Short Notes

2079 - LDT, DMA and Assemble Directives

2078- Program Counter, Von-Neumann Architecture, Interrupt Masking

2076- Harvard Architecture, GDT

2075 - Macro Assembler

Kurtosis

Five Number Summary

Sampling, Differences between stratified sampling and cluster sampling

Measurement scales and its types

Probability Sampling and Non-Probablity Sampling

Roles of Computer science in statistics or statistics roles in computer science

Differences between primary and secondary data

Types of sampling

Dispersion, a measure of dispersion, its roles in statistics

What do you mean by probability density function? Write down its properties

Marginal Probability Distribution write down its properties

Normal Distribution with properties, Standard Normal Distribution

Poisson Distribution with properties

Binomial Distribution with properties,

DMA:

- DMA stands for Direct Memory Access.
- It is a method that allows devices to transfer data directly to and from memory without involving the CPU.
- DMA is controlled by hardware and can transfer data to or from a specific memory address, independent of the CPU.
- DMA is used to increase the data transfer rate and reduce CPU usage.
- DMA is commonly used in devices such as hard drives, network interface cards, and sound cards.

LDT:

- LDT stands for Local Descriptor Table.

- It is a **data structure** in the x86 architecture that is **used to store segment descriptors.**
- The LDT provides additional memory protection and segmentation in the x86 architecture.
- The LDT is a per-process data structure, meaning each process has its own LDT.
- The LDT can be accessed using the LSL and LAR instructions.

Assembler Directives:

- Assembler directives **are instructions used by the assembler to control the assembly process.**
- Assembler directives are typically used to define data, allocate memory, define labels, and control the flow of code.
- Assembler directives are specific to the assembler being used and are not part of the instruction set of the target architecture.

Program Counter:

- The Program Counter (PC) is a register in the CPU that stores the address of the next instruction to be executed.
- After each instruction is executed, the PC is automatically incremented by the CPU to point to the next instruction in memory.
- The PC is used to fetch and execute instructions in the correct sequence.
- The PC is an essential part of the instruction cycle, which is the process by which the CPU retrieves, decodes, and executes instructions.
- The PC is crucial for the execution of programs and is used to control the flow of instructions and operations in a computer.

Interrupt Masking

- Interrupt masking is a technique used in computer systems to prevent interrupts from being processed during critical operations.
- Interrupts are signals sent by devices to the CPU to request attention or service.
- Interrupt masking temporarily disables these signals to ensure that critical operations are not interrupted.
- Interrupt masking is typically achieved by setting a flag or a mask bit in the CPU's interrupt control register.
- Interrupt masking is commonly used in real-time systems and other time-sensitive applications where interruptions can cause serious problems.

Macro Assembler

- A macro assembler is a type of assembler program that allows the use of macros in assembly language.
- A macro is a set of instructions that can be defined once and then used multiple times in a program.
- Macros simplify coding, reduce program size, and improve readability.
- A macro assembler translates macro instructions into machine language instructions.
- Macro assemblers enable programmers to define their own instruction sets, making it easier to write complex programs.
- Macro assemblers are commonly used in low-level programming applications such as embedded systems programming.

GDT

- The GDT is a data structure used by the x86 architecture to define the memory segmentation scheme.
- It contains descriptors for various segments of memory such as code, data, and stack segments.
- Each descriptor defines the segment's size, location, and access permissions.
- The GDT is loaded into memory during system startup and is an essential part of the memory management system in x86-based operating systems.
- It allows the operating system to protect memory regions from unauthorized access, provide memory isolation between different applications, and enable the use of virtual memory and paging mechanisms.

Von Neumann Architecture

Von Neumann Architecture is a computer architecture design that was proposed by John von Neumann in the late 1940s. Some important points about it are:

- It uses a single bus for both instructions and data, allowing for a simple and flexible design.
- The CPU fetches instructions and data from the same memory location, which can lead to slower processing.
- It is used in most modern computers, including personal computers, smartphones, and servers.
- It can modify its own instructions, which is known as self-modifying code.
- It is a popular architecture for general-purpose computers due to its simplicity, ease of implementation, and flexibility.

- It is also known as the stored-program computer architecture, as it stores both the program and data in the same memory space.

Harvard Architecture

Harvard Architecture is a computer architecture design that uses separate memory spaces for instructions and data. Some important points about it are:

- It uses separate buses for instructions and data, which allows for faster processing and more efficient use of memory.
- It was originally developed for use in digital signal processing and embedded systems.
- It is less flexible than Von-Neumann Architecture, as it cannot modify its own instructions.
- It is still used in some modern systems, such as microcontrollers and digital signal processors.
- It provides better security, as it prevents malicious code from being executed in the data memory space.
- It is named after the Harvard Mark I computer, which was one of the first computers to use this architecture.

