NEERAJ®

DATA ANALYSIS

BECS-184

Chapter Wise Reference Book
Including Many Solved Sample Papers

Based on

C.B.C.S. (Choice Based Credit System) Syllabus of

I.G.N.O.U.

& Various Central, State & Other Open Universities

By: *Anand Prakash Srivastava*

NEERAJ PUBLICATIONS
(Publishers of Educational Books)

Mob.: 8510009872, 8510009878 E-mail: info@neerajbooks.com
Website: www.neerajbooks.com

MRP ₹ 280/-

Content

DATA ANALYSIS

Sample Preview of the Solved Sample Question Papers

Published by:

NEERAJ PUBLICATIONS
www.neerajbooks.com

QUESTION PAPER

June – 2024

(Solved)

DATA ANALYSIS

BECS-184

Time: 3 Hours] ***[Maximum Marks : 100***

Note: *Attempt questions from each section as per instructions given.*

SECTION-A

Note : Attempt any two questions from this section.

Q. 1. ***(a)*** **Distinguish between quantitative data and qualitative data. Which techniques are used to summarise the quantitative data? Explain with examples.**

Ans. Ref.: See Chapter-11, Page No. 136, 'Qualitative vs. Quantitative Research'.

(b) **What is harmonic mean? Find out the harmonic mean from the following data:**

Marks (X)	10	20	30	40	50
Frequency (F)	2	4	8	3	3

Ans. The harmonic mean (HM) is a type of average calculated as the reciprocal of the arithmetic mean of the reciprocals of a set of numbers. It is particularly useful when dealing with rates or ratios.

The formula for harmonic mean is:

HM = n / (∑(1/x))

where:

n = number of observations

x = individual observations

∑ = summation symbol

Now, let's calculate the harmonic mean from the given data:

Marks (X)	Frequency (F)	1/X	F(1/X)
10	2	0.1	0.2
20	4	0.05	0.2
30	8	0.0333	0.2664
40	3	0.025	0.075
50	3	0.02	0.06

∑F = 2 + 4 + 8 + 3 + 3 = 20

∑F(1/X) = 0.2 + 0.2 + 0.2664 + 0.075 + 0.06 = 0.8214

HM = n / (∑(1/x)) = 20/0.8214 ≈ 24.35

So, the harmonic mean is approximately 24.35.

Q. 2. ***(a)*** **What do you mean by the term 'measure of position'? Which measure of position is widely used and why?**

Ans. Ref.: See Chapter-2, Page No. 21, Q. No. 11 and Page No. 22, Q. No. 12.

(b) **A 20 point test is assigned to 10 students. The score attained by the students are given below**

18, 15, 12, 6, 8, 2, 3.5, 20, 10

Find the percentile rank of score of 15.

Ans. To find the percentile rank of a score, we need to first arrange the scores in order from lowest to highest:

2, 3, 5, 6, 8, 10, 12, 15, 18, 20

Next, we need to find the number of scores below 15:

There are 6 scores below 15 (2, 3, 5, 6, 8, 10)

Now, we can calculate the percentile rank:

Percentile Rank = (Number of scores below 15/Total number of scores) × 100

= (6/10) × 100

= 60

So, the percentile rank of the score 15 is 60. This means that 60% of the students scored below 15.

Q. 3. ***(a)*** **Correlation does not simply causation. Do you agree with this statement? Give reasons in support of your answer.**

Ans. I strongly agree with the statement "Correlation does not imply causation." This phrase is a fundamental concept in statistics and data analysis.

Correlation refers to the statistical relationship between two or more variables. When two variables are correlated, it means that as one variable changes, the other variable tends to change in a predictable way. However, correlation does not necessarily imply that one variable causes the other to change.

Here are some reasons why correlation does not imply causation:

1. Third-variable problem: A third variable might be causing both variables to change, creating the illusion of a causal relationship. For example, the correlation between ice cream sales and number of drownings might be due to the third variable "summer weather."

2. Reverse causality: The supposed cause might actually be the effect, and vice versa. For example, a correlation between hours spent watching TV and obesity might be due to obesity leading to more TV watching, rather than TV watching causing obesity.

3. Coincidence: Correlations can occur by chance, especially when dealing with large datasets. With enough data, it's possible to find correlations between unrelated variables.

