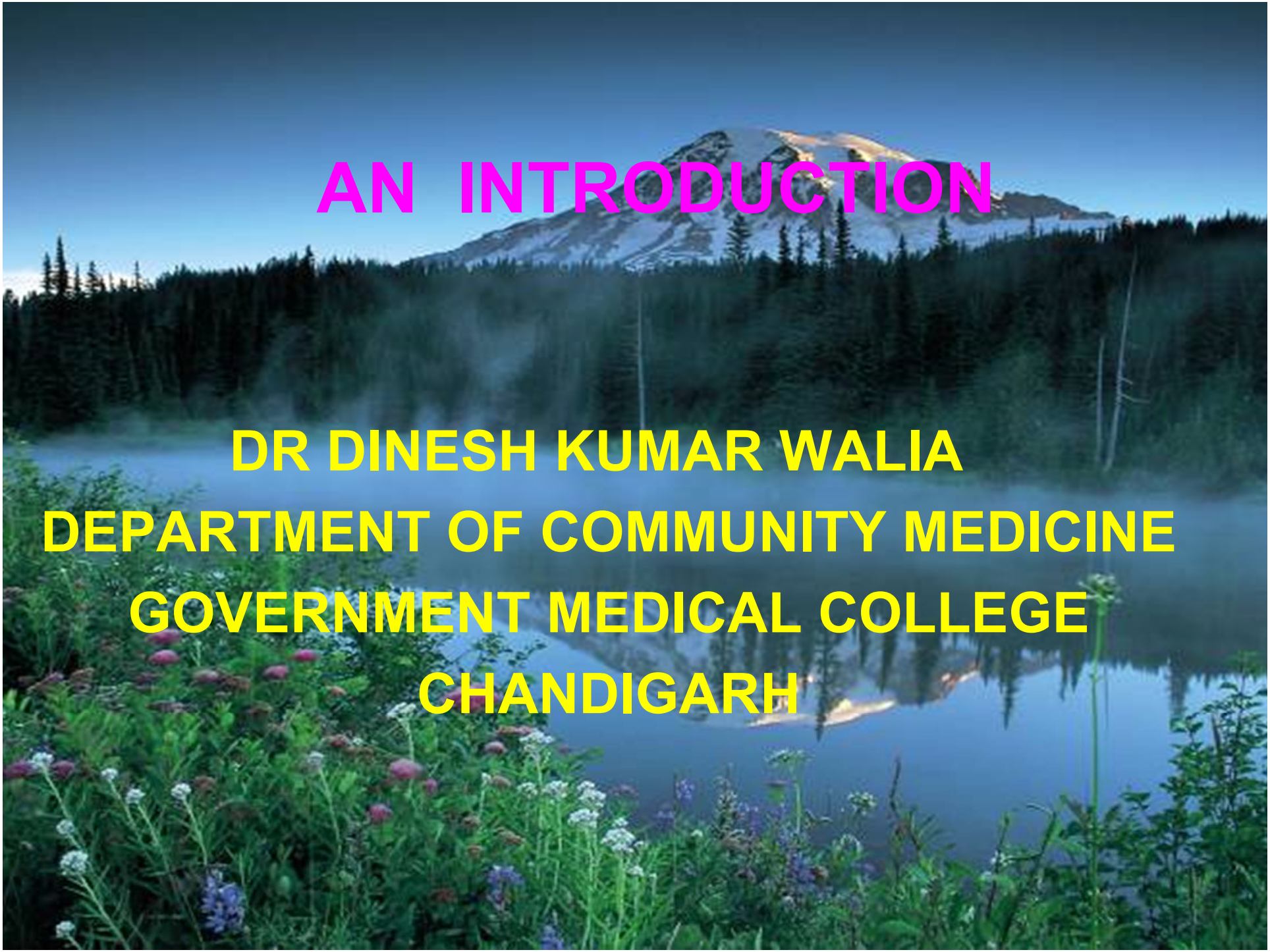
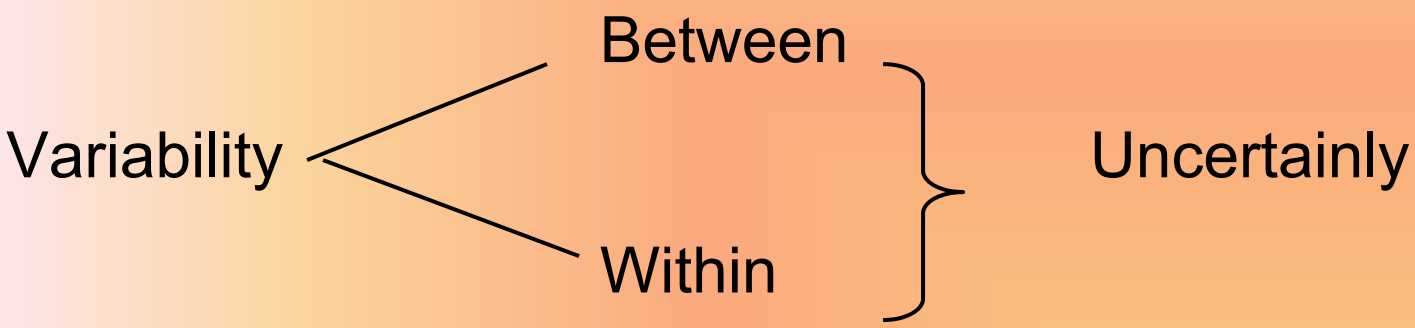

