RN Shetty Trust®

Estd : 2001

# RNS INSTITUTE OF TECHNOLOGY

**(An Autonomous Institution Affiliated to VTU, recognised by GOK, Approved by AICTE, NAAC 'A+ Grade'**

Dr. Vishnuvardhana Road, R R Nagar PO, Channasandra, Bengaluru - 560 098

## Master of Business Administration

# STATISTICS FOR MANAGERS

## Semester: I

## Subject code: MBA15

**December 2024 – April 2025**

| STATISTICS FOR MANAGERS | | | |
|---|---|---|---|
| **Course Code** | **MBA15** | CIE Marks | **50** |
| **Teaching Hours/Week (L:P:SDA)** | **3:0:2** | SEE Marks | **50** |
| **Total Hours of Pedagogy** | **50** | Total Marks | **100** |
| **Credits** | **04** | Exam Hours | **03** |

Course Learning objectives:

1.To impart conceptual knowledge of descriptive statistical theories and applications.

2. To facilitate usage of statistical techniques of relationship and prediction analysis for insightful decisions.

3. To orient learners towards the application of hypothesis and probability techniques for decision making.

4. To facilitate an understanding of the concept and applications of tests of hypothesis.

**Module-1 (12 Hours)**

**Statistics: Meaning, functions & applications; Descriptive vs. Inferential Statistics; Limitations of Statistics; Descriptive Statistics: Measures of Central Tendency – Mean, Median & Mode; Measures of Dispersion – Range, Quartile Deviation, Mean Deviation, Standard Deviation & Coefficient of Variation; Skewness (Theory & Problems)**

**Module-2: (8 Hours)**

**Correlation – Relevance for Business; Types of Correlation; Scatter diagrams; Karl Pearson's coefficient of correlation & Spearman's Rank Correlation (Theory & Problems); Regression – Meaning, Simple vs. Multiple regression; Lines of Regression; Problems on simple regression only. (Theory & Problems)**

**Module-3: (6 Hours)**

**Time Series Analysis: Meaning, objectives and components of time series; Measurement of Time Series – Secular trend (Least squares method, moving averages method and graphical method); Seasonal indices (Simple average method, ratio to moving averages method and ratio to trend method) (Theory & Problems)**.

**Module-4: (12 Hours)**

**Probability – Basic concepts & rules of probability; Bayes' theorem; Theoretical Distributions – Binomial Distribution, Poisson Distribution & Normal Distribution; (Theory & Problems**).

**Module-5 : (14 Hours)**

**Hypothesis Testing – Basic concepts; Process; Type I and Type II Errors; Parametric Tests and Non-Parametric tests – An overview; Z Test, T test, F test, ANOVA (One way and Two Way), Mann-Whitney U test and Kruskal – Wallis test – (Theory and Problems)**

**Assessment Details (both CIE and SEE)**

The weightage of Continuous Internal Evaluation (CIE) is 50% and for Semester End Exam (SEE) is 50%. The minimum passing marks for the CIE is 50% of the maximum marks. Minimum passing marks in SEE is 40% of the maximum marks of SEE. A student shall be deemed to have satisfied the academic requirements (passed) and earned the credits allotted to each course if the student secures not less than 50% in the sum total of the CIE (Continuous Internal Evaluation) and SEE (Semester End Examination) taken together.

**Continuous Internal Evaluation:**

There shall be a maximum of 50 CIE Marks. A candidate shall obtain not less than 50% of the maximum marks prescribed for the CIE.

**CIE Marks shall be based on:**

a) Tests (for 25Marks) and

b) Assignments, presentations, Quiz, Simulation, Experimentation, Mini project, oral examination, field work and class participation etc., (for 25 Marks) conducted in the respective course. Course instructors are given autonomy in choosing a few of the above based on the subject relevance and should maintain necessary supporting documents for same.

**Semester End Examination:**

The SEE question paper will be set for 100 marks and the marks scored will be proportionately reduced to 50.

- The question paper will have 8 full questions carrying equal marks.
- Each full question is for 20 marks with 3 sub questions.
- Each full question will have sub question covering all the topics.
- The students will have to answer five full questions; selecting four full question from question number one to seven in the pattern of 3, 7 & 10 Marks and question number eight iscompulsory.
- 40 percent theory and 60 percent problems.

**Suggested Learning Resources:**

**Books**


1. S C Gupta (2018), Fundamentals of Statistics, 7th edition Himalaya Publications.
2. J K Sharma (2020), Business Statistics 5th edition Vikas Publishing House.
3. S P Gupta (2021), Statistical Methods 46th edition Sultan Chand Publications.
4. C R Kothari (2015), Research Methodology- Methods and Techniques, Viswa Prakasam Publications.
5. William E. Wagner, III (2015), Using IBM SPSS- Statistics for Research Methods and Social Science Statistics 5th edition Sage Publications.


**Web links and Video Lectures (e-Resources):**

- Students should opt Swayam NPTEL Course on Business Statistics offered by Prof.M.K.Barua Dept. of Management studies IIT Roorkee.
- YouTube Videos are also available of the same https://www.youtube.com/watch?v=VDLyk6z8uCg
- Swayam NPTEL Course on Business Statistics by Dr. P. M. Shiv Prasad, Department ofCommerce, Teresian College, Mysuru

**Skill Development Activities Suggested**

- Role Play Techniques
- Quizzes
- Field Surveys
- Assignments

**Course outcome (Course Skill Set)**

At the end of the course the student will be able to:

| Sl. No. | Description | Blooms Level |
|---|---|---|
| CO1 | Understand the conceptual and practical dimensions of basics of statistics and data summarization | L2 |
| CO2 | Infer the relationships between data sets and discover hidden patterns in data leading to prediction. | L3 |
| CO3 | Apply tools of time series analysis as a measure to enhance statistical understanding | L4 |
| CO4 | Analyze statistical theories of uncertainties via probability measures. | L5 |
| CO5 | Evaluate best alternatives for decision problems via tests of hypotheses. | L6 |

**STATISTICS FOR MANAGERS (22MBA14)**

MODULE 1

**Introduction to Statistics; Measures of Central Tendency; Measures of Dispersion**

**Contents:**

- Statistics – Meaning, Stages, Functions, Scope/Importance & Limitations
- Descriptive vs. Inferential Statistics.
- Measures of central tendency.
- Measures of Dispersion.

**STATISTICS:** Statistics is a science which deals with the methods of collecting, classifying, presenting, comparing and interpreting the numerical data collected to throw some light on any sphere of enquiry.

## STATISTICAL INVESTIGATION – 5 STAGES

1. Collection – sources of data (primary/secondary)
2. Organization – editing, classification, tabulation.
3. Presentation – for analysis- through diagrams & graphs.
4. Analysis – done to dig out information useful for decision making.
5. Interpretation – drawing conclusions- in a methodical way.

## Functions of Statistics:

1. It helps in collecting and presenting the data in a systematic manner.
2. It helps to understand unwisely and complex data by simplifying it.
3. It helps to classify the data.
4. It provides basis and techniques for making comparison.
5. It helps to study the relationship between different phenomena.
6. It helps to indicate the trend of behaviour.
7. It helps to draw rational conclusions.

## Scope / Importance / Applications of Statistics

**1 Statistics in Planning:** The success of the planning depends upon the correct and sound analysis of statistical data. For example: the water supply corporation will be unable to solve the problem relating to the supply of water in Karnataka valley unless the population of Karnataka, the quantity of water required per day, the quantity of water to be supplied and by which source additional quantity of water can be supplied if necessary, is known. These are the necessary information to be collected by Water Supply Corporation. This problem can be solved through the powerful statistical tools by making use of statistical data.

**2 Statistics in Economics**: Economic problems almost always involve facts that can be expressed numerically such as production, consumption, distribution of incomes, wages, unemployment etc. The study of economic problems requires use of statistical methods.

**3 Statistics in Business:** For smooth functioning, the need of statistical information depends upon the size of the business. When the size of the business increases, a single person cannot contact the customers personally and look after all the business activities. He cannot get the information relating to business in the same manner as in case of small size business. Thus statistics is made use of when the size and scale of any business expands.

**4 Statistics in State:** For effective functioning of State, Statistics is indispensable. Today, the State authorities collect statistics through their agencies on multiple aspects like population, agriculture, defense, national income, oceanography, natural resources, space research, etc.

**5 Statistics in Social Sciences and Natural Science:** In social sciences, statistics are used in the field of demography for studying mortality, fertility, marriage, population, and growth. Also, in psychology and education, the intelligence quotient (IQ) is determined using statistics.

**6 Statistics in Biology and Medicine:** In biology and medical sciences, there is regular use of statistical tools for collecting, presenting, and also analyzing the observed data pertaining to the causes of the incidence of diseases. For example, the statistical pulse rate, body temperature, blood pressure, etc. of the patients helps the physician in diagnosing the disease properly.

**7 Statistics in Research:** Statistics help in the conduction of research in new areas and the opening of newer vistas of knowledge to mankind.

*Decision Situation and Corresponding Statistically Techniques*

| *Area* | *Decision Situation* | *Statistical Techniques Applicable* |
|---|---|---|
| Marketing | • Assessment / Forecast of Demand for the Product or a Service<br>• Customer Profiling Market Research | Time series<br>Correlation and Regression<br>Cluster Analysis<br>Conjoint Analysis<br>Multidimensional Scaling |
| Retail Management | Identifying Customer Buying Behaviours and patterns | Correlation and Regression<br>Cluster Analysis<br>Conjoint Analysis |
| Finance and Banking | Evaluation of Investment<br>Volatility of Stocks<br>Predicting EPS<br>Derivatives | Regression Analysis,<br>Decision Analysis<br>Beta analysis |
| Insurance | Determining the Premium | Profitability, Time series and Regression Analysis |
| Operations | Controlling and Improving Production Process and Quality<br>Inventory Management | Statistical Quality Control<br>Six Sigma<br>Sampling inspection<br>ABC Analysis |
| HRD | Performance Appraisal and Reward System | Normal Distribution<br>Percentiles |

Source: Statistics for management by T N Srivastava and Shailaja Rego, Published by the Tata McGraw- Hill Publishing Company Limited.

## Limitations of Statistics

1. **Statistics does not deal with individuals:** A single item or the isolated figure cannot be regarded as statistics. This is a serious limitation of statistics.

For example: the mark obtained by a student in English is 75 do not constitute statistics but the average of a group of students in English is 75 forms statistics.

2. **Statistics does not study qualitative phenomena:** Statistics cannot directly be used for the study of qualitative phenomena such as honesty, intelligence, beauty, poverty etc. however, some statistical techniques can be used to study such qualitative phenomena indirectly by expressing them into numbers.

3. **Statistical laws are not exact:** 100% accuracy is rare in statistical work because statistical laws are true only on the average. For example: the probability of getting a head in a single toss of a coin is ½. This does not imply that 3 heads will be obtained if a coin is tossed 6 times.

4. **Statistics is only a means:** Statistical methods provide only a method of studying problem. There are other methods also. These methods should be used to supplement the conclusions derived with the help of statistics.

5. **Statistics is liable to be misused:** Statistical methods are the most dangerous tools in the hands of inexpert. Since statistics deals with masses of figures, so it can easilybe manipulated by inexperienced and skilled persons.

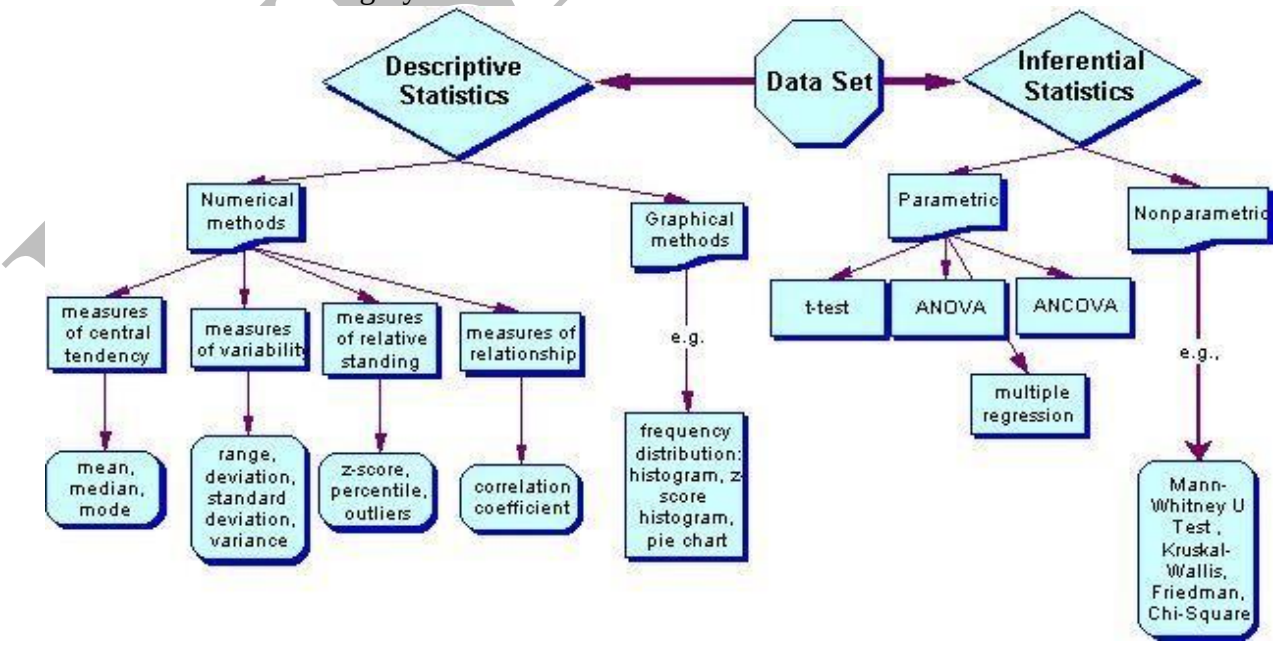
## TWO MAJOR TYPES – DESCRIPTIVE & INFERENTIAL STATISTICS

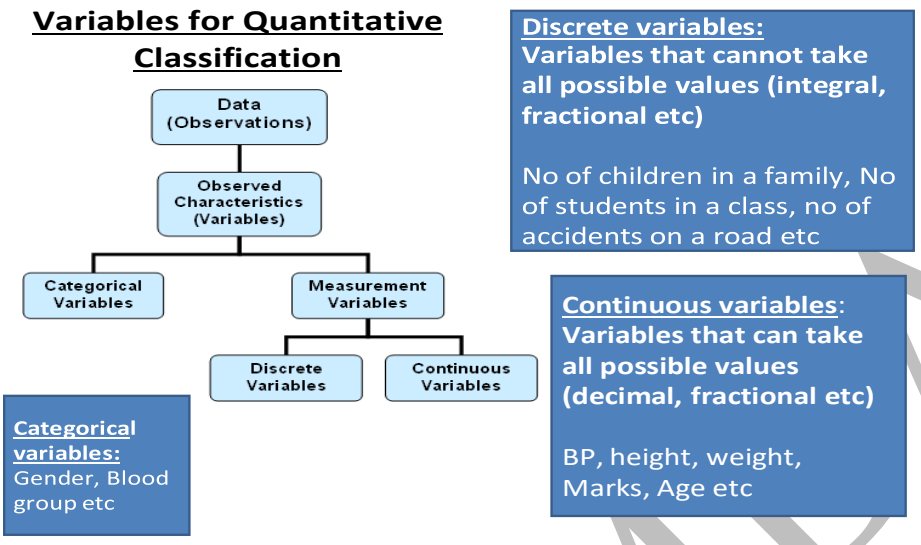
**1 Descriptive statistics** is the term given to the analysis of data that helps ***describe, show or summarize data*** in a meaningful way such that patterns might emerge from the data.

- Descriptive statistics ***do not, however, allow us to make conclusions beyond the data we have analyzed*** or reach conclusions regarding any hypotheses we might have made.
- If we simply present our raw data it would be hard to visualize what the data was showing, especially if there was a lot of it.
- ***For example***, if we had the results of 100 pieces of students' coursework, we may be interested in the overall performance of those students. We would also be interested in the distribution or spread of the marks. Descriptive statistics allow us to do this.

**2 Inferential Statistics:**

- Suppose you need to collect data on a very large population. ***For example***, suppose you want to know the average height of all the men in a city with a population of so many million residents. It isn't very practical to try and get the height of each of man.
- This is where inferential statistics comes into play. **Inferential statistics** makes inferences about populations using data drawn from population. Instead of using entire population to gather the data, the statistician will collect a sample or samples from the millions of residents and make inferences about the entire population using the sample.
- The **sample** is a set of data taken from the population to represent the population. Probability distributions, hypothesis testing, correlation testing and regression analysis all fall under the category of inferential statistics.






## Quantitative Data classification – 3 basic formats

1. **Raw Data** – Array – The set of individual observations (unorganized data): Marks of 10 students in a class: 4 5 6 7 8 9 1 12 34 56

2. **Discrete frequency distribution** – (without Class – Interval):

| No of accidents: (X) | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| No of days: (f) | 46 | 76 | 38 | 25 | 10 | 5 |

3. **Continuous frequency distribution** – (with C – I):

a. Inclusive type: 30 – 39, 40 – 49, 50 – 59 etc. [both upper & lower limits are included in the class]
b. Exclusive type: 10 – 20, 20 – 30, 30 – 40 etc. [Upper limit of the class is excluded from that class]
c. Open ended: Less than 10, 10 – 20, 20 – 30, 30 & above etc.
[*NOTE: Inclusive must always be converted to exclusive type for calculations.]

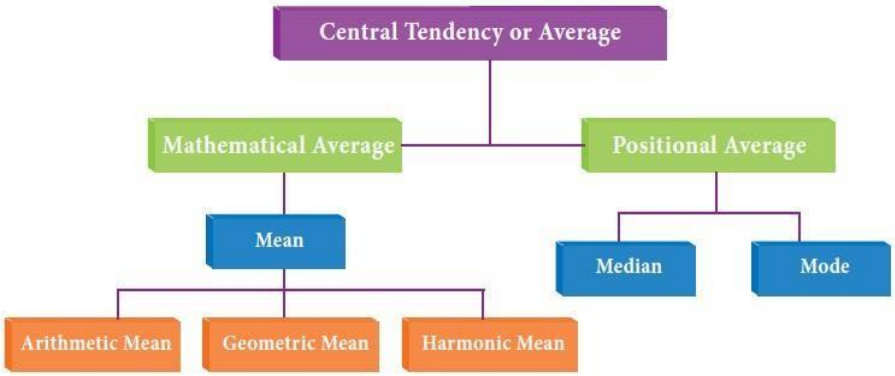
## MEASURES OF CENTRAL TENDENCY

**Introduction:** While distributions provide an overall picture of some data set, it is sometimes desirable to represent some property of the entire data set using a single statistic. The first descriptive statistics we will discuss are those used to indicate where the 'center' of the distribution lies.

**Meaning:** A measure of central tendency is a ***single value that attempts to describe a set of data*** by identifying the central position within that set of data.

### Different Measures of central tendency are as follows:

1. Arithmetic Mean (Mean)
2. Geometric Mean.
3. Harmonic Mean.
4. Median.
5. Mode.




**1 Arithmetic Mean:** The mean (or average) is the most popular and well known measure of central tendency. It can be used with both discrete and continuous data, although its use is most often with continuous data. The mean is equal to the sum of all the values in the data set divided by the number of values in the data set.

**2 Geometric Mean:** **[**Mathematical definition**:** The $n^{th}$ root of the product of n numbers.] The Geometric Mean is a special type of average where we multiply the numbers together and then take a square root (for two numbers), cube root (for three numbers) etc. The geometric mean is a type of average, usually used for growth rates, like population growth orinterest rates.

**3 Harmonic Mean:** The Harmonic Mean (HM) is defined as the reciprocal of the arithmetic mean of the given data values. It is based on all the observations, and it is rigidly defined. Harmonic mean is a type of average that is calculated by dividing the number of values in a data series by the sum of the reciprocals (1/x) of each value in the data series.

**4 Median:** The median is the middle score for a set of data that has been arranged in order of magnitude. The median is less affected by outliers and skewed data.

**5 Mode**: Mode is the most common number – the one that appears the maximum number of times in a given data set.

**Merits & Demerits**

| Merits | Demerits |
|---|---|
| **Mean:**<br>• Rigidly defined<br>• Based on all observations<br>• Easy to calculate & understand<br>• Least affected by sampling fluctuation, hence more stable | **Mean:**<br>• Can be used only for quantitative data<br>• Unduly affected by extreme observations |
| **Median:**<br>• Not affected by extreme observations<br>• Both for quantitative & qualitative data | **Median:**<br>• Affected more by sampling fluctuations<br>• Not rigidly defined<br>• Can be used for further mathematical calculation |
| **Mode:**<br>• Not affected by extreme observations<br>• Both for quantitative & qualitative data | **Mode:**<br>• Not rigidly defined<br>• Can be used for further mathematical calculation |

## Characteristics/prerequisites of an ideal measure of Central Tendency / Dispersion*.

1. It should be rigidly defined.
2. It should be easy to calculate & easy to understand.
3. It should be based on all the observations.
4. It should be capable of further mathematical treatment.
5. It should be affected as little as possible by effects of sampling.
6. It should not be affected much by extreme observations.

## MEASURES OF DISPERSION

**Introduction:**

An average does not tell the full story. It is hardly fully representative of a mass unless we know the manner in which the individual items scatter around it…..a further description of the series is necessary if are to understand how representative the average is. Thus emerges Dispersion.

### Defining Dispersion:

**"**It is the degree of the scatter or variation of the variables about a central value.**"**

### The various measures of dispersion are:

### 1. Range

The range is the simplest of all the measures of dispersion…it is defined as the diff b/w 2 extreme values of the distribution… (Difference b/w d maximum and minimum value of a distribution)

### 2. Quartile Deviation (QD) / Semi – inter – quartile – range

- It is a measure of dispersion based on upper quartile Q3 and lower quartile Q1.
- Inter quartile range = Q3 – Q1
- QD is obtained by dividing the inter quartile range with 2

### 3. Mean Deviation/Avg Deviation

This measure of dispersion is obtained by taking d Avg (usually AM) of deviations of given values from a measure of central tendency. It is the amt of scatter of the items in a distribution from either the mean or the median, ignoring the signs of distribution.

### 4. Standard Deviation

SD is defined as the +ve square root of the arithmetic mean of the squares of the deviations of the given observations from their arithmetic mean.

It satisfies all the characteristics of ideal measure of dispersion.

**Merits:**

a. SD is by far most vital & widely used measure of dispersion (capable of further arithmetic treatment)
b. The squaring of the deviations removes the drawback of ignoring signs as seen in MD.
c. SD is affected least by fluctuations of sampling.

**Demerits:**

The nature of extracting the square roots which s not readily comprehensible for a non – mathematical person.

## Absolute & Relative measures of Dispersion

**Absolute measures** – the measures of dispersion which are expressed in terms of original units of a series are termed as absolute measures; they are not suitable for comparing d variability of two distributions; the absolute measures are:

a. Range.
b. Quartile deviation (semi inter quartile range)
c. Mean deviation.
d. Standard deviation.

**Relative measures** – they are obtained as ratios or % …for comparing variability of the two distributions; we compute the relative measures of dispersion. The relative measures are:

a. Coefficient of range.
b. Coefficient of QD.
c. Coefficient of MD.
d. Coefficient of SD (C.V)

## Quartiles, Deciles & Percentiles – An overview:

**Quartiles:**

The values which divide data into four equal parts are called Quartiles. The first, second and third quartiles are denoted by $Q_1$, $Q_2$, $Q_3$ respectively. The first and third quartiles are also called the lower and upper quartiles respectively. The second quartile represents the median, the middle value.

**Deciles:**

The values which divide data into ten equal parts are called deciles. The first, second … ninth deciles are denoted by $D_1$, $D_2$, $D_3$, .....,$D_9$ respectively. The fifth Decile corresponds to median.

**Percentiles:**

The values which divide data into one hundred equal parts are called percentiles. The first, second,.....,ninety-ninth percentile are denoted by $P_1$, $P_2$, $P_3$,.....,$P_{99}$. The $50^{th}$ percentile corresponds to the median.

## Numerical Problems – Measures of Central Tendency

Q1. Compute mean, median, mode, $Q_1$, $Q_3$, $D_4$ and $P_{81}$: 10, 20, 30, 40, and 50.

Q2. Compute the arithmetic mean, median and mode; $Q_1$, $Q_2$, $Q_3$, $D_8$ and $P_{32}$.

| Marks | 20 | 30 | 40 | 50 | 60 | 70 |
|---|---|---|---|---|---|---|
| No of Students | 8 | 12 | 20 | 10 | 6 | 4 |

Q3. Compute arithmetic mean, median and mode, three quartiles, 6th Decile & 74th Percentile.

| Wages (in '000s) | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|---|---|
| No of workers | 5 | 10 | 25 | 30 | 20 | 10 |

Q4. Calculate Median & Mode. Using them, find Arithmetic Mean.

| Marks | 10 | 20 | 30 | 40 | 50 | 60 |
|---|---|---|---|---|---|---|
| No of Students | 8 | 23 | 45 | 65 | 75 | 80 |

Q5. The data on the profit (in lakhs) earned by 60 companies is given below:

| Profit (Rs. In lakhs) | Below 10 | 10-20 | 20-30 | 30-40 | 40-50 | Above 50 |
|---|---|---|---|---|---|---|
| No. of companies | 5 | 12 | 20 | 16 | 5 | 2 |

Find the three quartiles; calculate $D_6$ and $P_{42}$.