4. Confounding variables: Other variables might be influencing the relationship between the two variables, making it seem like there's a causal relationship when there isn't.

5. Non-linear relationships: Correlations can be non-linear, meaning that the relationship between the variables changes as the values of the variables change.

***(b)* Explain the characteristics of '*t*' distribution and its usage for solving statistical problems.**

Ans. The *'t'* distribution, also known as the Student's t-distribution, is a continuous probability distribution that plays a crucial role in statistical inference, particularly in hypothesis testing and confidence intervals.

Characteristics of *'t'* distribution:

1. Symmetric and bell-shaped: The *'t'* distribution is symmetric around the mean (0) and has a bell-shaped curve, similar to the normal distribution.

2. Heavy-tailed: The *'t'* distribution has heavier tails than the normal distribution, meaning it is more prone to outliers.

3. Dependent on degrees of freedom: The shape of the *'t'* distribution depends on the degrees of freedom *(df)*, which is typically the sample size minus 1 ($n - 1$).

4. Converges to normal distribution: As the degrees of freedom increase, the *'t'* distribution converges to the standard normal distribution.

Usage of *'t'* distribution:

1. Hypothesis testing: The *'t'* distribution is used to test hypotheses about the population mean when the sample size is small (typically < 30).

2. Confidence intervals: The *'t'* distribution is used to construct confidence intervals for the population mean when the sample size is small.

3. Regression analysis: The *'t'* distribution is used to test the significance of regression coefficients.

4. Analysis of variance (ANOVA): The *'t'* distribution is used to test the significance of differences between group means.

Advantages of *'t'* distribution:

1. Robustness: The *'t'* distribution is robust to outliers and non-normality.

2. Flexibility: The *'t'* distribution can be used for both small and large sample sizes.

3. Easy to compute: The *'t'* distribution is widely available in statistical software and is easy to compute.

Q. 4. What is composite index? Explain with example the various steps involved in construction of composite index.

Ans. Ref.: See Chapter-10, Page No. 127, Q. No. 1, Q. No. 2 and Page No. 128, Q. No. 3.

SECTION-B

Note: Attempt any five questions from this section.

Q. 5. What is meant by data? Distinguish between primary data and secondary data? State the various steps to be taken in organising the field work.

Ans. Ref.: See Chapter-2, Page No. 9, 'Statistics/ Data' and Chapter-4, Page No. 49, 'Methods of Data Collection', and Page No. 54, Q. No. 6.

Q. 6. To collect data for a study, an individual chooses to question people coming out of a mall directly, identify this technique of collecting data. Discuss the significance of employing such a sampling technique.

Ans. The technique of collecting data by questioning people coming out of a mall directly is called Convenience Sampling.

Convenience sampling is a non-probability sampling technique where participants are selected based on ease of access, proximity, or convenience. In this case, the individual is questioning people coming out of a mall because it is an easily accessible location.

Significance of employing convenience sampling:

1. Easy to implement: Convenience sampling is a straightforward and simple method to collect data, especially when time and resources are limited.

2. Low cost: This method eliminates the need for elaborate sampling frames, randomization, and recruitment strategies, making it a cost-effective option.

3. Quick data collection: Convenience sampling allows researchers to collect data rapidly, which is beneficial when dealing with time-sensitive topics or limited research timelines.

4. Exploratory research: Convenience sampling is suitable for exploratory research, where the goal is to gather preliminary insights or identify patterns.

However, convenience sampling also has some limitations:

1. Lack of representativeness: The sample may not be representative of the larger population, which can lead to biased results.

2. Sampling bias: The researcher's selection of participants may be influenced by personal biases or preferences.

3. Limited generalizability: The findings may not be generalizable to other populations or contexts.

To mitigate these limitations, researchers can use techniques like:

1. Stratified convenience sampling: Divide the population into subgroups and collect data from each subgroup using convenience sampling.

2. Quota sampling: Set quotas for specific characteristics, such as age or gender, to ensure a more representative sample.

