TOPIC 1

Introduction to IT & Computers (IT for Statistics)

1. Course Overview

This course introduces the fundamentals of Information Technology (IT) and how computers support statistical work using computational tools such as R, Python Excel, SPSS to analyze data.

2. Role of IT in Society

IT plays a major role in modern society, and statistics uses IT heavily.

Key roles include:-

a) Data Collection

Computers help collect data through:

Gathering accurate and up to date data from people, institution and systems:-

Examples of IT tools used include: -

• Online forms, google forms

• Excel- In Agriculture, Field officer use tables use tables to collect crop and livestock data

• SQL Database- In Ministry of Defense to extract, transforming and loading data from databases

• Sensors- In Education, Online school sensors forms collect enrollment and teacher data.

• Mobile apps- In Healthcare, collect data on diseases, immunization and clinic visits

b) Data Processing and cleaning- Organizing raw data, removing errors and preparing it for analysis.

Statisticians rely on IT tools to:

• Excel

• SPSS

• Panda

• Python

• R

• SQL Database

c) Data Storage and management- Keeping large volumes of data in secure easy to access systems.

IT tools used:-

• SQL

• Cloud Storage (Google Cloud)

• AWS provides on demand delivery of IT resources and application via the internet.

d) Data Analysis- Use of mat lab to analyze huge data

• R- In Education, analyzing performance and dropout rates

• Python- In Finance, analyzing revenue, inflation and economic drop

• Stata- IN Health Sector, Analyzing disease rends and predicting outbreaks

• Excel

• Access

• SciPY

e) Data Visualization

Presenting information in charts, graphs, dashboards and maps for easy interpretation

IT tools used include:

• Excel charts

• Tableau

• Power BI

• Python (Matplotlib)

• R(ggplot2)

• Dashboards- Transport showing roads in accidents and traffic laws

KLBS publishing graphs for sensors, results and economic surveys

f) Decision making and support

Helping leaders use statistical evidence to make policies

IT Tools used include:

• Tableau/ Power BI Dashboards

• Statistical models in R, Python and SPSS

g) Forecasting tools

In Treasury decisions on budget allocation and taxation based on economic data

h) Automation of Statistical work

Reducing manual work and increasing speed and accuracy

IT Tools Used:

• Python Script

• R automation

• SQL- Stored procedure

• Excel Macros

In KRA, automated tax calculation and verification using data systems

In Education, Automated nemis reporting

i)

Communication and Reporting- Sharing results with the public, stakeholders and decision makers

Tools used:

• Power point

• PDF reporting tools

• Emails and Web Portals

• Tableau online dashboards

In KLBS, publishing annual economic reports online

In Education, publishing KCPE and KSCE results

3. Basic Computer Terminology (Important for Statistics)

Hardware- Physical parts of a computer:

• CPU – brain of the computer; processes data

• RAM – temporary memory for running programs

• Storage – HDD/SSD used to store files

• Input devices – keyboard, mouse, scanner

• Output devices – monitor, printer

Software- Programs that run the computer:

• System software – OS like Windows, macOS, Linux

• Application software – Excel, SPSS, R, Python

• Operating- Word, Excel, Operating, Access

Data- Raw facts and figures (e.g., numbers from a survey).

Information-Processed data that has meaning.

Database- Organized storage of data (e.g., SQL database).

File formats used in statistics

CSV – comma-separated values- In Business and Finance, it helps companies organize and share important information quickly

XLSX – Excel files

SAV – SPSS

RDS – R data file-In Education to forecast school dropout rates.

4. History of Computing

1st Generation (1940s–1950s)

Used vacuum tubes

Very large and expensive

Example: ENIAC

2nd Generation (1950s–1960s)

Used transistors

Faster, smaller, more reliable

3rd Generation (1960s–1970s)

Integrated circuits

More computers used in business and statistics

4th Generation (1970s–Present)

Microprocessors

Personal computers (PCs), laptops

5th Generation (Future/Present Trend)

AI and machine learning

Cloud computing

Big data analytics

These advancements made statistical computing faster, easier, and more accurate.