Q6. Find the values of mean, median and mode from the following data:

| C – I | 93-97 | 98-102 | 103-107 | 108-112 | 113-117 | 118-122 | 123-127 | 128-132 |
|---|---|---|---|---|---|---|---|---|
| F | 3 | 5 | 12 | 17 | 14 | 6 | 3 | 1 |

Q7. For the following data given below, find the mean salary paid to the workers.

| Income Group | 800 – 1000 | 1000 – 1200 | 1200 – 1400 | 1400 – 1600 | 1600 – 1800 |
|---|---|---|---|---|---|
| No of firms | 40 | 32 | 26 | 28 | 42 |
| No of workers | 8 | 12 | 8 | 8 | 4 |

Q8. The following data gives the weekly wages of workers in a firm, their total working hours and average working hours per worker. Calculate average weekly wage per worker.

| Wage group (Rs.) | 80 – 100 | 100 – 120 | 120 – 140 | 140 – 160 | 160 – 180 | 180 – 200 |
|---|---|---|---|---|---|---|
| Total hrs worked | 168 | 170 | 225 | 272 | 126 | 91 |
| No of hrs worked per worker | 12 | 10 | 9 | 8.5 | 7 | 6.5 |

Q9. The following are the hourly salaries of 20 employees of a firm:
130 62 145 118 125 76 151 142 110 98 65 116 100 103 71 85 80 122 132 95

The firm gives bonuses of Rs. 10, Rs. 15, Rs. 20, Rs. 25 and Rs. 30 to the employees in the respective salary groups – Rs. 60 but not exceeding Rs. 80, exceeding Rs. 80 but not exceeding Rs. 100 and so on up to exceeding Rs. 140 but not exceeding Rs. 160.

Find the average bonus paid per employee.

Q10. The following is the age distribution of 1000 persons working in an organization:

| Age group | 20–25 | 25–30 | 30–35 | 35–40 | 40–45 | 45–50 | 50–55 | 55–60 | 60–65 |
|---|---|---|---|---|---|---|---|---|---|
| No of persons | 30 | 160 | 210 | 180 | 145 | 105 | 70 | 60 | 40 |

Due to continuous losses, it is desired to bring down the manpower strength to 30% of the present number according to the following scheme:

(a) Retrench the first 15% from the lower age group.
(b) Absorb the next 45% in other branches.
(c) Make 10% from the highest age group retire permanently.

Calculate the age limits of persons transferred to the other departments and those retained. Also find out the average age of those retained.

Q11. Given Mean = 28, find the missing frequency from the data given below:

| Profits per shop (in '000s) | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|---|---|
| No of shops | 12 | 18 | 27 | - | 17 | 6 |

Q12. Given total frequencies (N = 170) and Median = 35, find out the missing frequencies.

| C – I | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| f | 10 | 20 | ? | 40 | ? | 25 | 15 |

Q13. If Modal marks of 60 students are given to be 47.5, find the missing values in distribution.

| Marks | 30-35 | 35-40 | 40-45 | 45-50 | 50-55 | 55-60 | 60-65 |
|---|---|---|---|---|---|---|---|
| No of Students | 3 | 5 | ? | 18 | 14 | ? | 2 |

Q14. The median and mode of the following frequency distribution of marks are known to be 33.5 and 34 respectively. Find the missing frequencies. [Given N = 230]

| Marks | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| No of Students | 4 | 16 | ? | ? | ? | 6 | 4 |

Q15. There are two units of an automobile company in two different cities employing 760 and 800 persons. The arithmetic means of monthly salary paid in these two cities are 18750 and 16950 respectively. Find the combined arithmetic mean of salary of employees in both the units.

Q16. The average daily wages of all workers in a factory is 444. If average daily wages paid to male and female workers are Rs. 480 and Rs. 360 respectively, find the percentage of male and female workers employed in the factory.

Q17. The mean salary paid to 1500 employees of an organization was found to be Rs. 12500. Later on, it was discovered that salary of 2 employees was wrongly entered as Rs. 15760 and Rs.9590 instead of Rs. 17760 and Rs. 8590. Calculate the correct mean.

Q18. In a factory, there are 100 skilled, 250 semi skilled and 150 unskilled workers. It has been observed that on an average, a unit length of particular fabric is woven by a skilled worker in 3 hours, by a semi skilled worker in 4 hours and by an unskilled worker in 5 hours. Unskilled workers are expected to become semi skilled workers and semi skilled workers are expected to become skilled after a period of time. How much less time will be required after 2 years of training for weaving the unit length of fabric by an average worker?

## Numerical Problems – Measures of Dispersion

**Problems on Range**:

Q1. Find Range for the following data: 138 150 151 151 157 158 160 161 162 162 173 175

Q2. Calculate range and its coefficient for the data given below:

| C-I | 16 – 20 | 21 – 25 | 26 – 30 | 31 – 35 |
|---|---|---|---|---|
| F | 10 | 15 | 17 | 8 |

### Problems on Quartile Deviation (Semi Inter Quartile Range):

Q3. Calculate QD and its coefficient:

| C-I | Less than 35 | 35 – 37 | 38 – 40 | 41 – 43 | More than 43 |
|---|---|---|---|---|---|
| F | 14 | 62 | 99 | 18 | 7 |

### Problems on Mean Deviation:

Q4. Calculate mean deviation from mean and its coefficient for the following data:

| Wages (in '000s) | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| No of workers | 7 | 12 | 18 | 25 | 16 | 14 | 8 |

### Problems on Standard Deviation & CV:

Q5. Find the SD of the following observations: 1, 2, 4, 6, 8 and 9.

Q6. Calculate Mean, SD and CV:

| Wages (in '000s) | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 |
|---|---|---|---|---|---|---|---|
| No of workers | 5 | 12 | 30 | 45 | 50 | 37 | 21 |

Q7. Find the SD and variance:

| Age under | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|
| No of persons dying | 15 | 30 | 53 | 75 | 100 | 110 | 115 | 120 |

Q8. From the prices of shares of X and Y given below, find out which is more stable in value.

| X | 35 | 54 | 52 | 53 | 56 | 58 | 52 | 50 | 51 | 49 |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | 108 | 107 | 105 | 105 | 106 | 107 | 104 | 103 | 104 | 101 |

Q9. Two brands of tyres are tested with the following results:

| | No of Tyres (Brands) | |
|---|---|---|
| Life (in '000 miles) | X | Y |
| 20 – 25 | 1 | 0 |
| 25 – 30 | 22 | 24 |
| 30 – 35 | 64 | 76 |
| 35 – 40 | 10 | 0 |
| 40 – 45 | 3 | 0 |

(a) Which brand of tyres has greater average life?
(b) Compare the variability and state which brand of tyres you would use on your fleet of trucks.

Q10. The shareholders research centre of India as recently conducted a research study on price behavior of three leading industrial shares A, B and C for the period 1979 to 1985, the results of which are published in its quarterly journal:

| Share | Avg Price (Rs.) | S.D | Current Selling Price (Rs.) |
|---|---|---|---|
| A | 18.2 | 5.4 | 36.00 |
| B | 22.5 | 4.5 | 34.75 |
| C | 24.0 | 6.0 | 39.00 |

a. Which share in your opinion appears to be more stable in value?
b. If you are the holder of all the three shares, which one you would like to dispose off at present and why?

Q11. For a group of 50 male workers, the mean and standard deviation of their monthly wages are Rs. 6300 and Rs. 900 respectively. For a group of 50 female workers, these are Rs. 5400 and Rs. 600 respectively. Find SD of monthly wages for the combined group of workers.

**MISCELLANEOUS PROBLEMS**

Q1. Calculate SD of profit earned.

| Profit (in crores) | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|---|---|
| No of companies | 8 | 12 | 20 | 30 | 20 | 10 |

Q2. From analysis of monthly wages paid to workers in X & Y, following results were obtained:

| | Organization X | Organization Y |
|---|---|---|
| No of wage earners | 550 | 650 |
| Average monthly wages | 5000 | 4500 |
| Variance of distribution of wages | 900 | 1600 |

i) Which organization pays a larger amount as monthly wages?
ii) Find out which organization is more consistent in payment of wages.
iii) Find the combined SD of X and Y together.

Q3. Weekly sales of 2 two products are given below. Find which shows greater sale fluctuation.

| Product A | 59 | 75 | 27 | 63 | 27 | 28 | 56 |
|---|---|---|---|---|---|---|---|
| Product B | 150 | 200 | 125 | 310 | 330 | 250 | 225 |

Q4. The mean salary paid to 100 employees of a company was Rs. 5000. The mean salary paid to male and female workers are Rs. 5200 and Rs. 4200 respectively. Determine the percentage of male and female workers employed in the factory.

Q5. The average weight of a group of 30 friends increased by 1 kg when the weight of their football coach was added. If the average weight of the group after adding the weight of the football coach was 31 kg, find out the weight of the football coach.

Q6. A charitable organization decided to give old age pension to people over sixty years of age. The scale of pensions was fixed as follows.

| Age Group | Rs. Per Month | Frequency |
|---|---|---|
| 60 – 65 | 200 | 7 |
| 65 – 70 | 250 | 5 |
| 70 – 75 | 300 | 6 |
| 75 – 80 | 350 | 4 |
| 80 – 85 | 400 | 3 |

The ages of 25 persons who secured the bonus right are given. Calculate monthly average pension payable per person and the standard deviation.

Q7. Goals scored by 2 teams A and B in a football season are shown below. Find which team may be considered more consistent.

| | No of Matches | |
|---|---|---|
| No of goals scored in a Match | Team A | Team B |
| 0 | 27 | 17 |
| 1 | 9 | 9 |
| 2 | 8 | 6 |
| 3 | 5 | 5 |
| 4 | 4 | 3 |

Q8. For a group containing 100 observations, Mean & SD are 8 and $\sqrt{10.5}$ respectively. For 50 observations, selected from this, mean and SD are 10 and 2 respectively. Calculate the mean and SD of the other half.

Q.9. Twenty passengers were found ticketless in a bus. The sum of squares and the variance of the amount found in their pockets were Rs 2000 and Rs 36 respectively. If the total fine imposed on these passengers is equal to the amount recovered from them and the fine imposed is uniform, what is the average amount each one of them has to pay as fine?

Q.10. The average salary of 49 out of 50 employees in a firm is Rs 100. The salary of 50th employee is Rs 97.50 more than the average salary of all the 50 workers. Find the mean salary of all the employees in the firm.

* * * * * *

**STATISTICS FOR MANAGERS (24MBA15)**

MODULE 2

**Correlation & Regression Analysis**

**Introduction:** So far we have studied problems w.r.t one variable only; in practice we come across a large no of problems involving use of more than one variable; if two quantities vary in such a way that the movements in one are accompanied by movements in another, then we say that the quantities are **correlated.**

- Consider the following examples:

a. The series of sales revenue & advertisement expenditure.
b. Price of commodity & amt demanded.
c. Increase in rainfall up to a point & production of rice.

The degree of relationship b/w the above variables under consideration is **measured through correlation analysis.**

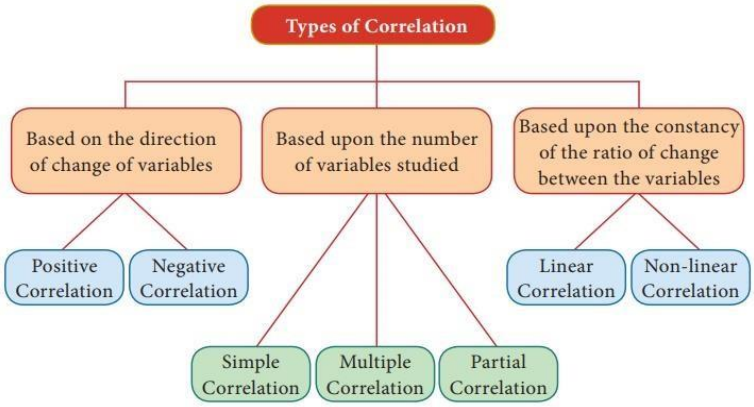
## Correlation - Meaning

**Correlation analysis attempts to determine the degree of relationship between variables**.

## Significance:

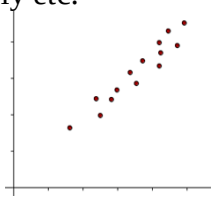
a. To measure degree of relationship between variables.
b. Correlation analysis contributes to understanding of economic behavior
c. In business, correlation analysis enables the executives to estimate cost, sales, prices on basis of some series which are functionally related.

## Types of Correlation

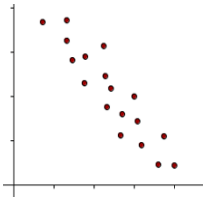



**I: Based on Direction of change of Variables:**

**1) Positive Correlation:** If two variables change in the same direction (i.e. if one increases the other also increases, or if one decreases, the other also decreases), then this is called a positive correlation. **For example** the training and performance of employees in a company; family income & expenditure on luxuries; price & supply etc.



**2) Negative Correlation:** In negative correlation, the two variables move in the opposite direction. When the value of a variable increases, the value of the other variable decreases. **For example**, the relationship between price and demand; volume & pressure of gas etc; T.V. registrations and cinema attendance.



## II: Based on the Number of Variables studied:

1) **Simple Correlation**: When only two variables are studied it is a problem of simple correlation.

2) **Multiple Correlation**: In multiple correlation, three or more variables are studied simultaneously. For example, when one studies the relationship between the yield of rice per acre and both the amount of rainfall and the amount of fertilizers used, it is a problem of multiple correlation.

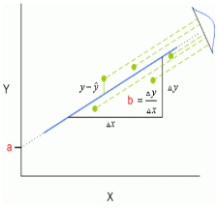
3) **Partial Correlation**: In partial correlation we recognize more than two variables, but consider only two variables to be influencing each other, the effect of other influencing variables being kept constant. For example, in the price problem taken above if we limit our correlation analysis of yield and rainfall to periods when a certain average daily temperature existed, it becomes a problem relating to partial correlation only.

## III: Based on Constancy of Ratio of change between Variables:

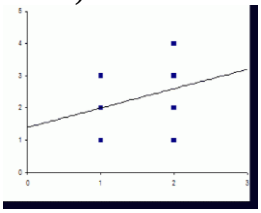
**1) Linear Correlation:** The correlation b/w two variables i**s said to be linear** if corresponding to a unit change in one variable, there is a constant change in other variable over entire data:

| | | | | | |
|---|---|---|---|---|---|
| X | 1 | 2 | 3 | 4 | 5 |
| Y | 5 | 7 | 9 | 11 | 13 |

Here x and y are linearly related as y = 2x + 3

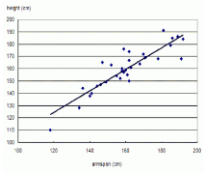



**2) Non Linear Correlation:** Correlation i**s said to be non linear when** the two variables do not move in a linear form (for e.g. – just because the amount of rainfall is doubled, we can't expect the yield of food crops to be doubled as well)

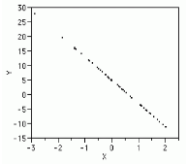


## IV: Other types:

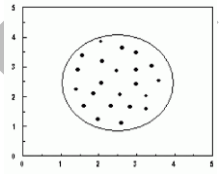
**1) Perfect Positive correlation:** When there is a change in one variable, and if there is equal proportion of change in the other variable say Y in the same direction, then these two variables are said to have a Perfect Positive Correlation.



**2) Perfect Negative correlation:** Between two variables X and Y, if the change in X causes the same amount of change in Y in equal proportion but in opposite direction, then this correlation is called as Perfectly Negative Correlation.



**3) Zero/No Correlation:** When the two variables are independent and the change in one variable has no effect in other variable, then the correlation between these two variables is known as Zero Correlation. No correlation occurs when there is no linear dependency between the variables.
**Example**: Age of your father and your shoe size.



## Interpreting "r" – Coefficient of determination ($r^2$)

When r = 1 or -1 or 0, the interpretation of r does not pose any problem – it means a perfect positive or perfect negative relationship. However, when it comes to other values of r, its interpretation is as follows:
Suppose we get a correlation of r = 0.9 b/w advertisement expenses and sales. The strength of "r" is judged by coefficient of determination. **($r^2$)**

Thus if r = 0.9, **($r^2$) = 0.81** (81%). It means that 81% of total variation in Y (Sales) is attributed to the relationship with X (Advertisement expenses) and the rest 19% of the variations are explained due **to other factors**.

Coefficient of determination [$r^2$] = Explained variation / Total variance

## Methods of studying Correlation

1. Scatter Diagram method.
2. Karl Pearson s coefficient of correlation
3. Two way frequency table (bivariate correlation)
4. Rank correlation method.
5. Concurrent deviation method.

# REGRESSION ANALYSIS

**Meaning:** Regression analysis in the general sense means the estimation or prediction of the unknown value of one variable from the known value of the other variable.

**Types of Variables**: In regression analysis, there are two types of variables ---the variable whose value is to be influenced or predicted called the ***dependent variable*** & the variable which influences the value called the ***independent variable***.

**Uses:** Regression analysis is used in business & economics to study the relationship b/w two or more variables that are related & for the estimation of demand/supply curves, cost/production, consumption functions etc.

## Simple vs. Multiple regressions

- When one independent variable is used in a regression, it is called a simple regression; when two or more independent variables are used, it is called a multiple regression.
- The general purpose of multiple regressions is to learn more about the relationship between several independent or predictor variables and a dependent or criterion variable.

## Multiple Regression - Application

- HR professionals customarily use multiple regression procedures to determine equitable compensation.
- You can determine a number of factors or dimensions such as "amount of responsibility" (*Resp*) or "number of people to supervise" (*No Super*) that you believe to contribute to the value of a job. This information can be used in a multiple regression analysis to build a regression equation of the form: **Salary = .5*Resp + .8*No_Super**
- Once this regression line has been determined, the analyst can now easily construct a graph of the expected (predicted) salaries and the actual salaries of job incumbents in his or her company. Thus, the analyst is able to determine which position is underpaid (below the regression line) or overpaid (above the regression line), or paid equitably.

*Regression Models**

## Differences b/w correlation & regression

1. Correlation literally means the relationship b/w 2 or more variables where movements in one tend to be accompanied by movements in another **but** regression literally means stepping back & is a measure expressing Avg relationship b/w two variables.
2. Correlation coefficient is a measure of d *linear* relationship b/w 2 variables (symmetric) **but** regression analysis is *functional* relationship b/w 2 variables where value of one variable can be predicted with d help of another
3. Correlation need not imply cause & effect relation b/w variables under study. **But** regression does explain the cause & effect relation (b/w independent & dependent variable)
4. There may be nonsense correlation (for e.g. – b/w size of shoe & height of an individual) **but** there is nothing as such in regression.
5. Correlation is confined to linear relationship b/w variables & thus has limited applications **but** as regression studies the functional relation, it has much wider applications.

## Numerical Problems – Correlation Analysis

Q1. The following data relate to age of employees and the number of days they were reported sick in a month: Calculate Karl Pearson's coefficient of correlation and interpret it.

| Employees | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Age | 30 | 32 | 35 | 40 | 48 | 50 | 52 | 55 | 57 | 61 |
| Sick Days | 1 | 0 | 2 | 5 | 2 | 4 | 6 | 5 | 7 | 8 |

Q2. Calculate Karl Pearson's coefficient of correlation b/w expenses on advertisement and sales:

| Advertisement Expenses | 39 | 65 | 62 | 90 | 82 | 75 | 25 | 98 | 36 | 78 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sales (in lakhs) | 47 | 53 | 58 | 86 | 62 | 68 | 60 | 91 | 51 | 84 |

Q3. Calculate coefficient of correlation for the ages of husbands and wives and interpret it:

| Ages of Husbands | 23 | 27 | 28 | 29 | 30 | 31 | 33 | 35 | 36 | 39 |
|---|---|---|---|---|---|---|---|---|---|---|
| Ages of Wives | 18 | 22 | 23 | 24 | 25 | 26 | 28 | 29 | 30 | 32 |

Q4. Calculate the coefficient of correlation and interpret the result.

| Year | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|---|---|
| Index of production | 100 | 102 | 104 | 107 | 105 | 112 | 103 | 99 |
| No unemployed | 15 | 12 | 13 | 11 | 12 | 12 | 19 | 26 |

Q5. Using rank correlation, determine relationship between debenture prices and share prices.

| Debenture price | 97.8 | 99.2 | 98.8 | 98.3 | 98.4 | 96.7 | 97.1 |
|---|---|---|---|---|---|---|---|
| Share price | 73.2 | 85.8 | 78.9 | 75.8 | 77.2 | 87.2 | 83.8 |

Q6. Calculate Spearman's rank correlation coefficient b/w X and Y:

| X | 39 | 65 | 62 | 90 | 82 | 75 | 25 | 98 | 36 | 78 |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | 47 | 53 | 58 | 86 | 62 | 68 | 60 | 91 | 51 | 84 |

Q7. Ten competitors in a beauty contest are ranked by 3 judges in the following order. Use rank correlation to determine which pair of judges has nearest approach to common tastes in beauty.

| 1st Judge | 1 | 6 | 5 | 10 | 3 | 2 | 4 | 9 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|
| 2nd Judge | 3 | 5 | 8 | 4 | 7 | 10 | 2 | 1 | 6 | 9 |
| 3rd Judge | 6 | 4 | 9 | 8 | 1 | 2 | 3 | 10 | 5 | 7 |

Q8. From following data, calculate rank correlation coefficient after adjustment for tied ranks:

| X | 48 | 33 | 40 | 9 | 16 | 16 | 65 | 24 | 16 | 57 |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | 13 | 13 | 24 | 6 | 15 | 4 | 20 | 9 | 6 | 19 |

Q9. A psychologist wanted to compare two methods of teaching – A & B. He selected a random sample of 22 students. He grouped them into 11 pairs. Scores obtained by students are given below. Find rank correlation coefficient and interpret the result.

| Pair | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | 24 | 29 | 19 | 14 | 30 | 19 | 27 | 30 | 20 | 28 | 11 |
| B | 37 | 35 | 16 | 26 | 23 | 27 | 19 | 20 | 16 | 11 | 21 |

Q10. A computer while calculating correlation coefficient between 2 variables X and Y from 25 pairs of observations obtained the following results. Find the correct "r".
$\Sigma x y = 508$, $n = 25$, $\Sigma x = 125$, $\Sigma x^2 = 650$, $\Sigma y = 100$ and $\Sigma y^2 = 460$.
It was later discovered that at the time of checking, two pairs of observations were not correctly entered. They were taken as (6, 14) and (8, 6) while the correct values were (8, 12) and (6, 8).

## Numerical Problems – Regression Analysis

Q1. Find the two regression equations and estimate the marks in statistics if marks in economics are 30. Also find r (correlation coefficient)

| Marks in Economics | 25 | 28 | 35 | 32 | 31 | 36 | 29 | 38 | 34 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks in Statistics | 43 | 46 | 49 | 41 | 36 | 32 | 31 | 30 | 33 | 39 |

Q2. Table below shows motor registrations in a certain territory & sale of motor tyres for that period. Estimate the sale of tyres when registrations are known to be 850.

| Year | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Motor registrations | 600 | 630 | 720 | 750 | 800 |
| No of tyres sold | 1250 | 1100 | 1300 | 1350 | 1500 |

Q3. Estimate:

i) Sales for ad expenses of Rs. 100 lakhs.
ii) Ad expenses for sales of Rs. 47 crores.

| Sales (in crores) | 14 | 16 | 18 | 20 | 24 | 30 | 32 |
|---|---|---|---|---|---|---|---|
| Ad expenses (in lakhs) | 52 | 62 | 65 | 70 | 76 | 80 | 78 |

Q4. The following table gives the aptitude test scores and productivity indices of 10 workers:

| Aptitude scores | 60 | 62 | 65 | 70 | 72 | 48 | 52 | 73 | 65 | 83 |
|---|---|---|---|---|---|---|---|---|---|---|
| Productivity Indices | 68 | 60 | 62 | 80 | 85 | 40 | 52 | 62 | 60 | 81 |

Calculate two regression equations and estimate:

(i) PI of a worker whose test score is 92.
(ii) Test score of a worker whose PI is 75.

Q5. A panel of judges A and B graded 7 debaters and awarded the following marks:

| Debater | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Marks by A | 40 | 34 | 28 | 30 | 44 | 38 | 31 |
| Marks by B | 32 | 39 | 26 | 30 | 38 | 34 | 28 |

An 8th debater was awarded 36 marks by Judge A while Judge B was not present. If Judge B was also present, how many marks would you expect him to award to the 8th debater assuming the same degree of relationship exists in their judgment?

