TOPIC 1 INTRODUCTION TO INFORMATION TECHNOLOGY FOR STATISTICS

History of Computing

- In the 1960s and 1970s, "IT" was a technical term used mainly within hospitals, banks and research institutions to describe early mechanical and electronic processes for storing and retrieving institutional data. These systems were designed primarily to handle structured records such as patient files, financial ledgers and laboratory logs.

- The rise of personal computers, relational databases and networked environments by the 1980s and 1990s transformed workplaces into "paperless" ecosystems. Due to expansion of computational capacity, it led to increased demand on IT to support increasingly complex analytical tasks, particularly in fields where data volume and decision stakes are high e.g financial markets.

ROLE OF IT IN THE SOCIETY

* INFORMATION TECHNOLOGY IS DEFINED AS -> The study, design, development, application / implementation, support or management of computer-based Information Systems.

1. Automate institutional work flows e.g in hospitals
2. Support high-stakes decision making through real-time information access.
3. Integrate data from heterogeneous sources i.e stock exchanges.
4. Provide productivity tools that enhance the speed and transparency of statistical operations.

ACADEMIC / PROFESSIONAL FIELD

5. Selecting appropriate hardware and software for an organisation.
6. Maintaining, upgrading and securing the computational infrastructure
7. Managing data capture tools for Cohort Studies
8. Configuring servers for R, Python and SAS Computing
9. Ensuring Security of trading algorithms and risk models
10. Managing distributed databases for tick-level or Sentimental-analysis data.
11. Deploying low-latency architectures for real time market feeds.

12. Store large datasets e.g Transactional
13. Transmit data securely across networks
14. Protect Sensitive information
15. Retrieve information rapidly for decision making

BASIC COMPUTER TERMINOLOGY

1. Central Processing Unit (CPU)

- It executes instructions organized into programs that determine computers actions
  Role in Biostatistics -> Execute statistical algorithms like survival analysis
  -> Enable Simulations i.e Epidemiological modelling

2. Memory (RAM)

- Used to temporarily store active programs, data currently being processed and intermediate results
  Role/Relevance in Biostatistics -> Holding large patient datasets for e.g during longitudinal analyses

NB -> Insufficient RAM causes disk Swapping which drastically slows computations in both health and financial simulations

3 Mass Storage Devices

- This is the slower, long-term memory that retains data between computational jobs e.g Hard Disk Drives (HDDs)
  - Solid-State Drives (SSD
  - Tape Drives (archival)

Role in Biostatistics -> · Genomic Sequences
· Longitudinal Surveys
· Historical Clinical trial data

SSDs -> They offer high read, crucial for large-scale risk simulations

Tape drives -> Relevant for long-term archival in hospitals or financial institutions requiring regulatory compliance.

4. Input Devices

- They deliver data and instructions into the system e.g Clinical trial data, lab measurements and Survey inputs in Biostatistics

5. Output Devices

- Used to visualize Computational results.
  e.g Monitors, printers, dashboards.
  Computational results such as Epidemic Curves and Gene expression heatmaps in Biostatistics.

6. Supporting Components: The Bus and Interconnects

- The bus is the communication backbone linking CPUs, memory storage and I/O systems to ensure efficient Computational performance.

Role in Biostatistics -> Influences large matrix operations & distributed database queries.

SUMMARY

Component	Purpose	Biostatistic Relevance
CPU	Execute Instructions	Simulations, Model fitting
RAM	Temporary Storage	Intermediate dataset
Mass Storage	Permanent data retention	longitudinal datasets, Genomic archives,
Input Devices	Data Entry	Clinical, lab, Survey data.
Output Devices	Data visualization	Survival Curves, heatmaps.
Bus	Data transfer	Efficient multi-site processing.

Hardware - Physical Component of a Computer that are tangible i.e keyboard, mouse

Software -> These are programs that run on a Computer

Operating Systems (OS) -> The main Software that manages Computer hardware and allows other programs to run i.e Windows, Android

Network - A group of Connected Computers that share resources

Internet - A global network that Connects Computers worldwide

IP Address -> A unique number that identifies a device on a network.

File -> A digital document or piece of information stored on a computer.

Backup -> A copy of data stored to prevent loss

T.2. FUNDAMENTALS OF COMPUTER OPERATIONS

DATA VS INFORMATION

Data: They are explained as plain facts.
-> When data are processed, organized, structured or presented in a given context so as to make them useful, they are called Information.

- When data is interpreted and processed to determine its true meaning, they become useful and can be called Information.

Data is the Computer's language. Information is our translation of this language.

