INTRODUCTION TO BIOSTATISTICS

Adil ZA *MBBS, DLSHTM, MSc, MPH, DrPH*
Department of Community Medicine,
Kulliyyah of Medicine, IIUM

وَعَلَّمَ ءَادَمَ ٱلْأَسْمَآءَ كُلَّهَا ثُمَّ عَرَضَهُمْ عَلَى ٱلْمَلَٰٓئِكَةِ فَقَالَ
أَنۢبِـُٔونِى بِأَسْمَآءِ هَٰٓؤُلَآءِ إِن كُنتُمْ صَٰدِقِينَ ﴿٣١﴾

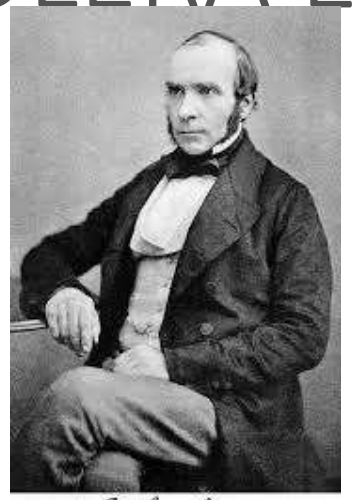
And He taught Adam the names - all of them. Then He showed them to the angels and said, "Inform Me of the names of these, if you are truthful."

(Al-Baqarah : 31)

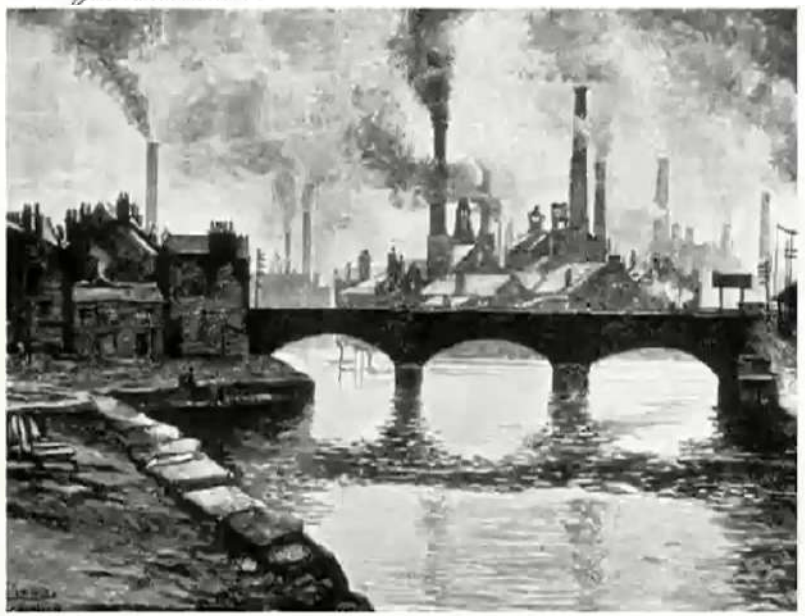
LEARNING OUTCOMES

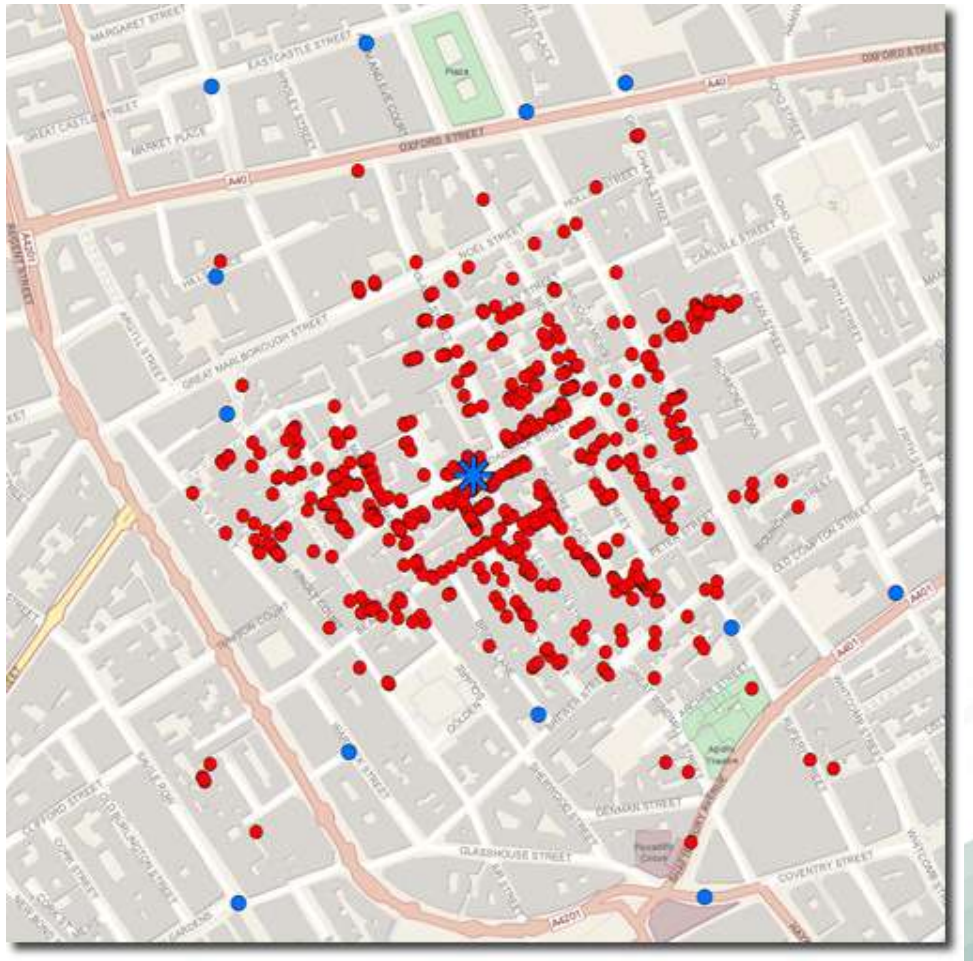
- Define statistics and biostatistics.
- Describe roles of statistics in public health.
- Identify and differentiate types of data, level of measurement.
- Type of probability distribution.