Q6. Use least squares regression line to estimate the increase in sales revenue expected from an increase of 7.5% in advertisement expenditure.

| Firm | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| Annual % increase in ad expense | 1 | 3 | 4 | 6 | 8 | 9 | 11 | 14 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Annual % increase in sales revenue | 1 | 2 | 2 | 4 | 6 | 8 | 8 | 9 |

Q7. Calculate the two regression coefficients given: r = 0.8, σx = 5 and σy = 7

Q8. You are given below the following data about advertisement and sales:

| | Ad expenses (Rs. Crores) | Sales (Rs. Crores) |
|---|---|---|
| Mean | 10 | 90 |
| S.D (σ) | 3 | 12 |

r = 0.8.

(i) Find the two regression lines.

(ii) What should be advertisement expenditure if the company wants to attain sales target of 150 crores?

Q9. If two lines of regression are: 20X – 9Y – 107 = 0 and 4X – 5Y + 30 = 0

i) Find the mean values of X and Y.

ii) Which of these lines is line of regression of y on x and which line is line of regression of x on y?

iii) Find the correlation coefficient between x and y.

Q10. The two regression equations are as follows: 8X – 10Y + 66 = 0 and 40X – 18Y = 214 Find:

i) Mean values of X and Y.

ii) r (correlation coefficient)

**MISCELLANEOUS PROBLEMS:**

Q1. From the following data, find the correlation between age and crime rates.

| Age group | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 | 80-90 | 90-100 |
|---|---|---|---|---|---|---|---|---|---|---|
| No of persons | 80 | 60 | 50 | 40 | 30 | 25 | 20 | 15 | 10 | 5 |
| No of crimes | 50 | 45 | 40 | 35 | 30 | 20 | 15 | 11 | 8 | 3 |

Q2. From the following data, calculate Karl Person's correlation:

| Particulars | X | Y |
|---|---|---|
| No of Items | 12 | 12 |
| Means | 15 | 10 |
| SD | 3.16 | 3.29 |
| Sum of squares of deviation from mean | 120 | 130 |

Sum of product of deviation of both the variables: 90

Q5. A Departmental store gives in service training to its salesmen which are followed by a test. It is considering whether it should terminate the service of any salesmen who do not do well in the test. Data regarding test scores and sales made by salesmen is given below.

| Age group | 14 | 19 | 24 | 21 | 26 | 22 | 15 | 20 | 19 |
|---|---|---|---|---|---|---|---|---|---|
| No of persons | 31 | 36 | 48 | 37 | 50 | 45 | 33 | 41 | 39 |

i) Is termination of services of salesmen with low test scores justified?

ii) If the firm wants a minimum sales volume of Rs. 3000, what is the minimum test score that will ensure continuation of service?

iii) Estimate probable sales volume of a salesman making a score of 28.

Q6. Find the two regression equations and estimate the sales value if ad expenses are known to be 15 lakhs. Given r = 0.6

| | AM | SD |
|---|---|---|
| Ad expenses | 16.5 | 8 |
| Sales | 72.4 | 11.1 |

Q7. An economist wanted to find out if there was any relationship between unemployment rate in country & inflation rate. Data gathered from 7 countries for 2004 are as follows:

| Country | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| Unemployment Rate | 4.0 | 8.5 | 5.5 | 0.8 | 7.3 | 5.8 | 2.1 |
| Inflation Rate | 3.2 | 8.2 | 9.4 | 5.1 | 10.1 | 7.8 | 4.7 |

Find degree of linear association b/w a country's unemployment rate and its level of inflation.

Q8. Given the two regression equations as $5x = 6y + 24$ and $1000y = 768x - 3708$, find "r".

Q10. Find Karl Pearson's correlation coefficient between size and defect in quality.

| Size group | 15 – 16 | 16 – 17 | 17 – 18 | 18 – 19 | 19 – 20 | 20 – 21 |
|---|---|---|---|---|---|---|
| No of items | 200 | 270 | 340 | 360 | 400 | 300 |
| No of defectives | 150 | 162 | 170 | 180 | 180 | 114 |

**STATISTICS FOR MANAGERS**
**(24MBA15)MODULE 3**
**TIME SERIES ANALYSIS**

**Contents:**

- Time Series – An overview; Uses.
- 4 Components of Time Series: Trend, Seasonal variations, Cyclical variations & Irregular variations.
- Measurement of Trend: Graphical Method, Method of Semi Averages, Method of Moving averages & Method of Least Squares
- Measurement of Seasonal Indices: Simple Avg Method, Ratio to Moving Avg Method & Ratio to trend Method.

## TIME SERIES – MEANING:

It is an arrangement of statistical data in a chronological order in accordance with its time of occurrence. It reflects a dynamic pace of movements of a phenomenon over a period of time.

## USES / OBJECTIVES of Time Series Analysis:

- It enables us to study the past behavior of phenomenon under consideration.
- The study of components of time series is of importance to a businessman in planning of future operations and formulation of executive/policy decisions.
- Comparisons of current performance with expected ones are possible.
- It helps us to compare the changes in values of different phenomenon at different times.

## COMPONENTS of Time Series:

1. Secular Trend.
2. Seasonal Variations.
3. Cyclical Variations.
4. Irregular Variations.

**1 SECULAR TREND:** The general tendency of time series data to increase/decrease/stagnate during a long period of time is called simple/secular trend. **Examples:** Upward tendency w.r.t population, production, sales, prices, incomes, money in circulation etc. Likewise downward tendency w.r.t issues relating to deaths, epidemics etc due to advances in technology, improved medical facilities etc. [**Note:** The concept of trend doesn't include short term oscillations but steady movements over a period of time.]

**2 SEASONAL VARIATIONS:** These are due to the forces that operate in a regular/periodic manner over a period less than a Year. (In the pattern of daily, weekly, quarterly etc). Examples: Sales & profits in a departmental store, bank clearings etc. The main objective of measurement of seasonal indices/variations is to isolate them from trend and study their variations so as to help businessmen develop plans, policies etc.

**Causes:**

a. Those resulting from natural sources: umbrella sales, ice cream sales, tourism demand.
b. Those resulting from manmade conventions: sales of ornaments go up during marriages, sale of cakes go up during Christmas & New Year etc.

## 3 CYCLICAL VARIATIONS:

The oscillatory movements in a time series with variation period more than a year is called as cyclical variation. These are due to ups and downs in an economic activity due to business cycles recurring after a period of more than a year and lasts up to 7 to 9 years. The usual phases are:

a. Prosperity.
b. Recession.
c. Depression.
d. Recovery.

Knowledge of cyclical component enables a businessman to have an idea about periodicity of booms; depressions etc and accordingly he can take timely steps for maintaining stable market for his product.

## 4 IRREGULAR VARIATIONS:

These are the result of unforeseen or unpredictable forces which operate in an absolutely erratic and irregular manner. They do not exhibit any definite patterns. They are caused by non recurring factors like earthquakes, strikes, lockouts, revolutions etc. Hence statistical forecasts based on them cannot be purely undertaken.

## Measurement of Secular Trend – Methods:

1. Graphical Method
2. Method of Semi Averages
3. Method of Moving averages
4. Method of Least Squares

## Measurement of Seasonal Indices – Overview & Methods:

Seasonal effects are measured in terms of an index called seasonal index attached to each period of the time series within a year. Hence if monthly data are considered, there are 12 separate indices, one for each month. Similarly for quarterly data, there are 4 separate indices. A seasonal index is an average that indicates the percentage deviation of actual values of the time series from a base value which excludes the short term seasonal influences. The following are the methods used to compute seasonal indices to measure seasonal effects in time series data:

1. Method of simple averages.
2. Ratio to trend method.
3. Ratio to moving average method.

## NUMERICAL PROBLEMS

**Method of Semi Averages**:

1. Apply the method of semi averages for determining trend for data given below. Also plot the actual and trend values graphically and estimate sales for the year 2021.

| Year | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|
| Sales | 20 | 24 | 22 | 30 | 28 | 32 |

### Method of Moving Averages:

2. Calculate 3 yearly and 5 yearly moving averages for the following data:

| Year | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sales | 242 | 250 | 252 | 249 | 253 | 255 | 251 | 257 | 260 | 265 | 262 |

3. Calculate trend values by method of moving averages (4 yearly cycle) from the following data w.r.t sugar production (X) in India:

| Year | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | 37.4 | 31.1 | 38.7 | 39.5 | 47.9 | 42.6 | 48.4 | 64.6 | 58.4 | 38.6 | 51.4 | 84.4 |

4. Compute 4 yearly moving & centered averages. Also graph deserved and trend values.

| Year | 2008 | 2009 | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 |
|---|---|---|---|---|---|---|---|---|---|
| Sales | 103 | 104 | 107 | 101 | 102 | 104 | 105 | 99 | 100 |

## Method of Least Squares: (Linear)

5. Fit a straight line trend for the following data by method of least squares and show trend values. What is the monthly increase in production?

| Year | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 | 2025 |
|---|---|---|---|---|---|---|---|
| Sales | 77 | 88 | 94 | 85 | 91 | 98 | 90 |

6. Fit a straight line trend for the following data by method of least squares and show trend values. Predict the earnings for the year 2020.

| Year | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|---|---|---|
| Sales | 38 | 40 | 65 | 72 | 69 | 60 | 87 | 95 |

7. The following table relates to the tourist arrivals (in millions) during 2004 to 2010 in India. Fit a straight line by method of least squares and estimate number of tourists to have arrived in the year 2014.

| Year | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|---|
| Sales | 18 | 20 | 23 | 25 | 24 | 28 | 30 |

## Seasonal Indices: Method of Simple Averages.

8. Use the method of monthly averages to determine the monthly indexes for the data of production of a commodity for the years 2016 to 2018.

| Month | 2016 | 2017 | 2018 |
|---|---|---|---|
| January | 15 | 23 | 25 |
| February | 16 | 22 | 25 |
| March | 18 | 28 | 35 |
| April | 18 | 27 | 36 |
| May | 23 | 31 | 36 |
| June | 23 | 28 | 30 |
| July | 20 | 22 | 30 |
| August | 28 | 28 | 34 |
| September | 29 | 32 | 38 |
| October | 33 | 37 | 47 |
| November | 33 | 34 | 41 |
| December | 38 | 44 | 53 |

## Seasonal Indices: Ratio to Trend Method

9. Quarterly sales data (Rs in million) are presented in following table for a 4 year period. Calculate seasonal index for each of the four quarters using ratio to trend method.

| | Quarters | | | |
|---|---|---|---|---|
| Year | I | II | III | IV |
| 2015 | 60 | 80 | 72 | 68 |
| 2016 | 68 | 104 | 100 | 88 |
| 2017 | 108 | 152 | 136 | 124 |
| 2018 | 160 | 184 | 172 | 164 |

## Seasonal Indices: Ratio to Moving Average Method

10. Calculate seasonal index by the ratio to moving average method for the following data:

| | Quarters | | | |
|---|---|---|---|---|
| Year | I | II | III | IV |
| 2015 | 75 | 60 | 53 | 59 |
| 2016 | 86 | 65 | 63 | 80 |
| 2017 | 90 | 72 | 66 | 85 |
| 2018 | 100 | 78 | 72 | 93 |

**STATISTICS FOR MANAGERS (24MBA15)**

Module 4

**PROBABILITY & PROBABILITY DISTRIBUTIONS**

**Contents of the Module**:

- Introduction to Probability; Basic Numerical Problems in Probability.
- Bayes' theorem of Probability – Concept & Numerical Problems.
- Introduction to Probability Distributions.
- Binomial Distribution – Concept & Numerical Problems.
- Poisson distribution – Concept & Numerical Problems.
- Normal Distribution – Concept & Numerical Problems.

## Introduction to Probability:

Probability is the foundation of statistical theory and applications. The term "Probability" in statistics refers to the chance of occurrence of an event among a large number of possibilities.
For example:

a. We will probably get the business contract that we negotiated last month.
b. A producer cannot ascertain the future demand of his product with certainty.
c. Team A has a better chance to win the throw ball match against Team B etc.

Thus a numerical measure of this uncertainty is provided by a very important branch of statistics called the theory of probability.

## Definitions of various terms used in Probability:

1. RANDOM EXPERIMENT: It is an experiment which gives different outcomes when repeated under identical conditions. For example – tossing a coin, rolling a dice etc.

2. SAMPLE SPACE: It is a set of all the outcomes of a random experiment. For example – In tossing two coins, the sample space (S) = {HH, TT, HT, TH}. In rolling a dice, the sample space (S) = {1, 2, 3, 4, 5, 6} etc.

3. EVENT: It is a sub set of the sample space. For example - In rolling a dice, the sample space (S) = {1, 2, 3, 4, 5, 6}; A: Getting prime number – A = {2, 3, 5}. Here A is an event.

4. UNION – Here either of the event occurs (at-least one); If A = (1, 2, 3, 4) and B = (3, 4, 5, 6); then A U B = (1, 2, 3, 4, 5, 6)

5. INTERSECTION – Both the events should happen simultaneously; If A = (1, 2, 3, 4) and B = (3, 4, 5, 6); then A ∩ B = (3, 4)

6. EXHAUSTIVE CASES: All the elements of a sample space are called exhaustive cases. For example, in rolling a dice, the sample space (S) = {1, 2, 3, 4, 5, 6}. The values 1, 2, 3, 4, 5 and 6 are called the exhaustive cases.

7. FAVORABLE CASES: These are the elementary events because of which an event occurs. For example: If the experiment is tossing two coins, then S = {HH, TT, HT, TH}.

Event A: Getting at-least one tail, then A = {TT, TH, HT} – These three observations are called the favorable cases.

8. MUTUALLY EXCLUSIVE EVENTS: The events are said to be mutually exclusive when the occurrence of one event will prevent the occurrence of another in same trial.
   For example: If the experiment is tossing two coins, then S = {HH, TT, HT, TH}.
   A: Want 2 heads – A = {HH}
   B: Want 2 tails – B = {TT}
   Here A and B are mutually exclusive events.

9. INDEPENDENT EVENTS: Events are said to be independent if the occurrence of one does not affect or depend on the occurrence of other event in a series of trials.
   For example: If we roll a die, the outcome could be 6 for the first time. The second time we roll a die; the outcome could be different and will not depend on the first trial.
   For example: A husband and wife appear in an interview for two vacancies in two posts. The probability of wife's selection has nothing to do with the probability of the husband's selection/rejection. Hence the events Selection of Wife and Selection of Husband are two independent events.

10. DEPENDENT EVENTS: Events are said to be dependent if the occurrence of one is dependent on the occurrence of another event.
    For example: A bag has 5 red and 6 blue balls. Another bag has 7 red and 8 blue balls. A bag is picked at random and then a ball is drawn. The probability that the drawn ball could be a red ball or a blue ball is dependent on the event that first the bag must be selected. Unless one of the two bags is selected, it would be tough to estimate that the probability of the drawn ball is red or blue. Hence picking a ball is dependent on the event selection of the bag.

11. EQUALLY LIKELY EVENTS: Events are said to be equally likely if they have equal chances of occurrence.
    For example – In throw of a single die,
    A: Want only even numbers. Hence A = (2, 4, 6)
    B: Want only odd numbers. Hence B = (1, 3, 5) Thus A and B are equally likely events.

12. CLASSICAL / MATHEMATICAL / APRIORI Definition of Probability:
    $P(A) = \frac{\text{No of favorable cases}}{\text{No of exhaustive cases}} = \frac{m}{n}$

13. ADDITION THEOREM OF PROBABILITY:
    Let A and B be any two events with respective probabilities P (A) and P (B). Then the probability of occurrence of at-least one of these two events is given by:
    $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

14. MULTIPLICATION THEOREM OF PROBABILITY:
    If A and B are two independent events, then their product is given by:
    $P(A \cap B) = P(A).\ P(B)$

15. BAYES' THEOREM (INVERSE PROBABILITY):

One of the important applications of probability is in the computation of unknown probabilities on the basis of information provided by the past records or experiments. For example – Suppose one event has occurred through one of the various other events. Then the conditional probability that it has occurred due to one of the events is called inverse probability. These probabilities are computed through the Bayes' rule or Bayes' theorem.

Let $E_1, E_2 \ldots E_n$ be mutually exclusive events defined on a sample space. Then the Probability of occurrence of event $E_i$ given event A has already occurred is given by

$$P(E_i/A) = \frac{P(A/E_i)\,P(E_i)}{\sum P(A/E_i)\,P(E_i)}$$

## PROBABILITY DISTRIBUTIONS:

A systematic presentation of values taken by the random variable and their corresponding probabilities is called a probability distribution. In the population, the values of the variables may be distributed according to some definite probability law which can be expressed mathematically and the corresponding distribution is known as theoretical probability distribution. These distributions are based on expectations on the basis of previous experience. Accordingly, three different types of theoretical probability distributions assume importance:

1. Binomial Distribution (BD)
2. Poisson distribution (PD)
3. Normal Distribution (ND)

## 1 BINOMIAL DISTRIBUTION

A Binomial Distribution (BD) is expected under the following **conditions**:

i) The number of trials "n" is finite (lesser); trials are independent.
ii) Each trial results in two mutually exclusive and exhaustive outcomes termed as success (p) and failure (q = 1 – p)
iii) The probability of happening of an event is fairly large.

**Definition**: A discrete random variable X is said to follow B.D if its probability mass function (p m f) is given by

$$pmf: \; p(x) = p^x q^{n-x} \, {}^nC_x$$

**Examples**:

a. Getting number of heads when 8 coins are tossed.
b. Getting number of 6s when 8 dice are rolled.
c. No of male children in a family of five.

### Properties of B.D:

1. Mean of BD = E (X) = n p
2. Variance of BD = V (X) = n p q
3. Recurrence relation:

$$E(X) = \frac{n - x + 1}{x} \cdot \frac{p}{q} E_{(x-1)}$$

## 2 POISSON DISTRIBUTION

Poisson distribution (PD) may be obtained as a limiting case of binomial distribution under the following **conditions**:

i) The number of trials "n" is large.
ii) The probability of success "p" is indefinitely small.
iii) n p = m [P.D in terms of B.D]

**Definition**: A discrete random variable X is said to follow P.D if its p m f is given by:

$$pmf :- \quad p(x) = \frac{e^{-m} m^x}{x!}$$

### Examples / Situations where P.D can be used:

a. The number of telephone calls arriving at a telephone switch board per unit of time.
b. The number of customers arriving at a supermarket per hour.
c. The number of defects of a manufactured product.
d. The number of accidents taking place per day on a busy road.
e. The number of errors per page in a typed material.

### Properties of P.D:

1. Mean = E (X) = m
2. Variance = V (X) = m
3. Recurrence relation:
   $E(X) = \frac{m}{x} E_{(x-1)}$

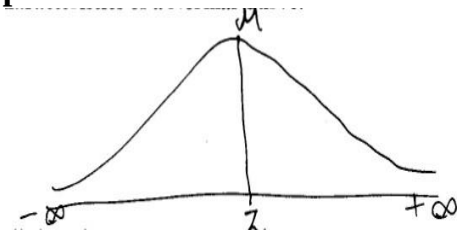
## 3 NORMAL DISTRIBUTION

The distributions (BD and PD) were of discrete random variables. Now we shall confine the discussion to continuous probability distributions which arises when underlying variable is a continuous one. [Here Mean = Median = Mode]

If x is a standard normal variate with mean μ and S.D ( ), then it can be converted to the form

$Z = \frac{x - \mu}{SD}$

### Properties / Characteristics of a Normal Curve:




1. It is a bell shaped curve.
2. Mean = Median = Mode.
3. The normal curve is symmetrical to the x axis.
4. The total are under the normal curve is 100% - 0.5 Area to the left and 0.5 area to the right. (total are is equal to 1)
5. The two tails of the curve extend indefinitely and hence they never touch the bases.

## Applications / Importance of N.D:

1. ND helps to draw inference about the universe by sample.
2. It is made use of in SQC – used for setting control limits for construction of control charts.
3. The whole theory of sample tests for hypothesis is based on the assumption that samples that are drawn follow normal distribution.
4. It is used in other areas such as central limit theorem, large sample theory etc.

## NUMERICAL PROBLEMS

**Problems on Basics of Probability**:

1. Find the probability that a throw of an unbiased die results in (a) even number (b) Multiple of 3.

2. Two fair dice are rolled. Find the P that (a) both dice show number 6 (b) sum of the number obtained is 7 or 10 (c) sum of the number obtained is divisible by 3.

3. Three unbiased coins are tossed. What is the P of getting (a) At-least one head (b) No tail?

4. There are 5 tennis balls and 6 cricket balls in a bag. 2 balls are picked at random. Find the P of getting (a) Balls of the same kind (b) Balls of different kinds. (c) Green marbles

**Problems on Addition theorem of Probability**:

5. A management executive committee consists of 5 members whose profiles are as follows – (a) Male 46 years (b) Female 39 years (c) Male 26 years (d) Male 35 years (e) Female 28 years. A Chairperson is to be selected for the committee among the 5 members. What is the P that the chairperson is a female or aged above 30 years?

6. The P that a contractor will get a plumbing contract is 2/3 and the P that he will not get an electric contract is 5/9. If the P of getting at-least one of the contracts is 4/5, what is the P that he will get both the contracts?

**Problems on Multiplication theorem of Probability**:

7. A husband and wife appear in an interview for two vacancies in two posts. The P of husbands' selection is 1/7 and that of wife's selection is 1/5. What is the P that (a) Both of them will be selected (b) Both of them will not be selected (c) Only one of them will be selected?

**Problems on Bayes' theorem**:

8. A factory produces bolts using 2 machines A and B with 60% and 40% output. Machine A and B produce 5% and 10% of the output with some defects. In the output if a bolt is drawn randomly, what is the P that (a) it is coming from Machine A (b) it is defective?

9. In a factory, 3 machines – A, B & C manufacture 25%, 35% and 40% output. Of them, 5%, 4% and 2% are defective. If an output unit is drawn at random, what is the P that (a)

the defective output comes from Machine A (b) that it comes from Machine B or Machine C?

10. A company has two plants to manufacture scooters. Plant I manufactures 80% of the scooters and Plant II manufactures 20%. At plant I, 85 out of 100 scooters are rated standard quality or better. At plant II, only 65 out of 100 scooters are rated standard quality or better.
    (i) What is the P that selected scooter at random came from Plant I if it is known that the scooter is of standard quality?
    (ii) What is the P that selected scooter at random came from Plant II if it is known that the scooter is of standard quality?

**Problems on Binomial Distribution**:

11. Ten unbiased coins are tossed simultaneously. Find the P of getting – (a) Exactly 6 heads (b) At-least 8 heads (c) No head (d) At-least 1 head.

12. The incidence of occupational disease in an industry is such that workers have 20% chance of suffering from it. What is the P that out of 6 workers, 4 or more will have the disease?

13. 95% of the bulbs manufactured by a company are good. The bulbs are supplied in packets of 5. In a consignment of 1000 packets, how many bulbs do you expect to be (a) All good (b) At-least 4 good (c) more than 3 good?

14. A brokerage survey reports that 30% of the individual investors have used a discount broker. In a random sample of 9 investors, what is the probability that
    i) Exactly two used a discount broker.
    ii) Not more than 3 used a discount broker.
    iii) At-least three of them used a discount broker?

15. Fit a Binomial Distribution (B.D) for the following data:

| X | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| f | 15 | 20 | 22 | 23 | 20 | 14 |

**Problems on Poisson distribution**:

16. It is known from past experience that in a certain plant, there are on an average 4 industrial accidents per month. Find the P that in a given year, there will be less than fouraccidents.

17. If 5% of the electric bulbs manufactured by a company are defective, use P.D to find P that in a sample of 100 bulbs, (a) None are defective (b) 5 bulbs will be defective.

18. A factory produces blades in packets of 10. The P of a blade to be defective is 0.2%. Find the number of packets having 2 defective blades in a consignment of 10000 packets.

19. The management of a photograph record company has discovered that the number of defects on record appears to follow P.D with mean equal to 0.4.
    (i) What is the P that a record selected at random will have 3 defects?
    (ii) If the management sets a policy that all photograph records sold to customers must not have any defects, what percentage of its records production will not be made available for sale because of defects?

20. Fit a Poisson Distribution (P.D) for the following data:

| X | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| f | 110 | 105 | 90 | 3 | 2 | 1 |

**Problems on Normal distribution**:

21. The weights of students are normally distributed with mean weight of 100 kg and S.D of 10 kg. What is the Percentage of students with weight:
    i) Between 100 and 110 kgs.
    ii) Above 110 kg.
    iii) Below 78 kgs.
    iv) 88.5 kgs to 117.5 kgs.

22. The daily wages of 1000 workmen are normally distributed around a mean of Rs. 70 with a standard deviation of Rs. 5. Estimate the number of workers whose daily wages are:
    i) Between Rs. 70 and Rs. 72.
    ii) Will not be between Rs. 69 and Rs. 72.
    iii) More than Rs. 75.
    iv) Less than Rs. 63.
    v) Between Rs. 65 and Rs. 68.
    vi) Also estimate the lowest wages of the 100 highest paid workers.

23. The life times of electronic devices have a mean life of 300 hours and a standard deviation of 25 hours. Assuming the distribution to be normal, calculate the following:
    i) The probability of life time more than 350 hours.
    ii) What percentage will have life time less than 300 hours?
    iii) What percentage will have life time between 220 hours to 260 hours?

24. In a Normal Distribution, 31% of the items are below 45 and 8% of the items are above 64. Find the mean and SD of the distribution.