TOPIC 2

FUNDAMENTALS OF COMPUTER OPERATION

1. Definition of a Computer

A computer is an electronic device that receives, processes, stores, and outputs data.

2. Basic Operations of a Computer

• Input: Entering data via keyboard, mouse, scanner or other input devices.

• Processing: The CPU performs calculations and logical operations on data.

• Storage: Temporary (RAM) or permanent (HDD, SSD) storage for data and programs.

• Output: Information presented via monitor, printer, speakers, or other output devices.

• Control: The Control Unit coordinates and sequences internal operations.

3. Main Components of a Computer

Hardware: Physical/tangible components such as CPU, monitor, keyboard, mouse, RAM.

Software: Programs and instructions that run on hardware. Two main types:

• System software (Operating System) — manages hardware and provides a platform for applications.

• Application software — user-facing programs (e.g., word processors, spreadsheets).

4. Characteristics of Computers

• High speed

• High accuracy

• Large storage capacity

• Automation

• Diligence (doesn't get tired)

5. Types of Computers

• Supercomputers

• Mainframe computers

• Personal computers (desktops, laptops)

• Mobile devices (smartphones, tablets)

6. Booting Process

Cold Booting: Starting the computer from a completely off state.

Warm Booting: Restarting or rebooting the system without powering off completely.

Typical steps: POST → Load Operating System → System ready for use.

7. Role of the Operating System

• Manage hardware and software resources

• Provide a user interface (graphical or command-line)

• Allocate memory and manage processes

• Manage files and storage

• Control input/output operations and device drivers

8. Data & Instruction Flow

• User provides input data using input devices or software.

• Software issues instructions; the CPU executes them following the instruction cycle.

• Processed results are sent to output devices or stored for future use.

• Storage can be temporary (RAM) or persistent (disk).

9. User Interaction

• Use of input devices to interact with the system.

• Application software supports user tasks: typing, browsing, calculations, etc.

• Basic troubleshooting: restart, check connections, verify software settings.

Advanced Concepts (CPU & Performance)

10. CPU Architecture — Overview

The CPU executes program instructions. Key components and design choices that affect performance:

• Cores — multiple cores enable parallel execution of threads or processes.

• Instruction Set Architecture (ISA) — defines supported instructions.

• Cache hierarchy — small fast caches (L1/L2/L3) reduce memory access latency.

• Microarchitecture optimizations — pipelines, branch prediction, out-of-order execution.

11. ALU, Control Unit, Registers, System Clock

• ALU: Performs arithmetic and logical operations (add, subtract, bitwise ops).

• Control Unit: Fetches, decodes, and sequences instruction execution.

• Registers: Small, very fast storage locations used for operands and control data.

• System Clock: Provides the timing heartbeat for instruction sequencing.

12. Machine Cycles and Instruction Execution

• Instruction cycle stages: Fetch → Decode → Execute → Memory access → Write-back.

• Pipelining overlaps stages for increased throughput; hazards (data, control, structural) can cause stalls.

• CPI (cycles per instruction) and cache hit/miss behavior crucially affect performance.

13. Parallel Processing — Basics

• Levels of parallelism: instruction-level (SIMD), thread-level (multi-core), process-level/distributed.

• Key concepts: concurrency vs parallelism, synchronization, shared vs distributed memory.

• Amdahl's Law bounds speedup given the serial fraction of a task; embarrassingly parallel tasks scale well.

14. Performance Evaluation for Statistical Computation

• Measure throughput (tasks/sec), latency (time per operation), and resource utilization (CPU, memory, I/O).

• Use profiling tools (time, perf, Rprof, cProfile) to identify hotspots.

• Benchmark representative tasks (matrix ops, sorting, model fitting); repeat runs and report mean±variance.

15. Practical Guidance — Benchmarking Exercises

• Inspect CPU specs: cores, base/turbo frequency, cache sizes, supported instruction sets.

• Run benchmarking tasks: matrix multiplication, sorting large vectors, linear model fitting.

• Tools: Excel for basic timings; R (microbenchmark, system.time(), profvis); Python (timeit, cProfile, numpy).