AN INTRODUCTION
DR DINESH KUMAR WALIA
DEPARTMENT OF COMMUNITY MEDICINE
GOVERNMENT MEDICAL COLLEGE
CHANDIGARH

STATISTICS IN MEDICAL SCIENCES

Why? & What is Statistics / Biostatistics



" Statistics may be defined as "a body of methods for making wise decisions in the face of uncertainty." ~W.A. Wallis.

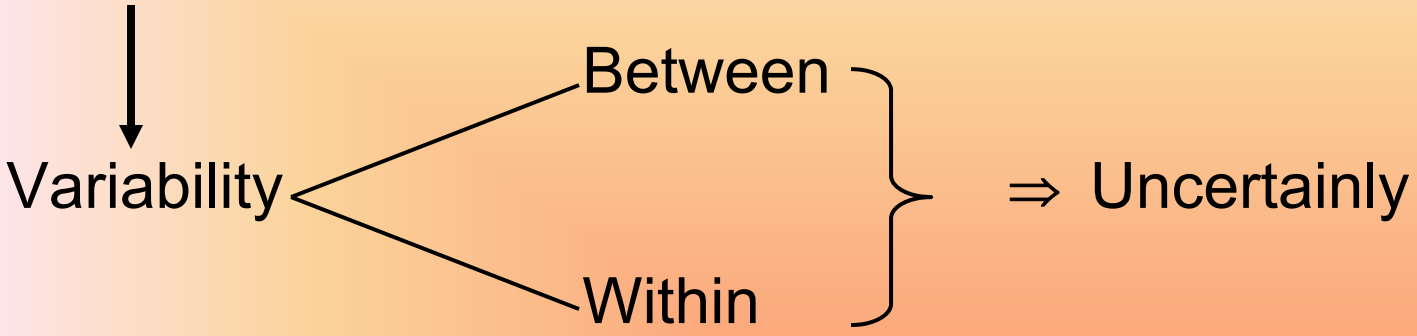
What is Statistics and Biostatistics/Medical Statistics

- *" Statistics may be defined as "a body of methods for making wise decisions in the face of uncertainty."*

  ~W.A. Wallis.

- *Statistics involve planning, designing, collecting, summarizing /reducing, analyzing, presenting, interpreting of data, and drawing inferences/ conclusions from data.*

Why? & What is Statistics / Biostatistics



STATISTICS IN MEDICAL SCIENCES

Uses in Epidemiology:

- Planning, designing and analysis of experiments / epidemiological studies and analyzing results.
- Studying natural history of diseases, occurrence and progression and finding causes / associated factors.
- To study morbidity, mortality and fertility patterns of community
- To determine met and unmet health needs of the community.
- To determine success / failure of specific health programmes and evaluating plan of action.