Sample Preview
of
The Chapter

Published by:

NEERAJ
PUBLICATIONS
www.neerajbooks.com

DATA ANALYSIS

Mathematical Concepts

1

INTRODUCTION

We've relied on numerical mathematics to tally our belongings and facilitate their distribution and accumulation. With the advent of the Industrial Revolution, our methods for producing, distributing, consuming, exchanging, and trading goods have grown markedly intricate. Consequently, we've found ourselves turning to more sophisticated mathematical tools and formulas to navigate these complexities.

Mathematics:

- **Numbers:** These include various types such as natural numbers (e.g., 3), integers (e.g., –12), rational numbers (e.g., 8.7, 2/3), irrational numbers (e.g., $\sqrt{112}$), and real numbers (e.g., 3, –12, 8.7, 2/3, $\sqrt{112}$), as well as complex numbers (e.g., $7 – 3i$).
- **Arithmetic Operations:** This involves basic operations like addition (+), subtraction (–), exponentiation/power (e.g., 5^3).
- **Arithmetic Relations:** Examples include greater than (>), less than (<), and equal to (=).
- **Variables, Constants, Equations, and Solutions:** These concepts are seen in equations like $x^2 – 16 = 0$, where x is a variable, 16 and 0 are constants, and the solutions are $x = 4$ and $x = –4$.
- **Sets:** Sets are collections of elements, such as {dog, cat, cow, buffalo, horse, sheep, goat} representing domestic animals, {4, –4} as solutions of an equation, and {0, 1, 2, 3, 4, 5, 6, 7, 8, 9} as decimal digits used in the decimal system. Sets are denoted using braces { }.

Symbol, Variable, Constant, and Parameter

- A symbol serves as a representation or substitute for something else within a system.
- A variable, denoted by letters such as *p, q, r, s, x, y, z,* is a symbol that can be measured and may vary.
- A constant is a symbol that represents only itself, like the number 5.67, which stands solely for that value.
- A parameter, like c, acts as a generalized constant, maintaining its value throughout a mathematical process or equation (e.g., $2x + c$ or $2x + c = 0$), often represented by letters such as *a, b, c* in mathematical contexts.

These concepts are crucial in math for formulating questions, discovering solutions using mathematical methods, and describing the solutions obtained.

Typically, the following five-step approach is utilized to solve economic problems using mathematics:

(i) Begin by defining the problem within the realm of economics, using economic concepts.

(ii) Next, translate the economic definition into mathematical terms, employing mathematical concepts.

(iii) Utilize mathematical tools such as techniques, algorithms, laws, principles, and constructions to calculate, compute, and represent the answer or logically derive conclusions. This results in a mathematical entity, which could be a number, an expression, a table, a graph, or a logical statement.

The importance of utilizing mathematics is highlighted by the following facts:

(a) A carefully and appropriately formulated mathematical model typically leads to a high-quality solution using the five-step procedure.

(b) Mathematics offers a diverse range of tools and techniques in step *(iii)* to tackle mathematical problems.

(c) Mathematics stands out as the most extensively tested and successful system for problem-solving, particularly in the realms of natural sciences and to some extent, social sciences.

CHAPTER AT A GLANCE

SET THEORY

A set is defined as a clearly specified grouping of unique objects.

Regarding this definition, some clarifications are necessary. Let's consider a set S according to the definition. The term "well-defined" when applied to a collection implies that for any given object, let's call it x, in the world, there should be a way, theoretically at least, to determine whether x belongs to S or not. Now, let's examine two collections named C_1 and C_2:

C_1 = collection of all students enrolled in IGNOU's undergraduate courses in the year 2017.

C_2 = collection of some students enrolled in IGNOU's undergraduate courses in the year 2017.

Concept and Representation of Set

The notation for representing sets is clarified through the following examples and explanations:

(i) Consider the set V representing the five vowels in the Latin alphabet: a, e, i, o, u. This set is denoted as:

$$V = \{a, e, i, o, u\}$$

Each element in this set, such as *a, e, i, o, u*, is referred to as a member or element of set V.

The notation for membership is expressed as follows:

If *e* is an element of set V, it is denoted as $e \in V$.

If *b* is not a member of set V, it is denoted as $b \in /V$.

