (X_i - X)^2}{n-1}$$

where s^2 is the sample variance, x is the sample mean, x_i is the i^{th} element from the sample and n is the number of elements in the sample. The formula for the variance of a population has the value 'n' as the denominator. The expression '$n-1$' is known as the degrees of freedom and is one less than the number of parameters. Each observation is free to vary, except the last one which must be a defined value. The variance is measured in squared units. To make the interpretation of the data simple and to retain the basic unit of observation, the square root of variance is used. The square root of the variance is the standard deviation (SD). The SD of a population is defined by the following formula:

$$\sigma = \sqrt{\left(\sum [X_i - X]^2 / N\right)}$$

where σ is the population SD, X is the population mean, X_i is the i^{th} element from the population and N is the number of elements in the population. The SD of a sample is defined by slightly different formula:

$$s = \sqrt{\left(\sum [X_i - x]^2 / n-1\right)}$$

where s is the sample SD, x is the sample mean, x_i is the i^{th} element from the sample and n is the number of elements in the sample. An example for calculation of variation and SD is illustrated in Table 2.

Table 2

Example of mean, variance, standard deviation

Example: The weight of five patients undergoing laparoscopic cholecystectomy was 90, 90, 70, 70, 80

$$\text{Mean weight} = \frac{90+90+70+70+80}{4} = 80$$

$$\text{Variance} = \frac{(90-80)^2+(90-80)^2+(70-80)^2+(80-80)^2}{5-1}$$

$$= \frac{100+100+100+100+0}{5-1}$$

$$= \frac{400}{5-1}$$

$$= 100$$

$$SD = \sqrt{100} = 10$$

SD – Standard deviation

Normal distribution or Gaussian distribution

Most of the biological variables usually cluster around a central value, with symmetrical positive and negative deviations about this point. The standard normal distribution curve is a symmetrical bell-shaped. In a normal distribution curve, about 68% of the scores are within 1 SD of the mean. Around 95% of the scores are within 2 SDs of the mean and 99% within 3 SDs of the mean [Figure 2].

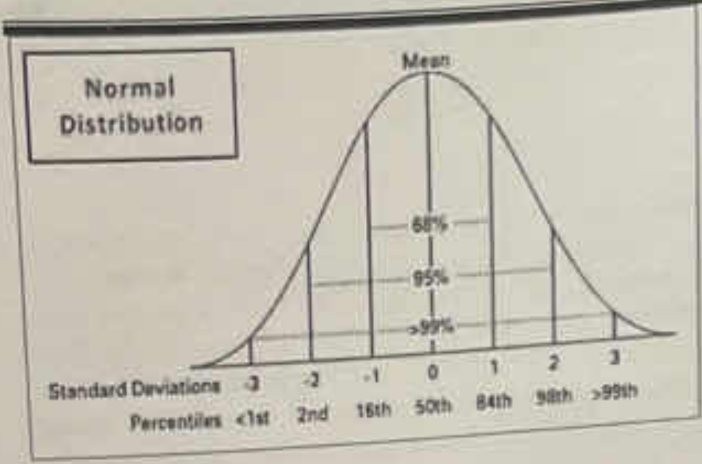

Figure 2

Normal distribution curve

Skewed distribution

It is a distribution with an asymmetry of the variables about its mean. In a negatively skewed distribution [Figure 3], the mass of the distribution is concentrated on the right of Figure 1. In a positively skewed distribution [Figure 3], the mass of the distribution is concentrated on the left of the figure leading to a longer right tail.

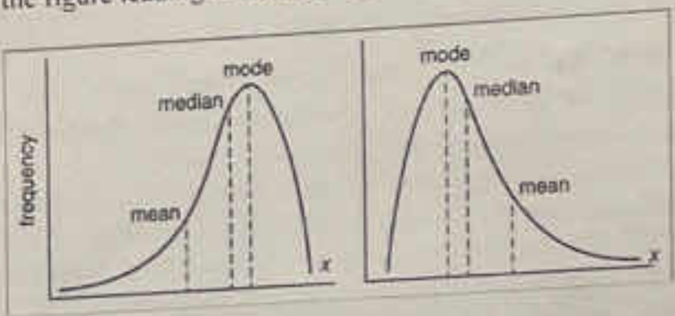

Figure 3

Curves showing negatively skewed and positively skewed distribution

Inferential statistics

In inferential statistics, data are analysed from a sample to make inferences in the larger collection of the population. The purpose is to answer or test the hypotheses. A hypothesis (plural hypotheses) is a proposed explanation for a phenomenon. Hypothesis tests are thus procedures for making rational decisions about the reality of observed effects.

