NJUGUNA SHALOM WANJIKU

Scm222-0837/2025

SUMMARY NOTES – IT FOR STATISTICS

Financial Engineering involves applying Mathematics, Statistics and Computer Science to solve problems in Finance.

Information Technology (IT) refers to the use of computers, software, networks, and related technologies to manage, process, store, and communicate information.

Computer: An electronic device that processes data into information using electronic signals.

TOPIC 1: ROLE OF INFORMATION TECHNOLOGY AND COMPUTERS IN TODAY'S WORLD

IT involves using computers and software to manage, process, and communicate information.

Applications

1. **Medical Research:** IT tools support medical analysis, research, and discovery.
2. **Education:** E-learning platforms and digital resources make information accessible and flexible.
3. **Agriculture:** IT tools help analyze weather patterns to guide planting and harvesting decisions.
4. **Environmental Sustainability:** IT supports remote work and digital systems that reduce resource consumption and carbon emissions.
5. **Communication & Collaboration:** Cloud platforms (e.g., Google Workspace, Microsoft Teams) and APIs allow seamless sharing of models and data globally.

Core Uses of IT in Finance

- **Data Processing & Analysis:** Computers clean, sort, and analyze large datasets (stock prices, economic indicators, trade feeds).
  Example: Running a Monte Carlo simulation to forecast the future price of a derivative.
- **Storage & Retrieval:** IT systems store portfolios, documents, and decades of market data in searchable databases.
- **Automation:** Computers perform repetitive tasks quickly and accurately.
  Example: High-Frequency Trading algorithms executing thousands of trades per second.

TOPIC 2: FUNDAMENTALS OF COMPUTER OPERATIONS

The core operation of a computer follows the **Input–Process–Output (IPO) Cycle**, guided by stored instructions.

1. **Input:** Entering commands or data through input devices such as the keyboard and mouse.
2. **Processing:** The CPU fetches instructions from memory and performs calculations on the data.
3. **Output:** The computer displays or returns the results of processing.
4. **Storage:** Data and results can be saved for later use on storage devices (e.g., hard disk, SSD, database).

The Fetch–Decode–Execute Cycle

The CPU repeatedly:

1. Fetches the next instruction from memory.
2. Decodes it to determine the required operation.
3. Executes the instruction.
   This cycle occurs billions of times per second.

TOPIC 3: COMPUTER HARDWARE AND SOFTWARE

a) Hardware

The physical, tangible components of a computer.

Key Components

1. **CPU (Central Processing Unit):** Executes instructions; supports parallel processing for faster computations in simulations and statistical tasks.
2. **RAM (Random Access Memory):** Short-term, fast memory used to store active data and programs. More RAM allows larger datasets to be loaded (e.g., 16GB minimum, 32GB+ for large models).

3. **Storage (HDD or SSD):** Long-term data storage. SSDs are faster and more reliable than HDDs.
4. **Motherboard:** Connects and coordinates all hardware components.
5. **Input and Output Devices:** Used for interaction with the computer (keyboard, mouse, monitor).

b) Software

The intangible instructions and programs that run on a computer.

System Software

- **Operating System (OS):** Windows, macOS, Linux. Manages hardware resources, files, processes, and provides the user interface.

Application Software

- **General-Purpose:** Excel, used for analysis and prototyping.
- **Specialized Statistical Software:** R, Python (Pandas, NumPy, SciPy), MATLAB, SAS.
- **Database Software:** SQL systems (MySQL, PostgreSQL) for managing large datasets.

UNIT 4: COMPUTER SOFTWARE BASICS

4.1 Operating System Functions

- **Kernel:** Core component that manages system resources and hardware interaction.
- **Virtual Memory:** Uses disk space (paging/swap file) to supplement physical RAM.
- **Multitasking:** Allows multiple processes to run by rapidly switching CPU time between them.

4.2 Software Licensing and Acquisition

- **Proprietary Software:** Requires purchase of a license (e.g., Microsoft Office).
- **Open-Source Software:** Freely available, modifiable source code (e.g., Linux, Python).
- **Software as a Service (SaaS):** Hosted online and accessed via the internet (e.g., Google Workspace, Salesforce).

4.3 Translating Software: Compilers vs Interpreters

- **Compiler:** Translates entire code into machine language before execution (e.g., C, C++).
- **Interpreter:** Translates and executes code line by line (e.g., Python, JavaScript).

TOPIC 5: CONSTRUCTING DATA FILES

A **data file** is a structured collection of information. Proper structuring ensures both humans and software can interpret it.

Common Data File Formats

1. **CSV (Comma-Separated Values):** Plain text, common for raw data.
2. **Excel (.xlsx):** Good for formatting and manual work.
3. **Plain Text (.txt):** Stores unstructured or simple textual data.

Methods of File Organization

1. **Serial (Sequential) Files:**
    - Records stored one after another as created.
    - **Pros:** Simple, fast to create.
    - **Cons:** Slow access.
2. **Random (Hashed) Files:**
    - Records stored at locations determined by a hash function applied to a key field.
    - Enables fast access to specific records.

TOPIC 6: DISK STORAGE

Physical data storage affects speed, durability, and cost.

Magnetic Storage (HDD – Hard Disk Drive)

- **How it works:** Uses spinning platters and a read/write head.
- **Pros:** Low cost, high capacity.
- **Cons:** Slower, mechanical failure risk.

Solid-State Storage (SSD – Solid-State Drive)

- **How it works:** Uses flash memory; no moving parts.
- **Pros:** Very fast, durable, energy-efficient.
- **Cons:** More expensive per GB.
- **Verdict:** Ideal for primary computing due to significant speed advantages.

Cloud Storage

- **How it works:** Data stored on remote servers accessed over the internet (AWS, Google Cloud, Azure).
- **Pros:** Scalable, accessible anywhere, reliable backups.
- **Cons:** Ongoing cost, internet-dependent.
- **Example:** A hedge fund may store petabytes of historical tick data on AWS S3 and use cloud computing services like EC2 for analysis.