25.
IQ of army volunteers in a given year are normally distributed with mean 110 and SD of 10. The army wants to give advanced training to 20% of those recruits with the highest score. What is the lowest IQ score acceptable for the advanced training?

26.
The daily wages of 1000 workmen are normally distributed with a mean of 70 and standard deviation of 5. Estimate the lowest daily wages of 100 highest paid workers.

## Miscellaneous Problems:

27.
Between hours 2 PM and 4 PM, the average number of phone calls per minute coming into the switch board of a company is 2.35. Find the probability that during one particular minute, there will be at-least 2 phone calls.

28.
An oil exploration firm plans to drill six holes. It is believed that the probability that hole will yield oil is 0.1. Since the holes are in quite different locations, the outcome of drilling one hole is statistically independent of that of drilling any of the other holes.

i) If the firm will able to stay in business only if two or more holes produce oil, what is the probability of its staying in business?
ii) Give the expected value of number of holes that result in oil.

29.
How many investigators would you expect to report that 3 people or less were vegetarian assuming that 100 investigators each take a sample of 10 individuals to see whether they are vegetarian?

30.
A hospital has 20 kidney dialysis machines and the chance of any of them malfunctioning during any day is 0.02. You are required to find the probability that exactly 3 machines will be out of service on the same day. Then,

i) Can we use binomial formula to find out the probability? If yes, calculate the probability.
ii) Can we use Poisson formula to find out this? If yes, calculate the probability.

31.
The life time (in hours) of a certain electrical equipment has the normal distribution with mean = 80 and standard deviation = 16.

i) What is the probability that the equipment lasts at least 100 hours?
ii) If the equipment has already lasted 88 hours, what is the conditional probability that it will last another 12 hours?

32.
A manufacturing firm produces pipes in two plants – I and II, the daily production being 1500 and 2000 pipes respectively. The fraction defectives of pipes being produced by plants I and II are 0.006 and 0.008 respectively. If a pipe selected at random from the day's production is found to be defective, what is the probability that it has come from plant I?

33.
Out of 320 families with 5 children each, what percentage would you expect to have at-least one boy?

34.
The average test marks in a class is 79. SD is 5. If the marks are distributed normally, how many students in a class of 200 did not receive marks between 75 and 82?

35.
A project yields an average cash flow of Rs. 500 lakhs with SD of 60 lakhs. Calculate the following probabilities:
i) Cash flow will be more than 560 lakhs.
ii) Cash flow will be less than 420 lakhs.

36.
If there are 200 typing mistakes randomly distributed in a 500 page book, find the probability that a given page has exactly 3 errors.

37.
Air Corporation Ltd had just 2 air crashes during its first 50 years of existence. The company wants to make the next decade air crash free. Assuming the same trend would continue, what is the probability that the company will meet its target?

38.
A merchant's file consisting of 20 accounts comprises of 6 processed and 14 unprocessed accounts. An auditor randomly selects five of these accounts for review. What is the probability that auditor finds exactly 2 processed accounts?

39.
Four cards are drawn at random from a pack of 52 cards. Find the probability that:
a. They are a king, queen, jack and an ace.
b. Two are kings are two are queens.
c. All are diamonds.

40.
6 dice are thrown 729 times. How many times do you expect at-least 3 dice to show a five or a six?

# STATISTICS FOR MANAGERS (24MBA15)

## MODULE 5
## TESTING OF HYPOTHESIS

### Contents of the Module:

- Hypothesis – Meaning, Types, Characteristics, Sources, Formulation & Errors.
- Parametric and Non Parametric tests – An overview.
- An overview of T-Test, Z-Test, F-Test, U-Test and K-W Test.
- Statistical Analysis: - Bivariate analysis & Multi-variate analysis.
- ANOVA – One way and two way classifications.

### HYPOTHESIS – MEANING & DEFINITIONS

**It is the presumptive statement of a proposition or a reasonable guess, based upon theavailable evidence, which the researcher seeks to prove through his study.**

Hypotheses reflect the research worker's guess as to the probable outcome of the experiments.

### EXAMPLES OF A HYPOTHESIS

1. Diet influences intelligence.
2. Attendance at lectures influences exam marks.
3. The average life times of the four brands of bulbs are the same.
4. Three fertilizers do not differ in yielding different productivity on same field.

### IMPORTANCE OF HYPOTHESIS

1. Hypotheses facilitate the extension of knowledge in an area. They provide tentative explanations of facts and phenomena, and can be tested and validated.
2. It sensitizes the investigator to certain aspects of situations which are relevant from the standpoint of the problem in hand.
3. Hypotheses provide the researcher with rational statements, consisting of elements expressed in a logical order of relationships which seek to describe or to explain conditions or events, that have not yet been confirmed by facts.
4. The hypotheses enable the researcher to relate logically known facts to intelligent guesses about unknown conditions. It is a guide to the thinking process and the process of discovery. It is the investigator's eye – a sort of guiding light in the work of darkness.
5. Hypotheses provide direction to the research. It defines what is relevant and what is not.

### SOURCES OF HYPOTHESIS:

- Review of similar studies in the area or of the studies on similar problems;
- Examination of data and records, if available, concerning the problem for possible trends, peculiarities and other clues;
- Discussions with colleagues and experts about the problem, its origin and the objectives in seeking a solution.
- Intuition is often considered a reasonable source of research hypotheses
- Prior empirical research findings are perhaps the most common source of new research hypotheses, especially when carefully combined using rational induction
- Thus hypothesis are formulated as a result of prior thinking about the subject, examination of the available data and material including related studies and the council of experts.

## CHARACTERISTICS OF A GOOD HYPOTHESIS:

1. **Clarity**: Hypothesis should be clear and precise. If the hypothesis is not clear and precise, the inferences drawn on its basis cannot be taken as reliable.
2. **Testability**: Hypothesis should be capable of being tested. Some prior study may be done by researcher in order to make hypothesis a testable one.
3. **Relevant to the Problem**: Hypothesis should state relationship between variables, if it happens to be a relational hypothesis.
4. **Specific:** Hypothesis should be limited in scope and must be specific. A researcher must remember that narrower hypotheses are generally more testable
5. **Simplicity**: Hypothesis should be stated as far as possible in most simple terms so that the same is easily understandable by all concerned.
6. **Closest to observable things**: Hypothesis should be consistent with most known facts i.e. it must be consistent with a substantial body of established facts.
7. **The hypotheses selected should be put to testing within a reasonable time**. The researcher should not select a problem which involves hypotheses that are not agreeable.
8. **Power of Prediction**: Hypothesis must explain the facts that gave rise to the need for explanation. It must actually explain what it claims to explain.
9. **Fruitful for new discoveries**: The hypothesis which is formulated must be able to draw insights into further research studies in the particular area.
10. **Relevant to available techniques**: The hypothesis set must be in accordance to the techniques (parametric / non parametric tests) available.

## TYPES OF HYPOTHESIS:

1. **Research hypothesis:** When a prediction or a hypothesized relationship is to be tested by scientific methods, it is termed as research hypothesis. The research hypothesis is a predictive statement that relates an independent variable to a dependent variable.
2. **Directional hypothesis:** The hypotheses which stipulate the direction of the expected differences or relationships are terms as directional hypotheses. For example, the research hypothesis: "Adolescent boys with high IQ will exhibit low anxiety than adolescent boys with low IQ" is a directional research hypothesis because it stipulates thedirection of the difference between groups.
3. **Non-directional hypothesis:** A research hypothesis which does not specify the direction of expected differences or relationships is a non-directional research hypothesis. For example, the hypotheses: "There will be difference in the adaptability of fathers and mothers towards rearing of their children" or "There is a difference in the anxiety level of adolescent girls of high IQ and low IQ" are non-directional research hypotheses.
4. **Statistical hypothesis:** When it is time to test whether the data support or refute the research hypothesis, it needs to be translated into a statistical hypothesis. Statistical hypotheses often are given in quantitative terms, for example: "The ***mean*** reading achievement of the population of third-grade students taught by Method A equals the ***mean*** reading achievement of the population taught by Method B." Therefore we can say that **statistical hypotheses** are, concerned with populations under study.
5. **Declarative hypothesis:** When the researcher makes a positive statement about the outcome of the study, the hypothesis takes the declarative form. For example, the hypothesis: "The academic achievement of extroverts is significantly higher than that of the introverts," is stated in the declarative form.

6. **Question form hypothesis:** In the question form hypothesis, a question is asked as to what the outcome will be instead of stating what outcome is expected. Suppose a researcher is interested in knowing whether programmed instruction has any relationship to test anxiety of children.

## BASIC TERMINOLOGIES:

### 1 NULL HYPOTHESIS – $H_0$

Null hypothesis is the statement about the population that is tested for possible rejection under the assumption that it is true. These are used when the researcher believes there is no relationship between two variables or when there is inadequate theoretical or empirical information to state a research hypothesis. Null hypothesis is a statement that reflects a status quo / no difference position – Why? Let us understand this logic through a **non statistical example** first.

E.g.: Suppose a person is facing a legal trial for committing a crime. The judge looks into all the evidence for and against it and then gives his verdict. Now the verdict can be:

a) That the person has not committed the crime.
b) That the person has committed the crime.

The point to be noted here is that the Judge would start off the hearings in the court by assuming that the person has not committed the crime because if he assumes that the person has committed the crime, there is no scope for further arguments to be heard. Thus statistically, point a) is referred to as Null hypothesis.

Hence while developing any hypothesis statement, care must be taken to ensure that the first statement that is framed (that is the null hypothesis) must be framed by defining the negative thought towards the outcome.

For example:

1. There is no difference between section A and section B in their test performance.
2. Diet does not influence intelligence.
3. The mean life times of all the four brands of bulbs are the same (i.e. No difference in their mean life times) etc.

Statistical example:
An automobile manufacturer claims that his battery will last on an average of at least 48 months. Now if there are some complaints by customers, the manufacturer's belief will have to be tested. Thus the statistical hypothesis to be tested here is called Null Hypothesis ($H_0$) which is written as follows:

$H_0$: $\mu \geq$ months $H_1$:
$\mu < 48$ months.

Thus the decision maker should always develop a null attitude towards the outcome of the sample.
[Note: If the Null Hypothesis gets rejected, the hypothesis that gets accepted is called **alternate hypothesis** – $H_1$].

**2 LEVEL OF SIGNIFICANCE ($\alpha$):** This is very important concept in the context of hypothesis testing. It is always some percentage (usually 5%) which should be chosen with great care, thought and reason. In case we take the significance level at 5 per cent, then this implies that $H_0$ will be rejected when the sampling result (i.e. observed evidence) has a less than 0.05 probability of occurring if $H_0$ is true. In other words, the 5 percent level of significance means that researcher is willing to take as much as a 5 percent risk of rejecting the null hypothesis when it ($H_0$) happens to be true. Thus the significance level is the maximum value of the probability of rejecting $H_0$ when it is true and is usually determined in advance before testing the hypothesis.

**3 TAILS OF A TEST:** In the context of hypothesis testing, these two terms are quite important and must be clearly understood. A two-tailed test rejects the null hypothesis if, say, the sample mean is significantly higher or lower than the hypothesized value of the mean of the population. Such a test is appropriate when the null hypothesis is some specified value and the alternative hypothesis is a value not equal to the specified value of the null hypothesis. In a two-tailed test, there are two rejection regions, one on each tail of the curve which can be illustrated as under:

But there are situations when only one-tailed test is considered appropriate. A one-tailed test would be used when we are to test, say, whether the population mean is either lower than or higher than some hypothesized value. We should always remember that accepting $H_0$, on the basis of sample information does not constitute the proof that $H_0$, is true. We only mean that there is no statistical evidence to reject it.

## STEPS IN HYPOTHESIS TESTING PROCEDURE:

1. **Formulate the Hypothesis:** Set up both the Null Hypothesis and Alternate Hypothesis.

2. **Set up a suitable level of Significance:** $\alpha$ = 1% or 5%

3. **Select the test criterion:** Define and compute the test statistic (type of test) to be used – Consider if it is parametric or non parametric after studying the nature of the research problem – for example (Parametric – T, Z, F, ANOVA etc; Non Parametric – Chi Square, U Test etc)

4. **Compute:**
   a. Calculated value: By using the formulae of the above tests.
   b. Tabulated Value: By looking into the corresponding values of the test tables.

5. **Make Decisions:** After comparing the calculated value with the tabulated value, a decision will have to be taken. The following decision rule could be made use of:
   a. If Calculated Value is < than Tabulated value, accept Null Hypothesis.
   b. If Calculated Value is > than Tabulated value, reject Null hypothesis and accept alternate hypothesis.

6. **Conclusion:** Write the conclusion of the test in simple language to draw the inference about the research problem being addressed through the hypothesis.

## HYPOTHESIS TESTS

**[**Note: By Parameter we mean, the measurable characteristic of the population such as a mean or a standard deviation**]**

### PARAMETRIC TESTS:

Parametric statistics is a branch of statistics that assumes that the data has come from a type of a probability distribution and makes inferences about the parameters of the distribution. Put in simple words, parametric tests are made use of to test a hypothesis if the nature of data is quantifiable in the general sense. They are said to have more statistical power.
**Common Parametric Tests are**: T, Z, F, ANOVA etc.

### NON PARAMETRIC TESTS:

Non parametric tests are distribution free methods which do not rely on the assumption that the data are drawn from a given probability distribution. Put in simple words, if the nature of the research data collected is of a non quantifiable nature, then non parametric tests are preferred. They are widely used for studying populations that take on a ranked order (such as movie reviews, beauty, intelligence etc). The use of these tests may be necessary when data has a ranking but no clear numerical interpretation (such as assessing preferences). As non parametric methods make fewer assumptions, their applicability is much wider than the corresponding parametric methods.
**Common non parametric tests are**: Chi-square, U test, K-W test, Smirnov test etc.

| | **Parametric** | **Non-parametric** |
|---|---|---|
| Assumed distribution | Normal | Any |
| Typical data | Ratio or Interval | Ordinal or Nominal |
| Data set relationships | Independent | Any |
| Usual central measure | Mean | Median |
| Benefits | Can draw more conclusions | Simplicity; Less affected by outliers |
| **Tests** | | |
| Choosing | Choosing parametric test | Choosing a non-parametric test |
| Correlation test | Pearson | Spearman |
| Independent measures, 2 groups | Independent-measures t-test | Mann-Whitney test |
| Independent measures, >2 groups | One-way, independent-measures ANOVA | Kruskal-Wallis test |

## TYPES OF PARAMETRIC TESTS1

### Z – TEST:

A ***Z*-test** is applied in cases where the sample size of the research is more than 30. The basic purpose / approach of a Z test is to establish the extent to which the sample mean or standard deviation differs significantly from the population mean or standard deviation. For each significance level, the *Z*-test has a single critical value (for example, 1.96 for 5% two tailed) which makes it more convenient than the Student's *t*-test which has separate critical values for each sample size. Therefore, many statistical tests can be conveniently performed as approximate *Z*-tests if the sample size is large or the population variance known.

### 2 T –TEST:

A ***t*-test** is any statistical hypothesis test in which the test statistic followsa Student's *t* distribution if the null hypothesis is supported. It is made use of in cases where the sample size is less than 30. It can be used to determine if two sets of data are significantly different from each other, and is most commonly applied when the test statistic would followa normal distribution if the value of a scaling term in the test statistic were known.

#### Independent (unpaired) samples

The independent samples *t*-test is used when two separate sets of independent and identically distributed samples are obtained, one from each of the two populations being compared. For example, suppose we are evaluating the effect of a medical treatment and we enroll 100 subjects into our study, and then randomly assign 50 subjects to the treatment group and 50 subjects to the control group. In this case, we have two independent samples and would use the unpaired form of the *t*-test.

#### Paired samples

Paired samples *t*-tests typically consist of a sample of matched pairs of similar units, or one group of units that has been tested twice (a "repeated measures" *t*-test). A typical example of the repeated measures *t*-test would be where subjects are tested prior to a treatment, say for high blood pressure, and the same subjects are tested again after treatment with a blood-pressure lowering medication. By comparing the same patient's numbers before and after treatment, weare effectively using each patient as their own control.

### 3 F – TEST:

At times, it is necessary to verify whether two independent populations have the same variability. For this purpose, we use the F test which measures the ratio of two sample variances. Suppose a company manufacturing light bulbs is using two different processes A and B. The life of the light bulbs of process A and process B with mean, number and SD are known. If the researcher wants to test the hypothesis that process A is more stable than process B or if he wants to test if there is any difference in the variability between both the processes, then the appropriate test to be used is the F test.

### 4 ANOVA (Analysis of Variance):

The previously discussed tests namely Z, T or F would be ideal if an inference is to be drawn in the research involving at the most two samples. But there could be a large number of cases where

the research study could be undertaken for more than two sample units. For example, if the

researcher is interested in finding out if the five machines used for producing a similar good have the same productivity or for example if the prices of a product is similar in five different cities etc. In such cases, a complex method for analyzing both the mean and standard deviations (variance) is called for. ANOVA helps achieve this objective.

***Meaning of ANOVA***: It is a statistical procedure for determining whether the means of several populations are equal or not.

**Assumptions of ANOVA:**

1. It is assumed that the universe (population) from which samples is drawn for the study is normally distributed.
2. It is assumed that the samples under study have been drawn at random and they are independent of each other.
3. Samples are homogeneous.
4. Various treatment and environmental factors are additive in nature.

**ONE-WAY ANOVA:** In simple words, One way ANOVA basically tries to study the hypothesis that only one criterion or a single independent variable is ascribed (used to explain) the effects on a dependent variable. For example if the marks of students of five different sections of a particular grade are known and if the researcher wishes to study if the mean marks of students in these five sections differ significantly or not, then this classification is referred toas one way ANOVA. Here the independent variable is the sections and the dependent variable is the marks secured by students.

**TWO-WAY ANOVA:** It tries to study the hypothesis that more than one (usually two) independent variables is used to explain the effects on a dependent variable. Continuing with the example given for one way ANOVA, now if the researcher wishes to test if the marks of the students is dependent on the type of instruction given (say in house vs. out house) and also on the number of hours of instruction given (say 1 hour and 4 hours), he can do so by using the technique of two way ANOVA as it tries to explain the interaction effects between the two independent variables namely type of instruction and hours of instruction on the dependent variable which is marks secured by the students.

**Applications of ANOVA:**

1. The origin of ANOVA technique lies in agricultural experiments – for example, if you have to test the difference between five plots of land and four types of fertilizers affecting the yield of a particular crop, you can use the two way ANOVA technique.
2. It is also put to use in different fields such as education, psychology, business etc.
3. In addition to testing the homogeneity of several sample means, ANOVA technique is now frequently applied in testing the linearity of the fitted regression line or the significance of the correlation ratio.

[**Degrees of Freedom** – By this we mean the number of observations which can freely estimate to get a set of expected data that satisfies relevant constraints – namely total of frequency, mean, SD etc. It implies the extent of independence (freedom) given by a set of observed frequencies]

## TYPES OF NON PARAMETRIC TESTS:

### 1 CHI SQUARE TEST:

As explained earlier, when the data is of a non quantifiable nature, the hypothesis test the researcher seeks to apply to establish relationships and thus seek to understand the difference between variables under consideration is referred to as a Chi Square test. Chi-square is the sum of the squared difference between observed (*o*) and the expected (*e*) data (or the deviation, *d*), divided by the expected data in all possible categories.

**The following scenarios demand an application of a Chi square test**:

1. To ascertain the association between ownership of computers and income levels.
2. To test the hypothesis that educated fathers have intelligent sons.
3. To test the hypothesis that the number of times a person shops per month at a grocery store is related to their satisfaction with the purchase.
4. To test the association between age group and liking towards a particular brand of car etc.

### 2 MANN –WHITNEY U TEST:

This test is used to compare the differences between two independent groups when the dependent variable is ordinal in nature.

**For example** – You could use U test to understand whether attitudes towards pay discrimination differ based on gender (Male & Female). Here attitudes are measured on an ordinal scale. While gender is the independent variable, an attitude is the dependent variable.

Hence in order to test the equality between the two populations / samples – that is trying to test the hypothesis that attitudes of men and women do not differ toward pay discrimination, U test.

Thus U test is considered the non parametric alternative to t or z test. Hence to test the equality of two population means which is ordinal in nature, U test is made use of.

### 3 KRUSKAR – WALLIS (KW) TEST:

K-W Test (One way ANOVA on ranks) is a rank based non parametric test that can be used to determine if there are statistically significant differences between two or more groups of an independent variable on an ordinal dependent variable. It is an extension of the U test to allow comparison of more than two groups.

**For example**: You could use a K-W test to understand whether exam performance measured on a continuous/ordinal scale from 0-100 differed based on test anxiety levels. Three categories of levels being – Low, Medium and High test anxiety levels. [Dependent variable – Exam performance; Independent variable – test anxiety levels]

## ERRORS IN HYPOTHESIS TESTING:

**Type I and Type II errors**: in the context of testing of hypotheses, there are basically two types of errors we can make. We may reject $H_0$ when $H_0$ is true and we may accept $H_0$ when in fact $H_0$ is not true. The former is known as Type I error and the latter as Type II error.

**In other words, Type I error means rejection of hypothesis which should have been accepted**

**And**

**Type II error means accepting the hypothesis which should have been rejected.**

Type I error is denoted by (alpha) known as (alpha) error, also called the level of significance of test; and Type II error is denoted by ß (beta) known as ß error. In a tabular form the said two errors can be presented as follows:

| | **Decision** | |
|---|---|---|
| | Accept $H_0$ | Reject $H_0$ |
| $H_0$ (true) | Correct decision | Type I error (alpha error) |
| $H_0$ (false) | Type II error (ß error) | Correct decision |

The probability of Type I error is usually determined in advance and is understood as the level of significance of testing the hypothesis. If type I error is fixed at 5 per cent, it means that there are about 5 chances in 100 that we will reject $H_0$ when $H_0$ is true. We can control Type I error just by fixing it at a lower level. For instance, if we fix it at 1 per cent, we will say that the maximum probability of committing Type I error would only be 0.01.

But with a fixed sample size, n, when we try to reduce Type I error, the probability of committing Type II error increases. Both types of errors cannot be reduced simultaneously.

There is a trade-off between two types of errors which means that the probability of making one type of error can only be reduced if we are willing to increase the probability of making the other type of error.

## BIVARIATE ANALYSIS

**Bivariate analysis** is one of the simplest forms of quantitative (statistical) analysis. It involves the analysis of two variables (often denoted as *X*, *Y*), for the purpose of determining the empirical relationship between them. In order to see if the variables are related to one another, it is common to measure how those two variables simultaneously change together.

**Types**: Common forms of bivariate analysis involve creating a percentage table or a scatter plot graph and computing a simple correlation coefficient. The types of analysis that are suited to particular pairs of variables vary in accordance with the level of measurement of the variables of interest (e.g. nominal/categorical, ordinal, interval/ratio).

## MULTI-VARIATE ANALYSIS:

**Multivariate analysis** (**MVA**) is based on the statistical principle of multivariate statistics, which involves observation and analysis of more than one statistical outcome variable at a time. In design and analysis, the technique is used to perform trade studies across multiple dimensions while taking into account the effects of all variables on the responses of interest.

### Techniques /Methods of Multi-variate analysis:

1. Cluster analysis.
2. Factor analysis.
3. Discriminant analysis.
4. Multi Dimensional Scaling.
5. Conjoint analysis.
6. Multiple analysis of variance (MANOVA)

## NUMERICAL PROBLEMS ON TESTING OF HYPOTHESISZ TEST:

### Z test for Single Proportion:

1. In a big city, 325 out of 600 men were found to be smokers. Does this information support the conclusion that majority of men in this city are smokers?

### Z test for Difference of Proportions:

2. In a random sample of 1000 persons from town A, 400 are found to be consumers of wheat. In a sample of 800 from town B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between town A and town B so far as the population of wheat consumers is concerned? [Assume $\alpha = 1\%$]

3. A machine produced 20 defective articles in a batch of 400. After overhauling, it produced 10 defectives in a batch of 300. Has the machine improved? [Assume $\alpha = 5\%$]

### Z test for Single Mean:

4. A stenographer claims that she can take dictation at 120 words per minute. Can we reject her claim on the basis of 100 trials in which she demonstrates a mean of 116 words with a standard deviation of 15 words? [Assume $\alpha = 5\%$]

5. A company produces tyres. The tyre life is normally distributed. The mean life is 40000 km and SD is 3000 km. A new process is introduced to produce better tyres. A sample of 100 tyres is taken with mean life of 40900 km. is the new tyre significantly better than the old tyre? [Assume $\alpha = 1\%$]

### Z test for Difference of Means:

6. An investigation of relative merits of 2 kinds of flashlight batteries showed that a random sample of 100 batteries of brand A lasted on average 36.5 hours with a SD of 1.8 hours. A second sample of 80 batteries of brand B lasted on average 36.8 hours with SD of 1.5 hours. Use % Level of significance to test whether the observed difference between average life times of both brands is significant.

7. Intelligence test on 2 groups of boys and girls gave the following results:

| | Mean | SD | Number |
|---|---|---|---|
| Girls | 75 | 15 | 150 |
| Boys | 70 | 20 | 250 |

Is the difference in mean scores of boys and girls statistically different? Test at 1% Level of Significance.

