Sampling, measures of central tendency & dispersion, Health economics

Dr. Md. Rahul Parvez
Lecturer, Community Medicine
Rajshahi Medical College

Population and sample

- **Population or target population:** The complete group of subjects in a research project. It is the population to be studied or to which investigator wants to generalize results.
- **Study population:** It is the population that is actually listed in the sampling frame.
- **Sample:** It is the representative part of the population from which data are actually collected. It is a subset of population.

Sampling

- **Sampling** is a method of gathering information about a larger population through the selection of a smaller subset of individuals or items from that larger population.
- **Sampling** is the process of **selecting representative** part of the population.
- **Sampling frame:** It is the listing of all elements of a population from which sample will be drawn.

Sampling technique

Probability sampling technique

- **Simple random sampling**
- **Systematic random sampling**
- **Stratified sampling**
- **Cluster sampling**
- **Multistage sampling**

Non-probability sampling technique

- **Convenience sampling**
- **Purposive sampling**
- **Quota sampling**
- **Snowball sampling**

Simple random sampling

- A simple random sample is one in which **every subject** has an **equal probability** of being selected for the study.
- It is the **most basic** sampling design.
- It can only be used if the researcher has a complete list of the target population which is called ***sampling frame***.
- It is done by assigning a number to each unit in the sampling frame, then a random number table, random number digit or lottery system is used to determine which units are to be selected.
- **Example:** Selecting apples from a basket.

Systematic random sampling

- Systematic random sample is one in which every k^{th} item is selected; k^{th} is determined by **dividing the number of items** in the sampling frame by the **desired sample size**.
- For example, if we want to take **20 students** from RMC 64 batch, then 200 divided by 20 is 10, so every 10^{th} student is sampled.
- In this approach, we must **select a roll randomly** between 1 and 10 first, and we then select every 10^{th} roll.
- Suppose, we randomly select the **roll 8** then, the systematic sample consists of roll numbers **8, 18, 28, 38, 48, 58** and so on.

Stratified random sampling

- A **stratified random sample** is one in which the population is first divided into **relevant strata (subgroups)**, and a random sample is then selected from each stratum.
- When a population is heterogenous, the study population is first **divided into homogeneous groups** or classes is called **strata**.
- In most public health research, criteria for strata are **race, income, education, gender, location** etc.
- Simple random sampling is done in each strata.

Cluster sampling

- A **cluster random sample** results from a **two-stage process** in which the population is **divided into clusters and a subset of the dusters** is randomly selected.
- Clusters are commonly **based on geographic areas or districts**, so this approach is used more often in epidemiologic research than in clinical studies.
- For example, the sample for a **household survey** taken in a city may be selected by using **city blocks as clusters**; a random sample of city blocks is selected, and **all households (or a random sample of households)** within the selected city blocks are surveyed.

Multistage sampling

- The selection of **clusters can be multistage** (e.g. **selecting districts** – the **primary sampling units, PSUs** – within a region for the sample, and within these sample **electoral wards** and finally within these a sample of **households**).
- This is known as **multistage sampling** and can be more economical, as it results in a concentration of fieldwork.

Convenience sampling

- Convenience sampling is also known as **opportunity or accidental or haphazard** sampling.
- It is a type of non-probability sampling that involves the sample being drawn from that part of the population which **is close at hand**.
- This is sampling of subjects for **reasons of convenience** (e.g. easy to recruit, near at hand, likely to respond).

Purposive sampling

- Purposive sampling is also called **judgmental sampling**.
- In this method, the researcher chooses the sample based on whom they think **would be appropriate** for the study.
- This is primarily done when there is **limited number of people** who have expertise in the area being researched.
- In other words, **judgmental or purposive sampling** is a sampling technique that **relies on the judgment of the researcher** when it comes to selecting the units that are to be studied.

Quota sampling

- According to this sampling technique, the **population is first segmented** into mutually exclusive subgroups, **just as in stratified sampling**.
- This classification is done based on characteristics such as **gender, age etc.**
- Subsequently, **judgement is used** to select subjects or units.
- For example, an interviewer may be told to sample 200 females and 300 males between ages of 45 and 60. in quota sampling, the **selection of the sample is non-random**.

Snowball sampling

- Snowball sampling is a technique, where **existing study subjects are used to recruit more similar subjects** into the sample.
- It is particularly appropriate when the population you are interested in is **hidden and or hard-to-reach**.
- Additional respondents are selected based on **referrals from the initial respondents**.
- When the study subjects are **I/V drug abusers or sex workers**, this sampling method is used.

