2024 SUMMARY

COMMUNITY MEDICINE

DR. AHMAD AL-QAWASMI

ARKAN ACADEMY

Biostatistics

- **Tabulation**
  - *Simplifies* data for easier understanding and *saves space* by condensing data meaningfully
  - Facilitates *comparison* between observations, Provides a reference for future studies
  - Aids statistical analysis by *preparing data* systematically
- Classification of Data:
  - Geographical Classification: Based on location (population by regions)
  - Chronological Classification: Based on time (yearly statistics)
  - Qualitative Classification: Based on attributes like gender, education,…
    - ✓ Simple Classification: Based on **one** attribute
    - ✓ Manifold Classification: Based on **multiple** attributes with sub-categories

- Methods Summarizing **Qualitative** Data include:
  - Frequency Distribution: Organizing data into non-overlapping classes.
  - Relative Frequency Distribution: Fraction of total observations in each class.
  - Percent Frequency Distribution: Relative frequency multiplied by 100
  - Bar Graphs & Pie Charts: Graphical representation of categorical data

- Summarizing **Quantitative** Data involves:
  - Recategorization and frequency tabulation
  - Construction of frequency tables for numeric data.
  - Use of graphs such as **histograms, cumulative distributions (ogives), and dot plots**
- Steps in Data Tabulation: Find minimum and maximum values, then Calculate the range and decide number and width of classes and lastly construct frequency distribution tables.
- Graphical Representation
  - **Bar Graphs:** Display categorical data with separated bars
  - **Pie Charts:** Show proportional data as segments of a circle
  - **Histograms:** Show quantitative data with adjacent bars (no natural separation between bars)
  - **Ogives:** Represent cumulative frequency data
  - **Dot Plots:** Represent frequency using dots

- Guidelines for Creating Tables, should include *title*, *column headings*, *footnotes*
- Guidelines for Creating Graphs, should include: Be clear with labeled axes and appropriate scales
  - Avoid 3D effects that may mislead interpretation

- Importance of Tables and Graphs
  - Summarize data concisely for easier interpretation
  - Save time and reduce word count in research.
  - Poorly constructed visuals can lead to misinterpretation

- **Probability experiment** refers to a process with uncertain outcomes (e.g., coin flips, dice rolls)
- **Probability distribution** shows the possible values of a random variable can take and their associated probabilities
  - Each probability must be between 0 and 1
  - The sum of probabilities must equal 1

- Random Variables include:
  - Discrete Random Variables: **Countable values** (e.g., number of heads in 10 coin flips).
    - ✓ Examples: Number of calls to a help center, number of defective items
    - ✓ Common discrete distributions:
    1. **Binomial Distribution:** Fixed trials with two possible outcomes (success/failure)
    2. **Poisson Distribution:** Random events over a fixed time/space (hospital emergencies)

  - Continuous Random Variables: **Infinite possible values** within a range (e.g., height, weight).
    - ✓ Continuous probability distributions are described using smooth curves, such as histograms and probability density functions (PDFs)
    - ✓ The **area under the curve** represents total probability (equals 1)
    - ✓ **Normal Distribution (Gaussian Curve):** Symmetrical, unimodal bell curve centered at the mean (μ) and Defined by mean (μ) and standard deviation (σ)
      - 68% : Covers data within $\pm 1\sigma$
      - 95% : Covers data within $\pm 1.96\sigma$
      - 99.7% : Covers data within $\pm 3\sigma$
      - 99.99% : Covers data within $\pm 4\sigma$

- **Central Limit Theorem (CLT)** states that for a sufficiently **large sample size**:
  - The sample mean approximates the population mean.
  - Sample distribution approaches normality, regardless of the population's original distribution
  - Enables making population inferences using sample data.
  - Works with any distribution if sample size is large enough (typically **$n \geq 30$**)
    - ✓ If data distribution is symmetric: $n \geq 15n$
    - ✓ If data is skewed: $n \geq 30n$
    - ✓ If highly skewed: $n \geq 40n$
  - Standard error can be used to calculate sample standard deviation in normal distribution

- The **standard normal distribution (Z-distribution)** : **Mean (μ) = 0 , Standard deviation (σ) = 1**
  - Z-score allow **comparison** between datasets
  - Standardization (Z-normalization): is done by this formula $$z = \frac{(x - \mu)}{\sigma}$$
    - ✓ A Z-score of 0 indicates the value is equal to the mean
    - ✓ Positive Z-scores indicate values above the mean.
    - ✓ Negative Z-scores indicate values below the mean
    - ✓ It determines how does the value away from the mean
      - A Z-score of **-0.8** indicates a value is 0.8 standard deviations below the mean
      - A Z-score of **2** indicates a value is 2 standard deviations above the mean.