### Z test for Difference of Standard Deviations:

8. The mean yield of 2 sets of plots and their variability are as given below. Examine whether the difference in the variability in yields is significant at $\alpha$=5%

| | Set of 40 plots | Set of 60 plots |
|---|---|---|
| Mean yield per plot | 1258 lb | 1243 lb |
| SD per plot | 34 | 28 |

## T TEST:

### T test for Single Mean:

9. The mean weekly sales of chocolate bar in candy stores were 146.3 bars per store. After an ad campaign, the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed SD of 17.2. Was the ad campaign successful?

### T test for Difference of Means (Independent Samples):

10. A random sample of 20 daily workers of State A were found to have average daily earning of Rs. 44 and sample variance of 900. Another sample of 20 daily workers from State B was found to earn on average Rs. 30 per day with variance of 400. Test whether workers in State A are earning more than those in State B.

11. A group of 5 patients treated with medicine A weigh 42 39 48 60 & 41 kg. A second group of 7 patients from the same hospital treated with medicine B weigh 38 42 56 64 68 69 & 62 kg. Do you agree with the claim that medicine B increases weight significantly?

### T test for Difference of Means (Paired Samples):

12. The following are the weights of ten persons before and after they underwent a weight reduction programme. Can we conclude that weights were reduced after persons underwent weight reduction programme? (Take $\alpha = 5\%$)

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Before (Kg) | 86 | 92 | 100 | 93 | 88 | 80 | 88 | 92 | 95 | 106 |
| After (Kg) | 77 | 84 | 92 | 87 | 80 | 74 | 80 | 85 | 95 | 96 |

13. The following are the BP's of 8 persons before and after meditation. Can we conclude that meditation reduces BP? ( Given for df $t_{0.05}$=1.90)

| Person | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Before meditation | 92 | 90 | 86 | 92 | 88 | 94 | 90 | 90 |
| After meditation | 86 | 88 | 80 | 86 | 86 | 84 | 84 | 90 |

## F TEST:

14. Suppose a company manufacturing light bulbs is using two different processes A and B. The data pertaining to two processes are given below. Test the hypothesis that the variability of the two processes is the same or not.
    Sample A: $n_1 = 16$, Mean = 1200 hrs SD = 60 hrs
    Sample B: $n_2 = 21$, Mean = 1300 hrs SD = 50 hrs

15. The following figures relate to the number of units of an item produced per shift by 2 workers A and B. Can it be inferred that worker A is more stable than worker B?

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Worker A | 19 | 22 | 24 | 27 | 24 | 18 | 20 | 19 | 25 | | |
| Worker B | 26 | 37 | 40 | 35 | 30 | 30 | 40 | 26 | 30 | 35 | 45 |

## CHI SQUARE TEST:

16. In a survey of 200 boys of which 75 were intelligent, 40 had educated fathers; while 85 of the unintelligent boys had uneducated fathers. Do these figures support the hypothesisthat educated fathers have intelligent boys?

17. An automobile company gives you the following information about age groups and liking for a particular model of car that it plans to launch:

| Persons | Below 25 | 25 – 30 | Above 30 |
|---|---|---|---|
| Liked the car | 45 | 30 | 25 |
| Disliked the car | 55 | 20 | 25 |

On the basis of the above data can it be concluded that the model appeal is independent of age group?

## U TEST (MANN WHITNEY U TEST):

18. It is generally believed that as people grow older, they find it harder to go to sleep. To test if there was a difference in time in minutes before people actually went after lying in the bed, a sample of 10 young persons (aged 21 to 25) and 10 old persons (aged 65 to 70)was randomly selected and their sleeping habits were monitored. The data below shows the number of minutes these 20 persons were awake in bed before going to sleep:

| Young Men | 58 | 42 | 68 | 20 | 15 | 35 | 26 | 40 | 47 | 28 |
|---|---|---|---|---|---|---|---|---|---|---|
| Old Men | 100 | 152 | 147 | 70 | 40 | 95 | 68 | 90 | 112 | 58 |

Is there evidence that young men significantly take more time to get to sleep than old men?

## KRUSKAL WALLIS (K-W) TEST:

19. Use K-W test to find whether there is a significant difference in the following populations. [Use 5% α]

| Population 1 | 17 | 19 | 27 | 20 | 35 | 40 | |
|---|---|---|---|---|---|---|---|
| Population 2 | 28 | 36 | 33 | 22 | 27 | | |
| Population 3 | 37 | 30 | 39 | 42 | 28 | 25 | 31 |

## ANOVA (Analysis of Variance)

### One Way ANOVA:

20. The lifetime of 3 electric bulbs of 4 brands measured in 100 hours are presented below. Perform One – Way ANOVA to test the hypothesis that mean life times of four brands of bulbs are the same:

**BRANDS**

| **A** | **B** | **C** | **D** |
|---|---|---|---|
| 20 | 25 | 24 | 23 |
| 19 | 23 | 20 | 20 |
| 21 | 21 | 22 | 20 |

21. A truck company wishes to test the average life of each of the 4 brands of tyres. The company uses all brands on randomly selected trucks. The records showing the lives (in '000s) of miles of tyres are shown in the table

| Brand I | Brand II | Brand III | Brand IV |
|---|---|---|---|
| 20 | 19 | 21 | 15 |

| | | | |
|---|---|---|---|
| 23 | 15 | 19 | 17 |
| 18 | 17 | 20 | 16 |
| 17 | 20 | 17 | 18 |
| | 16 | 16 | |

Test the H that average life for each brand of tyres is the same @ 1%

22. To test the size of variation in the retail prices of a commodity in 3cities- Mumbai, Kolkata, Delhi, 4 shops were chosen at random in each city and prices were as follows

| | | | | |
|---|---|---|---|---|
| Mumbai | 16 | 8 | 12 | 14 |
| Kolkata | 14 | 10 | 10 | 6 |
| Delhi | 4 | 10 | 8 | 8 |

Do the data indicate that the price in 3cities are significantly different @5%

Two way Anova

23. Five doctors each test five treatments for a certain disease and observe the no of days each patient takes to recover. Discuss the differences between
A) The Doctors B) the treatments

| | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| | 10 | 14 | 23 | 18 | 20 |
| | 11 | 15 | 24 | 17 | 21 |
| | 9 | 12 | 20 | 16 | 19 |
| | 8 | 13 | 17 | 17 | 20 |
| | 13 | 15 | 19 | 15 | 22 |

* * * * * *

**RNSIT – MBA**

**Subject Viva – Business Statistics & Analytics (18MBA14)**

**Q & A**

1. **Define statistics.**
   Statistics is a science which deals with the methods of collecting, classifying, presenting, comparing and interpreting the numerical data collected to throw some light on any sphere of enquiry

2. **State the stages in statistical investigation.**

A. Collection – sources of data (primary/secondary)
B. Organization – editing, classification, tabulation.
C. Presentation – for analysis- through diagrams & graphs.
D. Analysis – done to dig out information useful for decision making.
E. Interpretation – drawing conclusions- if not done, the objective of investigation is defeated.

3. **Define population.**
   A population is the total set of individuals, groups, objects, or events that the researcher is studying.

4. **Define sample.**
   A sample is a relatively small subset of people, objects, groups, or events that is selected from the population.

5. **What is descriptive statistics?**
   Descriptive statistics is the term given to the analysis of data that helps describe, show or summarize data in a meaningful way such that patterns might emerge from the data.

6. **What are the three ways of classifying quantitative data?**
   a. Raw Data.
   b. Discrete Data.
   c. Continuous Data.

7. **What is inferential statistics?**
   Inferential statistics makes inferences about populations using data drawn from the population.

8. **Define discrete variable.**
   Variables that cannot take all possible values (integral, fractional etc)
   Examples: No of children in a family, No of students in a class, no of accidents on a road etc

9. **Define continuous variable.**
   Variables that can take all possible values (decimal, fractional etc)
   Examples: BP, height, weight, Marks, Age etc

10. **What are the types of class intervals?**
    a. Inclusive type: 30 – 39, 40 – 49, 50 – 59 etc. [Both upper & lower limits are included in the class]
    b. Exclusive type: 10 – 20, 20 – 30, 30 – 40 etc. [Upper limit of the class is excluded from that class]
    c. Open ended: Less than 10, 10 – 20, 20 – 30, 30 & above etc.

11. **Define central tendency.**
    A measure of central tendency is a single value that attempts to describe a set of data by identifying the central position within that set of data.

12. **List the measures of central tendency.**

a. Arithmetic Mean (Mean)
b. Median.
c. Mode.
d. Geometric Mean.
e. Harmonic Mean.

13. **Define Mean.**
    The mean is equal to the sum of all the values in the data set divided by the number of values in the data set.

**14. Define Median.**
The median is the middle score for a set of data that has been arranged in order of magnitude

**15. Define Mode.**
Mode is the most common number – the one that appears the maximum number of times in a given data set.

**16. State the empirical relationship between mean, median and mode.**
Mode = 3 Median – 2 Mean

**17. Why is mean considered the ideal/best measure of central tendency?**
a. It is simple and easy to understand.
b. It considers all the values in the given data.
c. It is capable of further mathematical treatment. (like finding combined mean etc)

**18. Define dispersion.**
Dispersion is the measure of variation of the items.

**19. State the various measures of dispersion.**
a. Range.
b. Quartile deviation (semi inter quartile range)
c. Mean deviation.
d. Standard deviation.

**20. Define Range.**
It is defined as the diff b/w 2 extreme values of the distribution... (Difference b w the maximum and minimum value of a distribution)

**21. What are quartiles?**
Quartiles divide the data into four parts.

**22. What is Quartile Deviation?**
- ✓ It is a measure of dispersion based on upper quartile Q3 and lower quartile Q1.
- ✓ QD is obtained by dividing the inter quartile range with 2

**23. Define mean deviation.**
It is the amount of scatter of the items in a distribution from either the mean or the median, ignoring the signs of distribution.

**24. What is Standard Deviation?**
It is the square root of the difference between arithmetic mean and sum of squares of the deviations from the arithmetic mean.

**25. Differentiate between absolute and relative variation.**
Absolute measures – the measures of dispersion which are expressed in terms of original units of a series are termed as absolute measures; they are not suitable for comparing the variability of two distributions
Relative measures – they are obtained as ratios or % ...for comparing variability of the two distributions; for example coefficient of Range/QD/MD, Coefficient of Variation etc.

**26. When do we use the empirical relationship formula while calculating mode?**
a. If the highest frequency is repeated twice.
b. If highest frequency appears in the starting or at the ending of the given data.

**27. What is CV?**
Coefficient of Variation = Mean / SD expressed as a %

**28. Define correlation.**
Correlation analysis attempts to determine the degree of relationship b/w variables.

**29. Define positive correlation.**
If two variables change in the same direction (i.e. if one increases the other also increases, or if one decreases, the other also decreases), then this is called a positive correlation. For example. the training and performance of employees in a company; family income & expenditure on luxuries; price & supply etc.

**30. Define negative correlation.**
In negative correlation, the two variables move in the opposite direction. When the value of a variable increases, the value of the other variable decreases. For example, the relationship between price and demand; volume & pressure of gas etc; T.V registrations and cinema attendance.

**31. Define linear correlation.**
The correlation b/w two variables is said to be linear if corresponding to a unit change in one variable, there is a constant change in other variable over entire range of data:

**32. Define non linear correlation.**
Correlation is said to be non linear when the two variables do not move in a linear form (for e.g. – just because the amount of rainfall is doubled, we can't expect the yield of food crops to be doubled as well)

**33. Define zero correlation.**
When the two variables are independent and change in one variable has no effect in other variable, then the correlation between these two variables is known as Zero Correlation. No correlation occurs when there is no linear dependency between the variables. Example. Age of your father and your shoe size.

**34. State the uses of correlation analysis.**
a. To measure degree of relationship b/w variables.
b. Correlation analysis contributes to understanding of economic behavior
c. In business, correlation analysis enables the executives to estimate cost, sales, prices on basis of some series which are functionally related.

**35. What is coefficient of determination?**
Square of correlation coefficient ($r^2$) used to judge the strength of the value of r.

**36. What are the methods of studying correlation?**
a. Scatter Diagram method.
b. Karl Pearson s coefficient of correlation
c. Rank correlation method.
d. Concurrent deviation method.

**37. Define regression.**
Regression analysis in the general sense means the estimation or prediction of the unknown value of one variable from the known value of the other variable.

**38. Define dependent and independent variables.**
In regression analysis, there are two types of variables —the variable whose value is to be influenced or predicted called the *dependent variable* & the variable which influences the value called the *independent variable*.

**39. Differentiate between simple regression and multiple regression.**
Usually, more than one independent variable influences the dependent variable.
When one independent variable is used in a regression, it is called a simple regression; when two or more independent variables are used, it is called a multiple regression.

**40. Define probability.**
The term "Probability" in statistics refers to the chance of occurrence of an event among a large number of possibilities. [It can also be referred to as a numerical measure of uncertainty]

**41. Define random experiment.**
It is an experiment which gives different outcomes when repeated under identical conditions. For example – tossing a coin, rolling a dice etc.

**42. Define sample space.**

It is a set of all the outcomes of a random experiment. For example – In tossing two coins, the sam space (S) = {HH, TT, HT, TH}. In rolling a dice, the sample space (S) = {1, 2, 3, 4, 5, 6} etc.

**43. Define mutually exclusive events.**

The events are said to be mutually exclusive when the occurrence of one event will prevent the occurrence of another in same trial.

For example: If the experiment is tossing two coins, then S = {HH, TT, HT, TH}.

A: Want 2 heads – A = {HH}

B: Want 2 tails – B = {TT}

Here A and B are mutually exclusive events.

**44. Define Independent events.**

Events are said to be independent if the occurrence of one does not affect or depend on the occurrence of other event in a series of trials.

**45. Define Dependent events.**

Events are said to be dependent if the occurrence of one is dependent on the occurrence of another event.

**46. Define equally likely events.**

Events are said to be equally likely if they have equal chances of occurrence.

For example – In throw of a single die,

A: Want only even numbers. Hence A = (2, 4, 6)

B: Want only odd numbers. Hence B = (1, 3, 5). Thus A and B are equally likely events.

**47. What is Baye's theorem?**

One of the important applications of conditional probability is in the computation of unknown probabilities on the basis of information provided by the past records or experiments. Then the conditional probability that it has occurred due to one of the events is called inverse probability. These probabilities are computed through the Bayes' rule or Bayes' theorem.

**48. What are the three probability distributions?**

a) Binomial.
b) Poisson.
c) Normal.

**49. What are the conditions for Binomial distribution?**

a. The number of trials "n" is finite (lesser).
b. The probability of happening of an event is fairly large.

**50. What are the conditions for Poisson distribution?**

a. The number of trials "n" is large.
b. The probability of success "p" is indefinitely small.

* * * * * *

**RNSIT – MBA**

SUBJECT VIVA QUESTIONS

**BUSINESS STATISTICS**

**(24MBA15)**

**1.** Define statistics.
**2.** What are the five stages in a statistical investigation?
**3.** What are the functions of statistics?
**4.** What are the limitations of statistics?
**5.** Differentiate between discrete variable and continuous variable.
**6.** Define central tendency.
**7.** What are the measures of central tendency?
**8.** Define mean.
**9.** Define median.
**10.** Define mode.
**11.** What is harmonic mean?
**12.** Define quartiles
**13.** What is a percentile?
**14.** Define Dispersion.
**15.** What are the measures of dispersion?
**16.** Define range.
**17.** What is coefficient of variation (CV)?
**18.** What is descriptive statistics?
**19.** What is inferential statistics?
**20.** What is a Decile?
**21.** Define correlation.
**22.** What is the use of correlation analysis?
**23.** Define regression.
**24.** List the importance of regression analysis.
**25.** Differentiate between positive and negative correlation.
**26.** What are the types of correlation?
**27.** Differentiate between simple and multiple regression.
**28.** What is time series?
**29.** State the 4 components of a time series.
**30.** What is a secular trend?
**31.** What are seasonal variations?
**32.** What are cyclical variations?
**33.** What are irregular variations?
**34.** What are the methods of measuring trend?
**35.** List the objectives of time series analysis.
**36.** Define probability
**37.** What is random experiment?
**38.** What is a sample space?
**39.** What is Bayes' theorem?
**40.** Name the 3 major probability distributions.
**41.** State the two conditions necessary for binomial distribution.
**42.** State the two conditions necessary for poisson distribution.
**43.** What is hypothesis?
**44.** What is null hypothesis?
**45.** What is alternate hypothesis?
**46.** What are the steps in testing a hypothesis?
**47.** Differentiate between Type I error and Type II error.
**48.** Differentiate between parametric tests and non parametric tests.
**49.** Mention the common parametric tests and non parametric tests.
**50.** Differentiate between one way ANOVA and two way ANOVA.

******

**ASSIGNMENT 1**

1. The mean salary paid to 100 employees of a company was Rs. 5000. The mean salary paid to male and female workers are Rs. 5200 and Rs. 4200 respectively. Determine the percentage of male and female workers employed in the factory.

2. Calculate the Mode value pertaining to dividend paid by companies in a particular financial year:

| Dividend (in %) | 5-7.5 | 7.5-10 | 10-12.5 | 12.5-15 | 15-17.5 | 17.5-20 | 20-22.5 | 22.5-25 |
|---|---|---|---|---|---|---|---|---|
| No of companies | 182 | 75 | 59 | 127 | 280 | 236 | 378 | 331 |

3. The length of time taken by each of the 18 workers to complete a specific job was observed as follows:

| Time in (mins) | 5-9 | 10-14 | 15-19 | 20-24 | 25-29 |
|---|---|---|---|---|---|
| No of workers | 3 | 8 | 4 | 2 | 1 |

Calculate the median time, $Q_1$ and $Q_3$.

4. Given N = 685 and Median = 42.6, find the missing frequencies:

| Class | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 | 60-70 | 70-80 |
|---|---|---|---|---|---|---|---|
| Frequency | 185 | - | 34 | 180 | 136 | - | 50 |

5. Calculate the mean, median and mode for the following data pertaining to the marks in statistics.

| Marks (more than) | 0 | 20 | 40 | 60 | 80 | 100 | 120 |
|---|---|---|---|---|---|---|---|
| No of students | 80 | 76 | 50 | 28 | 18 | 9 | 3 |

6. The number of cars sold by each of the 10 car dealers during a particular month is as follows: 17, 14, 12, 20, 20, 20, 22, 24, 22, 25. Determine mean, median, mode, $Q_1$ and $P_{50}$.

7. The arithmetic mean of height of 50 students in a class is 5'8". The height of 30 of these is given in the frequency distribution below. Find the mean height of the remaining 20 students.

| Height (in inches) | 5'4 | 5'6 | 5'8 | 5'10 | 6'0 |
|---|---|---|---|---|---|
| No of students | 4 | 12 | 4 | 8 | 2 |

8. Use an appropriate measure to evaluate the variation in the following data:

| Farm size (acre) | Less than 40 | 41-80 | 81-120 | 121-160 | 161-200 | 201-24 | 241 and above |
|---|---|---|---|---|---|---|---|
| No of farms | 394 | 461 | 391 | 334 | 169 | 113 | 148 |

9. The following sample shows the weekly number of road accidents in a city during the two year period:

| No of accidents | 0-4 | 5-9 | 10-14 | 15-19 | 20-24 | 25-29 | 30-34 | 35-39 | 40-44 |
|---|---|---|---|---|---|---|---|---|---|
| f | 5 | 12 | 32 | 27 | 11 | 9 | 4 | 3 | 1 |

Find the inter quartile range and coefficient of Q.D.

10. The number of patients seen in an emergency ward of a hospital for a sample of 5 days in the last month was – 153, 147, 151, 156 and 153. Determine the mean deviation and interpret.

11. Find the mean absolute deviation and its coefficient for the following frequency distribution of sales (Rs. In'000s) of a cooperative store:

| Sales | 50-100 | 100-150 | 150-200 | 200-250 | 250-300 | 300-350 |
|---|---|---|---|---|---|---|
| No of days | 11 | 23 | 44 | 19 | 8 | 7 |

12. A study of 100 engineering companies gives the following information. Calculate SD of profit earned.

| Profit (in crores) | 0-10 | 10-20 | 20-30 | 30-40 | 40-50 | 50-60 |
|---|---|---|---|---|---|---|
| No of companies | 8 | 12 | 20 | 30 | 20 | 10 |

13. The weekly sales of two products A and B are given below. Find which shows greater sale fluctuation.

| Product A | 59 | 75 | 27 | 63 | 27 | 28 | 56 |
|---|---|---|---|---|---|---|---|
| Product B | 150 | 200 | 125 | 310 | 330 | 250 | 225 |

14. From the analysis of monthly wages paid to employees in two service organizations X and Y, the following results were obtained:

| | Organization X | Organization Y |
|---|---|---|
| No of wage earners | 550 | 650 |
| Average monthly wages | 5000 | 4500 |
| Variance of distribution of wages | 900 | 1600 |

i) Which organization pays a larger amount as monthly wages?
ii) Find out which organization is more consistent in payment of wages.
iii) Find the combined SD of X and Y together.

15. Two automatic filling machines A and B are used to fill tea in 500g cartons. A random sample of 100 cartons on each machine showed the following:

| Tea contents | Machine A | Machine B |
|---|---|---|
| 485-490 | 12 | 10 |
| 490-495 | 18 | 15 |
| 495-500 | 20 | 24 |
| 500-505 | 22 | 20 |
| 505-510 | 24 | 18 |
| 510-515 | 4 | 13 |

Comment on the performance of two machines based on average filling and dispersion.

16. Compute median from the following data:

| Mid value | 115 | 125 | 135 | 145 | 155 | 165 | 175 | 185 | 195 |
|---|---|---|---|---|---|---|---|---|---|
| f | 6 | 25 | 48 | 72 | 116 | 60 | 38 | 22 | 3 |

17. In a moderately asymmetrical distribution, the values of mode and median are 20 and 24 respectively. Find the value of mean.

18. Calculate mean, SD, $D_4$ and $P_{23}$ from the following data:

| Value | 90-99 | 80-89 | 70-79 | 60-69 | 50-59 | 40-49 | 30-39 |
|---|---|---|---|---|---|---|---|
| f | 2 | 12 | 22 | 20 | 14 | 4 | 1 |

19. Given N = 10, Mean = 12 and $\sum x^2 = 1530$, find S.D

20. The refrigerators of model A have an average life of 5.12 and standard deviation of 2.8 years. The life (in years) and number of refrigerators of model B are given as follows:

| Life [No of years] | 0-2 | 2-4 | 4-6 | 6-8 | 8-10 | 10-12 |
|---|---|---|---|---|---|---|
| No of refrigerators [model B] | 2 | 7 | 12 | 19 | 9 | 1 |

Which model of refrigerator has greater uniformity?

**ASSIGNMENT 2**

1. From a sales force of 150 persons, one will be selected to attend a special sales meeting. If 52 o them are unmarried, 72 are college graduates and 3/4$^{th}$ of the 52 that are unmarried are college graduates, find the probability that the sales person selected at random will be neither single nor a college graduate.

2. Find Karl Pearson's correlation coefficient between size and defect in quality.

| Size group | 15 – 16 | 16 – 17 | 17 – 18 | 18 – 19 | 19 – 20 | 20 – 21 |
|---|---|---|---|---|---|---|
| No of items | 200 | 270 | 340 | 360 | 400 | 300 |
| No of defectives | 150 | 162 | 170 | 180 | 180 | 114 |

3. Obtain the two regression lines and estimate price if supply is known to be 143.

| Price | 36 | 40 | 32 | 21 | 42 | 43 | 27 | 32 | 46 |
|---|---|---|---|---|---|---|---|---|---|
| Supply | 121 | 132 | 126 | 118 | 141 | 146 | 125 | 120 | 145 |

4. Out of 320 families with 5 children each, what percentage would you expect to have at-least one boy?

5. An economist wanted to find out if there was any relationship between unemployment rate in the country and inflation rate. Data gathered from seven countries for the year 2004 are as follows:

| Country | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| Unemployment Rate | 4.0 | 8.5 | 5.5 | 0.8 | 7.3 | 5.8 | 2.1 |
| Inflation Rate | 3.2 | 8.2 | 9.4 | 5.1 | 10.1 | 7.8 | 4.7 |

Find the degree of linear association between a country's unemployment rate and its level of inflation.

6. Compute the Mode and Standard Deviation for the following data:

| C – I | 0 – 10 | 10 – 20 | 20 – 30 | 30 – 40 | 40 – 50 | 50 – 60 |
|---|---|---|---|---|---|---|
| F | 7 | 9 | 3 | 11 | 8 | 6 |

7. Given the two regression equations as $5x = 6y + 24$ and $1000y = 768x – 3708$, find "r".

8. Calculate the missing frequencies from the following data: (N = 60 and AM = 11.09)

| C – I | 9.3-9.7 | 9.8-10.2 | 10.3-10.7 | 10.8-11.2 | 11.3-11.7 | 11.8-12.2 | 12.3-12.7 | 12.8-13.2 |
|---|---|---|---|---|---|---|---|---|
| F | 2 | 5 | - | - | 14 | 6 | 3 | 1 |

9. In a post office, 3 clerks are assigned to process incoming mail. Clerk A processes 40%, Clerk B processes 35% and Clerk C processes 25% of the mail. The first clerk has an error rate of 0.04, the second has an error rate of 0.06 and the third has an error rate of 0.03. A mail selected at random from a day's output is found to have an error. The post

master wishes to know the probability that it was processed by Clerk A or Clerk B or Clerk C. Help him.