- Fixing priorities in adoption of future health measures.
- Remodelling and strengthening of health services.
- To assess impact of intervention programmes.
- To study hazards of different toxicants and studying dose – response relations.
- Model building for different diseases of public health importance.

Uses in Clinical Medicine:

- Decisions related to clinical diagnosis predicting likely outcomes of intervention programmes, selection of appropriate treatments.
- Classifying individuals as N & AN, Accuracy of diagnostic / surgical procedures, testing efficacy of drugs and vaccine

Uses in Clinical Medicine:

- Decisions related to clinical diagnosis predicting likely outcomes of intervention programmes, selection of appropriate treatments.
- Classifying individuals as N & AN, Accuracy of diagnostic / surgical procedures, testing efficacy of drugs and vaccine.

STATISTICS IN MEDICAL SCIENCES

Uses in Epidemiology:

- Planning, designing and analysis of experiments / epidemiological studies and analyzing results.
- Studying natural history of diseases, occurrence and progression and finding causes / associated factors.
- To study morbidity, mortality and fertility patterns of community
- To determine met and unmet health needs of the community.
- To determine success / failure of specific health programmes and evaluating plan of action.

Uses in Epidemiology:

- Fixing priorities in adoption of future health measures.
- Remodelling and strengthening of health services.
- To assess impact of intervention programmes.
- To study hazards of different toxicants and studying dose – response relations.
- Model building for different diseases of public health importance.

STSTISTICS helps in

- Quantification of the scientific facts
- Substantiating the findings in clinical research with magnitude and degree of precision.
- Comparability of the results
- Standardization of the results


ELEMENRARY STATISTICAL METHODS
DR DINESH KUMAR WALIA
DEPARTMENT OF COMMUNITY MEDICINE
GOVERNMENT MEDICAL COLLEGE
CHANDIGARH

Definition Of Statistics:

Scientific study of methods dealing with collection, processing, reducing, presenting, analyzing, interpreting results, making inferences and drawing conclusions from data.

Mainly: Collection, description and analysis to draw conclusions.

Distrusts???

Data
- Primary
- Secondary

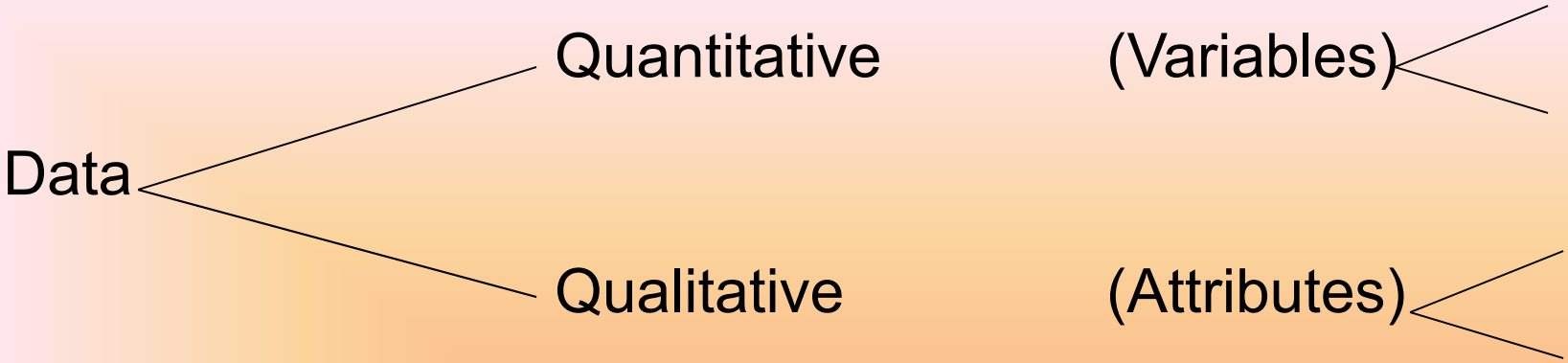
Data Collection
- Routine / Regular
- Surveys with specific objectives

