INTERNATIONAL JOURNAL FOR INNOVATIVE RESEARCH IN MULTIDISCIPLINARY FIELD ISSN: 2455-0620 Volume - 7, Issue - 5, May – 2021
Monthly, Peer-Reviewed, Refereed, Indexed Journal with IC Value: 86.87 Impact Factor: 6.719
Received Date: 06/05/2021 Acceptance Date: 21/05/2021 Publication Date: 31/05/2021

STATISTICAL DATA ANALYSIS IN CORPORATE RESEARCH: A CRUCIAL PERCEPTIONS

DR. SRINIVASA A. T.
Assistant Professor in Statistics, BLDEA's New Arts College, Tikota, Vijayapura District. Pin code:586130,
Email - srinivasaat1974@gmail.com

Abstract: *The data may be adequate, valid and reliable to any extent; it does not serve any worthwhile purpose unless it is carefully analyzed.There are a number of techniques can be used while analyzing data. These techniques fall into two categories; descriptive and inferential, constituting descriptive and inferential analysis. They can serve many purposes: to summarize the data in a simple manner, to organize it so it's easier to understand and to use the data to test theories about a larger population. Given the ready availability of computer software, tedious formulae and calculations can be avoided today. But, there is no substitute for having a good understanding of the conceptual basis of analytic methodologies that one applies in order to draw inferences about hard-ownresearch data. Hence, an effort has been made in this paper to provide theoretical introduction of few but most widely used analyticaltools, will allow one to produce meaningful data analysis in business research.*

Key Words: *Descriptive analysis, Inferential analysis, Hypothesis Testing, Estimation, Measures of Central Tendency, Measures of Dispersion,Relationship etc.*

1. INTRODUCTION:

Data Analysis is used in many industries to allow companies andorganization to make better business decisions. Data Analysis is one of the many steps that must be completed when conducting a research but it assumes special significance. Data when collected from various primary and secondary sources is in its raw form areincredibly useful but also overwhelming. It is almost impossiblefor researcher to deal with all this data in its raw form. Throughdata analysis such data is presented in a suitable and summarizedform without any loss of relevant information so that it can be efficiently used for decision making. Data can be presented in the tabular or graphic form. The tabular form (tables) implies numerical presentation of data. The graphical form (figure) involves the presentation of data in terms of structure which canbe visually interpreted, e.g., Bar charts, Pie charts, Histograms, Line charts etc.

Analysis of data means studying the tabulated material in order to determine inherent facts or meaning. A plan of analysis can and should be prepared in advance before the actual collection of material.

Processing and analysis of data is always found to be interwoven.Many experts are of the view that analysis of data is different from processing of data. Prof. John Gating had made distinctionbetween analysis of data and processing of data. According to himprocessing of data refers to concentrating, recasting and dealingwith the data so that they are as amenable to analysis as possible,while analysis of data refers to seeing the data in the light of hypothesis of research questions and the prevailing theories anddrawing conclusions that are as amenable to theory formation aspossible. (Gupta 2010). But there are experts who do not like to make difference between processing and analysis. Technically speaking, processing implies editing, coding, classification and tabulation of collected data so that they are amenable to analysis.

2. OBJECTIVES:

The main objective of this paper is to provide a detailed summaryof data analysis, and its uses in understanding the concept of dataanalysis for behavioral research.

3. METHODOLOGY:

For this study, data is collected from Secondary sources and available literature has been reviewed and analyzed for understanding the concept and use of data analysis for behavioral research.

4. DISCUSSION ON DATA ANALYSIS:

A. Definition of Data Analysis

The term analysis refers to the computation of certain measures(such as measures of central tendency, variation etc)

alongwith searching of patterns of relationship (such as correlation, regression) exist among data groups. Apart from that, in the process of analysis, relationships or differences, supporting or conflicting with original or new hypothesis should be subjectedto statistical tests of significance to determine with what validitydata can be said to indicate any conclusions. Analysis, therefore,be categorized as ***descriptive*** and ***inferential analysis*** (inferentialanalysis is also known as ***statistical analysis***). Descriptive analysisdeals with computation of certain indices from raw data with andestablishing relation between two or more variables. Whereas, inferential analysis is concerned with the: (a) the estimation of population parameters, and (b) the testing of statistical hypothesisor test of significance.