CHOLERA EPIDEMIC IN LONDON 1854



John Snow





Water Company	# Houses	# Cholera Deaths	Death/10,000 Houses
Southwark & Vauxhall	40,046	1,263	315
Lambeth	26,107	98	37
Rest of London	256,423	1,422	59

WHAT IS STATISTICS

Statistics is a field of science concerned with :
1. Collection, organization, summarisation and analysis of data.
2. Drawing of inferences about a body of data when only a part of the data is observed. (Daniel W.)

Biostatistics- Application of statistics in biological sciences and medicine.

WHY WE NEED STATISTICS IN MEDICINE

- Evidence based medicine
- A tool for research
- Communicating
- Manage uncertanties in medicine (biological variation)

MAIN BRANCHES IN STATISTICS

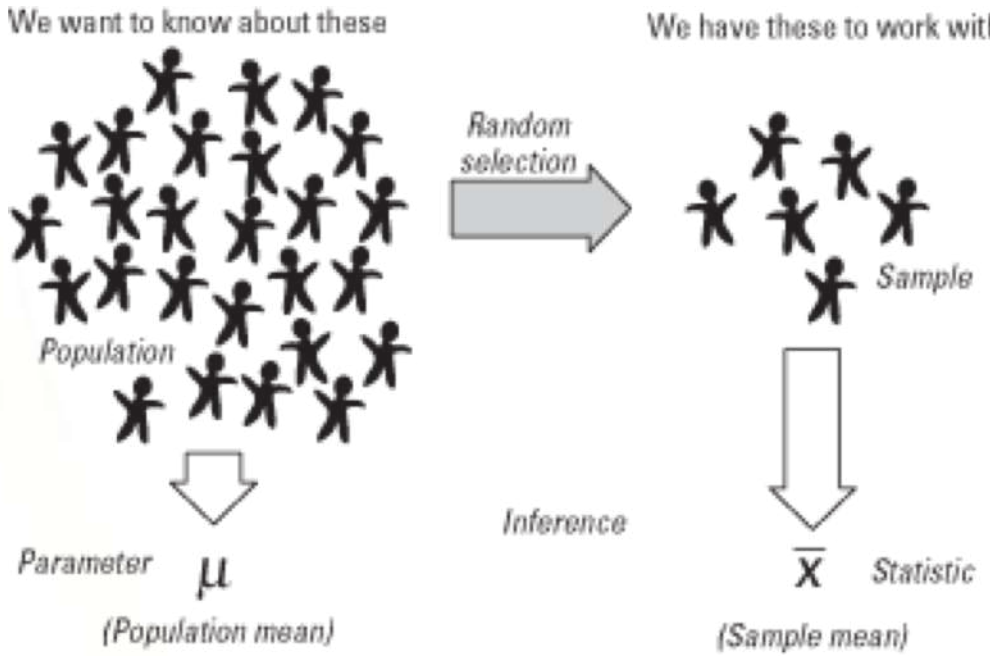
Descriptive statistics

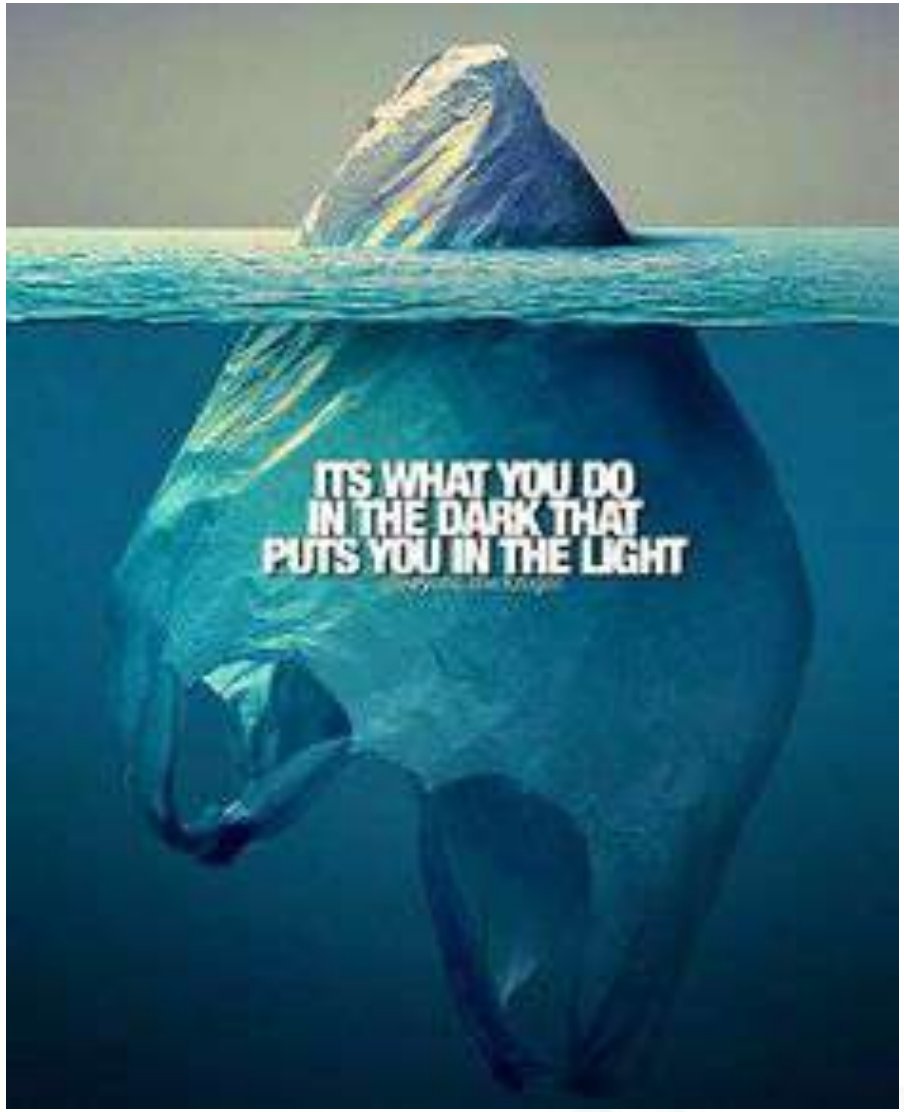
- Summarization of data describing in statistical idea