Quality of Data:

- Reliability (Consistency)
- Validity (indicative of the condition)

  [Data should be capable to distinguish those having the condition from those not having]

Validity
- Sensitivity
- Specificity



Quantitative: Ht, Wt, BP, Hb, Age, Calorie, Consumed, Serum Cholestrol, S. Bilrubin, S. Urea, S. Cretnine, BPS score etc.

Qualitative: Gender, sickness, religion, SES, caste, nationality, social background, anemic condition, severity of disease, blood groups, smoking status, nutritional grades, stages of disease (cancer etc), severity of the disease.

Variables

- Continuous (age, Ht, Wt, BMI)
- Discrete

Parity, calorie, WBC/RBC, no. of HH, No. of family members, age in years.

BP (due to = limitations of Measurements)

Scales:

Data

- Qualitative / Categorical
  - Nominal
  - Ordinal
- Quantitative/Measurement data
  - Interval
  - Ratio

VARIABLES TYPES

1. Categorical variables (e.g., Sex, Marital Status, income category)
2. Discrete variables (e.g.,Number of Children in a family)
3. Continuous variables (e.g., Age, income, weight, height, time to achieve an outcome)
4. Binary or Dichotomous variables (e.g., response to all Yes or No type of questions)

Same Variable : Different Scales

- The variable "Education" is only Nominal when measured by type of education, such as private or public
- It is Ordinal when measured by level of education such as high school, college or postgraduate
- It is Interval when measured by number of years i.e. 5, 10, 15 or 20 years of education.

Measures of Central Tendency:

- Definition
- Why called measures of central tendency?
- Different Measures:
    - Mean
    - Median
    - Mode

Calculations:

Mean:

$$\bar{x} = \frac{x_1 + x_2 + \text{----------} + x_n}{n}$$

$$= \frac{1}{n} \sum x$$

$$= \frac{\text{Sum of observations}}{\text{No. of observations}}$$

Calculations:

Discrete Classification:

Value: x_1 x_2 ---------- x_n

Frequency: f_1 f_2 f_n

$$\bar{x} = \frac{f_1x_1 + f_2x_2 + \text{---------} + f_nx_n}{f_1 + f_2 + \text{--------} + f_n}$$

$$= = \frac{\sum fX}{N},$$

Median: Mid value in an arranged set of observations.

$\equiv$ Value which divides the distribution in two equal parts.

Steps: (I) Arrange observations in ascending or discending order.

(II) Find the middle value.

(a) When N is odd:

$\Rightarrow$ Only one middle value

ie $\frac{(N+1)}{2}$ the observation

Md = Size of $\frac{(N+1)}{2}$ the observation in arrange set of observations eg: 2, 1, 11, 9, 10, $\Rightarrow$ 1, 2, 9, 10, 11

(b) When N is Even: There will be two middle values.

$\frac{(N)}{2}$th and $\frac{(N + 1)}{2}$th observations

Average of these two middle observations:

eg. 2, 1, 11, 9, 10, 15

Arrange: 1, 2, 9, 10, 11, 15

Md = (9 + 10) / 2 = 9.5

Mode: The value occurring most frequently.

$\Rightarrow$ The value whose frequency is maximum.

Relationship between Mean, Md, Mode.

$M_0 = 3\,Md - 2\,Mean$

Calculation of Measures of Central Tendency from grouped data:

Mean: $\bar{x} = A + \frac{h \sum fu}{N}$

Where A = Assumed Mean

h = Width of class interval

N = Total frequency

f = Frequency of the class

u = Working unit

Median:

$$Md = L + \frac{h}{f}\left(\frac{N}{2} - C\right)$$

L = Lower limit of Median class

h = Width of class interval

f = Frequency of median class

N = Total frequency

C = Cumulative frequency of the class just before Median Class.