• When reporting: include system specs, commands/code used, dataset descriptions, and multiple runs.

Submission Checklist

• Title and author (add your name and course details if required).

• Include system specifications if you performed benchmarking (CPU, RAM, OS).

• Ensure figures/tables (if any) are labeled and referenced.

• Proofread for clarity and formatting consistency

TOPIC 3

1. CPU (Central Processing Unit)

- The CPU is the brain of the computer.
- It processes instructions, performs calculations, and controls all other hardware components.
- Its speed and efficiency determine how fast a computer operates.

2. Memory (RAM)

- RAM stores data and instructions that the CPU is currently using.
- It is temporary storage—contents are lost when the computer is turned off.
- More RAM allows a computer to run more programs smoothly.

3. Input/Output Devices

- Input devices allow users to send data into the computer (e.g., keyboard, mouse, scanner).
- Output devices display or deliver results from the computer (e.g., monitor, printer, speakers).
- They act as communication bridges between the user and the computer system.

4. Storage Devices

- These store data permanently for future use (e.g., hard drives, SSDs, flash drives).
- Unlike RAM, storage keeps data even when power is off.
- Varies in speed, capacity, and durability depending on the type.

UNIT 4: BASICS OF COMPUTER SOFTWARE

INTRODUCTION

 Software is an organized collection of computer data and instructions, often broken into two

major categories;

A} SYSTEM SOFTWARE

 It provides basic non-task specific functions of the computer and applications software, which is

used to accomplish tasks.

It has 4 branches:

1. Operating systems {Supervisor program}

o The operating system manages the overall functioning and execution of the application

programs in the computer; assisted by computer resources such as the CPU and the I/O devices.

It acts as an interface between the user and the computer hardware.

Basic purpose of Operating Systems {OS}

- It is an interface between the user and the computer hardware increasing the efficiency in theprogram system e.g Codding and Debugging.
- It manages computer's resources i.e CPU. Memory etc
- It increases efficiency of system resources.
- It controls the allocation of resources among different tasks.
-

Functions of the OS

a) User Interface- It is used as a means of communication to the computer from the user to accomplish

tasks.

There are 2 types of the user interfaces:

- The Graphic User Interface - It is also referred to as WIMP (Windows, Icons, Menus,

Pointers). It simplifies the navigation and work on the computer through the mouse

It is the most commonly used.

- The Command Line Interface - It is used by professionals to perform specific functions.

b) Process Management- A process is a program under execution which requires resources such as CPU. - It has some functions:

 Allocation of resources to processes

 Creating and deleting the processes

 Sharing and exchanging information through processes

 Creating synchronization within the processes

c) File Management- It entails keeping track of the location and organization of files into specified directories for easy accessibility.

d) Memory management- This keeps track of the amount of memory and the time taken by processes.

e) Utility- It provides software used for: Finding files, Antiviruses, File conversion etc

TYPES OF OPERATING SYSTEMS

- Batch Operating Systems- This OS sorts the data through pre-defined sequences of commands,then collects the data in the batches and processes it for the execution.
- Real time Operating Systems- This OS processes data.

It has a very low response time and has a time constrain for the functions to be performed. It is applied in; Medical imaging systems, Scientific experiments etc

There are types, namely;

 Hard real time OS - Where the completion time

for the task is guaranteed.

 Soft real time OS – Where completion time for

the task isn't guaranteed, but is prioritized,

unless it was already competed.

Time Sharing Operating Systems – This OS allows several people to use the same computer

resources simultaneously.

iv.

v.

Distributed Operating Systems- This OS that uses multiple central processors (nodes), allowing

multiple users with multiple real time applications to use it at the same time. -

Its main objective is to share remote resources in a controlled

manner. -

One of its limitations is security issues and overloading issues.

Network Operating Systems- This OS allows activities of multiple computers to be coordinated

via a network. -

Examples: Lovable Intellect Not using XP(LINUX), Macintosh

OS (Mac OS), Uniplexed Information Computing Service(UNIX)

2. Language Processers

 This branch of System software is used to translate the programmer's instructions by the user

into a form which can be interpreted and executed by the computer system.

