Determination of Optimal Number of Class Interval and Class Size upon Certain Frequency Distribution in Medical Research

Authors: Dharmendra Kumar Dubey[1], Pramod Kumar Mishra[2]

Affiliation:[1]Associate Professor, School of Allied Health Sciences, Sharda University, Greater Noida, Uttar Pradesh, India.Email: dubey.dharm@gmail.com[2]Associate Professor, Symbiosis Institute of Health Sciences, Symbiosis International (Deemed University), Maharashtra, Pune, India.Email: drpramod@sihspune.org

Corresponding Author: Dr. Dharmendra Kumar Dubey, [1]Associate Professor, School of Allied Health Sciences, Sharda University, Greater Noida, Uttar Pradesh, India.

Email: dubey.dharm@gmail.com, dharmendra.kumar@shrda.ac.in

Abstract

Introduction: Calculation of exact class interval and class size is an important part of data visualization and data summarization. **Objective:** To determine the accurate class interval and its size. **Methods:** It is very important to understand that different frequency distributions are given and need to determine accurate class interval and its size; it cannot be used randomly in all frequency data set. IQ level of 100 hospital and health care management students were considered to find the appropriate number of the class and size of the class. **Results:** Here, author tried to educate researcher regarding the method of class interval (it may be exclusive, inclusive or open-ended) and its size calculation for different data set. **Conclusion:** In this study, number of class interval and size calculated with help of frequency distribution for given data set and for public health and other studies, readers may explore to read other sources as well.

Keywords: Class interval, Class size, Inclusive, Exclusive.

Introduction

An appropriate descriptive statistics need to apply correctly on different kind of the frequency distribution. As we know, evidence-based research required in this era and application of 'descriptive statistics' increased the scrutiny of the data. Research is as good as the evidence referred and it's justified the research, which claims to make the statistical accuracy. Calculation of the class interval and class size for ungrouped data set, during the analysis phase of the study, medical researcher faced the issues, which is the important point. In the published studies (National and International Journal), calculation of the class intervals and size of the class intervals are randomly reported, which needs to apply appropriately. The optimal number of the class denoted by 'K' for closed number of frequency distribution was conducted based on the 'maximum entropy concept' **(1).**

Calculation of the optimal number of the class and size of the class for given frequency distribution is very important to understand the appropriate technique. Equal size of theclass cannot be used for all kind of the data set; it may vary according to frequency distribution. In

this article, appropriate techniques for analysing the “number of class” and “size of the class” are discussed based on frequency distribution of given data set. Here, medical researchers are advised to understand the basics and prerequisites for calculation of the number and size of the class **(2)**. Objective of this study is to find the appropriate number of the class andfind the appropriate size of theclassin medical research forthe collected data set for the research purpose.

Methods

For determining the desirable number of group or class interval, a methods suggested by ‘Sturges’ for given frequency distribution to obtain appropriate groups or classification. Here, aim of the study was to obtain the appropriate number of class and size of the class. When the frequency distribution of observed data or sampled datawascollected for the research study. Therefore, for obtaining the authentic number and size of the classhad discussed in details.

The number of class intervals can vary and it may be 5, 10 and 15. Ideally, minimum 5 and maximum 15. To obtain the appropriate ‘number of classes or group’ for given frequency distribution it can be determined by the formula, which is

$$K = 1 + 3.322 \log_{10} N \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

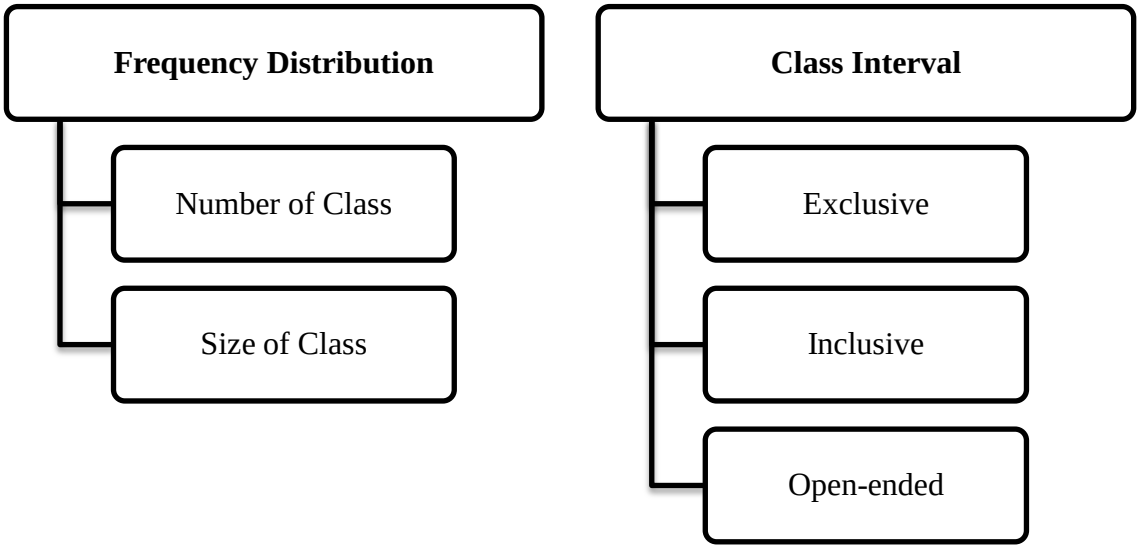
Where, N is known as number of sample size, ‘log’ is ‘logarithm’ of number and ‘K’ is number of class interval **(3)**.

