Data Analysis of Students Marks with Descriptive Statistics

Dr. Manju Kaushik [#1], Bhawana Mathur [#2]
Associate Professor [#1], Research Scholar [#2]
Department of Computer Science[#1], Department of Computer Science [#2]
JECRC University, jaipur[#1], JECRC University, jaipur[#2]
India[#1], India [#2]
[1] manju.kaushik@jecrcu.edu.in,
[2] bhawanamathur19@gmail.com

Abstract— Improving student's academic performance is not an easy task for the academic community of higher learning. Descriptive statistics is the discipline of quantitatively describing the main features of a collection of information, or the quantitative description itself.
We will be collecting the marks of 100 Students as population and here only marks of 20 students are shown as a sample and apply some measures that can be used like are commonly used to describe a data set. There are different measures of central tendency and measures of variability and dispersion. Measures of central tendency include the mean, median and mode, while measures of variability include the standard deviation (or variance), the minimum and maximum values of the variables, kurtosis and skewness.

Keywords— *Descriptive statistics, central tendency, measures of variability, dispersion, kurtosis and skewness.*

*****

I. INTRODUCTION

Descriptive statistics provide simple summaries about the sample and about the observations that have been made. Such summaries may be either quantitative, i.e. Summary statistics, or visual, i.e. Simple-to-understand graphs. These summaries may either form the basis of the initial description of the data as part of a more extensive statistical analysis, or they may be sufficient in and of themselves for a particular investigation.

Statistics is concerned with the scientific method by which information is collected, organized, analyzed and interpreted for the purpose of description and decision making.

Examples using statistics are: Hang Seng Index, Life or car insurance rate, Unemployment rate, Consumer Price Index, etc.

There are two subdivisions of statistical method.

(a) Descriptive Statistics - It deals with the presentation of numerical facts, or data, in either tables or graph form, and with the methodology of analyzing the data.

(b) Inferential Statistics - It involves techniques for making inferences about the whole population on the basis of observations obtained from samples.

Some Basic Definitions

(a) Population - A population is the group from which data are to be collected.

(b) Sample - A sample is a subset of a population.

(c) Variable - A variable is a feature characteristic of any member of a population differing in quality or quantity from one member to another.

(d) Quantitative variable - A variable differing in quantity is called quantitative variable, for example, the weight of a person, number of people in a car.

(e) Qualitative variable - A variable differing in quality is called a qualitative variable or attribute, for example, the color, the degree of damage of a car in an accident.

(f) Discrete variable - A discrete variable is one which no value may be assumed between two given values, for example, the number of children in a family.

(h) Continuous variable - A continuous variable is one which any value may be assumed between two given values, for example, the time for 100-meter run.

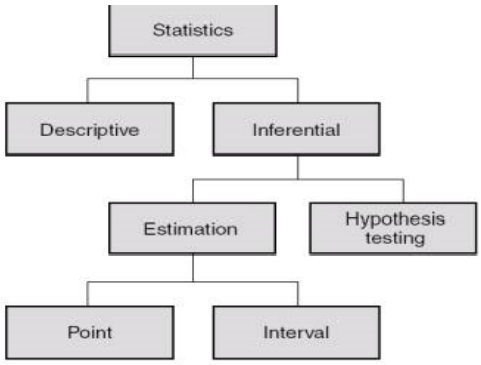

Fig.1 Major divisions in the field of Statistics

2. LITERATURE REVIEW

Experiential learning philosophy underlies simulation education, in which students are able to develop and refine knowledge. Simulation debriefing guides students through a reflection on what occurred during a simulation scenario, with the goal of developing the knowledge and skills [12] (Mark A. Neill, Karen Wotton).