Set Representation Notation

Roster or Tabular Form: In this format, all members of a set are listed within braces { }, separated by commas. For example, the set of all integers can be represented as follows: Z = {...,−3, −2, −1, 0, 1, 2, 3,...} Z = {..., −3, −2,−1, 0, 1, 2, 3,...}

Set-Builder Form: Let's take the set Sq_in_less_10001. Every element in this set possesses two properties: (a) it is the square of an integer, and (b) it is less than 10001.

Cardinality of a set S denoted by |S|: The notation $|x||x|$ is commonly used to express the absolute value of a number xx. However, when used with a set, denoted as XX, |X||X| signifies the cardinality of XX, i.e., the count of elements within XX. For instance, if X = {1, 2, 3}X = {1, 2, 3}, then |X| = 3 |X| = 3. Another example is the set of decimal digits, denoted as |Decimal_digits| = 10| Decimal_digits| = 10.

Infinite Set: The symbol for "infinite number" or infinity is ∞∞. By definition, |∞| = ∞|∞| = ∞. A set with an unlimited or uncountably infinite number of elements is termed an infinite set.

Set Relations

The formal definitions of "relation between sets" and "relation in a set" will be discussed later. For now, let's define some relations between two sets, denoted as X and Y, namely 'is a subset', 'equality of two sets', and 'is a proper subset'.

Subset of a Set: Sets represent collections of objects, and it's possible that every member of one set also belongs to another set. In such cases, the former set is termed a subset of the latter set.

This relationship is denoted as $x \subseteq Y$.

Proper Subset: If every element *x* in set *x* is also an element of set Y (i.e., if X belongs to X then X also belongs to Y), and there exists at least one element y in set Y such that y does not belong to set X, then *x* is considered a proper subset of Y. This relationship is denoted as $x \subset Y$, and it signifies that *x* is a proper subset of Y or that *x* is properly contained within Y.

Equality of two sets: Sets *x* and Y are considered equal if every element in *x* is also in Y and vice versa, meaning *x* is a subset of Y ($X \subseteq Y$) and Y is a subset of *x* ($Y \subseteq X$). This equality is denoted as $x = Y$.

Special Sets: Definitions and Properties

Universal Set: The Universal set, denoted as U, encompasses a comprehensive collection of entities within a problem domain. It comprises all possible entities needed at any point during the discussion or solution process of addressing problems within a specific type.

The Empty or Null set: It might initially sound peculiar that a set can have no members, or that the number of elements in a set can be zero.

A null set is typically denoted by the standard notation ∅ or by {}.

The Intervals as Sets

The notation for intervals, whether closed, open, or semi-open, has been previously discussed. An interval is defined as a subset of real numbers where, if *x* and *y* are two members of the interval I and $x < y$, then every real number between *x* and *y* is also a member of the interval I.

However, in interval notation, brackets '[ ]' and parentheses '( )' are used instead of braces '{ }' as follows:

For real numbers *a* and *b* where $a < b$

Number Sets (With Standard Notations)

Let's discuss some commonly used number sets and their names/notations briefly:

N = {1, 2, 3,…} is known as the set of natural or counting numbers.

W = {0, 1, 2, 3,…} is known as the set of whole numbers. Note that 0 is now considered a natural/counting number, so N = W, eliminating the need for a separate concept and notation for whole numbers.

Z = {…,−3, −2, −1, 0, 1, 2, 3,…} is known as the set of integers.

$Q = \{p/q | q \neq 0,\ p,\ q \in Z\}$ is known as the set of rational numbers.

R, the set of real numbers: The set of real numbers is denoted by R, where $R = \{x | x$ is a real number}.

It's evident that the length of the hypotenuse of a square with a side length of one unit is units.

C, the set of complex numbers: To further extend the real numbers in the context of solving equations like $x + 1 = 0$, or equivalently, $x = -1$, a new set of numbers was developed called the complex numbers.

Set Operations

We're familiar with how new numbers are derived from existing numbers through various number operations. For instance, given the number 2, we obtain its negative (or additive inverse) as (–2), its multiplicative inverse as (1/2), and its square root as $\sqrt{2}$. Each of these operations—additive inverse, multiplicative inverse, and square root—is a unary operation on numbers. They are termed unary operations because they operate on a single number (in this case, 2) and yield a new number (such as (1/2) in the case of the multiplicative inverse).