Inferential statistics

- Making inference from the selected information
- Drawing conclusion about the population from the sample evidence
- Estimation- point estimation, interval estimation
- Hypothesis testing

POPULATION AND SAMPLE




ITS WHAT YOU DO
IN THE DARK THAT
PUTS YOU IN THE LIGHT

POPULATION AND SAMPLE

Population - all members of a defined group
Sample - subset of the populations
Parameter - descriptive summary measures from population
Statistics - descriptive summary measures from sample

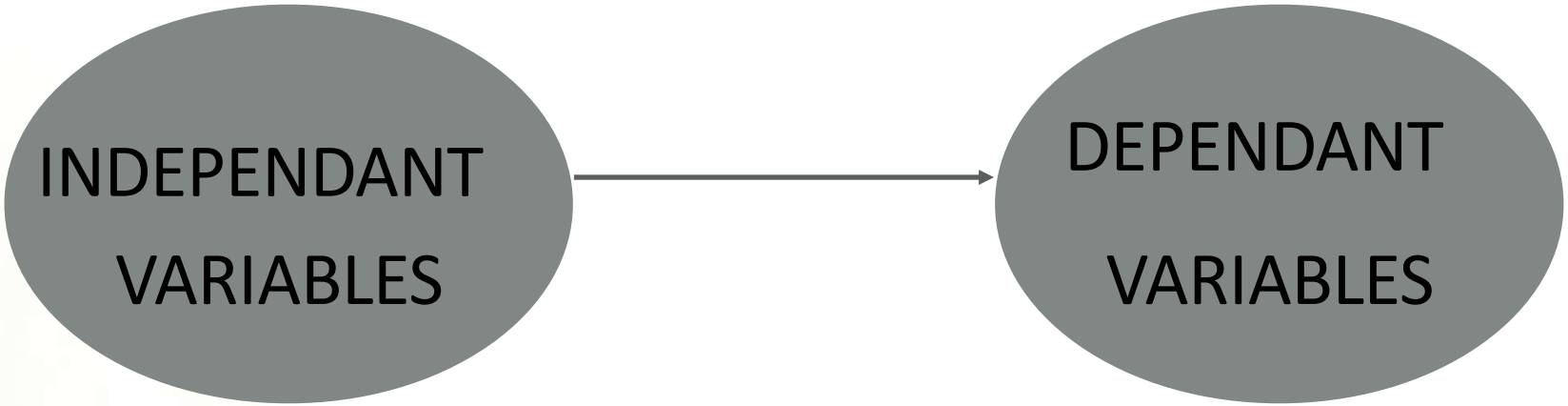
VARIABLES

- Term for an observation or reading giving information on the study question to be answered.