Median Class: Class containing $\frac{N}{2}$ th observation.

Mode:

$$M_0 = L + \frac{h\,(f_1 - f_0)}{(2f_1 - f_0 - f_2)}$$

L = Lower limit of modal class

f_0= Frequency just before modal class

f_1= Frequency of modal class

f_2= Frequency just after modal class

Modal Class: The Class having maximum frequency.

Short Cut Method:

Protein Intake	f	Mid value (x)	x – A	$u = \frac{x-A}{10}$	fu
20 – 30	4	25	-20	-2	-8
30 – 40	8	35	-10	-1	-8
40 – 50	12	45	0	0	0
50 – 60	18	55	10	1	18
60 – 70	5	65	20	2	10
70 – 80	3	75	30	3	9
	N = 40				**$\sum$ fu = 21**

Mean:

$$x = A + \frac{h \sum fu}{N}$$

$$= 45 + \frac{10 \times 21}{40}$$

$$= 45 + 4.2$$

$$= 49.2$$

Ex.

Protein Intake (gm/day)	No. of Individuals f	Mid value x	f . x
20 – 30	4	25	100
30 – 40	8	35	280
40 – 50	12	45	540
50 – 60	18	55	990
60 – 70	5	65	325
70 – 80	3	75	225
	$\sum f = N = 40$		**$\sum f.x = 2460$**

$$\text{Mean } \bar{x} = \frac{\sum f.x = 2460}{\sum f x / N}$$

$$\text{Mean Protein Intake (gm/day)} = \frac{2460}{40} = 49.2$$

Calculation of Median:

Protein Intake (gm/day)	No. of Individuals f	Cumulative x
20 – 30	4	4
30 – 40	8	12
40 – 50	12	24
→ 50 – 60	18	42 ←
60 – 70	5	47
70 – 80	3	50

$$\text{Median} = L + \frac{h\,(N - C)}{\frac{f}{2}}$$

$L = 50$ as $N = 25$

$h = 10$ $\quad \frac{N}{2}$

$f = 18$

$C = 24$

$$\text{Median} = 50 + \frac{10\,(25 - 24)}{18}$$

$$= 50 + \frac{10}{18} = \text{---------}$$

Calculation of Mode:

$$\text{Mode} = L + \frac{h\,(f_1 - f_0)}{(2f_1 - f_0 - f_2)}$$

$h = 1_0$, $f_0 = 12$

$f_1 = 18$

$f_2 = 5$

$$M_0 = 50 + \frac{10\,(18 - 12)}{(2 \times 18 - 12 - 5)}$$

$$= 50 + \frac{10 \times 6}{19}$$

Measures of Dispersion or Measures of Variability

- Meaning
- Need:
  - Comparison
  - Reliability of Measures of CT.

eg. → 7, 8, 9, 10, 11
→ 3, 6, 9, 12, 15
→ 1, 5, 9, 13, 17

$\bar{x} = 9$

- Separation of normal & Abn.
- Finding SE of estimates
- Probable Ranges / CI
- Tests of Significance

Different Measures of Variability:

- Range
- Inter-quartile Range
- Mean Deviation
- Standard Deviation
- Coefficient of Variation (CV)

(1) Range $= x_{max} - x_{min}$

(2) Quartile Deviation or Inter-quartile Range

$= Q_3 - Q_1$

(3) Mean Deviation (MD)

$$MD = \frac{\sum f|x - M|}{N}$$

(4) Standard Deviation (SD)

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{N}}$$

Short-cut formula (grouped data)

$$SD = h\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2}$$

Ex.

		TB		Total
BCG Vaccine ↓		**Yes**	**No**	
	Yes	6	12	18
	No	8	3	11
	Total	**14**	**15**	**29**

Coefficient of Assoc.

$$Q = \frac{6 \times 3 - 8 \times 12}{6 \times 3 + 8 \times 12}$$

$$= \frac{18 - 96}{18 + 96} = \frac{-78}{114}$$

$$= \text{---------}$$


THANKS