2079
Nominal and ordinal scale
Kurtosis
Five number summary
Nominal and ordinal scale:

Nominal and ordinal scales are two types of categorical data measurement scales. A nominal scale is used to categorize data into distinct categories without any specific order or ranking. For example, colors, gender, or zip codes are nominal data.

On the other hand, an ordinal scale is used to categorize data with a specific order or ranking. Examples of ordinal data include education levels (e.g., elementary, middle, high school), socioeconomic status (e.g., low, medium, high), or ratings (e.g., poor, fair, good).

Kurtosis:

Kurtosis is a statistical measure that describes the shape of a probability distribution curve. It measures the degree of peakedness or flatness of the distribution relative to a normal distribution. A normal distribution has a kurtosis of 0, meaning that the curve has neither a peak nor is it flat. Positive kurtosis (greater than 0) indicates a more peaked curve, whereas negative kurtosis (less than 0) indicates a flatter curve with more spread-out tails.

Five-number summary:

The five-number summary is a descriptive statistics technique used to summarize the distribution of a dataset. The five numbers include the minimum value, the first quartile (Q1), the median, the third quartile (Q3), and the maximum value. The first quartile (Q1) is the value below which 25% of the data falls, and the third quartile (Q3) is the value below which 75% of the data falls. The five-number summary is often used in box and whisker plots to visualize the distribution of the data.

2078

What do you mean by sampling? Explain the difference between stratified sampling and cluster sampling.

Sampling refers to the process of selecting a representative subset of individuals or units from a larger population in order to make inferences about the population as a whole. It is commonly used in research and statistical analysis when it is not feasible or practical to study the entire population.

Stratified Sampling	Cluster Sampling
Divides population into homogeneous subgroups	Divides population into clusters

Ensures proportional representation of each stratum	All units within selected clusters are included
Requires knowledge of population	Easier and more cost-effective to implement
More precision and accuracy	Useful for geographically dispersed populations
Useful when the population is heterogeneous	Useful when the population is homogenous within clusters

2076

What do your mean by measurement scale? Describe the different types of measurement scales used in statistics.

Measurement scale, also known as a level of measurement or scale of measurement, refers to the type of data generated by a measurement process. In statistics, there are four main types of measurement scales:

1. Nominal scale: This scale is used for qualitative data that can be placed into categories, such as gender, race, or religion. Nominal data can be counted and summarized using frequency tables, but mathematical operations cannot be performed on it.
2. Ordinal scale: This scale is used for qualitative or quantitative data that can be ranked in order, such as educational level or customer satisfaction ratings. Ordinal data can be summarized using median or mode, but mathematical operations are not possible.
3. Interval scale: This scale is used for quantitative data where the distance between values is meaningful, such as temperature or IQ scores. Interval data can be summarized using mean or standard deviation, but it does not have a true zero point.

4. Ratio scale: This scale is used for quantitative data where there is a true zero point, such as weight or height. Ratio data can be summarized using mean or standard deviation, and mathematical operations can be performed.

State with suitable examples the role played by computer technology in applied statistics and the role of statistics in information technology.short

The role of computer technology in applied statistics:

1. Data storage and management: Computers are used to store and manage large volumes of data, making it easier for statisticians to access and analyze the data.
2. Data analysis: Computers are used to perform complex statistical analyses on data, making it easier to identify patterns, relationships, and trends.
3. Visualization: Computers are used to create graphical representations of data, making it easier to understand and communicate statistical findings. Examples include scatter plots, bar charts, and histograms.

The role of statistics in information technology:

1. Data analysis: Statistics is used to analyze data in information technology, providing insights into customer behavior, website traffic, and system performance. Examples include A/B testing and user engagement analysis.
2. Machine learning: Statistics provides the foundation for machine learning algorithms, which are used to train models to make predictions and identify patterns in data. Examples include decision trees and neural networks.
3. Quality control: Statistics is used in information technology to monitor and control the quality of software and hardware products. Examples include statistical process control and acceptance sampling.
4. Risk management: Statistics is used to assess and manage risks in information technology, such as cyber-attacks, data breaches, and system failures. Examples include risk assessment and risk modeling.

Probability sampling	Non-probability sampling
The sample is selected at random.	Sample selection based on the subjective judgment of the researcher.
Everyone in the population has an equal chance of getting selected.	Not everyone has an equal chance to participate.
Used when sampling bias has to be reduced.	The researcher does not consider sampling bias.
Useful when the population is diverse.	Useful when the population has similar traits.
Used to create an accurate sample.	The sample does not accurately represent the population.
Finding the right respondents is not easy.	Finding respondents is easy.