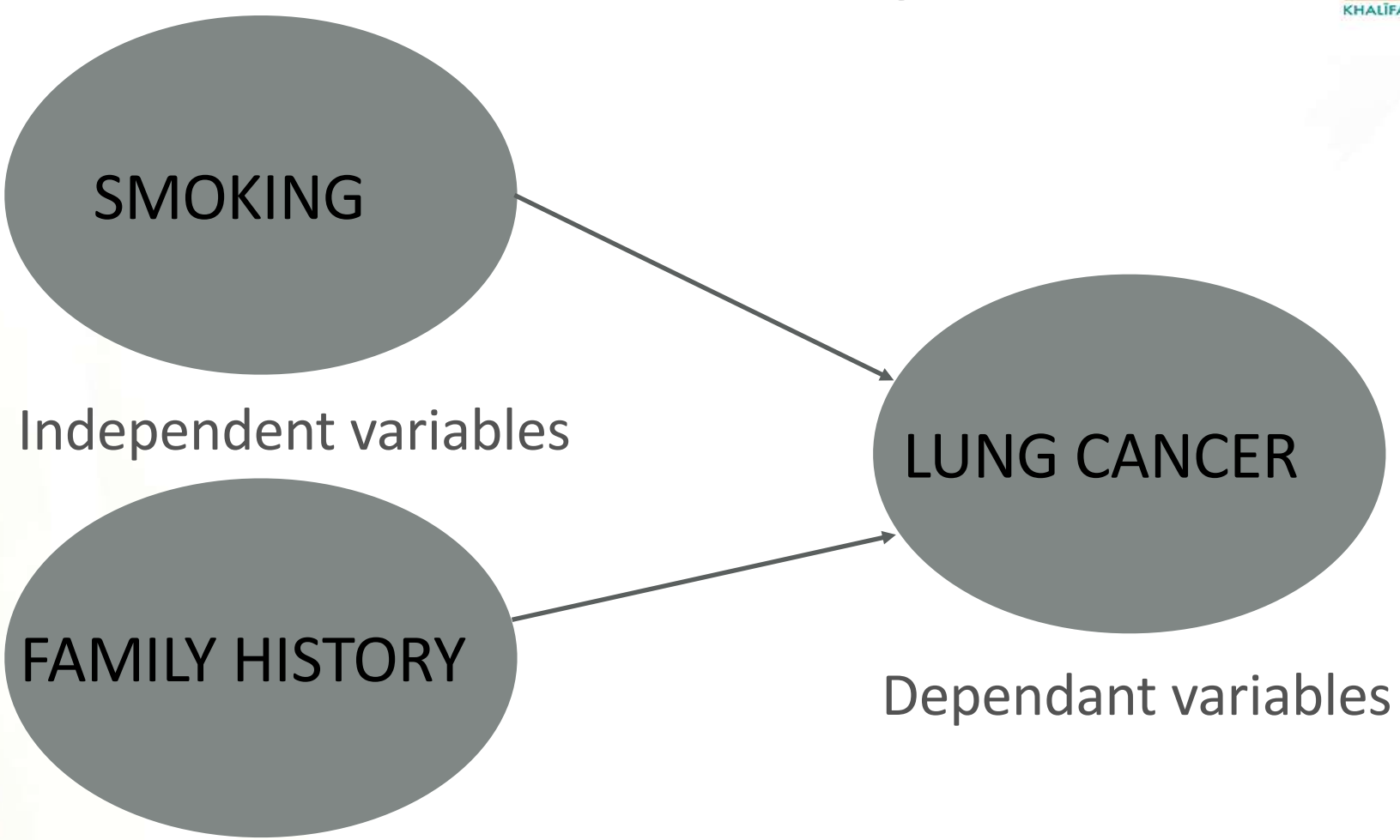
- Variables are characteristics under observation measurable with varying degree or accuracy. Example. blood pressure, blood glucose reading

- **Dependant** – outcome variables that depend on / influenced by eg lung cancer
- **Independant -** variables that independently of the effect beings being studied -e.g. smoking status, no of cigarettes

