

PREVIOUS YEAR SOLVED QUESTION BANK

— for —
BITS WILP MTECH AIML

INTRODUCTION TO STATISTICAL METHODS

(UNTIL MID SEMESTER - EC2)

Rohith Raj Reganti | crackbitswilp.in

ISM Previous Year Solved Question Papers

Until Mid Semester - For BITS WILP M.Tech

crackbitswilp.in

1st Edition | 2026

ISM Previous Year Solved Question Papers

Until Mid Semester - For BITS WILP M.Tech

Copyright © 2026 crackbitswilp.in

All rights reserved. No part of this publication may be reproduced without prior written permission of the publisher.

Printed in India

Contents

1	Basic Probability & Statistics	1
2	Conditional Probability & Bayes theorem	17
3	Probability Distributions	53
4	Hypothesis Testing	117

Basic Probability & Statistics

..... crackbitswilp.in

Q 1.1 7 Marks, 2026 January

- A) Discuss at least three statistical concepts which helps to understand the data as a part of pre-processing.
- B) Consider two events such that $P(A) = 0.6$ and $P(B) = 0.7$, $P(A \text{ and } B) = 0.3$. Then find the following probabilities of the events mentioned
- i) Only A
 - ii) Only B
 - iii) A and B^c
 - iv) A^c and B

(Where superscript c indicates the complimentary event)

Explanation

A) Data Cleaning:

Data cleaning is a crucial step in the data preprocessing pipeline, involving the identification and correction of errors or inconsistencies in a dataset. This process ensures that the data is accurate, complete, and ready for analysis. Common tasks include handling missing values, removing duplicates, and addressing outliers that may distort the overall picture.

Data Integration:

Data integration is the process of combining data from different sources to provide a unified view. It involves merging datasets with varying structures, formats, and semantics into a cohesive and comprehensive dataset. Successful data integration enhances the overall quality of the data and enables more meaningful analysis and insights.

Data Transformation:

Data transformation involves converting and restructuring data to meet the requirements of the analytical process or to comply with a specific data model. This step may include normalization, encoding categorical variables, or scaling numerical features. Data transformation aims to enhance the compatibility and usability of the data for subsequent analysis.

Data Reduction:

Data reduction involves simplifying and summarizing large datasets while retaining essential information. Techniques like dimensionality reduction, aggregation, or sampling are employed to reduce the dataset's size without significantly sacrificing its representativeness. This step

is essential for improving computational efficiency and reducing the risk of overfitting in machine learning models.

Importing Crucial Libraries:

To streamline the data preprocessing tasks, various libraries in programming languages like Python are essential. Commonly used libraries include Pandas for data manipulation, NumPy for numerical operations, and Scikit-learn for machine learning functionalities. These libraries provide powerful tools to efficiently handle and analyze data.

Identifying and Handling Missing Values: Missing values are a common challenge in datasets and can adversely impact analysis and model performance. Identifying and handling missing values involves techniques such as imputation, deletion, or using advanced algorithms to estimate missing values. The goal is to maintain data integrity and avoid bias in subsequent analyses.

B) $P(A) = 0.6, P(B) = 0.7, P(A \text{ and } B) = 0.3$

- i)** only $A = P(A) - P(A \text{ and } B) = 0.6 - 0.3 = 0.3$ ans
- ii)** Only $B = P(B) - P(A \text{ and } B) = 0.7 - 0.3 = 0.4$ ans
- iii)** $A \text{ and } B^c = P(A) - P(A \text{ and } B) = 0.6 - 0.3 = 0.3$ ans
- iv)** $A^c \text{ and } B = P(B) - P(A \text{ and } B) = 0.7 - 0.3 = 0.4$ ans

Q 1.2 6 Marks, 2024-2025

- A) A sample of ten weights (in kg) is given: 30, 32, 35, 38, 40, 42, 45, 48, 50, 53.
- Calculate Q_1 , Q_2 , Q_3 , and IQR.
 - Determine potential outliers using the $1.5 \times \text{IQR}$ rule.
- B) Consider the following data from a manufacturing company. Let W be the set of workers in the company. Some subsets of W and their probabilities are provided below. Fill in the missing terms in the table, validate the consistency of the data, and provide justifications.