6.8 Applications of Binomial Distribution

Binomial distribution is applicable when the trials are independent and each trial has just two mutually exclusive outcomes success and failures. It is applied in coin tossing experiments, sampling inspection plan, genetic experiments, quality control and so on.

1.9 Primary Data

Primary data are those fresh and original data, which are collected and recorded by the investigator or researcher. They are the first hand data. Primary data are collected for specific purpose of study of the investigator or researcher. The source of this type of data is called primary source. Following are the methods of collecting primary data.

- Direct personal interview method
- Indirect oral interview method
- Information through correspondence
- Mailed questionnaire method
- Schedule sent through enumerators

DIFFERENCE BETWEEN PROBABILITY SAMPLING AND NON-PROBABILITY SAMPLING

PROBABILITY SAMPLING	NON-PROBABILITY SAMPLING
1. It is a method of sampling which gives the probability that a sample is representative of population.	1. In the absence of any idea of probability the method of sampling is known as NON- PROBABILITY SAMPLING.
2. Probability sampling is generally used in fundamental research in which the purpose is to generalize the results.	2. It is generally used in action researches in which one studies a class without any generalization purpose.
3. It refers from the sample as well as the population.	3. There is no idea of population
4. Every individual of the population has equal probability to be taken into the sample.	4. There is no probability of selecting any individual.

Simple Random Sampling

Every member of the population has an equal chance of being selected.

Convenience Sample

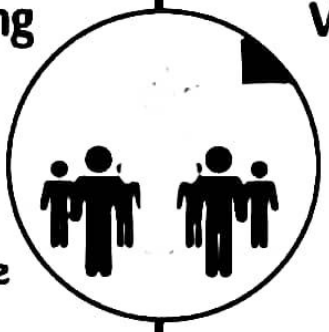
It includes the individuals who are most accessible to the researcher.

Systematic Sampling

Individuals of the population are chosen at regular intervals. It is easier to conduct than simple random method.

Voluntary Response

Here people volunteer themselves, instead of researchers choosing individuals.



Stratified Sampling

When the population shows mixed character then this method is us The population divides into subgroups.

Purposive Sampling

Researchers use judgements to select a sample that is most useful for research.

Cluster Sampling

Instead of sampling individuals from subgroups, the subgroups are randomly selected.

Snowball Sampling

In this sampling, the number of people who have access to "snowballs" as you come in contact with more people.

Data Types

Measurement Scales

There are four types of measurement scales used in statistics: nominal, ordinal, interval and ratio. Each scale has different properties and uses.

Nominal

The category labels are not ordered, so it doesn't matter which number comes first.

Ordinal

lin the ordinal scale of data, there is an order. However, the difference between them can not be quantified.

Interval

In the interval scale, we do have an order (just like ordinal data), and we can find the exact difference between the two values.

Ratio

The ratio scale has all the features of the Interval scale, and in addition, there is an absolute or true zero as well.

1.11 Difference between Primary and Secondary Data

The difference between primary and secondary data is basically depends on the mode of collection of data. The data which is primary for one agency is treated as secondary for the other and vice-versa. However, the major difference between primary and secondary data is as follows:

	Primary Data		Secondary Data
1.	Primary data are original in the sense that they are personally collected by the investigator or researcher involving himself/herself.	1.	Secondary data are not original in the sense that they are collected by some one other than the investigator or researcher.
2.	Primary data collection is more expensive and exhaustive.	2.	Secondary data are readily available at less expense.
3.	Primary data are collected as per requirement of the investigator.	3.	Secondary data might have been collected with different objectives.
4.	Primary data may be influenced by personal prejudice of the investigator etc.	4.	Secondary data may not be influenced by personal prejudice of the investigator etc.

1.10 Secondary Data

Any data that have been collected earlier for some purpose are secondary data for the individuals who are using them. The data is a primary for those persons or institutions that collect them but the same data become secondary for another. Actually secondary data are the data, which are borrowed from others who have collected them for some other purpose. The degree of accuracy of this type of data is comparatively less than that of the primary data. In order to collect secondary data, the following sources may be used.

Published Sources

- Reports and publications of ministries, departments of the government.
- Reports and publications of reputed INGO's such as UNDP, ADB, UNESCO, WHO, World Banks etc.
- Reports and publications of reliable NGO's, journals, periodicals, etc.

Note: *In recent years, the Internet has become one of the important sources of data. Almost all reputed companies, scientific institutions, business firms or institutions, government agencies, etc. have Internet web sites and provide public access.*

Unpublished Sources

- Records maintained by government offices.
- Records maintained by research institutions, research scholars etc.
- Records updated by the departments institutions for their internal purpose.

What is a marginal probability distribution? write down its properties

Marginal probability distribution refers to the probability distribution of a subset of random variables obtained by summing or integrating over the joint probability distribution of multiple random variables.