Probability is the measure of the likelihood that an event will occur. Probability is quantified as a number between 0 and 1 (where 0 indicates impossibility and 1 indicates certainty).

In inferential statistics, the term 'null hypothesis' (H_0 '*H-naught*,' '*H-null*') denotes that there is no relationship (difference) between the population variables in question.

Alternative hypothesis (H_1 and H_a) denotes that a statement between the variables is expected to be true.

The *P* value (or the calculated probability) is the probability of the event occurring by chance if the null hypothesis is true. The *P* value is a numerical between 0 and 1 and is interpreted by researchers in deciding whether to reject or retain the null hypothesis [Table 3].

Table 3

P values with interpretation

P	Result	Null hypothesis
<0.01	Result is highly significant	Reject (null hypothesis) H0
≥0.01 but <0.05	Result is significant	Reject (null hypothesis) H0
Value ≥0.05	Result is not significant	Do not reject (null hypothesis) H0

If *P* value is less than the arbitrarily chosen value (known as α or the significance level), the null hypothesis (H0) is rejected [Table 4]. However, if null hypotheses (H0) is incorrectly rejected, this is known as a Type I error. Further details regarding alpha error, beta error and sample size calculation and factors influencing them are dealt with in another section of this issue by Das S *et al.*

Table 4

Illustration for null hypothesis

A study was planned to evaluate if the use of intravenous dexmedetomidine attenuated the haemodynamic and neuroendocrine responses to fixation of skull pin head Holder in patients undergoing craniotomy [18] Sudy patients were randomly assigned, half in each group, to receive a single bolus dose of dexmedetomidine (1 µg/kg) intravenously over 10 min before induction of anaesthesia or normal saline (placebo) in the control group
It is possible for this study to be framed in a particular way that indicates competing beliefs about the drug to be studied in the patient population
First, we assume that dexmedetomidine makes no difference - has no effect. This is called the null hypothesis - the no change hypothesis. The symbol used is H0 – 'H' for hypothesis and '0' for zero change (the word 'null' is another way of saying 'zero')
Second, we set up an alternate hypothesis H_a which takes the opposite point of view - namely use of dexmedetomidine does make a difference (blunts the haemodynamic response)
The data are used to produce a test value - a test statistic - in this case, it measures the heart rate, arterial blood pressure and serial levels of cortisol, prolactin, insulin and blood glucose in each group (dexmedetomidine group and control group)
Then, the process evaluates whether the difference in these parameters between the two groups is significantly large

PARAMETRIC AND NON-PARAMETRIC TESTS

Numerical data (quantitative variables) that are normally distributed are analysed with parametric tests.

Two most basic prerequisites for parametric statistical analysis are:

- The assumption of normality which specifies that the means of the sample group are normally distributed
- The assumption of equal variance which specifies that the variances of the samples and of their corresponding population are equal.

However, if the distribution of the sample is skewed towards one side or the distribution is unknown due to the small sample size, non-parametric statistical techniques are used. Non-parametric tests are used to analyse ordinal and categorical data.

Parametric tests

The parametric tests assume that the data are on a quantitative (numerical) scale, with a normal distribution of the underlying population. The samples have the same variance (homogeneity of variances). The samples are randomly drawn from the population, and the observations within a group are independent of each other. The commonly used parametric tests are the Student's t-test, analysis of variance (ANOVA) and repeated measures ANOVA.

Student's *t*-test

Student's *t*-test is used to test the null hypothesis that there is no difference between the means of the two groups. It is used in three circumstances:

1. To test if a sample mean (as an estimate of a population mean) differs significantly from a given population mean (this is a one-sample *t*-test)

   The formula for one sample *t*-test is $t = \frac{X - u}{\mathrm{SE}}$

   where X = sample mean, u = population mean and SE = standard error of mean

2. To test if the population means estimated *by two independent samples* differ significantly (the unpaired *t*-test). The formula for unpaired *t*-test is:

   $$t = \frac{X_1 - X_2}{\mathrm{SE}_{x_1 - x_2}}$$

   where $X_1 - X_2$ is the difference between the means of the two groups and SE denotes the standard error of the difference.