There are 3 different language processors

1) The Compiler

2) The Interpreter

3) The Assembler

TRANSLATION PROCESS

SOURCE CODE ------ TRANSLATOR------ OBJECT CODE------- EXECUTION

3. Device Drivers (Hardware drivers)

 This branch helps the hardware devices to communication with the Operating System by acting

as a translator for the input\output devices.

4. Utilities

 This branch is built to perform specific tasks such as: Finding files, diagnosing and repairing

system problems, data recovery and virtually storing files…

APPLICATION SOFTWARE

 It is a set of program which enhances the overall functioning of the computer system

 The end users are being provided with presentation programs

1.GENERAL PURPOSE SOFTWARE

 These are created for the basic requirements of general users

 Examples are: Word, Excel, Power point

WINDOWS OS

 It is the OS of Microsoft Inc

 It is the interface between hardware and software of a computer

Main features of windows

 Interactive Package- Ensures smooth communication

 Menu Drive Packages- allows smoother navigation of the computer

 Program Manager-it controls, co-ordinates and manipulates the processing of other programs.

 Multi- taking packages-helps in multi-tasking hence reducing time taken and increasing productivity

ANDROID OS

 It is used in mobile handsets, PC, TV, wearable devices

 It is an open source operating system

 Its interface is touch interface

 It is popular because of the large pool of applications (on play store)

FREE AND OPEN SOFTWARE

 It means that this software has a freedom of use.

 Example: Linux, Android,Python

GOOGLE PLAY STORE

 It has an automated activities system to scan for malware ;though it still faces cases of apps with malware thus stealing users content

ACQUIRED SOFTWARE

This contains same features which are requested by business organizations to software developers

COMPILING SYSTEMS

 It is a software tool that converts source code written by humans (in languages like C, C++, Java) into machine code that a computer can understand and execute.

IMPORTANCE OF COMPILING SYSTEMS

1. They make statistical programs faster -Compiled languages (C, C++, Java) run very fast

2. Statistical tools rely on compilers internally -Even though R and Python are interpreted, many of their fast libraries (Pandas)

are written in Compiled languages.

3. They ensure accuracy and error checking -This helps maintain correctness in statistical models

4. They improve performance of heavy calculations -It helps them to run efficiently

TYPES OF COMPILING SYSTEMS

1.Compiler -Translates the whole program at once into machine code .For Example ; C, C++,Java.

2. Interpreter -Translates and executes code line-by-line. Examples ; PYTHON, R interpreter

3.Just-in-time (JIT) Compilers

-Combines both compiling and interpreting. Examples: Java Virtual

TOPIC 5

Different Means of Data Storage (Volatile / Non-Volatile)

✓ There are two types of storages in a computer:

1. Primary Storage

2. Secondary Storage

1) Primary Storage (Main Memory / Internal Memory)

✓ This is the computer's immediate working memory.

✓ It temporarily holds data that the CPU is currently using right now.

Examples:

i) RAM (Random Access Memory)

iii) Cache Memory – High-speed memory used by CPU

iv) Registers – Small storage units inside the CPU

Characteristics

- Volatile – Data is lost when power goes off (e.g. RAM)
- Very fast – They are fast since they are directly accessed by the CPU
- Small in size – They have small capacity compared to secondary storage
- More expensive per unit of storage

2) Secondary Storage

✓ It is used to store data permanently.

✓ It keeps data even when the computer goes off / is turned off.

Examples:

i) Hard Disk Drive (HDD) – Used for long-term data storage

ii) Solid State Drive (SSD) – It is quite faster than HDDs and they are used in modern computers

iii) USB Flash Drives – They are portable storage drives

iv) Memory Cards – They are used in phones and cameras

v) CDs / DVDs – Optical Storage

Feature	Primary Storage	Secondary Storage
Purpose	Temporary working memory	Permanent Data storage
Speed	Very fast	Slow
Volatility	Volatile (Loses data)	Non-Volatile
Capacity	Small	Large
Cost	Expensive	Cheaper
Examples	RAM, Cache, Registers	USB, SSD, HDDs, CD/DVD

2. Solid State Drive (SSD) / Solid State Storage

Meaning: Uses flash memory chips instead of moving parts (No moving parts).