Measures of central tendency

- **Measures of central locations/tendency are**
  i. **Mean**
     a. **Arithmetic mean**
     b. **Geometric mean**
     c. **Harmonic mean**
  ii. **Median**
  iii. **Mode**

Mean calculation

- To calculate mean, the individual observations are first added together, and then divided by the number of observations.
- For example, the diastolic blood pressure of 10 individuals are 83, 75, 81, 79, 71, 95, 75, 77, 84 and 90. Calculate the mean.
- Mean = $\frac{\text{summation of total observations}}{\text{number of observations}}$
- Mean = $\frac{83+75+81+79+71+95+75+77+84+90}{10} = \frac{810}{10} = \mathbf{81}$

Median calculation

- To calculate median, data are first arranged in an ascending or descending order and then find out the middle observation.
- If observations are **odd** in number, the **middle one** will be the median.
- For example, the diastolic blood pressure of 9 individuals are 83, 75, 81, 79, 71, 95, 75, 84 and 90. Calculate the median.
- Ascending order: 71, 75, 75, 79, **81**, 83, 84, 90, 95
- Total observations are 9 in number. So, median will be 5^{th} observation.
- Here, 5^{th} observation is **81**. So, median is **81**.

Median calculation

- If observations are **even** in number, the **average of the middle two** will be the median.
- For example, the diastolic blood pressure of 10 individuals are 83, 75, 81, 79, 71, 95, 75, 77, 84 and 90. Calculate the median.
- Ascending order: 71, 75, 75, 77, **79, 81**, 83, 84, 90, 95
- As observations are 10 in number. So, median will be the average of 5^{th} and 6^{th} observation.
- Median = $\frac{\text{summation of 5th and 6th observations}}{2}$
- Median = $\frac{79+81}{2}$ **= 80**

Mode calculation

- The mode is the commonly occurring value in a distribution of data.
- It is the most frequent item or the most fashionable value of a series of observations.
- For example, the diastolic blood pressure of 10 individuals are 83, **75**, 81, 79, 71, 95, **75**, 77, 84 and 90. Calculate the mode.
- Here, **mode or most frequent** occurring value is **75**.

Measures of dispersion/variability

- **Range**
- **Mean or average deviation**
- **Standard deviation**

Calculation of range

- Range is the **simplest measure** of dispersion.
- It is defined as the **difference between the highest and lowest figures** in a given sample.
- For example, the diastolic blood pressure of 10 individuals are 83, 75, 81, 79, 71, 95, 75, 77, 84 and 90. Calculate the range.
- Here, the highest value is 95 and the lowest value is 71.
- So, the range is **71 to 95 or (95-71)= 24**.

Calculation of mean or average deviation

- It is the average of the deviations from the arithmetic mean.
- It is calculated by the following formula:
- **Mean deviation** $= \sum \frac{(x-\bar{x})}{n}$
- For example, the diastolic blood pressure of 10 individuals are 83, 75, 81, 79, 71, 95, 75, 77, 84 and 90. Calculate the mean deviation.
- First, calculate the arithmetic mean of the observation. Then, calculate the deviation of each observation from the mean. Finally, divide the summation of deviation by the number of observation.

Calculation of mean or average deviation

Example : The diastolic blood pressure of 10 individuals was as follows : 83, 75, 81, 79, 71, 95, 75, 77, 84 and 90. Find the mean deviation.

Answer (Mean deviation)

Diastolic B.P. x	Arithmetic Mean $\bar{x}$	Deviation from the Mean $(x - \bar{x})$
83	81	2
75	81	−6
81	81	0
79	81	−2
71	81	−10
95	81	14
75	81	−6
77	81	−4
84	81	3
90	81	9
Total = 810		Total = 56(ignoring ± sign)

Mean $= \frac{810}{10} = 81$ The Mean Deviation $= \frac{56}{10} = 5.6$

Calculation of standard deviation

- It is the **most frequently used** measure of deviation.
- In simple terms, it is defined as root mean square deviation.
- It is calculated by the following formula:
- **Standard deviation or S.D.** $= \sqrt{\sum \frac{(x-\bar{x})^2}{n-1}}$
- First, calculate the arithmetic mean of the observation. Then, calculate the square of the deviation of each observation from the mean. Next, divide the summation of squared deviation by the number of observation. Finally, take the square root.