**10.** Find Spearman's rank correlation coefficient b/w X and Y:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| X | 17 | 23 | 32 | 43 | 16 | 23 | 32 | 45 |
| Y | 11 | 24 | 24 | 12 | 18 | 21 | 40 | 36 |

**11.** Compare simple comparative experiments and single factor experiments.

**12.** An aptitude test for selecting officers in a bank was conducted on 1000 candidates. The average score is 42 and standard deviation of the scores is 24. Assuming normal distribution for the scores, find the:

(a) Number of candidates whose scores exceed 58.

(b) Number of candidates whose scores lie between 30 and 66.

**13.** Fit a B.D for the following data:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| No of defectives in a sample | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| No of samples | 7 | 6 | 19 | 35 | 30 | 23 | 7 | 1 |

**14.** Find the two regression equations and estimate the sales value if ad expenses are known to be 15 lakhs.

| | AM | SD |
|---|---|---|
| Ad expenses | 16.5 | 8 |
| Sales | 72.4 | 11.1 |

Given r = 0.6

**15.** The following table gives the number of days in a 50 day period during which automobile accidents occur in a day. Fit a Poisson distribution (P.D) for the same.

| | | | | | |
|---|---|---|---|---|---|
| No of accidents | 0 | 1 | 2 | 3 | 4 |
| No of days | 21 | 18 | 7 | 3 | 1 |

**16.** In a certain factory manufacturing razor blades, there is a small chance of 1/50 for any blade to be defective. The blades are packed in packets of 10 each. Using an appropriate probability distribution, calculate the approximate number of packets containing not more than 2 defective blades in a consignment of 10000 packets.

* * * * * *

$$Z_{cal} = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$\text{Where } P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

$$P = \frac{500(0.62) + 400(0.59)}{500 + 400}$$

$$= 0.607$$

$$\therefore Q = 1 - 0.607 = 0.393$$

$$\therefore Z_{cal} = \frac{0.62 - 0.59}{\sqrt{(0.607)(0.393)\left(\frac{1}{500}+\frac{1}{400}\right)}}$$

$$= \underline{\underline{0.9155}}.$$

Conclusion:- As $Z_{cal} < Z_{tab}$, we accept $H_0$.

## Z test for sing diff of Means:-

(Q 2) A Company claims that its light bulbs are superior to those of a Competitor on the basis of a study which showed that a sample of 40 of its bulbs had avg life of 628 hrs with SD of 27 hrs, while sample of 30 bulbs made by a Competitor had an avg life of 619 hours with SD of 25 hrs. Test if Company's claim is Justified.

Soln

$H_0$: There is no difference in performance of bulbs of both the Companies

$H_1$: Co. A bulbs are superior to that of Co. B bulbs

$Z_{tab}$ @ 5% $\alpha$ (one tailed) = 1.645

$$Z_{cal} = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$= \frac{628 - 619}{\sqrt{\frac{(27)^2}{40} + \frac{(25)^2}{30}}}$$

$$= \frac{9}{6.24} = 1.44$$

Conclusion:- Since $Z_{cal} < Z_{tab}$, we accept $H_0$

## Z test for single mean:-

Q3) A stenographer claims that she can take dictation of 120 words per minute. Can we reject her claim on the basis of 100 trials in which she demonstrated a mean of 116 words with a SD of 15 words? ($\alpha = 5\%$)

Soln:-

$H_0$: There is no difference in her claim and actual words recorded.

$H_1$: There is a difference ——— " ———

$Z_{tab}$ :- 5% for 2 tailed = 1.96

$n = 100 \quad \mu = 120 \quad \bar{x} = 116$

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

$$= \frac{116 - 120}{15/\sqrt{100}}$$

$$= \frac{-4}{15/10} = \frac{-4}{1.5}$$

$$= |-2.67| = \underline{2.67}$$

Conclusion:- Since $Z_{cal} > Z_{tab}$, we reject $H_0$

→ ∴ Her claim is not true.

## Z test for difference of proportions

(Q4) A machine produced 20 defective articles in a batch of 400. After overhauling, it produced 10 defectives in a batch of 300. Has machine improved?

Soln:-

$H_0$: There is no difference in improvement of machine before and after overhauling.

$H_1$: Machine has improved after overhauling.

$Z_{tab}$:- 5% $\alpha$ (one tailed test) = 1.645

$p_1$ → proportion of defectives before overhauling → $\frac{20}{400} = 0.05$

$p_2$ → ——"—— after overhauling → $\frac{10}{300} = 0.03$

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{400(0.05) + 300(0.03)}{700} = \frac{20+9}{700}$$

$$P = \frac{29}{700} = 0.041$$

$$\therefore Q = 1 - 0.041 = 0.959$$

$$\therefore Z_{cal} = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.05 - 0.03}{\sqrt{(0.041)(0.959)\left(\frac{1}{400} + \frac{1}{300}\right)}}$$

$$= \frac{0.02}{\sqrt{0.000022}} = 1.42$$

Conclusion: Since $Z_{cal} < Z_{tab}$, we accept $H_0$ (machine hasn't improved after overhauling)

## Z test for single proportion:-

(Q5) A manufacturer claims that atleast 95% of the equipment which he supplied to a factory confirmed to specifications. An examination of a sample of 200 pieces of equipment revealed that 18 were faulty. Test his claim at 5% $\alpha$

Soln:-

$H_0$: There is no difference between claim of manufacturer and actual faulty pieces recorded

$H_1$: The number (%) of equipments confirming to specifications are more than those not confirming to it

$Z_{tab}$. 5% $\alpha$ (one tailed Test) $\rightarrow$ 1.645

WKT $Z_{cal} = \dfrac{p - P}{\sqrt{PQ/n}}$

$p \rightarrow$ proportion of pieces confirming to specifications

ie $\dfrac{200-18}{200} = 0.91$

$P$ = % confirming = 0.95

$Q = \therefore 1 - 0.95 = 0.05$

$$Z_{cal} = \frac{0.91 - 0.95}{\sqrt{\frac{(0.95)(0.05)}{200}}} = \frac{-0.04}{\sqrt{0.0002375}} = \frac{-0.04}{0.015}$$

$$= |-2.67|$$

$$= 2.67$$

Conclusion:- Since $Z_{cal} > Z_{tab}$, we reject $H_0$

Thus manufacturers' claim is validated

## t test for single mean:-

(Q.6) A machine is designed to produce insulating washers for electric devices of avg thickness of 0.025 cm. A random sample of 10 washers was found to have an avg thickness of 0.024 cm with SD of 0.002 cm. Test the significance of the deviation at 5% $\alpha$

Soln:-

$H_0$: There is no difference in avg thickness of the population and the sample

$H_1$: There is a difference ——————"——————

$t_{tab}$: (5% $\alpha$) → 2 tailed test ($df = 10 - 1 = 9$) is 2.26

$n = 10 \quad \bar{x} = 0.024 \quad \mu = 0.025 \quad \sigma = 0.002$

$$t_{cal} = \frac{\bar{x} - \mu}{s/\sqrt{n-1}}$$

$$= \frac{0.024 - 0.025}{0.002/\sqrt{10-1}}$$

$$= |-1.5| = 1.5$$

Conclusion:- Since $t_{cal} < t_{tab}$, we accept $H_0$

(Q.1) Type:- 1

[t test for single mean where sample mean and S.D are not given]

↳ A certain fertilizer is packed into bags by a machine. A random sample of 10 bags is drawn and their contents are found to weigh (in kgs) as follows:-

50 49 52 44 45 48 46 45 49 45

Test if the average packing can be taken to be 50 kgs.

Soln:-

$H_0$: There is no difference b/w population avg and the sample avg

$H_1$: There is a diff b/w ——— " ———

$t_{tab}$: [5% $\alpha$ 2 tailed (10-1) = 9 df] → 2.26

$$t_{Cal} = \frac{\bar{x} - \mu}{\sqrt{s^2/n}}$$

Note:- $\bar{x} = A + \left(\frac{\Sigma d}{n}\right)$

$$s^2 = \frac{\Sigma d^2 - \frac{(\Sigma d)^2}{n}}{n-1}$$

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| x: | 50 | 49 | 52 | 44 | 45 | 48 | 46 | 45 | 49 | 45 | Total = 473 |
| d = x-48 | 2 | 1 | 4 | -4 | -3 | 0 | -2 | -3 | 1 | -3 | -7 |
| $d^2$ | 4 | 1 | 16 | 16 | 9 | 0 | 4 | 9 | 1 | 9 | 69 |

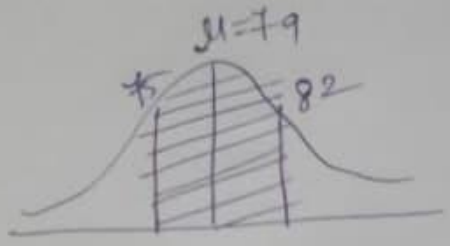
$$\bar{x} = A + \left(\frac{\Sigma d}{n}\right) = 48 + \left(\frac{-7}{10}\right) = 47.3$$

$$s^2 = \frac{\Sigma d^2 - (\Sigma d)^2/n}{n-1} = \frac{69 - (-7)^2/10}{10-1} = 7.12$$

$$\therefore t_{cal} = \frac{\bar{x} - \mu}{\sqrt{s^2/n}} = \frac{47.3 - 50}{\sqrt{7.12/10}} = |-3.2| = \boxed{3.2}$$

Conclusion:- As $t_{cal} > t_{tab}$, we reject $H_0$.
(ie avg packing can't be taken as 50 kgs)

(Q) The avg test marks of Students in a class is 79. The Std deviation is 5. If the marks are distributed normally, how many students in a class of 200 did not receive marks between 75 and 82?




B/w 75 & 82:-

$Z = \frac{\bar{x} - \mu}{\sigma}$ & $Z = \frac{\bar{x} - \mu}{\sigma}$

$Z = \frac{75 - 79}{5}$ & $Z = \frac{82 - 79}{5}$

$= \frac{4}{5} = 0.8$ & $= \frac{3}{5} = 0.6$

Area: 0.2881 Area: 0.2257

∴ Total Area:- 0.2881 + 0.2257 = 0.5138

∴ Required Area:- (Not b/w 75 and 82) = 1 − 0.5138

= 0.4862

∴ No not b/w 75 and 82 = (0.4862)200 = 97.24

≈ (97)

USN | | | | | | | | | | |  **20MBA14**

**R N S INSTITUTE OF TECHNOLOGY**
**DEPARTMENT OF MBA & RESEARCH CENTRE**
# FIRST SEMESTER MOCK EXAMINATION APRIL - 2021BUSINESS STATISTICS

Time: 3 hrs. Max. Marks: 100

**Instructions: Answer any FOUR full Questions from Q 1 to Q7. Q 8 is Compulsory.**

<table>
<tr><th>Q No</th><th>Question</th><th>Marks</th></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>1 a</td><td>Define Statistics and state four of its limitations.</td><td>3</td></tr>
<tr><td>1 b</td><td>Use an appropriate measure to evaluate dispersion in the following data:
<table>
<tr><td>C – I</td><td>Less than 35</td><td>35 – 37</td><td>38 – 40</td><td>41 – 43</td><td>More than 43</td></tr>
<tr><td>F</td><td>14</td><td>62</td><td>99</td><td>18</td><td>7</td></tr>
</table></td><td>7</td></tr>
<tr><td>1 c</td><td>(i) The incidence of occupational disease in an industry is such that workers have 20% chance of suffering from it. What is the Probability that out of 6 workers, 4 or more will have the disease?<br>(ii) The mean weekly sales of chocolate bar in candy stores were 146.3 bars per store. After an ad campaign, the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed SD of 17.2. Was the ad campaign successful?</td><td>10</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>2 a</td><td>Define Probability, Sample Space & Binomial distribution.</td><td>3</td></tr>
<tr><td>2 b</td><td>From the adjoining table, calculate the coefficient of correlation by Karl Pearson method.
<table>
<tr><td>X</td><td>6</td><td>2</td><td>10</td><td>4</td><td>8</td></tr>
<tr><td>Y</td><td>9</td><td>11</td><td>?</td><td>8</td><td>7</td></tr>
</table>
Arithmetic means of X and Y series are 6 and 8 respectively.</td><td>7</td></tr>
<tr><td>2 c</td><td>Calculate lower and upper quartiles, 3rd decile and 20th percentile from the following data:
<table>
<tr><td>Central Value</td><td>2.5</td><td>7.5</td><td>12.5</td><td>17.5</td><td>22.5</td></tr>
<tr><td>Frequency</td><td>7</td><td>18</td><td>25</td><td>30</td><td>20</td></tr>
</table></td><td>10</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>3 a</td><td>List the objectives of time series analysis.</td><td>3</td></tr>
<tr><td>3 b</td><td>Calculate Mode:
<table>
<tr><td>Marks(Above)</td><td>0</td><td>10</td><td>20</td><td>30</td><td>40</td><td>50</td><td>60</td><td>70</td><td>80</td><td>90</td></tr>
<tr><td>No of students</td><td>80</td><td>77</td><td>72</td><td>65</td><td>55</td><td>43</td><td>28</td><td>16</td><td>10</td><td>8</td></tr>
</table></td><td>7</td></tr>
<tr><td>3 c</td><td>The table below represents the number of bounced cheques in two banks – Bank A and Bank B – on randomly chosen 12 days for Bank A and 15 days for Bank B. use a Mann- Whitney U test to examine at a 5 percent level of significance whether Bank A has more bounced cheques as compared to Bank B.
<table>
<tr><td>Bank A</td><td>42</td><td>65</td><td>38</td><td>55</td><td>71</td><td>60</td><td>4</td><td></td><td>59</td><td>68</td><td>57</td><td>76</td><td>42</td><td></td><td></td><td></td></tr>
<tr><td>Bank B</td><td>22</td><td>17</td><td>35</td><td>19</td><td>8</td><td>24</td><td>4</td><td></td><td>14</td><td>28</td><td>17</td><td>10</td><td>15</td><td>20</td><td>45</td><td>50</td></tr>
</table></td><td>10</td></tr>
<tr><td></td><td></td><td></td></tr>
<tr><td>4 a</td><td>Why is mean considered the best measure of central tendency?</td><td>3</td></tr>
<tr><td>4 b</td><td>Find missing frequencies from the data given below, given that Mean = 1.46 and N = 200
<table>
<tr><td>No of accidents</td><td>0</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>No of Days</td><td>46</td><td>?</td><td>?</td><td>25</td><td>10</td><td>5</td></tr>
</table></td><td>7</td></tr>
</table>

**4 c**

(i) A research was conducted to understand whether women have a greater variation in attitude on political issues than men. Two independent samples of 31 men and 41 women were used for the study. The sample variances were 120for women & 80 for men. Test whether difference in attitude toward political issues is significant at 5% level of significance using F test.

(ii) You are given below the following data about advertisement and sales: **5**

| | Ad expenses (Rs. Crores) | Sales (Rs. Crores) |
|---|---|---|
| Mean | 10 | 90 |
| S.D (σ) | 3 | 12 |

r = 0.8.

Find the two regression lines. What should be advertisement expenditure if the company wants to attain sales target of 150 crores? **5**

**5 a** Differentiate between one tailed test and two tailed test. **3**

**5 b** Using Spearman's rank correlation, find the strength of association between inventory turnover and Earnings per share. Interpret your findings. **7**

| Company | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| Inventory turnover | 4 | 5 | 7 | 8 | 6 | 3 | 5 |
| EPS | 11 | 9 | 13 | 7 | 13 | 8 | 8 |

**5 c**

(i) In a factory, 3 machines – A, B & C manufacture 25%, 35% and 40% output. Of them, 5%, 4% and 2% are defective. If an output unit is drawn at random, what is the Probability that (a) the defective output comes from Machine A (b) that it comes from Machine B or Machine C? **5**

(ii) If there are 200 typing mistakes randomly distributed in a 500 page book, find the probability that a given page has exactly 3 errors. **5**

**6 a** State the significance of measuring variation. List the measures of variation/dispersion. **3**

**6 b** The mean and SD of a series of 20 items were calculated by a student as 20 cm and 5 cm respectively. But while calculating an item 13 were misread as 30. Find the correct mean and S.D. **7**

**6 c** Calculate seasonal index by the ratio to moving average method for the following data: **10**

| | Quarters | | | |
|---|---|---|---|---|
| Year | I | II | III | IV |
| 2015 | 75 | 60 | 53 | 59 |
| 2016 | 86 | 65 | 63 | 80 |
| 2017 | 90 | 72 | 66 | 85 |
| 2018 | 100 | 78 | 72 | 93 |

**7 a** Differentiate between simple regression and multiple regression. **3**

**7 b** In a Normal Distribution, 21% of the items are below 35 and 11% of the items are above 54. Find the mean and SD of the distribution. **7**

**7 c** The refrigerators of model A have an average life of 5.12 and standard deviation of 2.8years. The life (in years) and number of refrigerators of model B are given as follows: **10**

| Life [No of years] | 0-2 | 2-4 | 4-6 | 6-8 | 8-10 | 10-12 |
|---|---|---|---|---|---|---|
| No of refrigerators [model B] | 2 | 7 | 12 | 19 | 9 | 1 |

Which model of refrigerator has greater uniformity?

**8 a** Three products received the following performance ratings by a panel of 15 consumers.

| Product | | |
|---|---|---|
| A | B | C |
| 50 | 80 | 60 |
| 62 | 95 | 45 |
| 75 | 98 | 30 |
| 48 | 87 | 58 |
| 65 | 90 | 57 |

Use the Kruskal-Wallis test and $\alpha = 0.05$ to determine whether there is a significant difference in the performance ratings for the products. **10**

**8 b** Fit a Binomial Distribution for the following data: **10**

| X | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| F | 214 | 92 | 20 | 3 | 1 |



* * * * * *

**MODEL QUESTION PAPER – SET 1**
**Business Statistics (24MBA15)**

**Note: 1) Answer any 4 full questions from Q.No. 1 to Q.No. 7**
**2) Part – B, Q.No. 8 is compulsory**

**Part – A**

**1)** a. Explain the functions of statistics. **(3marks)**

b. Calculate Median, Upper quartiles, 3rd Decile from the following data: **(7 marks)**

| Heights in Cm | 145-150 | 150-155 | 155-160 | 160-165 | 165-170 | 170-175 | 175-180 |
|---|---|---|---|---|---|---|---|
| No. of persons | 2 | 4 | 12 | 22 | 30 | 25 | 10 |

c. From the prices of X & Y of shares A& B respectively given below, state which share is morestablein value.
(10marks)

| Price of Share A | 55 | 54 | 52 | 53 | 56 | 58 | 52 | 50 | 51 | 49 |
|---|---|---|---|---|---|---|---|---|---|---|
| Price of share B | 108 | 107 | 105 | 105 | 106 | 107 | 104 | 103 | 104 | 101 |

**2)** a. What do you mean by correlation? Mention any four uses of it? **(3marks)**

b. Discuss the difference between Parametric and Non-Parametric tests? **(7 marks)**

c. Calculate Spearman's rank correlation co-efficient between advertisement cost and sales from the following data

| Advertisement cost('000 Rs) | 39 | 65 | 62 | 90 | 82 | 75 | 25 | 98 | 36 | 78 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sales(lakhs Rs) | 47 | 53 | 58 | 86 | 62 | 68 | 60 | 91 | 51 | 84 |

**3)** a. Explain the significance in measuring dispersion? **(3marks)**

b. In a bolt factory, machines A, B and C manufacture 25%, 35% and 40% respectively of the total. Of their output 5, 4 and 2 percent are known to be defective bolts. A bolt is drawn at random from the product and is found to be defective. What are the probabilities that it was manufactured by machines B or C ? **(7marks)**

c. The following data related to the scores obtained 9 salesmen of a company in an intelligence test and their weekly sales in Rs ('ooo). **(10marks)**

| Salesman | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| Test scores | 50 | 60 | 50 | 60 | 80 | 50 | 80 | 40 | 70 |
| Weekly sales | 30 | 60 | 40 | 50 | 60 | 30 | 7 | 50 | 60 |

(i) Obtain the regression equation of sales on intelligence test scores of the salesman.

(ii) If the intelligence test score of a salesman is 65, what would be his expected weekly sales?

**4)** a. What is Poisson's distribution? Mention 2 application of Poisson distribution. **(3marks)**

b. calculate Rank Correlation co-efficient between marks assigned to 10 students by judges X & Y in a certain competitive test. **(7marks)**

| Marks given by judge X | 52 | 53 | 42 | 60 | 45 | 41 | 37 | 38 | 25 | 27 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks given by judge Y | 65 | 68 | 43 | 38 | 77 | 48 | 35 | 30 | 25 | 50 |

c. (i) A cyclist pedals from his house to his college at a speed of 10 km/hr and back from college to his house at 15 km/hr. Find the average speed. **(5 marks)**

(ii) If the A.M. of two observations is 127.5 & their G.M. is 60. Find (i) Their H.M. & (II) two observations. **(5 marks)**

**5)** a. What do you mean by correlation? Mention any four uses of it? **(3 marks)**

b. The mean and standard deviation of a set of 100 observations were worked out as 40 and 5 respectively by a computer which by mistake took value 50 in place of 40 for one observation. Find the correct mean and standard deviation **(7 marks)**

c. Goals scored by two teams A & B in a football season are given below. Find the more consistent team. **(10 marks)**

| No. of goals | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| Team A | 27 | 9 | 8 | 5 | 4 | 1 |
| Team B | 17 | 9 | 6 | 5 | 3 | 2 |

**6)** a. Define time series analysis? Mention the methods used for the study and measurement of trend in time series ? (3 marks)

b. Explain the Procedure of hypothesis test described various stages involved. **(7 marks)**

c. Mysuru Mahanagar Palike surveyed the travel preferences of people who travelled to work by train or bus. The initial analysis suggested that 1 in 5 people travelled by train to work. If 5 people are interviewed, what is the probability that,

i) Exactly 3 prefer travelling by train

ii) Three or more prefer travelling by train and

**iii)** Less than 3 prefer travelling by train. **(10marks)**

**7)** a. What is hypothesis? Mention the types of hypothesis testing? (3 marks)

b. From the following series of annual data, find the trend line by the method of semi-averages. Also estimate the value for 1999. (7 marks)

| Year | 1990 | 1991 | 1992 | 1993 | 1994 | 1995 | 1996 | 1997 | 1998 |
|---|---|---|---|---|---|---|---|---|---|
| Actual value | 170 | 231 | 261 | 267 | 278 | 302 | 299 | 298 | 340 |

c. Three products received the following performance ratings by a panel of 15 consumers.

| Product | | |
|---|---|---|
| A | B | C |
| 50 | 80 | 60 |
| 62 | 95 | 45 |
| 75 | 98 | 30 |
| 48 | 87 | 58 |
| 65 | 90 | 57 |

Use the Kruskal-Wallis test and $\alpha = 0.05$ to determine whether there is a significant difference in the performance ratings for the products. **(10 marks)**

## Part - B

**8)** a. A coaching centre claims that students will perform better in their exams after going through the coaching offered by their centre. The table given below shows the marks obtained by 6 students before and after the coaching course. Can you conclude that the students score has improved after the course with level of significance $\alpha = 0.05$?

| Student | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| Marks (before) | 85 | 96 | 70 | 76 | 81 | 78 |
| Marks (after) | 88 | 85 | 89 | 86 | 92 | 89 |

**(10 marks)**

b. The number of defects per unit in a sample of 330 units of manufactured product was givenbelow. Fit a Poisson distribution to the data ( Given e-0.439 = 0.6447)

| No. of defects | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| No. of units | 214 | 92 | 20 | 3 | 1 |

**(10 marks)**

**********

## MODEL QUESTION PAPER – SET 2
**Business Statistics (24MBA15)**

Note: 1) Answer any 4 full questions from Q.No. 1 to Q.No. 7
2) Part – B, Q.No. 8 is compulsory

Part – A

**1.** a.Mention the applications of statistics**.** **(3marks)**

b. Find the Mean & Standard Deviation from the following data**:** **(10marks)**

| X | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|
| Y | 15 | 30 | 53 | 75 | 100 | 110 | 115 | 125 |

**(7 marks)**

c. Fit a straight line trend by the method of least squares and estimate the arrival in the year 2013, for the data indicated in the table below: **(10marks)**

| Year | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|---|---|
| Tourists arrivals | 18 | 20 | 23 | 25 | 24 | 28 | 30 |

**2.** a. Discuss the advantages and limitations of Diagrams and Graphs. **(3marks)**

b. For a certain frequency table which has only been partly reproduced, here the Mean was foundto be 1.46 . Calculate the missing frequencies.

| No. of accidents | 0 | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|---|
| No. of days | 46 | ? | ? | 25 | 10 | 5 | N=200 |

**(7 marks)**

c. From the following data obtain the two regression equations. **(10 marks)**

| Sales | 91 | 97 | 108 | 121 | 67 | 124 | 51 | 73 | 111 | 57 |
|---|---|---|---|---|---|---|---|---|---|---|
| purchase | 71 | 75 | 69 | 97 | 70 | 91 | 39 | 61 | 80 | 47 |

