STATISTICS

Theory of Attributes-I

Instructor:

Dr. Nidhi Nagar

Assistant Professor,

Department of Statistics,

D.A.-V. (PG) College, Kanpur

Meaning of Attributes

- The measurement of physical characteristics of interest may be in terms of values expressed numerically on some reference scale such as an individual's weight, height or age, the length, width, number of objects, etc; these characteristics are known as VARIABLES.

- The measurement of abstract or qualitative characteristics of interest which cannot be expressed on some standard quantitative scale such as beauty, honesty, insanity, intelligence, deafness, vision impairment is made by grouping the objects according to their differences in quality with respect to these traits. For example, the objects possessing the particular characteristic (say, honesty) to a similar level are grouped together and counted as one category and the rest in different categories according to varying levels of the characteristic. Thus the objects may be classified as honest and dishonest and these two categories are expressed numerically as **0** and **1** respectively. Such a characteristic that yields different groups(according to its varied levels or presence/absence) rather exact measurements is called an ATTRIBUTES.

Classifying Attributes

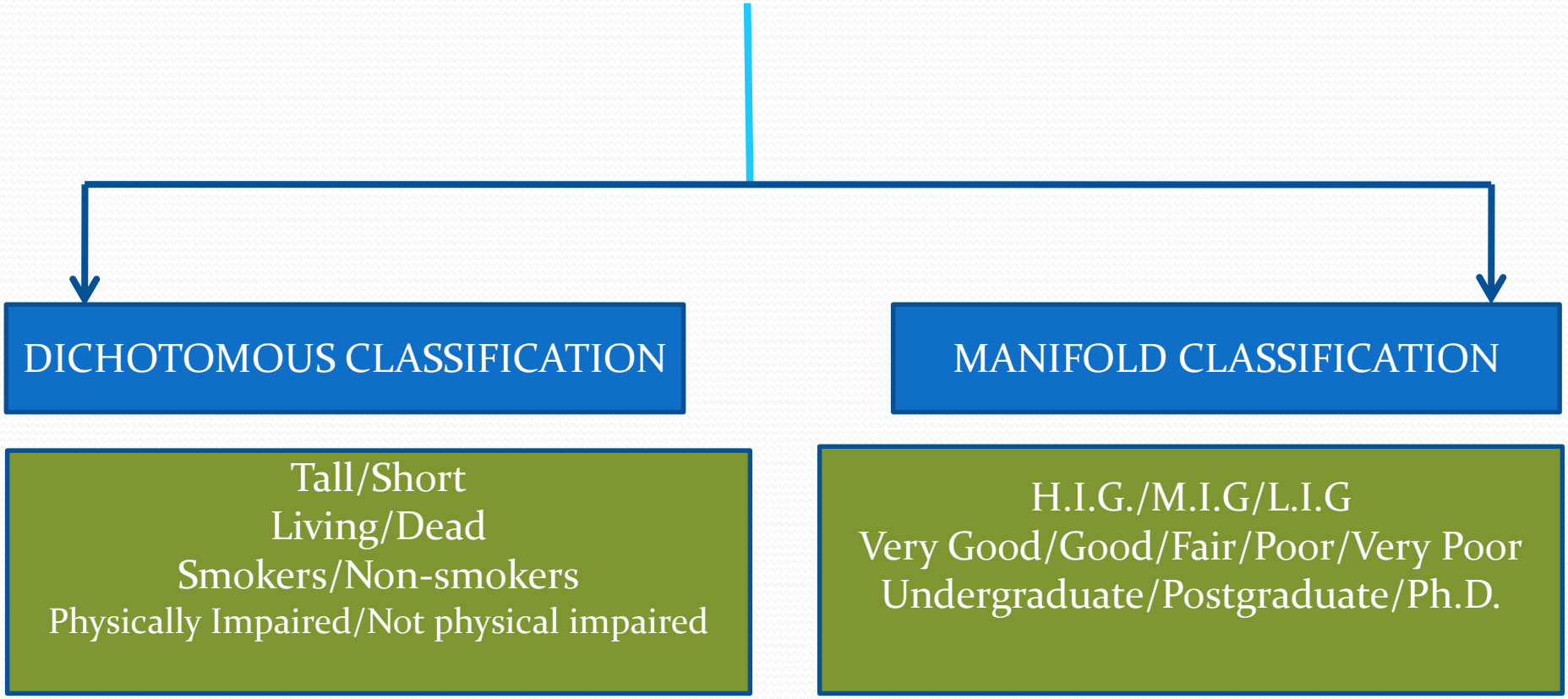
- The data related to attributes is classified on the basis of
  - the presence or absence of an attribute in the universe (thereby yielding a Dichotomous classification)

    OR
  - the varied levels of the attribute (thereby yielding a Manifold classification)
- It must be noted that such a classifications must yield mutually exclusive and exhaustive categories.

