STA 2102

IT FOR STATISTICS

1 Introduction to IT and computers

- Information technology is the use of computers, networks, software and other systems to store, process and transmit information.

Importance of IT in statistics & financial Engineering:

1. Helps in data collection, storage and analysis
2. Enables use of statistical softwares (R & Python)
3. Improves, speed, accuracy and reliability of financial models.
4. Supports automation of calculations like interest, risk analysis, forecasting and simulation

Computer

- An electronic device that accepts input, processes data, stores information and produces output.

Characteristics of computers

✓ Speed ✓ Accuracy ✓ storage ✓ Automation ✓ reliability

FUNDAMENTALS OF COMPUTER OPERATIONS

Computers work through input-process-output (IPO) cycle.

1) Input - entering data (keyboard, mouse, scanner)
2) Process - CPU converts data into information
3) Output - results displayed (monitor, printer).
4) Storage - saving data for later use (HDD, SSD, cloud).

Booting.

- The process of starting the computer.

Cold boot - turning on from power off
Warm boot - restarting

Types of software:

✓ System software - controls hardware (operating system like windows, linux)
✓ Application software - performs tasks (excel, spss, R, Python).

3. Computer Hardware Basics.

Computer hardware refers to the physical, tangible components of a computer system:- Components are:

a) Input devices, - keyboard, mouse, scanner, microphone, touchscreen, joystick, biometric devices, Barcode reader.
b) Output devices; monitor, printer, speaker, projector plotter.

Importance in statistics.
Performs calculations, graphs, simulations, regression, forecasting and data cleaning

5. Data and data files
Data are raw facts that have no meaning until processed.
Types of data
Qualitative - categories (gender)
Quantitative - numbers (age, income)
- Discrete (countable)
- Continuous (measurable)

Data files
This are ways data is stored in computers
- Text files (.txt)
- Spreadsheets (.xlsx)
- Databases (.db, .sql)
- CSV files (.csv)

Characteristics of good data
Accurate - free from errors
Complete - no missing values
Consistent - same format across files
Timely - up to data
Valid - follow rules
Reliable - trustworthy
Relevant - meets the purpose.

Data processing cycle.
1. Data collection - surveys, websites
2. Data entry - typing, scanning
3. Data cleaning - removing errors
4. Processing - calculations
5. Analysis - graphs, statistics.
6. Interpretation - conclusions.
7. Storage - saving in files
8. OUTPUT - reports

c) Storage devices - HDD, SSD, flash disk, memory card.
d) Processor central processing unit
- This is the brain of the computer
- It is a control unit - directs operations
- ALU - performs arithmetic and logical operations.
e) Memory
RAM - temporary, fast, volatile
ROM - permanent instructions

Computer categories.
- Microcomputers (PCs)
- Laptops
- Servers
- Supercomputers

4. Computer Software
Software is a set of instructions that tell the computer what to do.
① System software.
- Operating systems: windows, linux, macos.
- Utility programs: antivirus, disk clean up
- Drivers; allow hardware to communicate with the computer
② Application software
- Used for specific tasks
- Word processing (MS Word)
- Spreadsheets (Excel)
- Statistical software
- Database systems (

Types of Disk storage

1. Hard Disk Drive (HDD)
- Uses magnetic disks (platters)
- Has moving parts
- large storage capacity
- Slower than SSD
- Cheaper

use in finance
- storing large historical databases
- Archiving transaction files
- Backups

2. Solid State Drive (SSD)
- Uses flash memory
- No moving parts
- Very fast read
- More durable
- More expensive

Use in finance
- Running statistical software faster
- Speeding up large dataset processing
- Efficient for algorithmic trading systems

3. USB Flash Drives
- portable
- Uses flash memory
- Small capacity (GBs)

Use in finance
- Transfer of reports
- Temporary storage
- Sharing datasets in class/office

4. Memory cards

5 Optical discs

6 Cloud storage

Sources of data

1. Primary data
- Collected first-hand
e.g Investor surveys
- direct experimental financial models

2. Secondary data
- Already available.
e.g websites, world bank databases

Importance of data and data files in Financial Engineering

1 Foundation for risk modelling
2 Used in pricing derivatives
3 Supports forecasting
4 Enables portfolio optimization
5 Required for algorithmic and high-frequency trading
6 Helps measure volatility & market risk
7 Used for credit scoring and fraud detection

6 Disk storage fundamentals

Disk storage refers to the permanent, non-volatile storage used to save data, programs, files even when power is turned off
- It is secondary storage - persistent and large in capacity.

RAID storage.

Banks and financial institutions often use RAID for data reliability.

Common types:

- RAID 0 - fast (no redundancy)
- RAID 1 - mirroring (backup)
- RAID 5 - speed + redundancy.

Importance.

- ✓ Prevents data loss
- Enables continuous trading & operations
- ✓ Ensures risk models and market data are always available.

Importance of disk storage in BFE.

1. Storing big financial data.
2. Running statistical software.
3. Backup and recovery
4. Supports machine learning & modelling
5. Efficient Retrieval.

Difference btwn HDD and SSD.

Feature	HDD	SSD.
Speed	Slow	Very fast
Reliability	low	High
None	Noisy	Silent
Durability	lower	Higher
Cost	Cheaper	Expensive
Power use	more	less.

Examples of disk storage file in finance

- CSV files
- Excel files
- SQL databases
- Log files
- Parquet files

Key Disk Concepts.

1. Disk capacity.
- measured in GB / TB
2. Access time.
How quickly the disk can read/write data
SSDs - very low access time (fast)
HDDs - higher access time (slow)
3. File system.
e.g windows
4. Disk partitions.

Disk Organization & Structure

A disk is organized into:

a) Tracks - concentric circles
b) Sectors - sections on tracks
c) Clusters - group of sectors used to store a file.

Disk Management Activities

1. Formatting.
- Prepares a disk for storage by creating a file system
2. Partitioning
- Dividing a disk into independent sections.
3. Defragmentation (HDD)
- Rearranging files to speed up
4. Disk cleanup.
Removing temporary files
5. Backup
Saving copies of data in case failure.