For example, suppose we have two random variables X and Y with a joint probability distribution P(X, Y). The marginal probability distribution of X can be obtained by summing P(X, Y) over all possible values of Y. Similarly, the marginal probability distribution of Y can be obtained by summing P(X, Y) over all possible values of X.

The properties of marginal probability distribution:

1. It is defined for every real number.
2. The sum of the probabilities in a marginal probability distribution must always equal one, as it represents the total probability of all possible outcomes for a given variable.
3. It can be used to calculate conditional probabilities.
4. The properties of the marginal probability distribution depend on the properties of the joint probability distribution of the underlying variables.

• Binomial Probability distribution

let, $X \sim B(n, p)$

n = no. of trails
p = probability of success
q = " " Failure

1. P.m.f of binomial Probability distribution is,

$$P(X=x) = \begin{cases} {}^{n}C_x p^x q^{n-x} \\ {}^{n}C_x p^x (1-p)^{n-x} \\ 0 \quad ; \text{otherwise} \end{cases} \quad ; \quad x = 0, 1, 2, \ldots n \quad , \quad p+q = 1$$

> It's properties

2. It is bi-parametric distribution as it has two parameters n & p.
3. The mean of binomial distribution is given by

$$\text{Mean} = np = E(X)$$

4. The variance of binomial distribution is given by

$$V(X) = \sigma^2 = npq$$

5. The variance of a binomial is always less than mean.

$$np > npq$$

6. It is discrete probability distribution.

Definition :- It is the discrete probability distribution that gives only two possible results in an experiment, either success or failure.

• Normal distribution

Definition :- A continuous R.V X is said to follow normal distribution if its probability density function (pdf) is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < +\infty, \; -\infty < \mu < +\infty, \; 0 < \sigma < \infty$$

> It's properties

1. The mean, median and mode are all equal.
2. The curve is symmetric at the center.
3. The total area under the curve is 1.
4. It is determined by 'μ' and 'σ'.

⊠ Standard Normal distribution :

The normal distribution $N(0,1)$ is called Standard normal distribution whose pdf is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}; \quad -\infty < z < +\infty$$

where $z = \frac{x-\mu}{\sigma}$

It's mean = 0, variance = 1

What are the roles of the measure of dispersion in descriptive statistics?

The measure of dispersion plays an important role in descriptive statistics by providing information about the spread or variability of a set of data. Here are some of the roles of measure of dispersion in descriptive statistics:

1. Provides additional information: A measure of central tendencies, such as the mean or median, only provides information about the location or center of the data. The measure of dispersion complements this information by providing information about the variability or spread of the data.
2. Helps in decision-making: The measure of dispersion can be used to make decisions about the appropriateness of using a particular statistical model or method. For example, if the data has a high level of variability, a different statistical method may be more appropriate than one that assumes a low level of variability.
3. Comparing data sets: Measures of dispersion can be used to compare the variability of two or more data sets. This can be useful in identifying differences or similarities between groups, populations, or variables.
4. Detects outliers: The measure of dispersion can help in identifying outliers, which are data points that are significantly different from the other data points. Outliers can be important in understanding the behavior of the data and in detecting errors or unusual phenomena.

What do you mean by probability density function? Write down its properties

The probability density function (PDF) is a mathematical function that **describes the probability distribution of a continuous random variable**. It is used to model the probabilities of different outcomes for a continuous random variable **over a range of possible values.**

The PDF can be used to calculate the probability of the random variable taking on a value within a certain range or interval, and can also provide information about the shape, spread, and central tendency of the probability distribution.

1. It is a**lways non-negative,** meaning that it cannot take negative values.
2. The total area under the **probability density function curve i**s equal to 1,
3. **The shape of the probability** density function can be symmetrical or skewed,
4. It can be any **real positive number.**

• Poisson distribution

Definition :- It is a discrete probability distribution that measures the probability of a given number of events happening in a specified time period.

1. > pmf of poisson distribution

let, $X \sim P(\lambda)$ & its pmf is

$$P(X = n) = \begin{cases} \frac{e^{-\lambda}\lambda^n}{n!} & n = 0, 1, 2, \ldots \\ 0 & ; \text{otherwise} \end{cases}$$

> Properties

2. The mean of poisson distribution is given by

$$\text{Mean} = E(X) = \lambda$$

3. The variance of poisson distribution is given by

$$\text{Variance} = V(X) = \lambda$$

4. For poisson distribution mean & variance are equal.

i.e. $E(X) = V(X)$

> Conditions, Binomial distribution tends to poisson distribution :

i) n, the no. of trails is indefinitely large i.e. $n \to \infty$

ii) P, the probability of success is very small i.e. $P \to 0$

iii) $np = \lambda$ (say) is finite.