Classifying Attributes (continued)

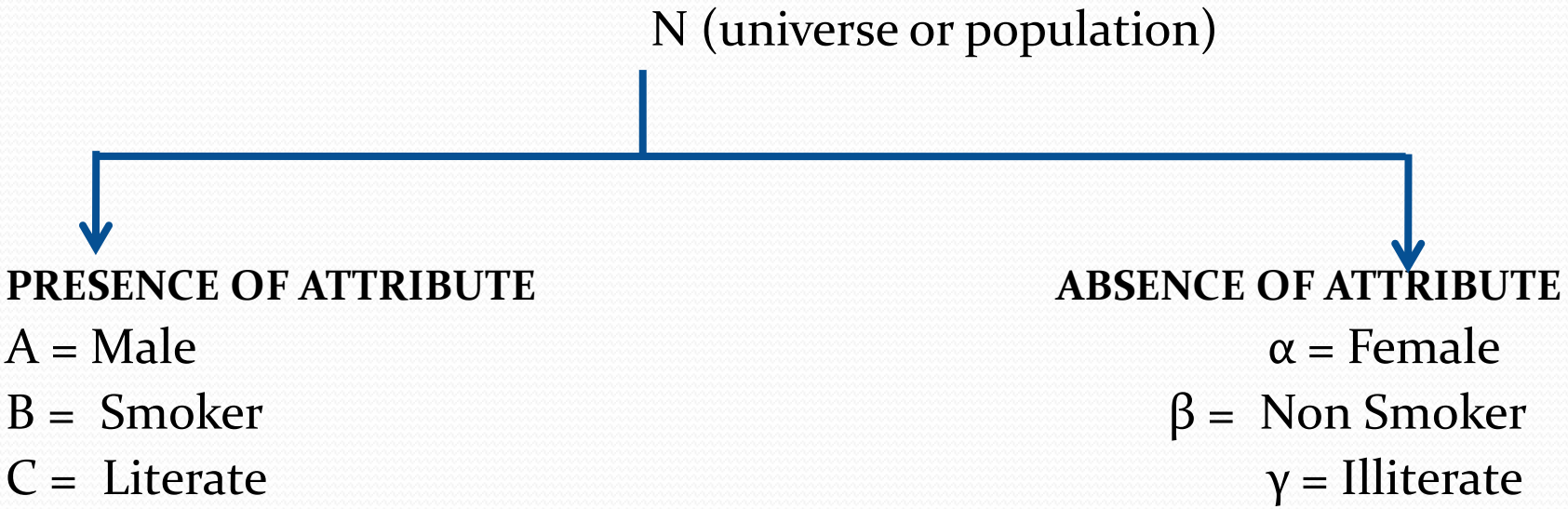
The universe(population) can be classified in two subgroups as dichotomy or more subgroups as manifold classification using one attribute.

Classification of Attributes



Notation & Terminology

For a dichotomous classification, the capital letters A,B,C… are used to denote the presence of attributes and the Greek letters α, β, γ.. are used to denote the absence of the respective attributes. In other words, α would imply not A; β implying not B and γ denoting not C and so on.