3. To test if the population means estimated *by two dependent samples* differ significantly (the paired *t*-test). A usual setting for paired *t*-test is when measurements are made on the same subjects before and after a treatment.

The formula for paired *t*-test is:

$$t = \frac{d}{\mathrm{SE}_d}$$

where d is the mean difference and SE denotes the standard error of this difference.

The group variances can be compared using the *F*-test. The *F*-test is the ratio of variances (var l/var 2). If F differs significantly from 1.0, then it is concluded that the group variances differ significantly.

Definition of Z-test

Z-test refers to a univariate statistical analysis used to test the hypothesis that proportions from two independent samples differ greatly. It determines to what extent a data point is away from its mean of the data set, in standard deviation.

The researcher adopts z-test, when the population variance is known, in essence, when there is a large sample size, sample variance is deemed to be approximately equal to the population variance. In this way, it is assumed to be known, despite the fact that only sample data is available and so normal test can be applied.

Assumptions of Z-test:

- All sample observations are independent
- Sample size should be more than 30.
- Distribution of Z is normal, with a mean zero and variance 1.

Key Differences Between T-test and Z-test

The difference between t-test and z-test can be drawn clearly on the following grounds:

1. The t-test can be understood as a statistical test which is used to compare and analyse whether the means of the two population is different from one another or not when the standard deviation is not known. As against, Z-test is a parametric test, which is applied

when the standard deviation is known, to determine, if the means of the two datasets differ from each other.

2. The t-test is based on Student's t-distribution. On the contrary, z-test relies on the assumption that the distribution of sample means is normal. Both student's t-distribution and normal distribution appear alike, as both are symmetrical and bell-shaped. However, they differ in the sense that in a t-distribution, there is less space in the centre and more in the tails.
3. One of the important conditions for adopting t-test is that population variance is unknown. Conversely, population variance should be known or assumed to be known in case of a z-test.
4. Z-test is used to when the sample size is large, i.e. $n > 30$, and t-test is appropriate when the size of the sample is small, in the sense that $n < 30$.

Analysis of variance

The Student's *t*-test cannot be used for comparison of three or more groups. The purpose of ANOVA is to test if there is any significant difference between the means of two or more groups.

In ANOVA, we study two variances – (a) between-group variability and (b) within-group variability. The within-group variability (error variance) is the variation that cannot be accounted for in the study design. It is based on random differences present in our samples.

However, the between-group (or effect variance) is the result of our treatment. These two estimates of variances are compared using the F-test.

A simplified formula for the *F* statistic is:

$$F = \frac{MS_b}{MS_w}$$

where MS_b is the mean squares between the groups and MS_w is the mean squares within groups.

Repeated measures analysis of variance

As with ANOVA, repeated measures ANOVA analyses the equality of means of three or more groups. However, a repeated measure ANOVA is used when all variables of a sample are measured under different conditions or at different points in time.

As the variables are measured from a sample at different points of time, the measurement of the dependent variable is repeated. Using a standard ANOVA in this case is not appropriate because it fails to model the correlation between the repeated measures: The data violate the ANOVA assumption of independence. Hence, in the measurement of repeated dependent variables, repeated measures ANOVA should be used.

Non-parametric tests

When the assumptions of normality are not met, and the sample means are not normally, distributed parametric tests can lead to erroneous results. Non-parametric tests (distribution-free test) are used in such situation as they do not require the normality assumption. Non-parametric tests may fail to detect a significant difference when compared with a parametric test. That is, they usually have less power.

As is done for the parametric tests, the test statistic is compared with known values for the sampling distribution of that statistic and the null hypothesis is accepted or rejected. The types of non-parametric analysis techniques and the corresponding parametric analysis techniques are delineated in Table 5.