**3.** a Mention the rules of probability with formula. **(3marks)**

b. Discuss : (i) Type I & Type II errors (ii) Different methods of estimating trend **(7marks**)

c. The table below represents the number of bounced cheques in two banks – Bank A and Bank B – on randomly chosen 12 days for Bank A and 15 days for Bank B. use a Mann-Whitney U test to examine at a 5 percent level of significance whether Bank A has more bounced cheques as compared to Bank B.

| Bank A | 42 | 65 | 38 | 55 | 71 | 60 | 47 | 59 | 68 | 57 | 76 | 42 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Bank B | 22 | 17 | 35 | 19 | 8 | 24 | 42 | 14 | 28 | 17 | 10 | 15 | 20 | 45 | 50 |

**(10 arks)**

4. a. Define time series analysis? Mention the methods used for the study and measurement of trend in time series ?(3marks)

b. Explain the Procedure of hypothesis test described various stages involved**.** **(7 marks)**

c. A financial analyst wanted to find out whether inventory turnover influences any company's earning per share (in percentage). A random sample of 7 companies listed in a stock exchange were selected and following data was recorded for each:

| Company | Inventory turnover(no. of times) | Earnings per share(percentage) |
|---|---|---|
| A | 4 | 11 |
| B | 5 | 9 |
| C | 7 | 13 |
| D | 8 | 7 |
| E | 6 | 13 |
| F | 3 | 8 |
| G | 5 | 8 |

Using Spearman's rank correlation, find the strength of association between inventory turnover and earnings per share. Interpret the findings. **(10 marks)**

5. a. What is Poisson's distribution? Write a formula for probability function of Poisson's distributions. (3marks)

b. Discuss the difference between Parametric &Non Parametric tests. **(7 marks)**

c. From the following table calculate the Co-efficient of correlation by Karl Pearson's method,

| | | | | | |
|---|---|---|---|---|---|
| X | 6 | 2 | 10 | 4 | 8 |
| Y | 9 | 11 | 5 | 8 | 7 |

**(10 arks)**

6. a. Define the following terms : i) Independent event; ii)Mutually exclusive event ; iii) Equally likelyEvent. **(3 marks)**

b. A machine is expected to produce nails of length 3 cm. A random sample of 25 nails gave an average length of 3.1 cm with SD 0.3 cm. can it be said that the machine is producing nails as per specification. (Given value of $t_{0.05}$ for 24 df is 2.064)

**(7marks)**

c. The quarterly sales for five years from 2008-2011 is given below. Use ratio to moving average method to determine the seasonal indexes.

| Quarter | Sales (Rs. In thousand) | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| 2008 | 77 | 62 | 56 | 61 |
| 2009 | 85 | 64 | 62 | 79 |
| 2010 | 91 | 73 | 67 | 86 |
| 2011 | 102 | 80 | 74 | 95 |

**(10 arks)**

**7.** a. Explain the difference between simple and multiple regression with example. **(3 marks)**

b. A research was conducted to understand whether women have a greater variation in attitude on political issues than men. Two independent samples of 31 men and 41 women were used for the study. The sample variances so calculated were 120 for women & 80 for men. Test whether the difference in attitude toward political issues is significant at 5% level of significance using F test. **(7marks)**

c. Three machines are used in the packaging of 10 kg of wheat flour. Each machine is designed to pack on an average 10 kg of flour per bag. Samples of six bags were selected from each machine and the amount of wheat packaged in each bag is shown below:

| | | | | | | |
|---|---|---|---|---|---|---|
| achine 1 | 15.8 | 15.9 | 16.2 | 15.7 | 16.3 | 15.8 |
| achine 2 | 16.5 | 16 | 15.4 | 15.9 | 16.2 | 16.1 |
| achine 3 | 15.7 | 16.4 | 16.2 | 15.9 | 15.7 | 16.3 |

Use a 5 percent level of significance to test the hypothesis that the amount of wheat packaged by the three machines is the same using K-W test. **(10 marks)**

## Part - B

2. a. The average daily sales of 500 branch offices was Rs. 150 thousand & the SD of Rs. 15 thousand. Assuming the distribution to be normal, indicate how many branches have sales between :(i) Rs. 120 thousand & Rs. 145 thousand

(ii) Rs. 140 thousand & Rs. 165 thousand **(10 marks)**

b. A manufacturing company has purchased three new machines of different makes and wishes to determine whether one of them is faster than the others in producing a certain output. Five hourly production figures are observed at random from each machine and the results are given in the table below:

| | Machine A1 | Machine A2 | Machine A3 |
|---|---|---|---|
| Observations | 25 | 31 | 24 |
| | 30 | 39 | 30 |
| | 36 | 38 | 28 |
| | 38 | 42 | 25 |
| | 31 | 35 | 28 |

Use Analysis of Variance technique and determine whether the machines are significantly different in their mean speeds. ( use $\alpha = 5\%$ value of F for (2,12) d.f. is 3.89).

**(10 marks)**

**********

Modified

CBCS SCHEME

USN | | | | | | | | | | |

20MBA14

# First Semester MBA Degree Examination, Jan./Feb.2021
## Business Statistics

Time: 3 hrs. Max. Marks:100

**Note:** ***1. Answer any FOUR full questions from Q1 to Q7.***
***2. Question No.8 is compulsory.***
***3. Use of statistical table is allowed.***

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8 = 50, will be treated as malpractice.

1 a. Discuss the importance of Business Statistics. **(03 Marks)**

b. From the prices of shares find out which is more stable in value. **(07 Marks)**

| X | 35 | 54 | 52 | 53 | 56 | 58 | 52 | 50 | 51 | 49 |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | 108 | 107 | 105 | 105 | 106 | 107 | 104 | 103 | 104 | 101 |

c. Find the three Quartiles 7th decide and 84th percentile from the following data:

| Wages(in Rs.) | 30 – 40 | 40 – 50 | 50 – 60 | 60 – 70 | 70 – 80 | 80 – 90 | 90 – 100 |
|---|---|---|---|---|---|---|---|
| No. of persons | 1 | 3 | 11 | 21 | 43 | 21 | 9 |

**(10 Marks)**

2 a. Mention any three properties of Arithmetic mean. **(03 Marks)**

b. Explain Pictorially scatter diagram and how is it used in predictions? **(07 Marks)**

c. The following data relate to age of employees and the number of days they reported sick in a month. Calculate Karl Pearson's co-efficient of correlation and interpret it.

| Age (years) | 30 | 32 | 35 | 40 | 48 | 50 | 52 | 55 | 57 | 61 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sick Days | 1 | 0 | 2 | 5 | 2 | 4 | 6 | 5 | 7 | 8 |

**(10 Marks)**

3 a. Define Mutually exclusive, equally likely and exhaustive events. **(03 Marks)**

b. A systematic sample of 100 pages was taken from the Oxford dictionary and the observed frequency distribution of foreign words per page was formed to be as follows:

| No. of foreign words (per page) | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Frequency | 48 | 27 | 12 | 7 | 4 | 1 | 1 |

Calculate the expected frequency use in Poisson distribution. **(07 Marks)**

c. From the following data obtain the two regression equation and calculate correlate co-efficient.

| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Y | 9 | 8 | 10 | 12 | 11 | 13 | 14 | 16 | 15 |

**(10 Marks)**

4 a. What do you mean by Regression Analysis? Give any two uses of it. **(03 Marks)**

b. You have been provided with the figures of production (in 000's tons) of sugar factory.

| Year | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|---|---|
| Production | 77 | 88 | 94 | 85 | 91 | 98 | 90 |

Fit a straight line by the method of least square and find trend values. **(07 Marks)**

c. The income of a group of 10,000 persons were found to be normally distributed with mean Rs.7500 p.m. and standard deviation is Rs.500/-. Show that of this group about 95% has income exceeding Rs.6680 and only 5% had income exceeding 8320. What was the lowest income among the richest 100? **(10 Marks)**

5 a. Define binomial distribution. **(03 Marks)**

b. Particulars of regarding the income of two villages are given below:

| | Village 'X' | Village 'Y' |
|---|---|---|
| Number of people | 600 | 500 |
| Average income (Rs.) | 175 | 186 |
| Variance | 100 | 81 |

(i) In which village the variate in income is greater?

(ii) What is the combined standard deviation of the village 'X' and village 'Y' put together? **(07 Marks)**

c. Calculate Spearman's Rank correlation co-efficient between advertisement cost and sales from the following data:

| Advertisement cost (000 Rs.) | 39 | 65 | 62 | 90 | 82 | 75 | 25 | 98 | 36 | 78 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sales (lakhs Rs.) | 47 | 53 | 58 | 86 | 62 | 68 | 60 | 91 | 51 | 84 |

**(10 Marks)**

6 a. What are the different measures of dispersion? Explain. **(03 Marks)**

b. Explain the different components of time series. **(07 Marks)**

c. One forth of the first year student admitted to a Bangalore college are out of state students. If the students are assigned at random to the dormitories, 3 to a room, what is the probability that in one room.

(i) At least 2 of the 3 roommates are out of state students.

(ii) At most 2 of the 3 roommates are out of state students. **(10 Marks)**

7 a. Define normal distribution. **(03 Marks)**

b. Explain the characteristics of Good Hypothesis. **(07 Marks)**

c. Define Hypothesis. Describe the formulation of Hypothesis with flow process chart. **(10 Marks)**

8 a. Write briefly on the following terms:

(i) Type I error.

(ii) Type II error. **(06 Marks)**

b. The daily wages of 1000 workmen are normally distributed around a mean of Rs.70 and with a standard deviation of Rs.5. Estimate the number of workers whose daily wages will be:

(i) Between Rs.70 and 72

(ii) Between Rs.69 and 72

(iii) More than Rs.75

(iv) Less than Rs.63

(v) More than Rs.80

(vi) Also estimate the lowest daily wages of the 100 highest paid workers. **(14 Marks)**

* * * * *

13:43:11 21-10-2

CBCS SCHEME

USN | 1 | R | X | 1 | 9 | M | B | A | 5 | 1 |

20MBA14

# First Semester MBA Degree Examination, July/August 2021
## Business Statistics

Time: 3 hrs. Max. Marks:100

**Note:** *1. Answer any FIVE full questions.*
*2. Table values given in the questions.*

1 a. Explain the functions of statistics. **(03 Marks)**

b. Find mean and standard deviation from the following data :

| X | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|
| Y | 15 | 30 | 53 | 75 | 100 | 110 | 115 | 125 |

**(07 Marks)**

c. From the prices of X and Y shares, A and B respectively given below state which share is more stable in value?

| Price of Share A | 55 | 54 | 52 | 53 | 56 | 58 | 52 | 50 | 51 | 49 |
|---|---|---|---|---|---|---|---|---|---|---|
| Price of Share B | 108 | 107 | 105 | 105 | 106 | 107 | 104 | 103 | 104 | 101 |

**(10 Marks)**

2 a. Discuss the advantages and limitations of diagrams and graphs. **(03 Marks)**

b. For a certain frequency table, which has only been partly reproduced, here the mean was found to be 1.46. Calculate the missing frequency. N = 200.

| No of accidents | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| No of days | 46 | ? | ? | 25 | 10 | 5 |

**(07 Marks)**

c. Calculate Spearman's rank correlation coefficient between advertisement cost and sales from the following data

| Advertisement cost (1000Rs.) | 39 | 65 | 62 | 90 | 82 | 75 | 25 | 98 | 36 | 78 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sales (lakhs Rs.) | 47 | 53 | 58 | 86 | 62 | 68 | 60 | 91 | 51 | 84 |

**(10 Marks)**

3 a. Explain the significance in measuring dispersion. **(03 Marks)**

b. Discuss the difference between parametric and non-parametric test. **(07 Marks)**

c. Mysuru Mahanagar Pal ke surveyed the travel preferences of people who travelled to work by train or bus. The initial analysis suggested that in 5 people travelled by train to work. If 5 people are interviewed, what is the probability that,
i) Exactly 3 prefer travelling by train
ii) Less than 3 prefer travelling by train. **(10 Marks)**

4 a. What is Type I and Type II errors. **(03 Marks)**

b. Define the following terms :
i) Independent event
ii) Mutually exclusive event
iii) Equally likely event. **(07 Marks)**

c. Below are given the figures of production (in thousand tons) of a sugar factory :

| Year | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|---|
| Production | 77 | 88 | 94 | 85 | 91 | 98 | 90 |

i) Fit a straight line by the method of 'Least Squares' and show the trend values
ii) What is the monthly increase in production? **(10 Marks)**

5 a. What do you mean by correction? Give any two examples of negative correction. **(03 Marks)**

b. The hourly wages of 1000 workmen are normally distributed around a mean of Rs.70 and with a standard deviation of Rs.5. Estimate the number of workers whose hourly wages will be : i) Between Rs.69 and Rs.72 ii) More than Rs.75 iii) Less than Rs.63.

Area under standard normal probability curve :

| Z | 0.1 | 0.2 | 0.3 | 0.4 | 1.0 | 1.4 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|
| AREA | 0.0398 | 0.0793 | 0.1179 | 0.1554 | 0.3413 | 0.4192 | 0.4332 | 0.4772 |

**(07 Marks)**

c. The sales data of an item in six shops before and after a special promotional campaign are as under : $t_5(0.05) = 2.02$.

| Shops | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Before campaign | 53 | 28 | 31 | 48 | 50 | 42 |
| After campaign | 58 | 29 | 30 | 55 | 56 | 45 |

**(10 Marks)**

6 a. Under what conditions binomial distribution tends to Poisson distribution. **(03 Marks)**

b. If 5% of the electric bulbs manufactured by a company are defective, use Poisson distribution to find the probability that in a sample of 100 bulbs.
i) None is defective ii) 5 bulbs will be defective? (Given : $e^{-5} = 0.007$). **(07 Marks)**

c. From the data given below find
i) The two regression coefficients
ii) The two regression equations
iii) The coefficient of correlation between.
Marks in economics and statistics :

| Marks in economics | 25 | 28 | 35 | 32 | 31 | 36 | 29 | 38 | 34 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks in statistics | 43 | 46 | 49 | 41 | 36 | 32 | 31 | 30 | 33 | 39 |

**(10 Marks)**

7 a. What is mode? Give two examples of mode. **(03 Marks)**

b. Intelligent test of two groups of boys and girls gave the following results :

| | Mean | S.D. | N |
|---|---|---|---|
| Girls | 75 | 15 | 150 |
| Boys | 70 | 20 | 250 |

Is there a significant difference in the mean scores obtained by boys and girls. **(07 Marks)**
(Test $\alpha = 1\%$).

c. The following data present the yield in quintals of JONS on ten sub-divisions of equal area of two agricultural plots :

| PLOT – 1 | 6.2 | 5.7 | 6.5 | 6.0 | 6.3 | 5.8 | 5.7 | 6.0 | 6.0 | 5.8 |
|---|---|---|---|---|---|---|---|---|---|---|
| PLOT – 2 | 5.6 | 5.9 | 5.6 | 5.7 | 5.8 | 5.7 | 6.0 | 5.5 | 5.7 | 5.5 |

Test whether two samples taken from two random population having the same variance. **(10 Marks)**
(At 5% level for $V_1 = 9$ and $V_2 = 9$ is 3.18).

8 a. Find : i) Interquartile range ii) Quartile deviation
iii) Coefficient of quartile deviation for the following :

| Class interval | 0 – 15 | 15 – 30 | 30 – 45 | 45 – 60 | 60 – 75 | 75 – 90 | 90 – 105 |
|---|---|---|---|---|---|---|---|
| Frequency | 8 | 26 | 30 | 45 | 20 | 17 | 4 |

**(10 Marks)**

b. Three products received the following performance :

| Product A | 50 | 62 | 75 | 48 | 65 | $\chi^2_2(0.05) = 5.991$ |
|---|---|---|---|---|---|---|
| Product B | 80 | 95 | 98 | 87 | 90 | |
| Product C | 60 | 45 | 30 | 58 | 57 | |

Use the Kruskal – Wallies test at $\alpha = 0.05$ to determine whether there is a significant difference in the performance rating of products. **(10 Marks)**

* * * * *

CBCS SCHEME

USN | | | | | | | | | | |

20MBA14

# First Semester MBA Degree Examination, Feb./Mar.2022
## Business Statistics

Time: 3 hrs. Max. Marks:100

**Note:** *1. Answer any FOUR full questions from Q1 to Q7.*
*2. Question No.8 is compulsory.*
*3. Use of Statistical tables is allowed.*



1 a. Find the average rate of increase in population which in the first decade has increased by 20%, in the second decade by 30% and in the third decade by 40%. **(03 Marks)**

b. Measure the 3rd quartile, 2nd decile and 95th percentile for the following data:

| Marks | 10 – 19 | 20 – 29 | 30 – 39 | 40 – 49 | 50 – 59 | 60 – 69 |
|---|---|---|---|---|---|---|
| Frequency | 12 | 27 | 34 | 41 | 23 | 3 |

**(07 Marks)**

c. The weekly sales of products A and B are recorded in the following table. Examine which one shows greater fluctuation in sales.

| Product A | 59 | 75 | 27 | 63 | 27 | 28 | 56 |
|---|---|---|---|---|---|---|---|
| Product B | 150 | 200 | 125 | 310 | 330 | 250 | 225 |

**(10 Marks)**

2 a. Interpret the values of r = 0, r = −1 and r = +1. **(03 Marks)**

b. Explain the scope of statistics. **(07 Marks)**

c. Calculate the coefficient of correlation by Karl Pearson's method from the following:

| Overheads (in '000 Rs.) | 80 | 90 | 100 | 110 | 120 | 130 | 140 | 150 | 160 |
|---|---|---|---|---|---|---|---|---|---|
| Cost (in '000 Rs.) | 15 | 15 | 16 | 19 | 17 | 18 | 16 | 18 | 19 |

**(10 Marks)**

3 a. Determine the two regression coefficients when r = 0.8, $\sigma_x = 5$ and $\sigma_y = 7$. **(03 Marks)**

b. The average percentage of failures in a certain examination is 40. What is the probability that out of a group of 6 candidates, at least 4 passed in the examination. What is the probability that at the most 2 passed. **(07 Marks)**

c. In an intelligence test administered to 500 students and data is normally distributed. The average score was 42 and standard deviation was 24. Find (i) The number of students whose score exceeded 50. (ii) The number of students who scored between 30 and 40. (iii) The number of students who scored above 60. **(10 Marks)**

4 a. Explain the uses of time series analysis. **(03 Marks)**

b. A sample of 200 bulbs made by a company gives a lifetime mean of 1540 hours? With a standard deviation of 42 hours. It is likely that the sample has been drawn from a population with a mean lifetime of 1500 hours. Evaluate at 5% level of significance. **(07 Marks)**

c. The information given below relates to the sales and advertisement expenditure of the firm,

| | Advertisement Expenses (Rs. lakhs) | Sales (Rs. lakhs) |
|---|---|---|
| Arithmetic mean | 20 | 100 |
| Standard deviation | 3 | 12 |

Coefficient of correlation r = 0.8

(i) Find the two regression equations.
(ii) What should be the advertisement expenditure if the firm proposes a sales target of Rs.120 lakhs. **(10 Marks)**

5 a. A uniform die is thrown at random. What is the probability that the number on it is greater than 4. **(03 Marks)**

b. Explain the components of time series. **(07 Marks)**

c. Calculate (i) Three yearly (ii) Five yearly, moving averages for the following data:

| Year | 2010 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | 242 | 250 | 252 | 249 | 253 | 255 | 251 | 257 | 260 | 265 | 262 |

**(10 Marks)**

6 a. Compare Type I error and Type II error. **(03 Marks)**

b. Explain ANOVA, K-W test and Mann-Whitney test. **(07 Marks)**

c. A systematic sample of 100 pages was taken from the concise Oxford Dictionary and the observed frequency distributon of foreign words per page was found to be as follows:

| No. of foreign words per page (X) | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Frequency (f) | 48 | 27 | 12 | 7 | 4 | 1 | 1 |

Calculate the expected frequencies using Poisson distribution. Also compute the mean and variance of fitted distribution. **(10 Marks)**

7 a. Discuss the applications of binomial distribution with example. **(03 Marks)**

b. A selects 8 salesmen at random and the sales figures for the previous month are recorded. They then undergo a training course. Their sales figure for the following month are recorded as shown in the table. Test if the training course, caused an improvement in the salesmen's ability? Choose 5% level of significance. The table value for V = 7 is 1.8975.

| Previous month | 75 | 90 | 94 | 95 | 100 | 90 | 70 | 64 |
|---|---|---|---|---|---|---|---|---|
| Following month | 77 | 101 | 93 | 92 | 105 | 88 | 76 | 68 |

**(07 Marks)**

c. Given below are the values of production ('000 tons) of a steel factory.
(i) Fit a straight line trend by the method of least squares.
(ii) Determine the monthly increase in production.

| Year | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|---|---|
| Production | 77 | 88 | 94 | 85 | 91 | 98 | 90 |

**(10 Marks)**

8 Case Study :
Using 'Ratio to trend' method, determine the quarterly seasonal indices for the following data:

Production of steel (in million tons)

| Year | $Q_1$ | $Q_2$ | $Q_3$ | $Q_4$ |
|---|---|---|---|---|
| 1 | 68 | 60 | 61 | 63 |
| 2 | 70 | 58 | 56 | 60 |
| 3 | 68 | 63 | 68 | 67 |
| 4 | 65 | 56 | 56 | 62 |
| 5 | 60 | 55 | 55 | 58 |

**(20 Marks)**

* * * * *

CBCS SCHEME

USN | | | | | | | | | | |

20MBA14

# First Semester MBA Degree Examination, July/August 2021
## Business Statistics

Time: 3 hrs. Max. Marks:100

**Note:** *1. Answer any FIVE full questions.*
*2. Table values given in the questions.*



1 a. Explain the functions of statistics. **(03 Marks)**

b. Find mean and standard deviation from the following data :

| X | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |
|---|---|---|---|---|---|---|---|---|
| Y | 15 | 30 | 53 | 75 | 100 | 110 | 115 | 125 |

**(07 Marks)**

c. From the prices of X and Y shares, A and B respectively given below state which share is more stable in value?

| Price of Share A | 55 | 54 | 52 | 53 | 56 | 58 | 52 | 50 | 51 | 49 |
|---|---|---|---|---|---|---|---|---|---|---|
| Price of Share B | 108 | 107 | 105 | 105 | 106 | 107 | 104 | 103 | 104 | 101 |

**(10 Marks)**

2 a. Discuss the advantages and limitations of diagrams and graphs. **(03 Marks)**

b. For a certain frequency table, which hrs only been partly reproduced, here the mean was found to be 1.46. Calculate the missing frequency. N = 200.

| No of accidents | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| No of days | 46 | ? | ? | 25 | 10 | 5 |

**(07 Marks)**

c. Calculate Spearman's rank correlation coefficient between advertisement cost and sales from the following data.

| Advertisement cost (1000Rs.) | 39 | 65 | 62 | 90 | 82 | 75 | 25 | 98 | 36 | 78 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sales (lakhs Rs.) | 47 | 53 | 58 | 86 | 62 | 68 | 60 | 91 | 51 | 84 |

**(10 Marks)**

3 a. Explain the significance in measuring dispersion. **(03 Marks)**

b. Discuss the difference between parametric and non-parametric test. **(07 Marks)**

c. Mysuru Mahanagar Palike surveyed the travel preferences of people who travelled to work by train or bus. The initial analysis suggested that in 5 people travelled by train to work. If 5 people are interviewed, what is the probability that,
i) Exactly 3 prefer travelling by train
ii) Less than 3 prefer travelling by train. **(10 Marks)**

4 a. What is Type I and Type II errors. **(03 Marks)**

b. Define the following terms :
i) Independent event
ii) Mutually exclusive event
iii) Equally likely event. **(07 Marks)**

c. Below are given the figures of production (in thousand tons) of a sugar factory :

| Year | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 |
|---|---|---|---|---|---|---|---|
| Production | 77 | 88 | 94 | 85 | 91 | 98 | 90 |

i) Fit a straight line by the method of 'Least Squares' and show the trend values
ii) What is the monthly increase in production? **(10 Marks)**

5 a. What do you mean by correction? Give any two examples of negative correction. **(03 Marks)**

b. The hourly wages of 1000 workmen are normally distributed around a mean of Rs.70 and with a standard deviation of Rs.5. Estimate the number of workers whose hourly wages will be : i) Between Rs.69 and Rs.72 ii) More than Rs.75 iii) Less than Rs.63.