Features / Characteristics a) Very fast.

b) More durable (No moving parts).

c) More expensive.

d) Silent operation.

Examples

SATA SSD.

NVMe SSD.

3. Hybrid Drive (SSHD)

Meaning:

Combines HDD + SSD in one device.

Frequently used files are stored on SSD portion for speed.

Remaining data goes to HDD portion.

Features / Characteristics

a) Faster than HDD.

b) Cheaper than SSD.

c) Stores frequently used data in SSD section.

4. Optical Disks

Meaning: Uses laser technology to read/write data.

✓ They are mostly used for media & backups

Features / Characteristics

a) Portable.

b) Good for media distribution.

c) Lower capacity compared to HDD / SSD.

Examples

CD (Compact Disk).

DVD (Digital Versatile Disk).

Blu-ray Disc.

5**) Flash Storage Devices**

Meaning:

Uses flash memory and are non-volatile.

Used for file transfer and temporary storage.

Features / Characteristics

a) Portable

b) Durable

c) Medium capacity

Examples

USB flash drives

6) Network Attached Storage (NAS)

Meaning: This is a storage device connected to a network for shared access.

Features / Characteristics

a) Used in office and homes

b) Uses multiple HDDs or SSDs

c) Not a disk type but a disk storage method

Storage Type	**Technology**	**Example**
HDD	Magnetic	Hard Disk
SSD	Flash Memory	SATA, NYMe, SSD
SSHD	Hybrid	Hybrid Disk
Optimal Disk	Laser	WB, SD cards
Flash Storage	Flash Memory	Networking Storage box
NAS	Network storage	

TOPIC 6: Use The Different Number Systems and Bases and Apply the Concepts in the Four Basic Operations

Number Systems

1. Decimal systems

The decimal system is a number system that uses ten digits 0-9 and has a base of ten. It is derived from the suffix deci meaning 1/10

These are the numbers we use every day.

Example of decimal numbers: 43,85,552

Where the decimal system is use;

- Everyday counting
- Money and Finance: Prices in shops, bank accounts
- Measurements
  Used in measurements eg meters, kilometers
- Education and Mathematics

2. Binary

Binary is a language of computers, to store and process data represented using 0s and 1s and is based on base 2.

Example of binary numbers; 00100, 1011,101001

Where is binary numbers used;

- Programming; low level programming language, machine code
- Used in computers to store and process data
- Digital electronics – phones and calculators
- Logical circuits – like the one we learn in discrete mathematics

3. Octal

The octal system is a number system that uses digits 0-8 and as a base of 8. Derived from the suffix octa meaning 8.

Example of octal numbers; $58_{8,}$ 73_{8}

Why is octal used;

- Easier to read than long binary strings
- Simple to convert from binary
- Shorter than binary

Where it is used;

- Old computer systems
- Digital electronics

4. Hexadecimal

The hexadecimal number system is a base-16 that uses 16 symbols to represent numbers: the digits 0-9 and the letters A-F.

Examples of hexadecimal number systems; 0,1,2,3,4,5,6,7,8,9,10=A, 11=B, 12=C, 13=D, 14=E, 15=F

Uses of hexadecimal;

- Computing and programming
- Memory addresses – hex is shorter than binary for memory locations
- Color codes and web design