Combining Attributes

The presence or absence of various attributes in one object together is represented as the combination of respective letters. For example: AB represents the presence of the both the attributes A & B together. Thus if A indicates Male gender and B identifies the smoker group then the following combinations are interpreted in the following manner:

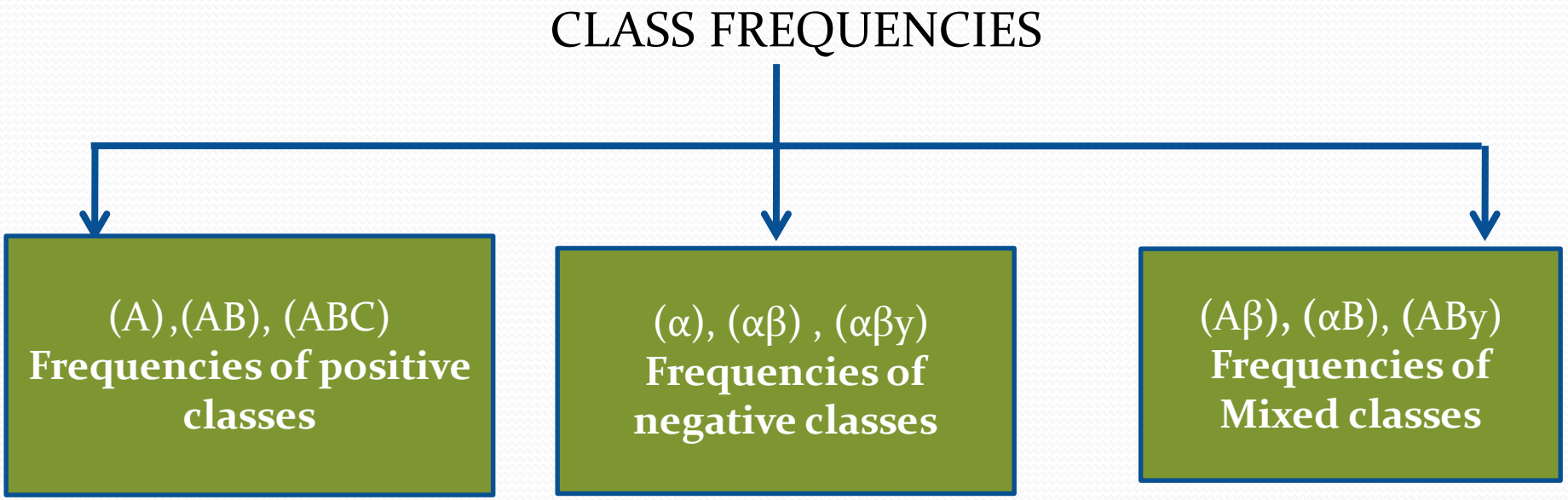
AB = Males who smoke.
αB = Females who smoke
$A\beta$ = Males who do not smoke
$\alpha\beta$ = Females who do not smoke

In case, another attribute C indicating Literates is added then it will have more classes such as ABC, ABγ, αBC, αB γ,, AβC, Aβγ,, αβC and αβγ, indicating Literate male smokers, illiterate male smokers, literate female smokers, illiterate female smokers and so on.

Class Frequency

Class frequency is the count of observations belonging to the specific class and it is denoted by enclosing the corresponding class symbol in parenthesis.

For example, (A) indicates the number of males in the population, (α) is the number of females, (B) gives the number of smokers, (β) the number of non-smokers, (AB) the number of male smokers, (αβ) the number of female non-smokers and so on. The classes that represent presence of attributes are called positive classes, the classes with absence of attributes are called negative classes while the classes with presence of some attributes and absence of others are called mixed classes and hence the corresponding class frequencies are termed accordingly.



ORDER OF CLASS

The number of attributes representing a class is said to be order of the classes.

Class N is of zero order, classes A, B, C... are classes of order 1, classes AB, AC, BC, αB are classes of order two and so on.

Total number of class frequencies of different orders are given below:

Order	0	1	2	r	n
No. of frequencies	1	2n	$nc_2 . 2^2$	$nc_r . 2^r$	2^n

TOTAL NUMBER OF CLASS FREQUENCIES

For n dichotomous attributes,

*the total number of class frequencies of all orders is,

$$\sum nc_r . 2^r = 1 + nc_1 . 2^1 + nc_2 . 2^2 + \ldots\ldots\ldots + nc_n . 2^n = (1+2)^n = 3^n$$

*total number of class frequencies of order r = $nc_r . 2^r$; r = 0, 1, 2 … n.