B. Goals of Data Analysis:

1. Giving a feel to research data: After data collection, the first steptowards understanding the huge mass of data has been gathered,is to arrange the materials in a concise and logical order. The procedure is referred to as classification and tabulation of data. However, these forms of presentation may not be very interestingto the common man. Too many figures are often confusing and may fail to convey the message effectively for which they are meant. Hence, another important convincing and easily understoodmethod of presenting the data is the use of graphs and diagrams.Constructing tables and graphs for the concerned data is a majorpart of analysis, which will facilitate in better understanding and comparison of data.

2. Identifying average values and variability: Most of the researchstudies result in a large volume of raw data which must be suitablereduced so that the same can be read easily and can be used for further analysis. One of the most important goals of data analysisis to get one single value that describes the characteristic of the entire mass of unwieldy data. Such a value is called central value or average values. The most important averages are mean,median and mode. The various measures of average values alone cannot adequately describe a set of observations, unless all the observations are same. To identify the measurement of scatteredness or variability of the mass of data in a series from the average is equally important to describe the data.

3. Identifying relation between variables: One of the ways that one can get better insights into the data is by discovering that variables are related to each other i.e. with increase in one variablethere is an increase in other and vice versa. Also effort is made to know the cause and effect relation between two or more thantwo variables.

4. To make inferences about population parameter: In most of theresearch studies, it is not possible to enumerate whole populationin the study. Hence, a part of the population i.e. sample is taken toconsider for the study. One of the goals of data analysis of thesesamples is to use information contained in sample of observation(such as sample mean, sample standard distribution) for drawingconclusion or making inference about the larger population (suchas population mean, standard deviation etc).

5. To test the hypothesis: A statistical hypothesis is an assumptionabout any aspect of a population. For e.g., there is no relationshipbetween compensation and job satisfaction (i.e. null hypothesis)Analysis of data is carried out to test a hypothesis on the basis of sample values, so that hypothesis can be accepted or rejected.Ultimate decisions are taken on the basis of the collected information and the result of the test.

C. Types of Analysis:

As mentioned earlier, in section 1, statistical analysis can be categorized into descriptive and inferential analysis.

1. Descriptive analysis: is mostly concerned with computation of certain indices or measures from the new data. Zikmund has quoted "…with descriptive analysis, the raw data is transformedinto a form that will make them easy to understand & interpret."It is largely the study of distributions of one variable. This sort ofanalysis can be analyzed data in three different ways: Univariate analysis: When a single variable is analyzed alone, of age group, it is known as bivariate analysis.\

2. Multivariate analysis: In multivariate analysis, three or morevariables are investigated simultaneously, allowing us to consider the effects of more than one variable at the same time. For example, identifying job satisfaction in terms of age, sex, salary and so on. Multivariate analysis includes techniques like multiple regression analysis, multiple discriminant analysis, multivariate analysis of variance (MANOVA) , factor analysis and canonical analysis. Someof these terms are briefly described in upcoming sections.

$y = a + bx$ *OR* $x = a + by$ are two possible regression equationsin case of two variables involved in regression analysis.

First equation said as *regression equation of y on x* and so on. Infirst equation y and in second equation x is dependent variable, whereas x in first equation and y in second equation is independentvariable. Here, 'a' and 'b' are constants, 'a' is *intercept* and 'b'is *slope* or *inclination* or most popularly known as ***regression coefficient.*** Regression coefficient gives the change in dependent variable when independent variable changes by 1 unit. To estimate the relationship between x and y it is vital to determine 'a' and 'b' respectively. This is done through the ***Principle of Least Squares Method***. Apart from that the Principles of Least Squares provide criterion to select "line of best fit" mentioned in the last paragraph.