COMPUTER SIZES AND POWER

1. Personal Computer -> A small, single-user Computer based on a microprocessor.

2. Workstation - A powerful, single-user Computer. A workstation is like a personal computer, but it has a more powerful microprocessor and, in general, a higher-quality monitor.

3. Minicomputer - A multi-user Computer capable of supporting up to hundreds of users simultaneously.

4. Mainframe :-> A powerful multi-user Computer capable of supporting many hundreds or thousand of users simultaneously.

5. Super Computer :-> An extremely fast Computer that can perform hundreds of millions of instructions per second.

T2. FUNDAMENTALS OF COMPUTER OPERATIONS

1. CPU

- The heart of the Computer System is the processor unit.
- It Consists of Arithmetic and Logic Unit and Control Unit.

(a) Arithmetic and logic Unit (ALU)

- This is where most arithmetical and logical data Computer operations are executed.
- E.g. Suppose two numbers (Operands) located in main memory are to be added. These operands are brought into arithmetic unit - actual addition is Carried. The result is then Stored in the memory or retained in the processor itself for immediate use.
- Note that all operands may not reside in the main memory. processor Contains a number of high speed Storage elements Called Registers, which may be used for temporary Storage of frequently used operands. Each register can Store one word of data.

(b) Control Unit

- The operations of all the units are Coordinated by the Control unit (Cast as the nerve Center that Sends Control Signal to other units).
- Timing Signal that governs the I/O transfers are generated by the Control unit.
- It generate Synchronization Signals
- It maintains order and direct the operation of the entire System by Selecting, Interpreting and executing the program instructions

The System loads instructions

2. REGISTERS

INTERNAL MEMORY ORGANIZATION OF PROCESSOR

- Processor Contains a number of registers used for temporary Storage of data other than ALU and Control Circuitry.

(a) Instruction Register (IR) - It holds the instruction that is Currently being executed. Its output is available to the Control Circuits which generate the timing Signals that Control the various processing elements involved in executing the instruction (System
- Generates clock Pulses
- Determines CPU speed
- Synchronizes Computer components

(b) Program Counter (PC) - It Contains the address of the instruction Currently being executed.

(c) General Purpose Registers (R_0 to R_{n-1}) - facilitates communication with the main memory.

(d) Memory Address Register (MAR) - holds the address of the location to or from which data are to be transferred

(e) Memory Data Register (MDR) - Contains the data to be written into or read out of the address location

3. Machine Cycles

- This is the Complete set of steps the CPU goes through to process a single instruction or a single piece of data.
- It usually has four main steps

(a) Fetch → The CPU gets (fetches) the instruction from memory

(b) Decode → The CPU interprets (decodes) the instruction to understand what action is required

(c) Execute → The CPU performs the operation e.g. addition, Comparison, reading/writing data e.t.c

(d) Store → The result of the operation is written back to memory or a register.

- A Clock Cycle is one pulse from the system clock.
- A machine cycle may take multiple clock cycles to complete depending on the CPU design and instruction complexity.

Why Machine Cycles Matter:

- They determine how fast a CPU processes instructions
- CPUs with fewer machine cycles per instruction are faster.

TOPIC 3 - COMPUTER HARDWARE I: INPUT AND OUTPUT DEVICES

INPUT DEVICES -> Any hardware Component that allows you to enter data, programs, commands and user responses into a Computer. e.g - Keyboard
- Mouse
- Touchpad
- Joystick
- Scanners

OUTPUT DEVICES -> They make the information resulting from processing available for use. e.g Printers
Display devices/monitors
Speakers
Head phones

(a) Bar Code reader → Input packaging information, used primarily in Stores
(b) Camera → Input video image (Still or moving)
(c) Goggles → Output video image, used primarily in virtual reality
(d) Joystick → Input motion/pointing information, used primarily for computer games.
(e) Keyboard → Input text information, primary means for most users
(f) Microphone → Input voice information, used in cases where either the user is unable to use hands or has a large amount of information to input.
(g) Monitor → Primary output device
(h) Mouse → Input user interactions with GUI windowing system.
(i) MIDI device → Input of musical instrument data
(j) Pen tablet → Input written information when keyboard is undesirable.
(k) Printer → Output text &/or images to paper.
(l) Scanner → Input text and/or images from paper.
(m) Speakers → Output music and sound
(n) Touchpad → Alternate pointing input device when mouse is not desired.
(o) Touch Screen → Alternate pointing input device when mouse and touch pad is not desired.
(p) Trackball → Alternate pointing input device when mouse is not desired. Sometimes used for computer games.