*The number of positive frequencies in a class of order r is nc_r; r = 0, 1, 2, … n. Hence the total number of positive frequencies is

$$\sum nc_r = nc_0 + nc_1 + nc_2 + \ldots\ldots\ldots + nc_n = (1+1)^n = 2^n$$

Example, in case of 3 dichotomous attributes A, B and C total number of class frequencies is $3^3 = 27$ and the total number of positive frequencies is $2^3 = 8$.

ULTIMATE CLASSES & ULTIMATE CLASS FREQUENCIES

For n dichotomous attributes, the ultimate classes are the ones that comprise of all these attributes and their respective frequencies are called ultimate class frequencies. For n attributes there are 2^n ultimate classes.

For example,

For two attributes A and B there are 2^2 = 4 ultimate class frequencies which are

(AB) (AB) (αB) (αβ).

For three attributes A, B and C there are 2^3 = 8 ultimate class frequencies i.e.

(ABC), (AβC), (αBC), (AβC), (ABγ) , (Aβγ), (αBγ) and (αβγ).

The set of ultimate class frequencies is also referred as fundamental set of frequencies.

References

Part A:

Goon, A.M., Gupta, M.K. and Dasgupta, B. (2013). Fundamental of Statistics, Vol I, World Press, Kolkata.

Goon, A.M., Gupta, M.K. and Dasgupta, B. (2011). Fundamental of Statistics, Vol II, World Press, Kolkata.

Gupta, S.C. and Kapoor, V.K. (2000). Fundamentals of Mathematical Statistics (10th ed.), Sultan Chand and Sons.

Hanagal, D. D. (2009). Introduction to Applied Statistics: A Non-Calculus Based Approach. Narosa Publishing Comp. New Delhi.

Miller, I. and Miller, M. (2006). John E. Freund's Mathematical Statistics with Applications, (7th Edn.), Pearson Education, Asia.

Mood, A.M. Graybill, F.A. and Boes, D.C. (2011). Introduction to the Theory of Statistics, 3rd Edn., Tata McGraw-Hill Pub. Co. Ltd.

Weatherburn, C.E. (1961). A First Course in Mathematical Statistics, The English Lang. Book Society and Cambridge Univ. Press.

Part B:

David, S. (1994) : Elementary Probability, Cambridge University Press.

Dudewicz, E.J. and Mishra, S.N. (2008). Modern Mathematics Statistics, Wiley.

Gupta, S.C. and Kapoor, V.K. (2000). Fundamentals of Mathematical Statistics (10th ed.), Sultan Chand and Sons.

Hanagal, D. D. (2009). Introduction to Applied Statistics: A Non-Calculus Based Approach. Narosa Publishing Comp. New Delhi.

Johnson, S. and Kotz, S. (1972). Distribution in Statistics Vol. I-II & III, Houghton and Mifflin.

Lipschutz, S., Lipson, M. L. and Jain, K. (2010). Schaum's Outline of Probability. 2nd Edition. McGraw Hill Education Pvt. Ltd, New Delhi.

Meyer, P. (2017). Introductory Probability and Statistical Applications (2nd ed.), New Delhi, Oxford & IBH Publishing Co. Pvt. Ltd.

Mood A.M., Graybill F.A. and Boes D.C. (2007). Introduction to the Theory of Statistics (3rd ed.), New Delhi , Tata McGraw Hill Publishing Co. ltd.

Mukhopadhyay, P. (1996). Mathematical Statistics, New Delhi, New Central Book Agency Pvt. Ltd.

Parzen, E.S. (1992). Modern Probability Theory and its Applications. Wiley Interscience.

Pitman, J. (1993). Probability. Narosa Publishing House.

Rao, C.R. (2009). Linear Statistical Inference and its Applications, 2nd Edition, Wiley Eastern.

Rohatgi, V.K. and Saleh, A.E. (2008). An introduction to Probability Theory and Mathematical Statistics, Wiley Eastern.

Books in Hindi Language may be included by the Universities.

Suggested Online Links/Readings:

http://heecontent.upsdc.gov.in/SearchContent.aspx
https://swayam.gov.in/explorer?searchText=statistics
https://nptel.ac.in/course.html
https://www.edx.org/search?q=statistics
https://www.coursera.org/search?query=statistics&

THANK YOU !

Dr. Nidhi Nagar
Assistant Professor,
Department of Statistics,
D.A.-V. (PG) College, Kanpur