Set in Words	Symbol	Probability
All machine operators	M	
All quality inspectors	Q	
All machine operators and/or quality inspectors		0.40
All machine operators who are not quality inspectors		0.15
All quality inspectors who are not machine operators		0.10
All machine operators and quality inspectors		

Explanation**A) Given Data:**

Weights: 30, 32, 35, 38, 40, 42, 45, 48, 50, 53 (already sorted).

i) Calculate Q_1 , Q_2 , Q_3 , and IQR**1) Q_1 (First Quartile):**

The first quartile (Q_1) is the median of the lower half of the data (excluding the overall median if the dataset size is odd).

Lower half: 30, 32, 35, 38, 40 . Median of this subset:

$$Q_1 = 35$$

2) Q_2 (Second Quartile or Median):

The overall median (Q_2) is the middle value of the dataset. Dataset: 30, 32, 35, 38, 40, 42, 45, 48, 50, 53.

Since there are 10 values (even number of observations):

$$Q_2 = \frac{40 + 42}{2} = 41$$

3) Q_3 (Third Quartile):

The third quartile (Q_3) is the median of the upper half of the data (excluding the overall median if the dataset size is odd).

Upper half: 42, 45, 48, 50, 53.

Median of this subset:

$$Q_3 = 48$$

4) Interquartile Range (IQR):

The IQR is the difference between Q_3 and Q_1 :

$$IQR = Q_3 - Q_1 = 48 - 35 = 13$$

ii) Determine potential outliers using the $1.5 \times IQR$ rule

The $1.5 \times IQR$ rule states:

- Lower bound: $Q_1 - 1.5 \times IQR$
- Upper bound: $Q_3 + 1.5 \times IQR$

i) Calculate bounds:

$$\text{Lower bound} = Q_1 - 1.5 \times IQR = 35 - 1.5 \times 13 = 35 - 19.5 = 15.5$$

$$\text{Upper bound} = Q_3 + 1.5 \times IQR = 48 + 1.5 \times 13 = 48 + 19.5 = 67.5$$

ii) Identify potential outliers:

Outliers are any data points outside the range $[15.5, 67.5]$.

Given data: 30, 32, 35, 38, 40, 42, 45, 48, 50, 53.

- No values are below 15.5.
- No values are above 67.5.

Conclusion: There are no potential outliers in this dataset.

Final Answers:**1) Quartiles and IQR:**

- $Q_1 = 35, Q_2 = 41, Q_3 = 48, IQR = 13$.

2) Potential Outliers: There are no potential outliers in the dataset.**B) Step 1: Relationship Between the Sets**

The union of the two sets $M \cup Q$ can be expressed as:

$$P(M \cup Q) = P(M - Q) + P(Q - M) + P(M \cap Q)$$

Substitute the given values:

$$0.40 = 0.15 + 0.10 + P(M \cap Q)$$

Solve for $P(M \cap Q)$

$$P(M \cap Q) = 0.40 - 0.25 = 0.15$$

So, $P(M \cap Q) = 0.15$

Step 2: Calculate $P(M)$

The set of all machine operators M includes those who are only machine operators $M - Q$ and those who are both machine operators and quality inspectors $M \cap Q$

$$P(M) = P(M - Q) + P(M \cap Q)$$

Substitute the values:

$$P(M) = 0.15 + 0.15 = 0.30$$

So, $P(M) = 0.30$

Step 3: Calculate $P(Q)$

Similarly, the set of all quality inspectors Q includes those who are only quality inspectors ($Q-M$) and those who are both machine operators and quality inspectors ($M \cap Q$)

$$P(Q) = P(Q - M) + P(M \cap Q)$$

Substitute the values:

$$P(Q) = 0.10 + 0.15 = 0.25$$

So, $P(Q) = 0.25$

Step 4: Validate Consistency

The probabilities must satisfy the following constraint:

$$P(M \cup Q) \leq P(M) + P(Q)$$

Substitute the values:

$$0.40 \leq 0.30 + 0.25$$

This is valid. Thus, there are no inconsistencies in the data.

Final Table with Missing Terms

Set in Words	Symbol	Probability
All machine operators	M	0.30
All quality inspectors	Q	0.25
All machine operators and/or quality inspectors	$M \cup Q$	0.40
All machine operators who are not quality inspectors	$M - Q$	0.15
All quality inspectors who are not machine operators	$Q - M$	0.10
All machine operators and quality inspectors	$M \cap Q$	0.15

Justification The calculations confirm that all probabilities align with set theory rules and satisfy constraints, ensuring no inconsistencies.