Table 5

Analogue of parametric and non-parametric tests

Parametric tests	Non-parametric tests
One sample	
One sample *t*-test	Sign test Wilcoxon's signed rank test
Two-sample	
Paired *t*-test	Sign test Wilcoxon's signed rank test
Unpaired *t*-test	Mann-Whitney U-test Kolmogorov-Smirnov test
K-sample	
ANOVA	Kruskal-Wallis test Jonckheere test
Two-way ANOVA (repeated measure ANOVA)	Friedman test
Pearson correlation coefficient (*r*)	Spearman rank order (ρ)

ANOVA – Analysis of variance

Median test for one sample: The sign test and Wilcoxon's signed rank test

The sign test and Wilcoxon's signed rank test are used for median tests of one sample. These tests examine whether one instance of sample data is greater or smaller than the median reference value.

Sign test

This test examines the hypothesis about the median θ0 of a population. It tests the null hypothesis H0 = θ0. When the observed value (Xi) is greater than the reference value (θ0), it is marked as+. If the observed value is smaller than the reference value, it is marked as − sign. If the observed value is equal to the reference value (θ0), it is eliminated from the sample.

If the null hypothesis is true, there will be an equal number of + signs and − signs.

The sign test ignores the actual values of the data and only uses + or − signs. Therefore, it is useful when it is difficult to measure the values.

Wilcoxon's signed rank test

There is a major limitation of sign test as we lose the quantitative information of the given data and merely use the + or – signs. Wilcoxon's signed rank test not only examines the observed values in comparison with θ0 but also takes into consideration the relative sizes, adding more statistical power to the test. As in the sign test, if there is an observed value that is equal to the reference value θ0, this observed value is eliminated from the sample.

Wilcoxon's rank sum test ranks all data points in order, calculates the rank sum of each sample and compares the difference in the rank sums.

Mann-Whitney test

It is used to test the null hypothesis that two samples have the same median or, alternatively, whether observations in one sample tend to be larger than observations in the other.

Mann–Whitney test compares all data (xi) belonging to the X group and all data (yi) belonging to the Y group and calculates the probability of xi being greater than yi: P (xi > yi). The null hypothesis states that P (xi > yi) = P (xi < yi) =1/2 while the alternative hypothesis states that P (xi > yi) ≠1/2.

Kolmogorov-Smirnov test

The two-sample Kolmogorov-Smirnov (KS) test was designed as a generic method to test whether two random samples are drawn from the same distribution. The null hypothesis of the KS test is that both distributions are identical. The statistic of the KS test is a distance between the two empirical distributions, computed as the maximum absolute difference between their cumulative curves.

Kruskal-Wallis test

The Kruskal–Wallis test is a non-parametric test to analyse the variance. It analyses if there is any difference in the median values of three or more independent samples. The data values are ranked in an increasing order, and the rank sums calculated followed by calculation of the test statistic.

Jonckheere test

In contrast to Kruskal–Wallis test, in Jonckheere test, there is an a priori ordering that gives it a more statistical power than the Kruskal–Wallis test.

Friedman test

The Friedman test is a non-parametric test for testing the difference between several related samples. The Friedman test is an alternative for repeated measures ANOVAs which is used when the same parameter has been measured under different conditions on the same subjects.

Tests to analyse the categorical data

Chi-square test, Fischer's exact test and McNemar's test are used to analyse the categorical or nominal variables. The Chi-square test compares the frequencies and tests whether the observed data differ significantly from that of the expected data if there were no differences between groups (i.e., the null hypothesis). It is calculated by the sum of the squared difference between observed (*O*) and the expected (*E*) data (or the deviation, *d*) divided by the expected data by the following formula:

$$\chi^2 = \sum \frac{(O-E)^2}{O}$$

A Yates correction factor is used when the sample size is small. Fischer's exact test is used to determine if there are non-random associations between two categorical variables. It does not assume random sampling, and instead of referring a calculated statistic to a sampling distribution, it calculates an exact probability. McNemar's test is used for paired nominal data. It is applied to 2 × 2 table with paired-dependent samples. It is used to determine whether the row and column frequencies are equal (that is, whether there is 'marginal homogeneity'). The null hypothesis is that the paired proportions are equal. The Mantel-Haenszel Chi-square test is a multivariate test as it analyses multiple grouping variables. It stratifies according to the nominated confounding variables and identifies any that affects the primary outcome variable. If the outcome variable is dichotomous, then logistic regression is used.