VARIABLES



VARIABLES

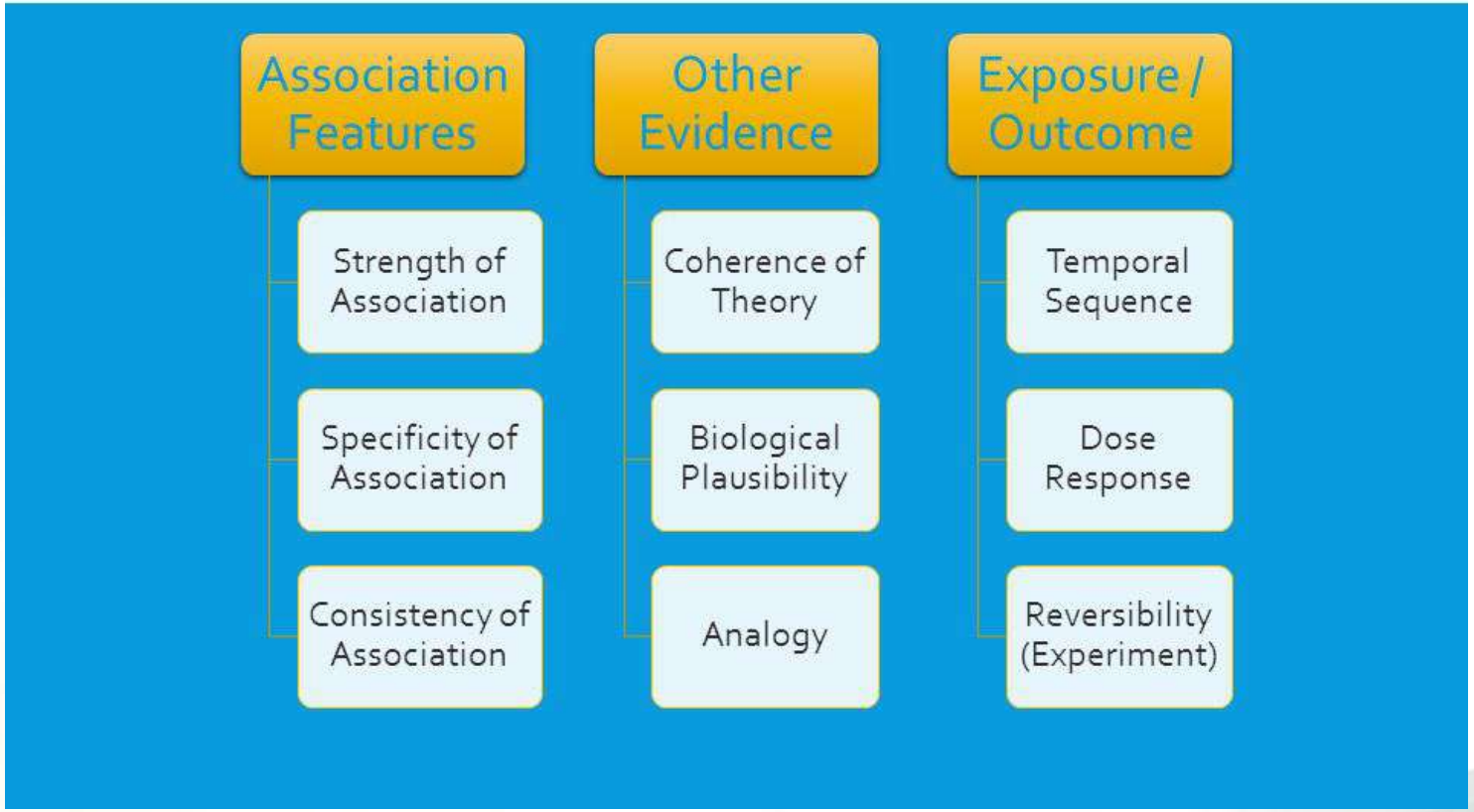


Independent variables

Dependant variables

BRADFORD HILL CRITERIA

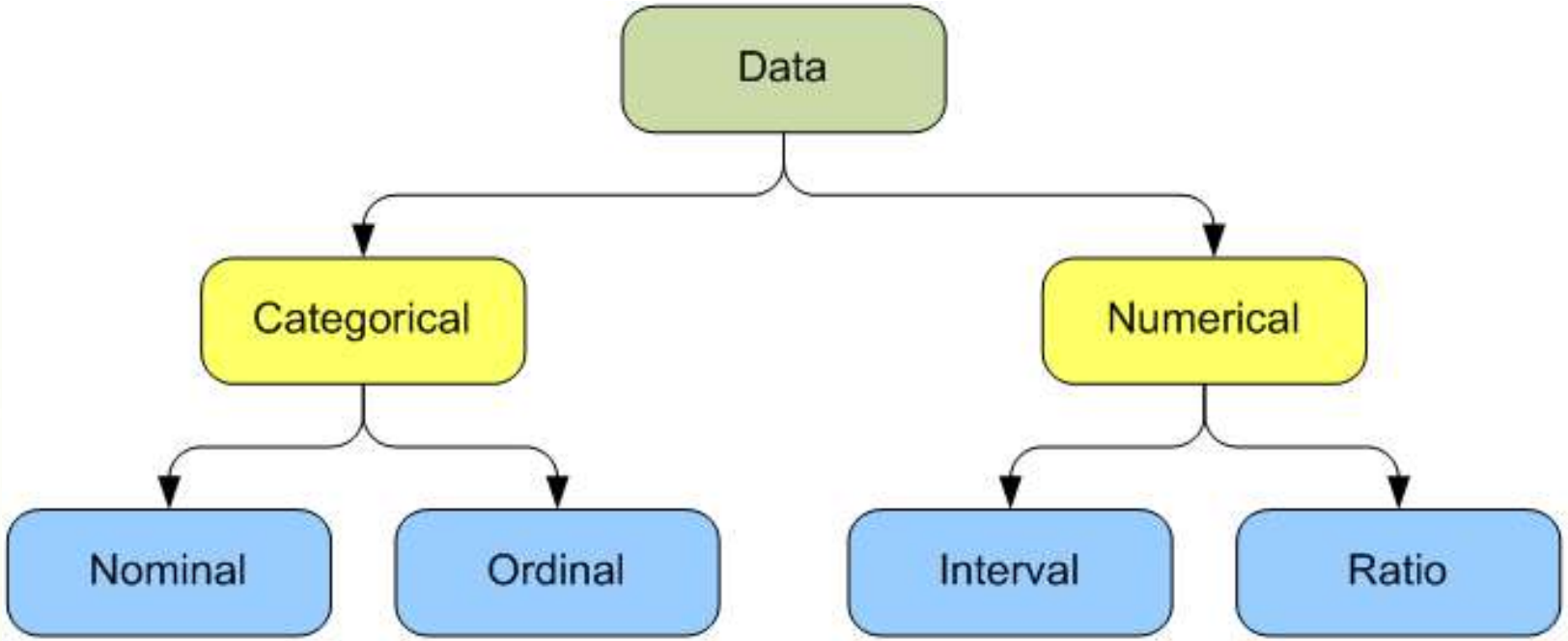
9 CRITERIA



SOURCES OF DATA

- Demographic data - census, population survey, birth registry, migration
- Mortality data - death registration, postmortem record,
- Morbidity - hospital records, cancer registration, infections disease notification, health survey
- Lifestyle data - The health survey for England (HSE)

DATA


Data
Categorical
Numerical
Nominal
Ordinal
Interval
Ratio

NOMINAL DATA

Nominal or also known as categorical
Are indicators of type or category and may be thought as counts
The categories occur in any sequences and are not order able
Eg. sex : male/female

ORDINAL DATA

Also known as rank order
Are indicators of some ordering characteristics
Smallest to biggest , most likely to survive
Lung cancer stages
But not necessary to be equal distances
Retain some of the information of continuous data.

INTERVAL DATA

Continous data are positions on a scale
Equal distances between each unit
Can include 0 but not ‘true’ 0
Eg Temperature
Subset of interval is discrete data.

RATIO DATA

Indicates actual amount
There is equal distances
It includes a true absolute zero
Eg body weight

EXAMPLES

Gender
Body weight
Height
BMI
Martial status
Temperature

PROBABILITY

PROBABILITY

Its not same with possibility.
Numerical measure of the likelihood that a possible events occur on a random opportunity to occur.

PROBABILITY

What is the probability of getting headache as a side effect of a certain medication ?

What is the probability of successful recovery following appendicectomy at HTAA ?