TOPIC 4
T4: COMPUTER HARDWARE II
STORAGE DEVICES AND MEMORY

(a) Primary Memory (Main Memory)
Memory unit

- The memory unit stores program and data. There are two classes of memory devices: Primary memory, Secondary memory

(A) Primary Memory (Main Memory)

- Contains a large no. of semiconductor cells each capable of storing one bit of information.
- For accessing data, a distinct address is associated with each word location.
- Data and programs must be in the primary memory for execution.
- Fast memory
- Expensive
- Time required to access one word is called Memory Access Time - 10ns to 100ns. This time is fixed and independent of the location.

- The primary memory allows the computer to store data for immediate manipulation and to keep track of what is currently being processed.
- It is volatile that is when the power is turned off, the contents of primary memory are lost forever hence to store the data permanently, a computer requires some non volatile storage medium like a hard disk.
- This kind of storage is known as Secondary memory. Such memories store all the data (files) and instructions even after the power is turned off.

(B) SECONDARY STORAGE

- They are used when large amount of data have to be stored.
- The Secondary Storage devices have a larger storage capacity; they are less expensive as compared to primary storage devices, but slow in comparison.

NB → When we talk about memory, we generally refer to the primary memory only, and when we talk about storage, we refer to Secondary memory.

Various units used to measure Computer memory areas follows:

Bit: It is the smallest unit of data on a machine and a single bit can hold only one of two values: 0 or 1

Byte: A unit of eight bits is known as byte. Hence a byte is able to contain any binary number btwn 00000000 and 11111111. It is represented by an upper case B.

Kibbyte: ~~In a decimal System~~ Kilo refers to 1024. Therefore, a kilobyte is equal to 1024 bytes. It is usually rep as 1KB

Megabyte: It Comprises 1024 kilobytes or 1,048,576 bytes. It is the standard unit of measurement for RAM & is rep as MB.

Gigabyte: It Consists of 1024 megabytes. It is the standard unit of measurement for hard disks and is often rep as GB.

Terabyte: It refers to 1024 gigabytes. Represented as T

MEMORY HIERARCHY

- The memory in a Computer System is of three fundamental types:

1. Internal Processor Memory - It is placed within the CPU or it is attached to a special fast bus
- It usually includes Cache memory and Special registers, both of which can be directly accessed by the processor.

a) Cache Memory → A buffer, Smaller and faster than main Storage used to hold a copy of instructions and data in main Storage that are likely to be needed by the processor and that have been obtained automatically from main Storage

2. Primary Memory → RAM
ROM

RAM → is the place where the Computer temporarily stores its operating System, application programs and Current data So that the Computer's processor Can reach them quickly and easily

Types of RAM

There are two types of RAM used in PCs:
Static RAM - The information remains Stable as long as power Supply is Supplied but data are lost as soon as the power goes down. Hence SRAM is volatile

Dynamic RAM - It is named So because it is unstable. The data Continue to move in and out of the memory as long as power is available. Information needs to be refreshed after every few miliseconds to avoid it being erased.

ROM → The Storage of program and data in the ROM is Permanent.
There are two types of ROM.

a) Programmable ROM (PROM) - It is programmable. However, when the chip has been programmed, the recorded information cannot be changed.

b) Erasable PROM (EPROM) - Information can be erased and the chip programmed to record different information

3. Secondary Memory / Storage devices
- Secondary memory provides back up Storage for instructions and data.

- Computers use disks for storage.
Hard disks located inside the PC and those that are attached to the PC externally.
- They are non-volatile.
- Can be easily removed and attached to other devices i.e Memory sticks, floppy disks, CDs.

Cloud Storage

- A disk drive track is a circular path on the surface of a disk or diskette on which information is magnetically recorded and from which recorded information is read.

- A ~~track~~ track - is a physical division of data in a disk drive, as used in the Cylinder-Head-Record (CCHHR) addressing mode of a CKD disk.
- Tracks are subdivided into blocks or Sectors, pages.

- A Sector - is a subdivision of a track on a magnetic disk or optical disk. Each sector stores a fixed amount of user data. The disk sector refers to the intersection of a track and mathematical sector.

- Clusters -> Also called grid storage. It is a new technology paradigm that pushes the scalability and efficiency boundaries of storage area networks to (SANs) to new levels. They are typically made up of network connected storage with an administrative function that manages a collection of physical ~~data~~ disks.