Studies have shown (demos, Garfield, Ooms, and Chance 2007) that students' abilities to describe and interpret a variable's distribution from a histogram, in the context of the data, is

quite high even before taking a first course in statistics. While qualitative descriptions of a Distribution are helpful for summarizing a data set, students eventually will be asked to use statistics to numerically describe a distribution in terms of center, variability, and shape. Without difficulty, they can see how the mean, median, and mode can indicate the center, and how standard deviation and range can describe variability. But the terms skewness and kurtosis are non-intuitive. Worse, skewness and kurtosis statistics and formulas are opaque to the average student, and lack concrete reference points.

A word about kurtosis is in order. Horswell and Looney (1993, p. 437) note that "The performance of skewness tests is shown to be very sensitive to the kurtosis of the underlying distribution." Few instructors say much about kurtosis, partly because it is difficult explains, but also because it is difficult to judge from histograms. Kurtosis is essentially a property of symmetric distributions (Balanda and MacGillivray 1988).

Data sets containing extreme values will not only be skewed, but also generally will be leptokurtic. We cannot therefore speak of non-normal Skewness as if it were separable from non-normal kurtosis. The best we can do is to focus on the skewness statistic simply as one test for departure from the symmetric normal

Distribution.

Cobb and Moore (1997, p. 803) note that "In data analysis, context provides meaning." Realizing this, over the past several decades, more and more instructors are using sample data arising from real (or realistic) scenarios. One result is that students are learning that perfectly symmetrical graphical displays are hard to find. Even with the ability to verbally describe a distribution from a visual display, researchers have found (delMas et al. 2007) that students cannot translate their understanding of shape when asked to compare numerical statistics such as the mean and median.

Hence, measures of skewness are becoming more important (although many instructors may reasonably conclude that kurtosis does not deserve extended discussion in a basic statistics class)

3. **Proposed Approach**

1. Univariate Analysis

We will use Univariate analysis which is involves describing the distribution of a single variable, including its central tendency (including the mean, median, and mode) and dispersion (including data-set, and measures of spread such as the variance and standard deviation). The shape of the distribution may also be described via indices such as skewness and kurtosis. Characteristics of a variable's distribution may also be depicted in graphical or tabular format, including histograms and stem-and-leaf display.

2. Bivariate Analysis

When a sample consists of more than one variable, descriptive statistics may be used to describe the relationship between pairs of variables. In this case, descriptive statistics include:

- Cross-tabulations and contingency tables
- Graphical representation via scatter plots
- Quantitative measures of dependence
- Descriptions of conditional distributions

The main reason for differentiating univariate and bivariate analysis is that bivariate analysis is not only simple descriptive analysis, but it also describes the relationship between two different variables. [5]

3. Advantages of Descriptive statistics

Descriptive statistics is a powerful beast of burden:

(1) It collects and summarizes vast amounts of data and information in a manageable and organized manner,

(2) A fairly straightforward process that can easily translate into results in a distribution of frequency, percents and overall averages.

(3) Establishes standard deviation,

(4) It's used when it may not be desirable to develop a complex Research models,

(5) Deals with immediate data and single variables rather than trying

to establish conclusions,

(6) Can identify further ideas of research,

(7) A good primer to learn about statistical processes, and

(8) Can lay the groundwork for more complex statistical analysis.

TABLE I:

FROM 100 STUDENTS MARKS AS A POPULATION, HERE ONLY MARKS OF 20 STUDENTS SAMPLE

Case	Students Marks (20)
1	20
2	19
3	17
4	18
5	17
6	17
7	17
8	17
9	17
10	18
11	18
12	19
13	17
14	19
15	18
16	17
17	19
18	18
19	19
20	16.8