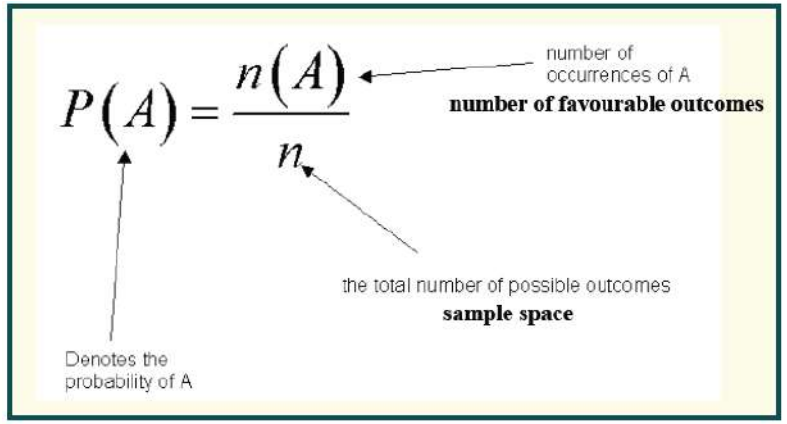
Probability is likelihood that a possible events occur on a random opportunity to occur.

Probability can vary from 0 to 1

an event cannot occur = 0

an event can occur =1

Thus lies between 0 (impossible event) and 1 (certain event)



PROBABILITY DISTRIBUTION

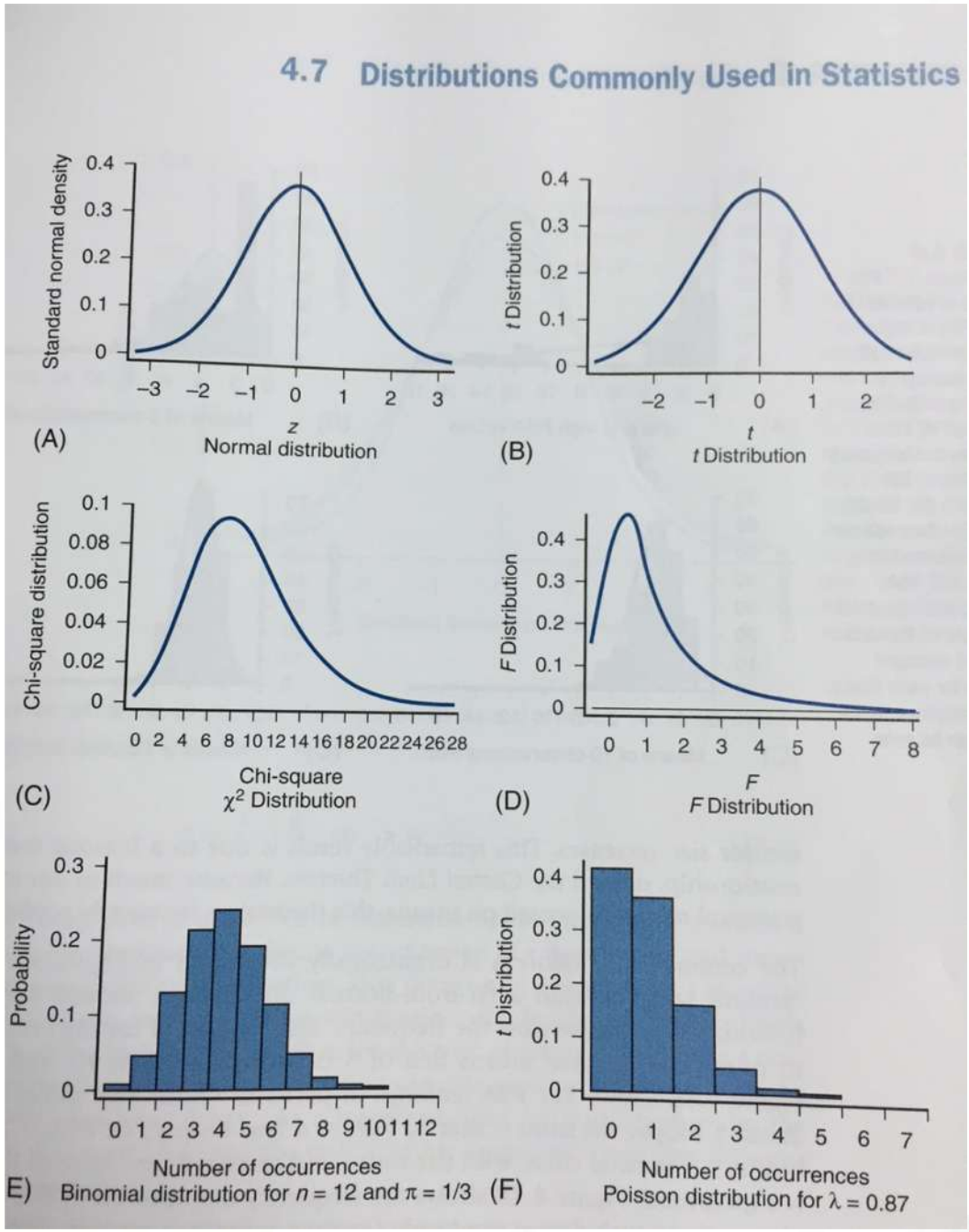
PROBABILITY DISTRIBUTION

Describe the range of possible values that a random variable can attain and the probability that the value of the random variable is within any (measurable) interval or subset of that range.