Conversions of number systems

1. Decimal to binary

Question: 348_{10}

348 ÷ 2 = 174 rem 0

174 ÷ 2 = 87 rem 0

87 ÷ 2 = 43 rem 1

43 ÷ 2= 21 rem 1

21 ÷ 2= 10 rem 1

10 ÷ 2= 5 rem 0

5 ÷ 2= 2 rem 1

2 ÷ 2= 1 rem 0

1 ÷ 1= 0 rem 1

So take the all the remainders in each series from bottom to top

Ans = 101011100

<u>Decimal to octal</u>

Question 348_{10}

$8 \times 0.5 = 4$

$348 \div 8 = 43.5 = 43$ rem 4

$43 \div 8 = 5.375 = 5$ rem 3

$5 \div 8 = 0.625 = 0$ rem 5

$= 534_8$

Decimal to hexadecimal

Question; 348_{10}

$16 \times 0.75 = 12$

$348 \div 16 = 21.75 = 21 \text{ rem } 12$

$21 \div 16 = 1.3125 = 1 \text{ rem } 5$

$1 \div 16 = \quad = 0 \text{ rem } 1$

$15 \ (12 \rightarrow C)$

$= 15C$

Octal to binary

Question; 58_8

5 6

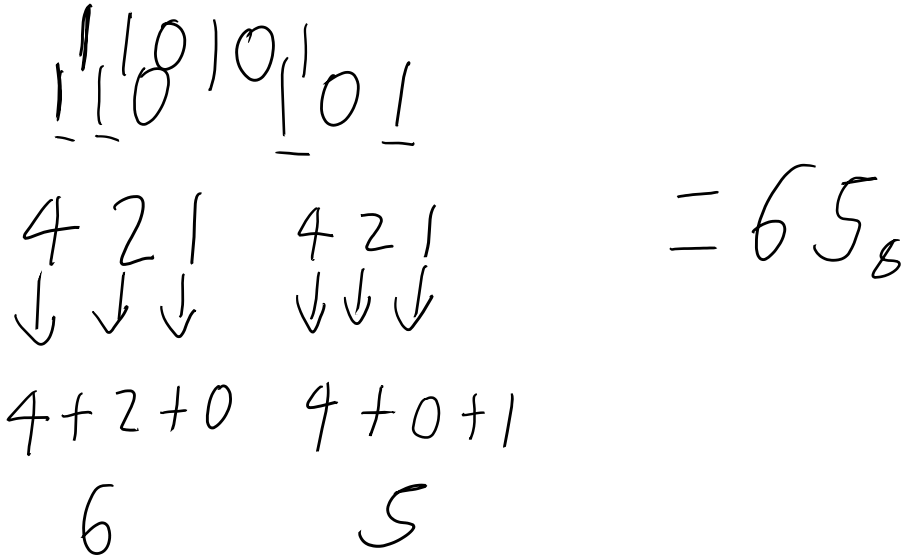
4 2 1 4 2 1

1 0 1 1 1 0

$56_8 \rightarrow 101110_2$