Calculation of standard deviation

Example : The diastolic blood pressure of 10 individuals was as follows : 83, 75, 81, 79, 71, 95, 75, 77, 84, 90. Calculate the standard deviation.

Answer

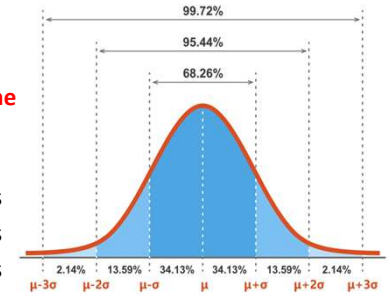
x	$(x - \bar{x})$	$(x - \bar{x})^2$
83	2	4
75	−6	36
81	0	–
79	−2	4
71	−10	100
95	14	196
75	−6	36
77	−4	16
84	3	9
90	9	81
$\bar{x} = 81$ $n = 10$		Total = 482

S.D. $= \sqrt{\frac{\Sigma (x - \bar{x})^2}{n-1}} = \sqrt{\frac{482}{10-1}} = \sqrt{\frac{482}{9}}$

$= \sqrt{53.55} = 7.31$

Normal distribution curve

- It is a **bell-shaped** curve
- It is **symmetrical**
- The curve never **touch the baseline**
- Mean, median and modes are **equal**
- Mean±1SD covers 68.3% of values
- Mean±2SD covers 95.5% of values
- Mean±3SD covers 99.7% of values



Economics & health economics

- **Economics** is the study of **study of scarcity** (**maximizing the utilization** of **scarce resource)**.
- Economics is the study of how people and society end up choosing to employ **scarce resources** that could have alternate use.
- The key objectives of healthcare system are to ensure **access to quality healthcare** and generate the **greatest health benefit** for the **greatest number of people** with a **finite set of resources**.
- Health economics is the study of how **scarce resources** are **allocated & distributed rationally** within the health economy.

Importance of health economics

- **Priority setting:** By estimation of disease burden
- **Policy formulation**/decision making
- Health care **production, distribution and consumption**
- **Demand assessment and supply analysis**
- **Cost analysis**
- **Economic evaluation**
- **Health service financing**
- **Health market regulation**

Importance of health economics

- **Analysis of investment** in health
- **Budget formulation**
- **Rationing healthcare**
- **Health insurance**
- **Managed care in health**
- **National income generation**

Cost in healthcare

- **Healthcare costs** are the costs that people, organizations, or a society have to pay directly or indirectly for **healthcare goods and services** that are meant to keep or improve a person's or a group's health.
- Healthcare costs can be of **two types**:

i. **Direct costs:** Direct healthcare cost or **medical cost** and direct non-healthcare cost or **non-medical cost**. Direct cost in healthcare includes- **physician's cost, cost of drug, laboratory cost, travel cost and cost of food**.

ii. **Indirect costs:** These include the cost related to productivity losses. Indirect cost in healthcare includes- **loss of income, cost of disability, social cost, family cost and unofficial payment.**

Demand

- **Demand** is an economic concept that describes a **consumer's desire** to **purchase** a good or service.
- The price that consumers are willing to pay for the good or service is used to **determine demand**.
- Demand is the **consumer's intention** to **buy** a specific product or service.
- Price appears to **increase** when there is a **greater demand** than there is **supply** for a certain good.
- Price **decrease** when a product's **supply exceeds its demand**.

Demand curve

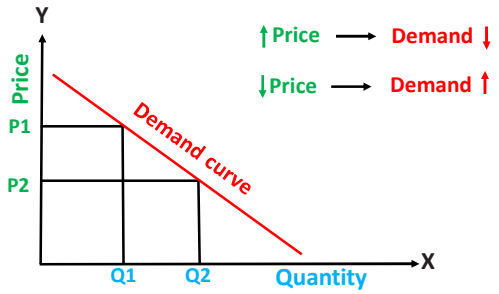


Fig: Demand curve

Supply

- **Supply** denotes willingness and ability to **sell commodity** at each and every price over a given period of time, subject to all else being constant.
- Industries or firms produce commodities subject to **capacity and profit maximization**.
- The number of things that a provider is willing to **manufacture and sell** is referred to as **supply**.
- The relationship between a **commodity's market price** and the amount of that commodity that producers are willing to **produce and sell** is stated to as the supply schedule of that commodity.