In case of *ordinal data* suchcorrelation can be enumerated by ***Kendall partial rank correlation*** & in case of *nominal data* ***discriminant analysis*** is used. (See Table 1)

Table 1: Choice of relationship analysis tool based on number ofvariables and scale of measurement

For two variables (i.e.simple correlation) For interval or ratio data For ordinal data For nominal data	Pearson product moment correlation coefficient Spearman rank ordercorrelation coefficient or Kendall Tau rank correlation Contingency coefficient
For more than two variables (i.e. multiple correlation) For interval or ratio dataFor ordinal data For nominal data	Multiple regression analysis Kendall partial rankcorrelation Discriminant analysis NA

Source: Compiled by Authors

Inferential Analysis: Inferential analysis is mainly concernedwith (a) estimation of population values such as population mean,population standard deviation, and (b) various tests of significance/testing of hypothesis. Inferential analysis plays a major role in statistics since mostly it is not possible to go for whole populationwhile conducting researches, hence, a sample is chosen and usinginferential analysis the sample values obtained are used to infer about the population. The objective of inferential analysis is to use the information contained in a small sample of observationsfor drawing a conclusion or making an inference about the largerpopulation. Such inference may be in the form of estimation or Testing of Hypothesis or assumptions. For example, either one could estimate population parameter based on sample statistic, like ' mean life of a car battery' , or one could test the claim of company that 'mean life of car battery is 3 years'. In both the cases an inference about population is made.

methods of estimation and testing of hypothesis.

a) Estimation: It deals with the estimation of parameters such aspopulation mean based on the sample values. The method or ruleof estimation is called an ***estimator*** like sample mean, the valuewhich the method or rule gives in a particular case is ***estimate*** ofpopulation parameter. In other words, estimator is a function ofsample values to estimate a parameter of population. With the helpof samples of observation, an estimate in the form of a specific number like 25 years can be given or in the form of an interval 23-27 years. In the former case it is referred as ***point estimate***, whereas in the latter case it is termed as ***interval estimate.***

i. Point estimate: It is used to estimate a population parameter, with the help of sample of observations. Apoint estimate is a single value, say 50. This number is taken as the best value of unknownpopulation parameter. An estimator is said to be efficient, if it has minimum variance such as sample arithmetic mean.

There are several methods of estimating the parameters of a distribution, such as, maximum likelihood, least squares, methodsof squares and minimum chi-square.

ii. Interval estimate: Point estimate gives a single value, taken asbest estimate of parameter. However, if another data is collectedfrom same population, the point estimate may change. In real life situation population parameter may not be exactly equal to sample statistic, and could be around this value. ***b). Testing of Hypothesis/Test of Significance***: In most of the cases, it is almost impossible to get knowledge about populationparameter, therefore, hypothesis testing or test of significance isthe often used strategy for deciding whether sample offers suchsupport for a hypothesis or assumptions that generalizations about population can be made. In other words, test can find the probability that a sample statistic would differ from a parameteror another sample.

Hypothesis testing typically begins with some assumptions or hypothesis or claim about a particular parameter

of a population.It could be the parameters of a distribution like mean, describingthe population; the parameters of two or more population, correlations or associations between two or more characteristics of a population. Hypothesis can be of two types, null and alternativehypothesis. Null hypothesis is considered to be a hypothesis of "norelationship". Such as 'there is ***no significant*** difference between sample means'. The term Null hypothesis is said to have been introduced by R. A. Fisher.

Once the null and alternative hypothesis has been set up, the next step is to decide on the level of significance. It is used asa criterion for rejecting the null hypothesis. It is expressed as a percentage like 5% or 1%, or sometimes as 0.05 or 0.01. It is that level, at which we are likely to reject null hypothesis evenif it is true. Now decision on the appropriate statistic such as t, z, f etc is taken. Based on the level of significance critical or tabulated value is found.