Only two categories of set operations will be covered: binary and unary set operations

The binary set operations for the two given sets, X and Y, are as follows: *(i)* Union of sets, represented by the symbol '*u*' as in *x u* Y; *(ii)* Intersection of sets, represented by the symbol *n* as in *x n* Y; *(iii)* Difference of (two) sets, represented by the symbol or as in *x* – Y (or *r* Y); and *(iv)* the symmetric difference operator A as in *x* A Y. Lastly, we will talk about the unary operation, which is the complement of a set, represented by the letter 'X'.

RELATION AND FUNCTIONS

Numerous (formal) notions in mathematics and other academic fields have their origins in elements, objects, or events that are frequently met, noticed in passing or even unintentionally. One such idea is "relation", which will be defined in formal terms in this section. Our acquaintances are "father", "daughter", and "friend." What role does any of these play in mathematics? The reason for this is that, using the procedure outlined below, each can be regarded as a (mathematical) set. In order to accomplish this, we first need to understand the concepts of "ordered set", "ordered pair", and "Cartesian product".

Ordered Pair/ Set: Coordinate geometry informs us that the points (1, 2) and (2, 1) are not the same; that is, (1, 2) (2, 1). Therefore, in the pairs that indicate the coordinates of points, the order in which the numbers occur is taken into account. Conversely, sets {1, 2} equal {2, 1}, since the representation of a universal set does not account for element order.

The set of all ordered pairs *(x, y)* is represented by the Cartesian product, or cross-product, of two non-empty sets, *x* and Y, or X×Y= {*(x, y)*: *x*∈X and *y*∈Y}.

If (at least) one of *x* and Y is Ø, the empty set, then, by definition X∈Y is Ø.

Image of an element, Domain, Range and Codomain of a relation: The element *b* is referred to as the image of element a for any relation R that maps a set *x* to a

set Y and for any ordered pair *(a, b)* in R.

The domain of a relation R is the set of all/first elements of the ordered pairs in a relation R from a set *x* to a set Y.

The range of a relation R is the set of all second elements in the relation R from a set *x* to a set Y. The relation R's co-domain refers to the entire set Y.

LOGIC

We start off our conversation about the subject by asking: Why do we learn logic? Furthermore, What Is Logic?

We begin by providing brief, basic responses to the questions, then we progressively extend them to fit our needs. We state at the outset that logic is the study of reasoning, more specifically, correct/proper reasoning. It explains and teaches us how to reason correctly so that:

(i) We always draw the correct conclusion from the initially assumed facts, every time; and

(ii) The process of drawing the conclusion from the initially assumed facts is sufficiently convincing, akin to mathematical calculations, to leave little room for disagreement between the involved (reasonable) parties.

Argument: Argumentation in logic is not the same as argumentation in everyday language. In logic, an argument is a set, or more accurately, a series of (n + l) statements, such as S_l, S_2. S_n, S_{n+l}, where *n* é N (or even O in some circumstances), so that S_{n+l} can be deduced from the other statements.

Typically, an argument of this kind is indicated (vertically) as

S_1
S_2
…
S_n
S_{n+1}

Or, (horizontally) as

$S_1, S_2, \ldots, S_n$
………..
S_{n+1}

to denote that the given or assumed statements S_l, S_2, and S_n (all above the horizontal line) are used to infer or conclude the statement S_{n+1} (the statement below the horizontal line). "Inferring" refers to the process of assuming the truth of each assertion above the horizontal line and using that assumption to infer the truth of each statement below it.

Symbolic Language and Logical Connectives

Why symbolic language?

(a) The symbols are concise and striking for lengthy sentences; for instance, we may use the sign 'CMRain', or the symbol CR, or even the symbol P: The village of Cherrapunji, also known as Chirapoonji, in northeastern India, experienced the world's highest average rainfall from 1901 to 2000.

(b) In a logical argument, we are typically more concerned with a statement's logical truth—whether it is "True" or "False"—than with its structure or substance. Compared to the similar English statements, the symbols emphasise the logical substance more clearly by abstracting away the other parts of the argument.

Symbolic Language: The following kinds of symbols make up a symbolic language for logic:

(i) Statements with symbolic names like P, Q, R, CMRain, and EF.