Implications for large-scale Biostatistical ~~&~~ ~~financial~~ data sets

1(a) - Scalability of performance i.e. - Enables faster computation of complex models
2(b) - Easy management of a single System Image - organizations also get more value from the hardware purchased
3(c) - Non-disruptive data movement
4(d) - Higher utilization rates -> This means that fewer servers or extra systems need to be bought hence cost efficiency
5(e) - Lower hardware acquisition costs
6(f) - Ability to sustain failure of multiple elements without affecting data access - Even if servers or disks fail, biostatisticians can still access their datasets with no downtime hence continuous disease surveillance and real-time analytics.

1(a) - Organizations can process complex analyses faster, run more models simultaneously and support real-time data processing without performance degradation.

2(b) - Biostatistical teams focus more on analysis and less on system administration complexity, improving efficiency and ~~and re~~

3(c) Researchers and analysts continue working uninterrupted, ensuring continuous access to datasets and maintaining workflow stability

5e - Organizations can process massive datasets using economical hardware options or scalable cloud services - making high-end biostatistical research accessible even with limited ...

T5- COMPUTER SOFTWARE

TOPIC 5 SYSTEM SOFTWARE

- It is a program that manages and supports the computer resources and operations of a computer system while it executes various tasks. Such as processing data and information, controlling hardware components and allowing users to use application software.

- That is System software functions as a bridge between computer system hardware and the application software.

- System software is made up of many control programs including Operating systems, Communication software and database manager.

- System software consists of three kinds of programs (a) System Management Programs
  (b) System Support programs
  (c) System Development Programs

(a) System Management Programs -> These are programs that manage the application software, computer hardware and data resources of the computer system.

(b) System Support Programs -> These are the programs that help the operations and management of a computer system.

(c) System Development Programs -> These are programs that help users develop information system programs and prepare user programs for computer processing.

OPERATING SYSTEMS

- An operating system is a collection of integrated computer programs that provide recurring services to other programs or to the user of a computer.
- These services consist of a disk and file management, memory management and device management.

- An operating system executes many functions to operate computer system efficiently. Among them, four essential functions are the following.

(a) Resource Management → An operating System manages a Collection of Computer (hardware resources) by using a variety of programs. It manages Computer System resources including its CPU, primary memory, virtual memory, Secondary Storage devices, Input/Output peripherals and other devices.

(b) Task Management:- It manages one program or many programs within a Computer System Simultaneously. A task Management program in an O.S provides each task and interrupts the CPU operations to manage tasks efficiently. Task management may involve multitasking Capability.

(c) File management:- → This is a function that manages data files. An Operating System Contains file management programs that provide the ability to create, delete, enter, change, ask and access of files of data. They also produce reports on a file.

(d) User Interface: It is a function of an operating System that allows users to interact with a Computer. i.e Screen design, keyboard Commands. A well-designed user interface is essential for an operating System to be popular.

UTILITY SOFTWARE

- Is a System Software designed to help analyze, Configure optimize, or maintain a Computer. A Single piece of utility Software is usually Called a Utility or tool.

- Utility Software usually focuses on how the Computer infrastructure operates.

Types of Utilities

- Anti-virus utilities Scan for Computer viruses
- Backup utilities can make a copy of all information Stored on a disk and restore either the entire disk or Selected files.
- Data Compression utilities output a Shorter Stream or a Smaller file when provided with a Stream or file.
- Disk Checkers Can Scan operating hard drive
- Disk Cleaners Can find files that are unnecessary to Computer operation, or take up Considerable amounts of Space

- Disk Compression utilities can transparently compress/uncompress the contents of a disk, increasing the capacity of the disk.
- Disk defragmenters: Can detect computer files whose contents are broken across several locations on the hard disk, and move the fragments to one location to increase efficiency.
- Disk partitions - Can divide an individual drive into multiple logical drives, each with its own file system which can be mounted by the operating system and treated as an individual drive.
- Disk space analyzers - for the visualization of disk space usage by getting the size for each folder and files in folder or drive showing the distribution of the used space.
- Archive utilities output a stream or a single file when provided with a directory or a set of files.
- File managers: Provide a convenient method of performing routine data management tasks such as moving, copying, merging among others (creating and modifying data sets, renaming, deleting).
-
- Cryptographic utilities encrypt and decrypt streams and files.
- Hex editors directly modify the text or data of a file.
- Memory testers check for memory failures.
- Network utilities analyze the computer's network connectivity, configure network settings, check data transfer or log events.
- Registry Cleaners: Clean and Optimize the windows registry by removing old registry keys that are no longer in use.
- System monitors - for monitoring resources and performance in a computer system.
- System profilers provide detailed information about the software installed and hardware attached to the to the computer.