Whenever we take a decision about population based on sample,the decision cannot be 100% reliable. The possibilities can be, we would reject null hypothesis even if it is true, termed as ***TypeI error, denoted as α*** or we could accept the null hypothesis evenif it is false, termed as ***Type II error, denoted as β.***

Type I error is also referred as *level of significance*, as discussedabove. The quantity ***1- β*** is called the ***'power' of test,*** signifyingthe test ability to reject null hypothesis when it as false, and ***1- α***is called ***confidence coefficient***.

Various tests of significance have been developed to meet various types of requirements. They may be heavily classified into, parametric and non-parametric tests. ***Parametric tests*** are basedon the assumptions that the observations are drawn from a normal distribution. Since the testing procedure requires assumptions about the type of population or parameters values these tests areknown as 'parametric tests'. The test of significance developed for situations when this condition is not satisfied, known as '***non- parametric tests***' or '***distribution-free tests'.*** As a matter of fact,parametric tests are more powerful test than non-parametric tests.

Table 2: Choice of parametric/non-parametric test based onfunction to perform & scale of measurement

Function	**Non-Parametric Tests (Interval/RatioData)**	**Parametric tests (Ordinal/ NominalData)**
Test of Significance ofone sample test	*'t'(mean known,S.D.* unknown)* *'z'(mean known, S.D. known)*	*Sign test*
Test of Significance for differencebetween two independent sample	*Independent samples 't' test (S.D. unknown)* *'z' test (S.D.known)*	*Kolmogorov- Smirnov two sampletest* ***or*** *Mann Whitney U Test* ***or*** *Wilcoxon Sum of Rank Test (Ordinal Data)* ***&*** *Chi-Square test (Nominal Data)*
Test of Significance for difference between two paired samples (series of samples taken from same population)	Paired 't' test	*Wilcoxon matched-pairs signed ranks test (ordinal data)* ***&*** *Mc Nemar test for the significance of changes.(nominaldata)*
Test of Significance for difference between seriesof independentsamples	*'F' Test & Oneway ANOVA (Analysis of Variance)*	*Kruskal-Wallis Rank SumTest – HTest* ***or*** *Wilcoxon-Wilcox multiple comparison test (ordinal data)* ***&*** *chi-sqaure for k independent samples (nominaldata)*

Source: Compiled by Authors
(*S.D. is standard deviation of population)

Apart from above functions Chi-square test , denoted as usedas test of goodness of fit i.e. how well observed values fits withof expected values and test of independence i.e. it tests existing association between two categorical variables. Non- parametricalternative for □□ test as goodness of fit is *Kolmogorov-Smirnovtest.* Apart from that, 't' test is used for testing significance of correlation & regression coefficient or slope; 'F' test & ANOVA isused for testing significance of multiple regression coefficient.

5. CONCLUSIONS:

For any successful study, analysis of data is one of the most crucialstep. It is always advisable that it should be designed before thedata are actually collected. Otherwise there is always danger of being too late and the chances of

missing out relevant facts. Thereare a number of analytical tools can be used for summarizing thedata and inferring about the population based on sample values. But to use any tools some of the assumptions have to be fulfilled,therefore, this assumptions always have to be kept in mind by theresearcher before applying any analytical tool. That is why it is said that analysis requires a lot of experience and knowledge in the field of data analysis.

REFERENCES:

1. Bhattacharyya, D. K.(2006). Research Methodology, New Delhi, Excel Books, pp. 127- 166
2. Dunn, O.J. & Clark, V.A.(2009). Basic Statistics-A Primerfor the Biomedical Sciences, New Jersey, John Wiley & Sons,pp. 50-58 & 189-198.
3. Gupta, S.(2010). Research Methodology, New Delhi, Deep& Deep Publications Pvt. Ltd, pp. 192-209
4. Kumari, A.(2008). An Introduction to Research Methodology,Udaipur, Agrotech Publishing Academy, pp. 79-112.
5. Kumar,C.R.(2008). Research Methodology, New Delhi, APHPublishing Corporation, pp. 113-131
6. Kothari, C.R.(2009). Research Methodology-Methods & Techniques, New Delhi, New Age International (P) Limited,pp. 122-149.
7. Levine, D. M., Krehbiel, T.C., Be