(Only atomic statements—defined as statements lacking both *(i)* any logical connective and *(ii)* any quantifier*—are given symbolic names.)

(ii) In logic, a symbolic name is a logical variable in that it can always be assigned to precisely one of the two possible values for a logical constant, T or F (explained under *(iii)* below).

***(iii)* Constants of Symbolism:** In symbolic logic, there are two truth values: True and False. These are known as constants. In addition, the two values can be written as "T" and "F", or even as "1" and "0". Generally, we will refer to the two logical constants as "T" and "F".

(iv) Well-formed formulas, quantifiers, symbolic expressions, and logical connectives, or simply wff, will all be covered in detail.

(v) Statement forms (Symbolic parameters), which are typically represented by letters in an unusual font or even by Greek letters O, W, etc. for identities such as De Morgan's Laws:~ () ≡ ~∅ Λ~ W, where wff may be used in place of each of and W.

Logical Connectives: Logical connectives are symbols used in logic and mathematics to connect propositions or statements and form compound statements. They are essential in constructing logical arguments and reasoning. Here are some common logical connectives:

Conjunction (AND): Denoted by ∧ (caret symbol) or sometimes by the word "and". It connects two statements and is true only when both statements are true. The truth table for conjunction is:

P	**Q**	**P Λ Q**
True	True	True
True	False	False
False	True	False
False	False	False

Disjunction (OR): Denoted by ∨ (vee symbol) or sometimes by the word "or." It connects two statements and is true if at least one of the statements is true. The truth table for disjunction is:

P	**Q**	**P ∨ Q**
True	True	True
True	False	True
False	True	True
False	False	False

Negation (NOT): Denoted by ¬ (tilde symbol) or sometimes by the word "not." It negates or reverses the truth value of a statement. The truth table for negation is:

P	**¬P**
True	False
False	True

Implication (→): Denoted by → (right arrow) or sometimes by phrases like "if...then" or "implies." It connects two statements and asserts that if the first statement (antecedent) is true, then the second statement (consequent) must also be true. The truth table for implication is:

P	**Q**	**P → Q**
True	True	True
True	False	False
False	True	True
False	False	True

Biconditional (↔): Denoted by ↔ (double arrow) or sometimes by phrases like "if and only if" or "iff." It connects two statements and asserts that they are both true or both false. The truth table for biconditional is:

P	**Q**	**P ↔ Q**
True	True	True
True	False	False
False	True	False
False	False	True

Logical connectives are fundamental in formal logic, mathematics, computer science, and various fields where precise reasoning and argumentation are required.

Well-formed Formula (in short, wff, or statement): A Well-Formed Formula (WFF), also known as a statement or a formula, is a syntactically correct expression in a formal language or logical system. In logic and mathematics, WFFs are constructed using symbols and rules that define the syntax of the language. These symbols typically include logical connectives (such as AND, OR, NOT) and variables representing propositions.

Interpretation and Evaluation of a wff: As was previously stated, a logical symbol can always take exactly one of the two potential values, "T" or "F". There are four possible values that can be assigned to the two symbols in a logical expression E if it contains two symbols. Every one of the four potential assignments is referred to as an interpretation of the symbols that appear in E, and it is then feasible to determine the logical value of E that corresponds to this interpretation. For instance, two symbols, A and B, are involved in the wff E provided by. Let's give A and B the truth values F and T, respectively.

$$E: \sim((A \vee (\sim B)) \rightarrow (A \leftrightarrow B))$$

Logically Equivalent Formulas (or Logical Identities): When the truth values of two formulas, G_l and G_2, are the same for every interpretation (i.e., truth assignment to every atom that occurs in either G_l or G_2), then the two formulas are said to be (logically) equivalent. Stated otherwise, two formulas, G_l and G_2, are considered (logically) similar if and only if G_l is True for every interpretation of G_2 and False for every interpretation of G_l.

PROOF TECHNIQUES

The well-known methods for mathematical proof are as follows:

(i) Direct proof and counter examples,

(ii) Proof by contradiction and Proof by contra positive,

(iii) Proof by cases,

(iv) Existence proof,

(v) Proof by Mathematical Induction, and

(vi) Proof of equivalences.

CHECK YOUR PROGRESS

Q. 1. Which of the following collections are sets? Justify your answer.