COMPILING PROGRAM

- A Compiler is a computer program that transforms source code written in a programming language into another computer language.

- The most common reason for wanting to transform source code is to create an executable program.

- A program that translates from a low level language to a higher level one is a decompiler.

TG: DATA FILES AND FILE MANAGEMENT

Meaning of Data

- Data refers to raw facts, Symbols, figures or observations that have not yet been processed.
- It has no meaning until it is interpreted or organized.

Examples → Numbers (Toso, 3.14)
- Words or letters
- Images, audio, video
- Sensor readings.

Characteristics of Good Data

1. Accurate - free from errors
2. Complete - Contains all necessary details
3. Relevant - Relates to the subject matter
4. Reliable - Can be trusted
5. Timely - Available when needed
6. Consistent does not contradict itself.

Types of Data

(a) According to Nature

1. Numeric Data - Numbers (integers, decimals)
2. Textual data - letters, words, Sentences
3. Boolean data - True/false, Yes/No
4. Audio data - Recordings/Music
5. Image data - photos, Scans
6. Video) data - Moving pictures
7. Graphical data - Graphs, diagrams

(b) According to Structure

1. Structured data - Neatly organized (tables, Spread sheets)
2. Unstructured data - No fixed format (videos, PDFs)
3. Semi-Structured data - Mix of both (Emails)

INFORMATION VS DATA

Data	Information
Raw facts	Processed data
No meaning	Has meaning

Information = Data + Processing

Data processing

These are steps involved in converting data into information.

- Collection - Gathering raw data
- Input - Entering data into a system
- Processing - Sorting, Calculating, Summarizing
- Output - Presenting results (reports, charts)
- Storage - Saving for future use
- Distribution - Sharing information with users

Data Representation in Computers

- Computers store everything using binary digits (0 and 1)

Units -> Bit - Smallest unit of data
Byte - 8 bits
Kilobyte (KB) = 1024 bytes
Megabyte (MB) - 1024 KB
(GB) Gigabyte = 1024 MB
Terabyte (TB) - 1024 GB

Meaning of a Data file

- A data file is a collection of related data stored together under one name.
- It has a filename and an extension (e.g. report.docx, image)
- Files help organize and retrieve information quickly.

Types of Data files

(a) Text files
- Contain readable characters
- e.g. .txt, .docx
(b) Binary files
- Data stored in binary form.
- e.g. .exe, .bin, .dat
(c) Image files
- .jpg, .png, .gif
(d) Audio files
- .MP3, .wav, .aac
(e) Video files
- .MP4, .mkv, .avi
(f) Program files
- Executable or System files
(g) Database files
- .db, .mdb, .sql
(h) Compressed files
- .zip, .rar, .7z

File Organization Method

- Used especially in databases

1. Sequential file Organization
- Records Stored one after another in order
- Best for batch processing

2. Direct/Random file Organization
- Records accessed using a key field
- Fast for large files

3. Indexed-Sequential Organization
- Combines Sequential and direct access
- Uses an index for quick Searches

FILE MANAGEMENT

- Activities involved in handling files include:

(a) Creating files
(b) Naming files
(c) Opening and closing files
(d) Saving and retrieving files
(e) Renaming files
(f) Moving files btwn folders
(g) Deleting files
(h) Backing up important data

File Extensions

Common file extensions and their meanings:

-> .txt - text file
- .docx - Word document
- .pdf - Portable document
- .jpg/.png - Image
- .mp3 - Sound
- .mp4 - Video
- .exe - executable
- .xlsx - Spreadsheet

File Security

- Methods of protecting files include:
- Passwords - restrict access
- Encryption - Scramble data so only authorized users can read it
- Access rights/permissions -> Control who can read edit or delete files.
- Firewalls - block unauthorized network access
- Antivirus software - protects files from malware

File Storage Devices

- HDD (Hard Disk Drive) - large capacity, cheaper
- SSD (Solid State Drive) - faster, more durable
- USB flash drives - Portable & Convenient
- Memory Cards - Used in phones/cameras
- Optical discs - Compact discs, DVDs
- Cloud Storage - Google Drive, OneDrive, iCloud

Data Backup

Purpose - to prevent loss of data

- Backup methods:
  - -> Full backup - Copies everything
  - Incremental backup - Copies only changes since last backup.
  - -> Differential backup -> Copies changes since last full backup
  - -> Cloud backup - remote servers
  - -> External drive backup - Offline backups

Importance of Data and File Management

- Prevents data loss
- Ensures quick access to information
- Reduces errors and duplication
- Improves decision-making
- Enhances security & organization
- Increases efficiency