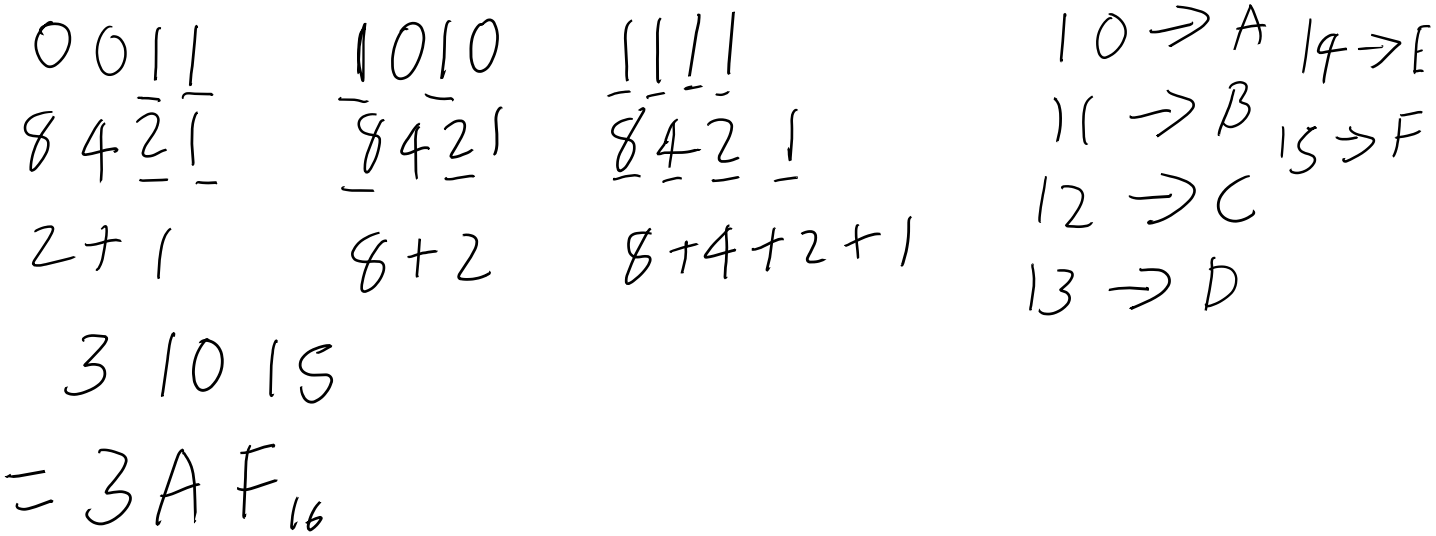
Binary to octal

Question; 110101



Octal to hexadecimal

Question; 1657_8



Application of concepts in the four basic operations

Addition of binary numbers

Question; 100110 + 110101

$$\begin{array}{r} {}^{1}1\,{}^{1}00\,{}^{1}10 \\ +\,1\,1\,0\,1\,0\,1 \\ \hline 1\,0\,1\,1\,0\,1\,1 \end{array} = 1011011$$

$$\begin{array}{r} {}^{1}\,1\,{}^{1} \\ +\,1\quad 1 \\ \hline 1\,0\quad +1 \\ \hline 1\,1 \end{array}$$

Subtraction of binary numbers

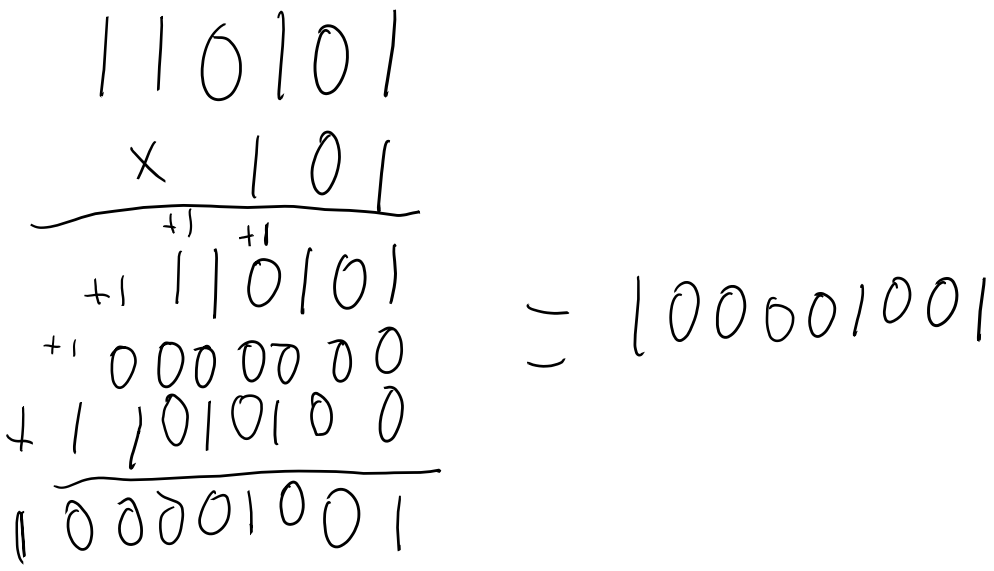
Question; 100 - 10

$$\begin{array}{r} 0\,{}^{2}\!\not{1}\,0\,0 \\ -\,1\,0 \\ \hline 0\,1\,0 \end{array} = 010$$

$$\begin{array}{r} 100-1 \\ 0\,{}^{\not 1}\!\not{0}\,{}^{1}0\,{}^{2} \\ -\quad 1 \\ \hline 0\,1\,1 \end{array} = 011$$

Multiplication of binary numbers

Question; 110101×101



Decimal of binary numbers