 ISSN: 2321-8169
1188– 1190

Analyze -> Descriptive Statistics -> Descriptives

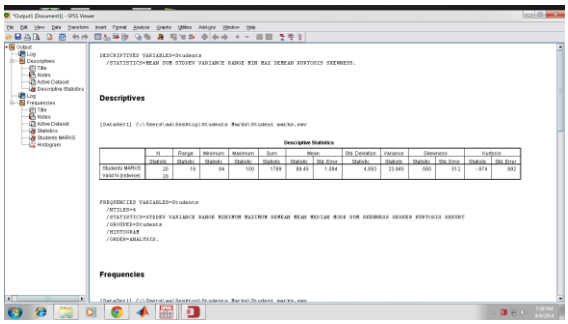

Fig.2 Descriptives

Analyze - > Descriptive Statistics - > Frequencies

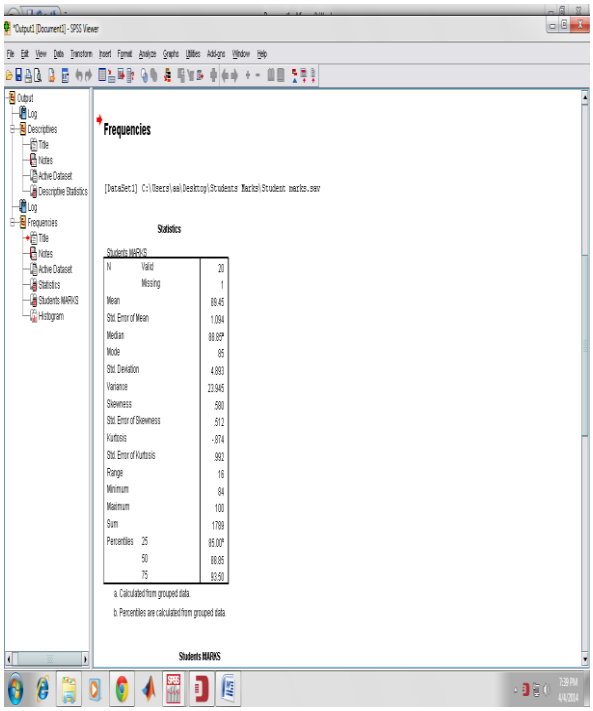

Fig.3 Frequencies

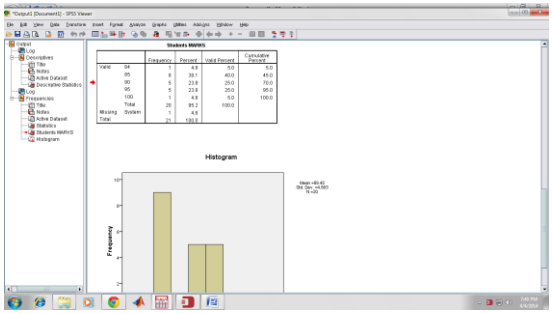

Fig.4 Students Marks

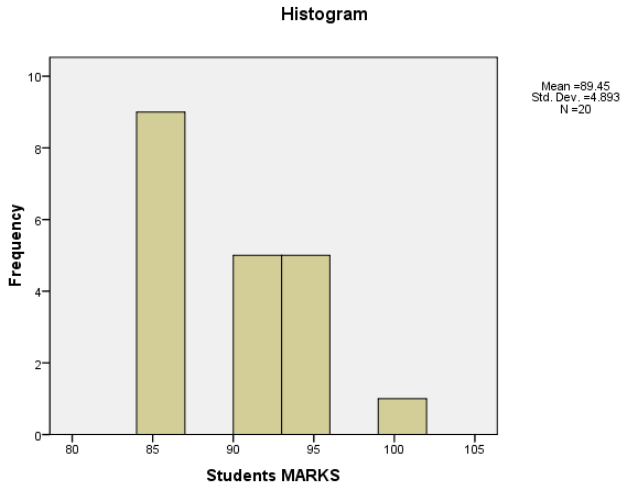

Fig.5 Histogram: Students Marks vs Frequency

TABLE III

Mean and Standard Deviation and Number of Students

Mean	Std. Deviation	N
89.45	4.893	20

TABLE IIIII

Frequencies of 20 Students Results

Mean	89.45
Std. Error of Mean	1.094
Median	88.85
Mode	85
Std. Deviation	4.893
Variance	23.945
Skewness	0.580
Standard Error of Skewness	0.512
Kurtosis	-0.874
Standard Error of Kurtosis	0.992
Range	16
Minimum	84
Maximum	100