- **Research Hypothesis** is a clear, testable statement predicting the relationship between variables.
  - It must involve at least two variables
  - Should suggest a relationship ("more than," "different from")
  - Must be testable, specific, and predictive
- **Null Hypothesis (H_0):** Assumes **no** effect or difference (no relationship, independent)
- **Alternative Hypothesis (H_1):** Assumes an effect or difference exists (relationship, dependent)
  - If sample data significantly differ from H_0, it is rejected in favor of H_1.
  - If differences are not significant, H_0 cannot be rejected.

- Types of Errors in Hypothesis Testing
  1. **Type I Error (α):**
  - Occurs when **H_0 is wrongly rejected** (false positive)
  - Example: Concluding smoking causes cancer when it does not
  2. **Type II Error (β):**
  - Occurs when **H_0 is wrongly accepted** (false negative)
  - Example: Concluding smoking has no effect when it actually does.

- Parametric Tests (Assumptions required): Observations must be independent
  - The *dependent variable should be continuous* (interval/ratio)
  - Data should follow a normal distribution (large sample >30)
  - Groups should have equal variances (homogeneity)
  - Common parametric tests: **T-test, ANOVA (Analysis of Variance)**
  - Advantages: More powerful and flexible, Allow studying multiple variables

- Non-Parametric Tests (Fewer assumptions): Used for nominal or ordinal data.
  - **Distribution-free** methods (unknown distribution method) and suitable for **small sample** sizes
  - Common non-parametric tests: **Chi-square test, Mann-Whitney U test, Kruskal-Wallis test**
    - ✓ In Chi-square test, data must be in raw frequencies, any observation must have at *least a frequency of 5* and the total observations must be more than 20
  - Advantages: No assumptions about population distribution

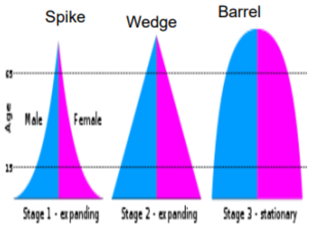
- If the probability (P-value) of the test statistic is **less than or equal** to the probability of the alpha error rate (usually 0.05), we reject the null hypothesis and conclude that there is a **relationship** between the variables (**dependent**)
- If the probability of the test statistic is **greater than** the probability of the alpha error rate, we fail to reject the null hypothesis. We conclude that there is **no relationship** between the variables, i.e. they are **independent**.

Research (Standardization)

- Demography: Study of population size, density, fertility, mortality, and growth
  - Formula for population size dynamics: **P2 = P1 + B – D + IM – EM**
- **Population doubling time (PDT):** Number of years it will take for the population to double in size
  - PDT = 70 / annual percent
  - The population can only increase if the number of births exceeds the number of deaths

- Sources of Demographic Information
  - **Census:** costly and slow
    - ✓ Decennial: poll count on 100% sample held every 10 years
    - ✓ Midcensus: poll count on 10% sample held every 10 years between censuses
  - Population **registers** and vital event registrations
  - Sample household **surveys** and **governmental/private records**

- Population Data Types: Population size, mortality, birth/fertility rates, mobility, and composition
- Age and sex distribution analyzed via population pyramids
  - **Spike:** High BR. High DR, low growth rate (under-developed country)
  - **Barrel shape:** low BR and low DR at younger ages (developed country)
  - **Wedge shape:** high BR and low DR, high growth rate



- Influencing factors od mortality: age structure (main), environment, economic and technological development, medical services
  - Common causes: infectious diseases (primitive societies), chronic diseases (modern societies)
  - Population health can be compared by death rate, age-specific death rate, standardized death rate (SDR), and standardized mortality ratio (SMR)

- Standardization is used to eliminate demographic differences when comparing populations
  - **Direct Standardization:** Uses a reference population to adjust age structure differences
    - ✓ Requires age-specific rates and large populations
  - **Indirect Standardization:** Uses a standard population's mortality rates to expect deaths.
    - ✓ Easier to apply when age-specific rates are unavailable
    - ✓ Produces **SMRs** for comparison across different populations.
- Standardized Mortality Ratio (SMR) Calculation
  - SMR > 100: Higher mortality compared to reference
  - SMR < 100: Lower mortality compared to reference
  - SMR=120 means that the mortality in the study population is 20% higher than in the reference population
  - SMR=50 means that the mortality in the study population is 50% less than in the reference population.

$$SMR = \frac{Observed\ deaths}{Expected\ deaths} \times 100$$

	Country A	Country B
	Expected deaths	Expected deaths
0-29	0.0012 x 6,000,000 = 7,200	0.0012 x 1,500,000 = 1,800
30-59	0.0036 x 5,500,000 = 19,800	0.0036 x 550,000 = 1,980
60+	0.048 x 2,500,000 = 120,000	0.048 x 120,000 = 5,760
Total expected deaths (E)	147,000	9,540
Total observed deaths (O)	147,000	15,300
Standardised Mortality Ratio O/E x 100	100	160

ARKAN
ACADEMY

علم في كل مكان

Arkan academy

Arkanacademy

www.arkan-academy.com

+962 790408805