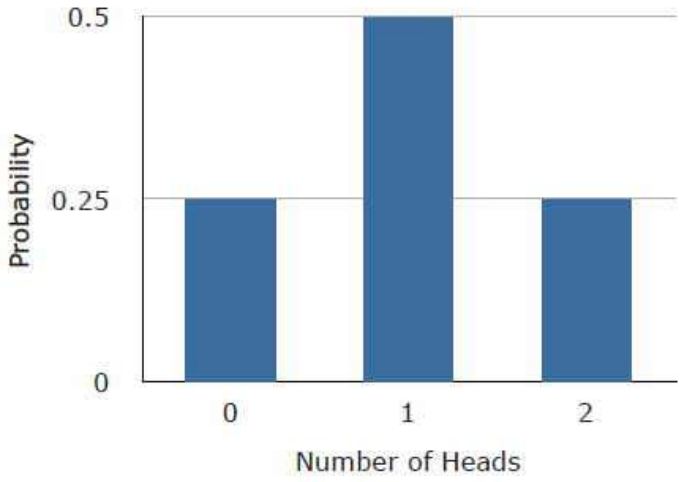
E.g What is the possible value for haemoglobin level for each resident in a given population ?
Probability distribution - table and graph

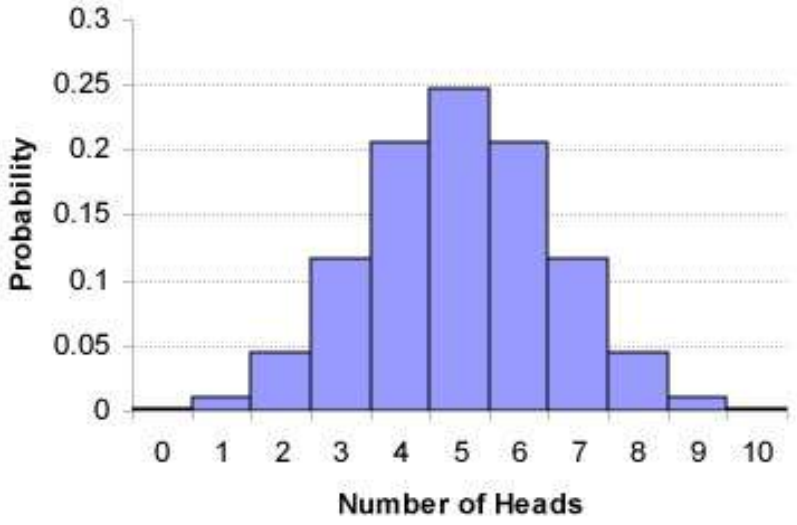
TYPES OF DISTRIBUTION

Discrete probability distribution
finite probability - binomial
eg probability that a randomly selected patient is male
Continuous probability distribution
Infinite probability - z, t, F
Probability for patient's Haemoglobin after surgery

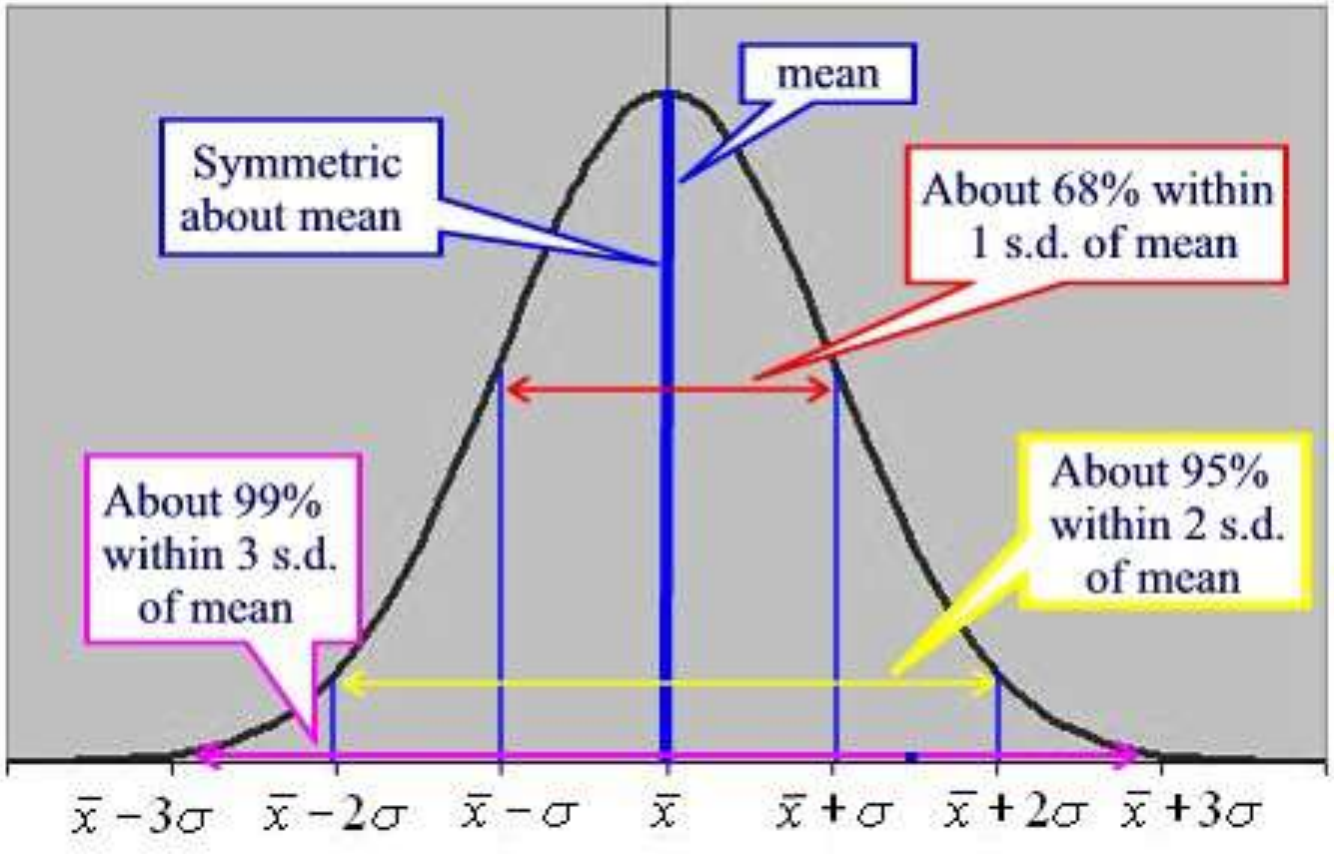


BINOMIAL DISTRIBUTION





NORMAL DISTRIBUTION



NORMAL DISTRIBUTION

A smooth bell shaped curve that is symmetric about the population mean μ , unimodal
Mean = median = mode
The two tails never touch the base, $-\infty < x < \infty$
Total Area under curve (AUC) = 1
The AUC between 2 points is the probability of that range
within ±1 SD = 68 %, ± 2 SD = 95%, ± 3SD = 99.7%