Supply curve

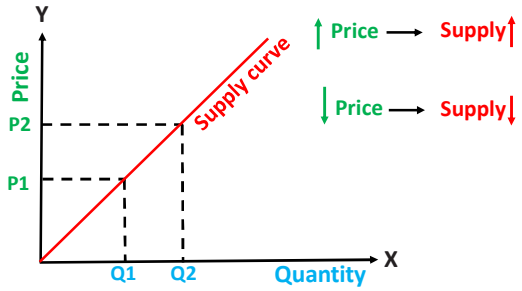


Fig: Supply curve

Methods of disease burden estimation

- Disease burden estimation is a crucial aspect of public health research and policy-making, providing insights into the impact of various health conditions on a population.
- Disease burden can be estimated using **three** methods:
  i. **Disability burden**
  ii. **Economic burden**
  iii. **Burden of consequences**

Measuring unit of disease burden

- Frequently disease burden is measured using **two widely accepted** indicators that facilitates comparison of disease burden and accounts for both **morbidity and mortality** in a single measure:
- **Quality Adjusted Life Years (QALY):** Measurement of life expectancy that takes into account the quality of life lost due to diseases and disabilities. QALY **'0'** means **death** and **'1'** means **perfect health**.
- **Disability Adjusted Life Years (DALY):** Represents the potential years of life lost due to premature death and the equivalent years of healthy life lost due to poor health or disability. DALY **1 year** is equivalent to **1 year** of **healthy life lost**.

Disability Adjusted Life Years (DALY)

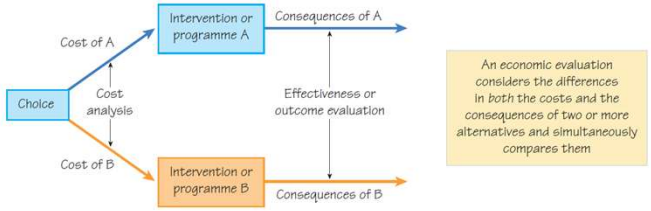
- **DALY** is a quantitative indicator of burden of disease that reflects the total amount of healthy life lost whether from **premature mortality** or from some degree of **disability** during the period of life.
- To calculate DALY, first we need to calculate the **Y**ears of **L**ife **L**ost due to premature death (**YLL**) and **Y**ears of **L**ife lived with **D**isability (**YLD**).
- DALYs lost are due to-
  i. **Immediate death**
  ii. **Death following disability**
  iii. **Permanent disability**
  iv. **Disability followed by complete recovery**

Quality Adjusted Life Years (QALY)

- **QALY** is a generic measure of disease burden, including both the quality and the quantity of life lived.
- It is used in **economic evaluation** to assess the value of **medical intervention**.
- The **outcome of health intervention** for a specific disease is estimated by **QALY**.
- **QALY= Life years gain + Quality of life**
- **1 QALY** equates to one year in **perfect health** and **0 QALY** means one year of **total functional impairment (death).**

What is an economic evaluation?

- Economic evaluation is simply the **comparative evaluation** of *both* the costs and the effects of ***two or more alternatives***.
- An **economic evaluation** is a comparison of the **costs and outcomes** of health care interventions thus it provides a measurement of **economic efficiency**.



3 main types of economic evaluation

Cost-benefit analysis	Outcomes expressed as health and other benefits are measured, valued in monetary terms, and costs are then deducted (e.g. £ or $)	Results expressed as net benefit, in £s or $s (if positive, benefits exceed costs A then intervention A should be implemented)
Cost-effectiveness analysis	Outcomes are expressed in 'clinical units'	£ or $ per case detected, per life-year saved, or per acute episode avoided
Cost-utility analysis	Outcomes are expressed in terms of quality-adjusted life-years (QALYS) or Disability-Adjusted-Life- Years (DALYs)	£ or $ per QALY gained

References

- Park, K. (2023) Parks Textbook of Preventive and Social Medicine. 27th Edition, M/S Banarsidas Bhanot Publishers, Jabalpur.
- Rashid, Khabir, Hyder's Textbook of Community Medicine and Public Health, 6th Edition, MAHS Publishers, Dhaka.
- Fundamentals of biostatistics by Prof. M.A. Hafez.
- Health economics by Prof. Md. Ziaul Islam PhD
- Public health and epidemiology at a glance / Margaret Somerville, Kalyanaraman Kumaran, Robert Anderson