The size of the class varies due to number of observations or sample and number of the class interval. It is completely dependent on the collected data under the proposed study with scientific and appropriate statistical methodology. Below given the formula, which will give us actual size of class.

$$\text{Size of the class } = c = \frac{\text{Range}}{\text{Class interval}} \ldots\ldots\ldots\ldots\ldots\ldots (2)$$

The above given equation (1) and (2) have been applied on given hypothetical data set which gives us actual number and size of the class after that we could classified exclusive class interval and inclusive class interval for proposed study during data analysis. In this study, a hypothetical IQ level of 100 hospital and health care management students were considered to find the appropriate number of the class and size of the class.

Flow Diagrams



Results and discussion

To determine the number of classes, class size and types of class intervalthe following tabulated data set considered.

Table-1 shows the IQ level.

93	127	82	114	118	116	93	100	93	118
118	113	92	80	105	110	127	140	140	118
139	94	110	98	130	130	135	130	102	140
136	98	114	126	101	111	140	135	119	121
92	110	120	100	122	134	119	117	136	97
115	86	85	80	111	101	115	137	109	111
81	90	130	133	119	129	110	135	124	87
107	103	135	117	113	84	100	100	100	95
97	85	114	92	133	94	94	90	95	92
115	140	127	84	135	82	108	100	139	106

Therefore, in given frequency distribution totalsample size was 100. To determine the number of classes, the requiredformula was given in equation (1) and it was used to determine for number of classes, which can be calculated as for putting the value of 'N'

$$'K' = 1 + 3.322 \log 100 = 7.644 \approx 8 \text{ (approx.)}$$

Therefore, if the size of the frequency distributionwasvaryingthen the number of classes will be varying accordingly. Now, questions came, when we have calculated class interval and it was 8approximately, then what will be the class size for the obtained number of class intervals for the same frequency distribution, for thatequation (2) was applied to determine the size of class (C), which is

$$C = \frac{\text{Range}}{\text{Number of the class interval}}$$

$$= \frac{(140-80)}{8}$$

$$= \frac{60}{8} = 7.5 = 8 \text{ (approx)}$$

After finding the "number of the class" and "size of the class", researcherneeds to put the frequency for each class and the data can be in grouped from ungrouped frequency distribution, which can be exclusive class interval, inclusive class interval and open-ended class intervals, all the points are given in the table -2.

Table-2

Exclusive Class-interval	Frequency	Inclusive Class-interval	Frequency
80-88	11	80-88	11
88-96	14	89-96	14
96-104	14	97-104	14
104-112	12	105-112	12
112-120	18	113-120	19
120-128	08	121-128	07
128-136	13	129-136	15
136- 144	10	137- 144	08
Total	**100**	**Total**	**100**

There were 'strict rule' to decide the exact "number of classes" and its size. The number of class intervals can vary and it may be 5, 10 and 15. Ideally, minimum 5 and maximum 15. For getting, 5, 10, 15, we may take nearest number of calculated number of class or size of the class.

Exclusive class interval,

In the given frequency distributions, there were eight class intervals determined with help of appropriate defined techniques. For example, if we take first class interval (80 – 88), in this interval two points are given, first 80 was known as "lower limit" of the first class interval and 88 upper limit of first class interval. Therefore, if the number of classes were fixed, in that case "upper limit" of the first class would be "lower limit" of next class, this process was known as "exclusive class interval".

Inclusive class interval

In given frequency distributions with help of appropriate techniques eight number of classes and size of the classeswere obtained. As given in table 2, first class interval was (80-88) and second was (89- 96), here upper and lower "limit of the class" belongs in the same class, in

other context ‘overlapping was avoided and “lower and upper limits” were included in the same class, which was known as inclusive class interval of given frequency distribution.

Open-ended class interval

If the first class interval lower limit of the class is missing and last class interval upper limit is missing or group data set minimum and maximum both points are missing. In given frequency distribution may be some values are extremely small or large in comparison to others data set. In that case we do not use “lower limit” of the first class and “upper limit” of the last class, such process known as the open ended class interval.

Conclusion

The present study was found that “number of classes” and “size of classes” neither be ‘too large’ nor ‘too small’. It is preferred between 5 and 15. However, the “number of classes” may be greater than 15, it was depending on required details of the ‘frequency distribution’. If the ‘number of classes’ were fixed and it was divided by the difference between minimum and maximum value of the frequency distribution (Range), it’s nearest integer value gives the magnitude of the class interval which was known as the “size of class”. If the “size of class” was broader then number of classes lesser, it gives us a rough estimate while for high degree of accuracy if class intervals are small then number of classes would be larger. Therefore, class interval does not tell the essential characteristics of given frequency distribution of the certain population.

Conflict of interest

No

Funding

No

Reference

1. Perillo, G.M.E., Marone, E. Determination of optimal numbers of class intervals using maximum entropy. Math Geol; 1986; **18;** 401–407.
2. Ronald c. Ray and williammcgill. Effects of class-interval size upon certain frequency distributions of inter-response times. Journal of the Experimental Analysis of Behavior; 1964.
3. S. C. *Gupta*, V. K. *Kapoor. Book. Fundamentals of Mathematical Statistics*. Sultan Chand and Sons; 1980.