Why normal distributions? many medical , psychological variables are distributed following normal distributions.
Normal intervals for haemoglobin, cholesterol level etc
Many statistical tests have been derived for a normal distributions

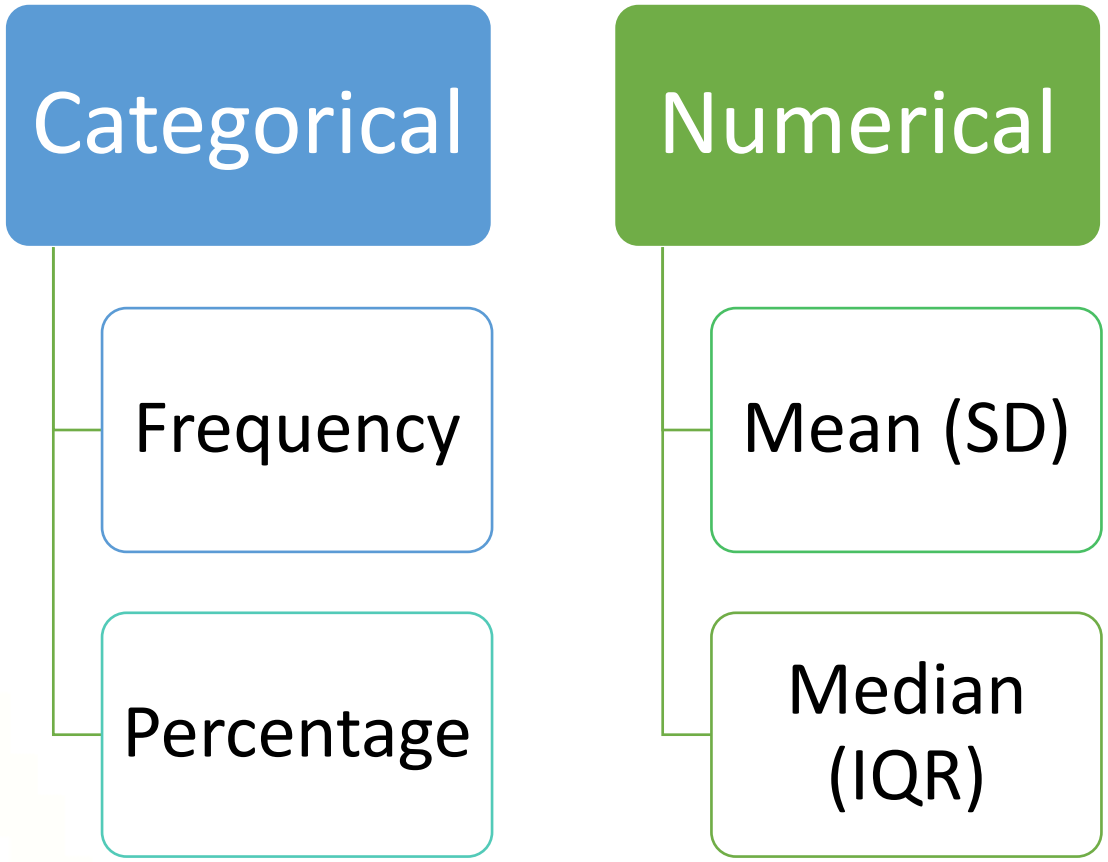
CENTRAL LIMIT THEOREM

The central limit theorem states that even if a population distribution is strongly non-normal, its sampling distribution of means will be approximately normal for large sample sizes (over 30). The central limit theorem makes it possible to use probabilities associated with the normal curve to answer questions about the means of sufficiently large samples.

Data Analysis Tips

- Must answer the objectives of the study
- Must be planned before data is collected
- Must decide whether it is descriptive or inferential, or a combination
- Must be guided by clinical judgment/biological plausibility

Descriptive statistics



Inferential statistics (Unpaired)

Dependent variable	Independent variable	Test
Categorical	Categorical	Chi square test or Fisher's exact test
Numerical (normal)	Categorical (2 cat)	Independent sample t-test
Numerical (not normal)	Categorical (2 cat)	Mann-Whitney U test
Numerical (normal)	Categorical (>2 cat)	One-way ANOVA
Numerical (not normal)	Categorical	Kruskal-Wallis test
Numerical (normal)	Numerical (normal)	Pearson correlation coefficient test
Numerical (not normal)	Numerical (not normal)	Spearman correlation coefficient test
Numerical (normal)	Numerical (not normal)	
Numerical (not normal)	Numerical (normal)	

Jamalludin AR (2015). Brief guidelines for methods and statistics in medical research

Inferential statistics (Paired)

Dependent variable	Independent variable	Test
Numerical (normal)	Numerical (normal)	Paired t-test
Numerical (not normal)	Numerical (not normal)	Wilcoxon signed rank test

Jamalludin AR (2015). Brief guidelines for methods and statistics in medical research

THANK YOU