RESULT

Measures of central tendency include the mean is 89.45, median is 88.85 and mode are 85, while measures of variability include the standard deviation is 4.893 and the variance is 23.945, the minimum is 84 and maximum values of the variables are 100, kurtosis are -0.874 and skewness. are 0.580

CONCLUSIONS AND FUTURE WORK

In this study, we make use of descriptive statistics in the student's database this technique to predict the student's learning activities. The information generated after the implementation of descriptive statistical technique may be helpful for the instructor as well as for the students. This work may improve student's performance; reduce failing ratio by taking appropriate steps at the right time to improve the quality of education.

For future work, we will refine our technique in order to get more valuable and accurate outputs that would be, useful for instructors to improve the students learning outcomes.

REFERENCES

[1] R. Serfling, "Robust, Affine Invariant, Computationally Easy Nonparametric Multivariate Outlyingness Functions", Session on "Data Depth Methods in Data Mining, Clustering, and Beyond", *International Conference on Nonparametric Statistics and Statistical Learning*, Ohio State University, May 2010.

[2] L Hartling, A. Chisholm, D. Thomson, DM Dryden, "A Descriptive Analysis of Overviews of Reviews ", Health Care Literature, Published between 2000 and 2011.

[3] A. Field Miles, J., & Z. Field, "Discovering statistics using R.", *American Psychological Association*, Washington, DC, 2012.

[4] De Veaux, RD, Velleman, PF & Bock, DE, *Intro Stats*, 4th edn, Pearson Addison Wesley, Boston. , 2009.

[5] Argyrous, G. Statistics for Research: With a Guide to SPSS, SAGE, London, ISBN 1-4129-1948-7, 2005

[6] R. Levesque, *SPSS Programming and Data Management: A Guide for SPSS and SAS Users*, Fourth Edition, SPSS Inc., Chicago, Ill. ISBN 1-56827-390-8, 2007.

[7] SPSS 15.0 Command Syntax Reference, SPSS Inc., Chicago, Ill, 2006.

[8] Wellman, B. "Doing It Ourselves: The SPSS Manual as Sociology's Most Influential Recent Book." pp. 71–78 in *Required Reading: Sociology's Most Influential Books*, edited by Dan Clawson. Amherst: University of Massachusetts Press, ISBN 978-1-55849-153-3, 2006.

[8] R. Serfling, "Advances in Nonparametric Multivariate Statistics in Bootstrap and Data Depth", *International Indian Statistical Association (IISA) Conference*, Riverside, July 2014

[9] J. B. Davis, and J. W. McKean, "Rank based methods for multivariate linear models", *Journal of the American Statistical Association*, Vol. 88, pp. 245-251, 1993.

[10] T.P. Hettmansperger, and R. H. Randles," A practical new equivalent multivariate median", *Journal of Biometrika*, Vol. 89, pp. 851-860, 2002.

[11] R. A. Johnson, and D. W. Wichern, Applied Multivariate Statistical Analysis, 4th Ed., Upper Saddle River, New Jersey: Prentice Hall, 2002.

[12] Mark A. Neill, " High-Fidelity Simulation Debriefing in Nursing Education: A Literature Review", *Clinical Simulation in Nursing*, Volume 7, Issue 5, Pages e161–e168, September 2011

[13] A Haarala-Muhonen, M Ruohoniemi**,** **"**Comparison of students' Perception of their teaching –learning environment in three professional Academic disciplines: A valuable tool for quality enhancement- Learning Environments ", *International Journal of Springer*, 2011.

[14] David P. Doane, Lori E. Seward, "Measuring Skewness: A Forgotten Statistic", *Journal of Statistics Education*, Volume 19, Number 2, 2011,