VARIOUS SOFTWARES AVAILABLE FOR STATISTICS, SAMPLE SIZE CALCULATION AND POWER ANALYSIS

Numerous statistical software systems are available currently. The commonly used software systems are Statistical Package for the Social Sciences (SPSS – manufactured by IBM corporation), Statistical Analysis System ((SAS – developed by SAS Institute North Carolina, United States of America), R (designed by Ross Ihaka and Robert Gentleman from R core team), Minitab (developed by Minitab Inc), Stata (developed by StataCorp) and the MS Excel (developed by Microsoft).

There are a number of web resources which are related to statistical power analyses. A few are:

- StatPages.net – provides links to a number of online power calculators
- G-Power – provides a downloadable power analysis program that runs under DOS
- Power analysis for ANOVA designs an interactive site that calculates power or sample size needed to attain a given power for one effect in a factorial ANOVA design
- SPSS makes a program called Sample Power. It gives an output of a complete report on the computer screen which can be cut and paste into another document.

CONCLUSION

It is important that a researcher knows the concepts of the basic statistical methods used for conduct of a research study. This will help to conduct an appropriately well-designed study leading to valid and reliable results. Inappropriate use of statistical techniques may lead to faulty conclusions, inducing errors and undermining the significance of the article. Bad statistics may lead to bad research, and bad research may lead to unethical practice. Hence, an adequate knowledge of statistics and the appropriate use of statistical tests are important. An appropriate knowledge about the basic statistical methods will go a long way in improving the research designs and producing quality medical research which can be utilised for formulating the evidence-based guidelines.

Financial support and sponsorship

Nil.

Conflicts of interest

There are no conflicts of interest.

REFERENCES

1. Winters R, Winters A, Amedee RG. Statistics: A brief overview. *Ochsner J.* 2010;10:213–6.

2. Nahm FS. Nonparametric statistical tests for the continuous data: The basic concept and the practical use. *Korean J Anesthesiol.* 2016;69:8–14.

2. Sprent P. Statistics in medical research. *Swiss Med Wkly.* 2003;133:522–9.

3. Kaur SP. Variables in research. *Indian J Res Rep Med Sci.* 2013;4:36–8.

4. Satake EB. *Statistical Methods and Reasoning for the Clinical Sciences Evidence-Based Practice.* Ist ed. San Diego: Plural Publishing, Inc; 2015. pp. 1–19.

5. Wilder RT, Flick RP, Sprung J, Katusic SK, Barbaresi WJ, Mickelson C, et al. Early exposure to anesthesia and learning disabilities in a population-based birth cohort. *Anesthesiology.* 2009;110:796–804.

6. Manikandan S. Measures of central tendency: Median and mode. *J Pharmacol Pharmacother.* 2011;2:214–5.

7. Myles PS, Gin T. *Statistical Methods for Anaesthesia and Intensive Care.* Ist ed. Oxford: Butterworth Heinemann; 2000. pp. 8–10.

8. Binu VS, Mayya SS, Dhar M. Some basic aspects of statistical methods and sample size determination in health science research. *Ayu.* 2014;35:119–23.

9. Nickerson RS. Null hypothesis significance testing: A review of an old and continuing controversy. *Psychol Methods.* 2000;5:241–301.

10. Mushtaq K, Ali Z, Shah NF, Syed S, Naqash I, Ramzan AU. A randomized controlled study to compare the effectiveness of intravenous dexmedetomidine with placebo to attenuate the hemodynamic and neuroendocrine responses to fixation of skull pin head holder for craniotomy. *North J ISA.* 2016;1:16–23.

11. Bajwa SJ. Basics, common errors and essentials of statistical tools and techniques in anesthesiology research. *J Anaesthesiol Clin Pharmacol.* 2015;31:547–53.

12. Das S, Mitra K, Mandal M. Sample size calculation: Basic principles. *Indian J Anaesth.* 2016;60:652–6.

13. Altman DG, Bland JM. Parametric v non-parametric methods for data analysis. *BMJ.* 2009;338:a3167.

14. Bewick V, Cheek L, Ball J. Statistics review 10: Further nonparametric methods. *Crit Care.* 2004;8:196–9.