Area under standard normal probability curve :

| Z | 0.1 | 0.2 | 0.3 | 0.4 | 1.0 | 1.4 | 1.5 | 2.0 |
|---|---|---|---|---|---|---|---|---|
| AREA | 0.0398 | 0.0793 | 0.1179 | 0.1554 | 0.3413 | 0.4192 | 0.4332 | 0.4772 |

**(07 Marks)**

c. The sales data of an item in six shops before and after a special promotional campaign are as under : $t_5(0.05) = 2.02$.

| Shops | A | B | C | D | E | F |
|---|---|---|---|---|---|---|
| Before campaign | 53 | 28 | 31 | 48 | 50 | 42 |
| After campaign | 58 | 29 | 30 | 55 | 56 | 45 |

**(10 Marks)**

6 a. Under what conditions binomial distribution tends to Poisson distribution. **(03 Marks)**

b. If 5% of the electric bulbs manufactured by a company are defective, use Poisson distribution to find the probability that in a sample of 100 bulbs.
i) None is defective ii) 5 bulbs will be defective. (Given : $e^{-5} = 0.007$). **(07 Marks)**

c. From the data given below find :
i) The two regression coefficients
ii) The two regression equations
iii) The coefficient of correlation between.
Marks in economics and statistics :

| Marks in economics | 25 | 28 | 35 | 32 | 31 | 36 | 29 | 38 | 34 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks in statistics | 43 | 46 | 49 | 41 | 36 | 32 | 31 | 30 | 33 | 39 |

**(10 Marks)**

7 a. What is mode? Give two examples of mode. **(03 Marks)**

b. Intelligent test of two groups of boys and girls gave the following results :

| | Mean | S.D. | N |
|---|---|---|---|
| Girls | 75 | 15 | 150 |
| Boys | 70 | 20 | 250 |

Is there a significant difference in the mean scores obtained by boys and girls, (Test $\alpha = 1\%$). **(07 Marks)**

c. The following data present the yield in quintals of IONS on ten sub-divisions of equal area of two agricultural plots :

| PLOT – 1 | 6.2 | 5.7 | 6.5 | 6.0 | 6.3 | 5.8 | 5.7 | 6.0 | 6.0 | 5.8 |
|---|---|---|---|---|---|---|---|---|---|---|
| PLOT – 2 | 5.6 | 5.9 | 5.6 | 5.7 | 5.8 | 5.7 | 6.0 | 5.5 | 5.7 | 5.5 |

Test whether two samples taken from two random population having the same variance. (At 5% level for $V_1 = 9$ and $V_2 = 9$ is 3.18). **(10 Marks)**

8 a. Find : i) Interquartile range ii) Quartile deviation
iii) Coefficient of quartile deviation for the following :

| Class interval | 0 – 15 | 15 – 30 | 30 – 45 | 45 – 60 | 60 – 75 | 75 – 90 | 90 –105 |
|---|---|---|---|---|---|---|---|
| Frequency | 8 | 26 | 30 | 45 | 20 | 17 | 4 |

**(10 Marks)**

b. Three products received the following performance.

| Product A | 50 | 62 | 75 | 48 | 65 | $\chi^2_2(0.05) = 5.991$ |
|---|---|---|---|---|---|---|
| Product B | 80 | 95 | 98 | 87 | 90 | |
| Product C | 60 | 45 | 30 | 58 | 57 | |

Use the Kruskal – Wallies test at $\alpha = 0.05$ to determine whether there is a significant difference in the performance rating of products. **(10 Marks)**

* * * * *

CBCS SCHEME

USN | | | | | | | | | | |

20MBA14

# First Semester MBA Degree Examination, Jan./Feb.2021
## Business Statistics

Time: 3 hrs. Max. Marks:100

**Note:** *1. Answer any FOUR full questions from Q1 to Q7.*
*2. Question No.8 is compulsory.*
*3. Use of statistical table is allowed.*



1 a. Discuss the importance of Business Statistics. **(03 Marks)**

b. From the prices of shares find out which is more stable in value. **(07 Marks)**

| X | 35 | 54 | 52 | 53 | 56 | 58 | 52 | 50 | 51 | 49 |
|---|---|---|---|---|---|---|---|---|---|---|
| Y | 108 | 107 | 105 | 105 | 106 | 107 | 104 | 103 | 104 | 101 |

c. Find the three Quartiles 7th decide and 84th percentile from the following data:

| Wages(in Rs.) | 30 – 40 | 40 – 50 | 50 – 60 | 60 – 70 | 70 – 80 | 80 – 90 | 90 – 100 |
|---|---|---|---|---|---|---|---|
| No. of persons | 1 | 3 | 11 | 21 | 43 | 21 | 9 |

**(10 Marks)**

2 a. Mention any three properties of Arithmetic mean. **(03 Marks)**

b. Explain Pictorially scatter diagram and how is it used in predictions? **(07 Marks)**

c. The following data relate to age of employees and the number of days they reported sick in a month. Calculate Karl Pearson's co-efficient of correlation and interpret it.

| Age (years) | 30 | 32 | 35 | 40 | 48 | 50 | 52 | 55 | 57 | 61 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sick Days | 1 | 0 | 2 | 5 | 2 | 4 | 6 | 5 | 7 | 8 |

**(10 Marks)**

3 a. Define Mutually exclusive, equally likely and exhaustive events. **(03 Marks)**

b. A systematic sample of 100 pages was taken from the Oxford dictionary and the observed frequency distribution of foreign words per page was formed to be as follows:

| No. of foreign words (per page) | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Frequency | 48 | 27 | 12 | 7 | 4 | 1 | 1 |

Calculate the expected frequency use in Poisson distribution. **(07 Marks)**

c. From the following data obtain the two regression equation and calculate correlate co-efficient.

| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| Y | 9 | 8 | 10 | 12 | 11 | 13 | 14 | 16 | 15 |

**(10 Marks)**

4 a. What do you mean by Regression Analysis? Give any two uses of it. **(03 Marks)**

b. You have been provided with the figures of production (in 000's tons) of sugar factory.

| Year | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 |
|---|---|---|---|---|---|---|---|
| Production | 77 | 88 | 94 | 85 | 91 | 98 | 90 |

Fit a straight line by the method of least square and find trend values. **(07 Marks)**

c. The income of a group of 10,000 persons were found to be normally distributed with mean Rs.7500 p.m. and standard deviation is Rs.500/-. Show that of this group about 95% has income exceeding Rs.6680 and only 5% had income exceeding 8320. What was the lowest income among the richest 100? **(10 Marks)**

5 a. Define binomial distribution. **(03 Marks)**

b. Particulars of regarding the income of two villages are given below:

| | Village 'X' | Village 'Y' |
|---|---|---|
| Number of people | 600 | 500 |
| Average income (Rs.) | 175 | 186 |
| Variance | 100 | 81 |

(i) In which village the variate in income is greater?

(ii) What is the combined standard deviation of the village 'X' and village 'Y' put together? **(07 Marks)**

c. Calculate Spearman's Rank correlation co-efficient between advertisement cost and sales from the following data:

| Advertisement cost (000 Rs.) | 39 | 65 | 62 | 90 | 82 | 75 | 25 | 98 | 36 | 78 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sales (lakhs Rs.) | 47 | 53 | 58 | 86 | 62 | 68 | 60 | 91 | 51 | 84 |

**(10 Marks)**

6 a. What are the different measures of dispersion? Explain. **(03 Marks)**

b. Explain the different components of time series. **(07 Marks)**

c. One forth of the first year student admitted to a Bangalore college are out of state students. If the students are assigned at random to the dormitories, 3 to a room, what is the probability that in one room.

(i) At least 2 of the 3 roommates are out of state students.

(ii) At most 2 of the 3 roommates are out of state students. **(10 Marks)**

7 a. Define normal distribution. **(03 Marks)**

b. Explain the characteristics of Good Hypothesis. **(07 Marks)**

c. Define Hypothesis. Describe the formulation of Hypothesis with flow process chart. **(10 Marks)**

8 a. Write briefly on the following terms:

(i) Type I error.

(ii) Type II error. **(06 Marks)**

b. The daily wages of 1000 workmen are normally distributed around a mean of Rs.70 and with a standard deviation of Rs.5. Estimate the number of workers whose daily wages will be:

(i) Between Rs.70 and 72

(ii) Between Rs.69 and 72

(iii) More than Rs.75

(iv) Less than Rs.63

(v) More than Rs.80

(vi) Also estimate the lowest daily wages of the 100 highest paid workers. **(14 Marks)**

* * * * *

CBCS SCHEME

20MBA14

USN | 1 | R | N | 2 | 1 | 6 | A | 1 | S | 0 |

# First Semester MBA Degree Examination, July/August 2022
## Business Statistics

Time: 3 hrs. Max. Marks:100

**Note:** *1. Answer any FOUR full questions from Q1 to Q7.*
*2. Question No.8 is compulsory.*
*3. Use of statistical table is allowed.*



1 a. Why statistics is required for business and management. **(03 Marks)**

b. You are working as a purchase manager for company. The following information has been supplied to you by two manufactures of electric bulbs.

| | Company A | Company B |
|---|---|---|
| Mean life (in hrs) | 1300 | 1248 |
| SD (in hrs) | 82 | 93 |
| Sample size | 100 | 100 |

**(07 Marks)**

c. The following data relate to age of employees and the number of days they reported sick in a month. Calculate Karl Pearson's coefficient of correlation and interpret it.

| Age (years) | 30 | 32 | 35 | 40 | 48 | 50 | 52 | 55 | 57 | 61 |
|---|---|---|---|---|---|---|---|---|---|---|
| Sick (days) | 1 | 0 | 2 | 5 | 2 | 4 | 6 | 5 | 7 | 8 |

**(10 Marks)**

2 a. What is Bayesian Decision Rule? **(03 Marks)**

b. The following distribution gives the pattern of overtime work done by 100 employees of a company. Find the mean and SD.

| Overtime (hrs) | 10 – 15 | 15 – 20 | 20 – 25 | 25 – 30 | 30 – 35 | 35 – 40 |
|---|---|---|---|---|---|---|
| No. of employees | 11 | 20 | 35 | 20 | 8 | 6 |

**(07 Marks)**

c. The following table relates to the tourist arrivals (in millions) during 2001 to 2007 in India.

| Year | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|---|---|---|
| Tourist arrive (in million) | 18 | 20 | 23 | 25 | 24 | 28 | 30 |

Fit a straight line trend by the method of least squares an estimate the number of tourist that would arrive in the year 2011. **(10 Marks)**

3 a. Define Correlation and Regression. **(03 Marks)**

b. What are the Components of Time Series Analysis? **(07 Marks)**

c. The score of two batsman A and B in 10 innings during a certain season are given below. Find which of the two batsmen A and B is more consistent in scoring.

| A | 32 | 28 | 47 | 63 | 71 | 39 | 10 | 60 | 96 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|
| B | 19 | 31 | 48 | 83 | 67 | 90 | 10 | 62 | 40 | 80 |

**(10 Marks)**

4 a. Define Hypothesis? (03 Marks)

b. The following table given the numbers days in a 50 day period during automobile accidents occurred in a city. Find a Poisson distribution.

| Number of Accidents | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| Number of days | 21 | 18 | 7 | 3 | 1 |

c. From the following data find : (07 Marks)

i) The two regression coefficient?
ii) Two regression equations?
iii) The coefficient of correlation between the marks in economics and statistics?
iv) Most likely marks in statistics when marks in economics are 30.

| Marks in Economics | 25 | 28 | 35 | 32 | 31 | 36 | 29 | 38 | 34 | 32 |
|---|---|---|---|---|---|---|---|---|---|---|
| Marks in Statistics | 43 | 46 | 49 | 41 | 36 | 32 | 31 | 30 | 33 | 39 |

(10 Marks)

5 a. Mention various measures of central tendency. (03 Marks)

b. Construct a scatter plot for a data obtained in a study of age and blood pressure of 6 randomly selected people. The data are shown in the table.

| Subject | Age | Blood pressure |
|---|---|---|
| A | 43 | 128 |
| B | 45 | 120 |
| C | 56 | 135 |
| D | 61 | 143 |
| E | 67 | 141 |
| F | 70 | 152 |

c. A hospital has 20 kidney dialysis machines and the chance of any one of them malfunctioning during any day is 0.02. You are required to find the probability that exactly 3 machines will be out of service on the same day then. (07 Marks)

i) Can we use the binominal formula to find out this probability?
If yes, calculate the probability?
ii) Can we use the Poisson formula to find out this probability?
If yes, calculate the probability?

(10 Marks)

6 a. What is Time Series Analysis? (03 Marks)

b. Explain the steps in formulating of Hypothesis. (07 Marks)

c. Out of a sample of 120 persons in a village, 76 persons were administered a new drug for preventing influenza and out of them 24 persons were a attacked by influenza out of those who were not administered the new drugs, 12 persons were not affected by influenza. Prepare 2 × 2 table showing and expected frequencies. Chi square test for finding the new drug is effective or not. [At 5% level of one degree of freedom, the value of $X^2$ is 3.84]. (10 Marks)

7 a. What is Normal distribution? (03 Marks)

b. Discuss the various types of correlation with example. (07 Marks)

c. A brokerage survey reports that 30% of individual investors have used a discount broker, i.e. one which does not charge the full commission. In a random sample of 9 individuals, what is the probability that :

i) Exactly two of the sampled individuals have used a discount broker?

ii) Not more than 3 have used a discount broker

iii) Atleast 3 of them have used a discount broker. (10 Marks)

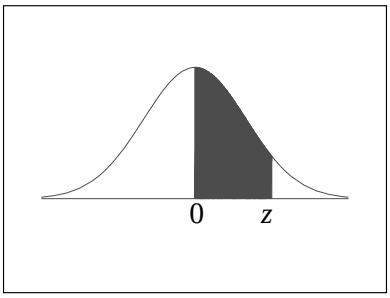
8 To study the performance of three detergents and three different water temperatures the following whiteness readings were obtained with specially designed equipment :

| Water temp | Detergent A | Detergent B | Detergent C |
|---|---|---|---|
| Cold water | 57 | 55 | 67 |
| Warm water | 49 | 52 | 68 |
| Hot water | 54 | 46 | 58 |

Perform a two way analysis of variance, using 5% level of significance. (20 Marks)
[Given : F 5% = 6.940].

* * * * *

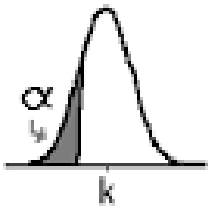
# Standard Normal Distribution Table




| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | .0000 | .0040 | .0080 | .0120 | .0160 | .0199 | .0239 | .0279 | .0319 | .0359 |
| 0.1 | .0398 | .0438 | .0478 | .0517 | .0557 | .0596 | .0636 | .0675 | .0714 | .0753 |
| 0.2 | .0793 | .0832 | .0871 | .0910 | .0948 | .0987 | .1026 | .1064 | .1103 | .1141 |
| 0.3 | .1179 | .1217 | .1255 | .1293 | .1331 | .1368 | .1406 | .1443 | .1480 | .1517 |
| 0.4 | .1554 | .1591 | .1628 | .1664 | .1700 | .1736 | .1772 | .1808 | .1844 | .1879 |
| 0.5 | .1915 | .1950 | .1985 | .2019 | .2054 | .2088 | .2123 | .2157 | .2190 | .2224 |
| 0.6 | .2257 | .2291 | .2324 | .2357 | .2389 | .2422 | .2454 | .2486 | .2517 | .2549 |
| 0.7 | .2580 | .2611 | .2642 | .2673 | .2704 | .2734 | .2764 | .2794 | .2823 | .2852 |
| 0.8 | .2881 | .2910 | .2939 | .2967 | .2995 | .3023 | .3051 | .3078 | .3106 | .3133 |
| 0.9 | .3159 | .3186 | .3212 | .3238 | .3264 | .3289 | .3315 | .3340 | .3365 | .3389 |
| 1.0 | .3413 | .3438 | .3461 | .3485 | .3508 | .3531 | .3554 | .3577 | .3599 | .3621 |
| 1.1 | .3643 | .3665 | .3686 | .3708 | .3729 | .3749 | .3770 | .3790 | .3810 | .3830 |
| 1.2 | .3849 | .3869 | .3888 | .3907 | .3925 | .3944 | .3962 | .3980 | .3997 | .4015 |
| 1.3 | .4032 | .4049 | .4066 | .4082 | .4099 | .4115 | .4131 | .4147 | .4162 | .4177 |
| 1.4 | .4192 | .4207 | .4222 | .4236 | .4251 | .4265 | .4279 | .4292 | .4306 | .4319 |
| 1.5 | .4332 | .4345 | .4357 | .4370 | .4382 | .4394 | .4406 | .4418 | .4429 | .4441 |
| 1.6 | .4452 | .4463 | .4474 | .4484 | .4495 | .4505 | .4515 | .4525 | .4535 | .4545 |
| 1.7 | .4554 | .4564 | .4573 | .4582 | .4591 | .4599 | .4608 | .4616 | .4625 | .4633 |
| 1.8 | .4641 | .4649 | .4656 | .4664 | .4671 | .4678 | .4686 | .4693 | .4699 | .4706 |
| 1.9 | .4713 | .4719 | .4726 | .4732 | .4738 | .4744 | .4750 | .4756 | .4761 | .4767 |
| 2.0 | .4772 | .4778 | .4783 | .4788 | .4793 | .4798 | .4803 | .4808 | .4812 | .4817 |
| 2.1 | .4821 | .4826 | .4830 | .4834 | .4838 | .4842 | .4846 | .4850 | .4854 | .4857 |
| 2.2 | .4861 | .4864 | .4868 | .4871 | .4875 | .4878 | .4881 | .4884 | .4887 | .4890 |
| 2.3 | .4893 | .4896 | .4898 | .4901 | .4904 | .4906 | .4909 | .4911 | .4913 | .4916 |
| 2.4 | .4918 | .4920 | .4922 | .4925 | .4927 | .4929 | .4931 | .4932 | .4934 | .4936 |
| 2.5 | .4938 | .4940 | .4941 | .4943 | .4945 | .4946 | .4948 | .4949 | .4951 | .4952 |
| 2.6 | .4953 | .4955 | .4956 | .4957 | .4959 | .4960 | .4961 | .4962 | .4963 | .4964 |
| 2.7 | .4965 | .4966 | .4967 | .4968 | .4969 | .4970 | .4971 | .4972 | .4973 | .4974 |
| 2.8 | .4974 | .4975 | .4976 | .4977 | .4977 | .4978 | .4979 | .4979 | .4980 | .4981 |
| 2.9 | .4981 | .4982 | .4982 | .4983 | .4984 | .4984 | .4985 | .4985 | .4986 | .4986 |
| 3.0 | .4987 | .4987 | .4987 | .4988 | .4988 | .4989 | .4989 | .4989 | .4990 | .4990 |
| 3.1 | .4990 | .4991 | .4991 | .4991 | .4992 | .4992 | .4992 | .4992 | .4993 | .4993 |
| 3.2 | .4993 | .4993 | .4994 | .4994 | .4994 | .4994 | .4994 | .4995 | .4995 | .4995 |
| 3.3 | .4995 | .4995 | .4995 | .4996 | .4996 | .4996 | .4996 | .4996 | .4996 | .4997 |
| 3.4 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4997 | .4998 |
| 3.5 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 | .4998 |

## Z Test Table Values

**One Tail (Left)***

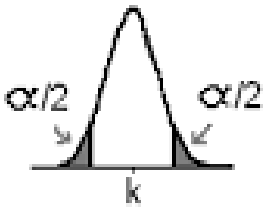



$H_0: \mu = k$

$H_1: \mu < k$

| α | z critical |
|---|---|
| 0.10 | -1.28 |
| 0.05 | -1.65 |
| 0.01 | -2.33 |

**Two Tail✓**

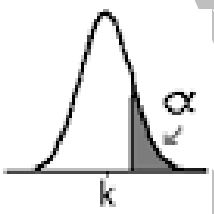



$H_0: \mu = k$

$H_1: \mu \neq k$

| α | z critical |
|---|---|
| 0.10 | ± 1.65 |
| 0.05 | ± 1.96 |
| 0.01 | ± 2.58 |

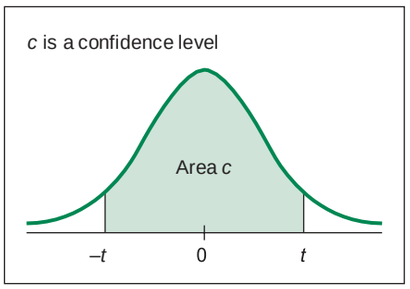
**One Tail (Right)✓**

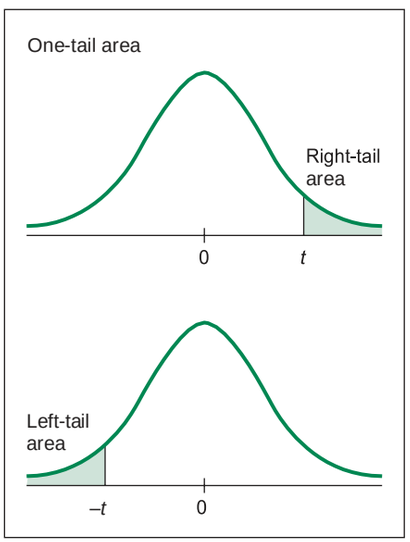


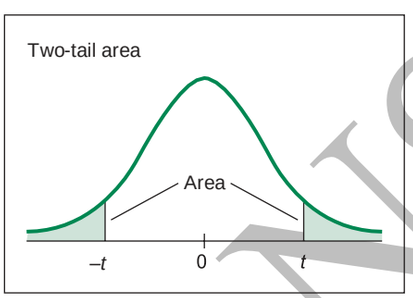

$H_0: \mu = k$

$H_1: \mu > k$

| α | z critical |
|---|---|
| 0.10 | 1.28 |
| 0.05 | 1.65 |
| 0.01 | 2.33 |




Critical Values for Student's *t* Distribution

| one-tail area | 0.250 | 0.125 | 0.100 | 0.075 | 0.050 | 0.025 | 0.010 | 0.005 | 0.0005 |
|---|---|---|---|---|---|---|---|---|---|
| two-tail area | 0.500 | 0.250 | 0.200 | 0.150 | 0.100 | 0.050 | 0.020 | 0.010 | 0.0010 |
| *d.f.* *c* | 0.500 | 0.750 | 0.800 | 0.850 | 0.900 | 0.950 | 0.980 | 0.990 | 0.999 |
| 1 | 1.000 | 2.414 | 3.078 | 4.165 | 6.314 | 12.706 | 31.821 | 63.657 | 636.619 |
| 2 | 0.816 | 1.604 | 1.886 | 2.282 | 2.920 | 4.303 | 6.965 | 9.925 | 31.599 |
| 3 | 0.765 | 1.423 | 1.638 | 1.924 | 2.353 | 3.182 | 4.541 | 5.841 | 12.924 |
| 4 | 0.741 | 1.344 | 1.533 | 1.778 | 2.132 | 2.776 | 3.747 | 4.604 | 8.610 |
| 5 | 0.727 | 1.301 | 1.476 | 1.699 | 2.015 | 2.571 | 3.365 | 4.032 | 6.869 |
| 6 | 0.718 | 1.273 | 1.440 | 1.650 | 1.943 | 2.447 | 3.143 | 3.707 | 5.959 |
| 7 | 0.711 | 1.254 | 1.415 | 1.617 | 1.895 | 2.365 | 2.998 | 3.499 | 5.408 |
| 8 | 0.706 | 1.240 | 1.397 | 1.592 | 1.860 | 2.306 | 2.896 | 3.355 | 5.041 |
| 9 | 0.703 | 1.230 | 1.383 | 1.574 | 1.833 | 2.262 | 2.821 | 3.250 | 4.781 |
| 10 | 0.700 | 1.221 | 1.372 | 1.559 | 1.812 | 2.228 | 2.764 | 3.169 | 4.587 |
| 11 | 0.697 | 1.214 | 1.363 | 1.548 | 1.796 | 2.201 | 2.718 | 3.106 | 4.437 |
| 12 | 0.695 | 1.209 | 1.356 | 1.538 | 1.782 | 2.179 | 2.681 | 3.055 | 4.318 |
| 13 | 0.694 | 1.204 | 1.350 | 1.530 | 1.771 | 2.160 | 2.650 | 3.012 | 4.221 |
| 14 | 0.692 | 1.200 | 1.345 | 1.523 | 1.761 | 2.145 | 2.624 | 2.977 | 4.140 |
| 15 | 0.691 | 1.197 | 1.341 | 1.517 | 1.753 | 2.131 | 2.602 | 2.947 | 4.073 |
| 16 | 0.690 | 1.194 | 1.337 | 1.512 | 1.746 | 2.120 | 2.583 | 2.921 | 4.015 |
| 17 | 0.689 | 1.191 | 1.333 | 1.508 | 1.740 | 2.110 | 2.567 | 2.898 | 3.965 |
| 18 | 0.688 | 1.189 | 1.330 | 1.504 | 1.734 | 2.101 | 2.552 | 2.878 | 3.922 |
| 19 | 0.688 | 1.187 | 1.328 | 1.500 | 1.729 | 2.093 | 2.539 | 2.861 | 3.883 |
| 20 | 0.687 | 1.185 | 1.325 | 1.497 | 1.725 | 2.086 | 2.528 | 2.845 | 3.850 |
| 21 | 0.686 | 1.183 | 1.323 | 1.494 | 1.721 | 2.080 | 2.518 | 2.831 | 3.819 |
| 22 | 0.686 | 1.182 | 1.321 | 1.492 | 1.717 | 2.074 | 2.508 | 2.819 | 3.792 |
| 23 | 0.685 | 1.180 | 1.319 | 1.489 | 1.714 | 2.069 | 2.500 | 2.807 | 3.768 |
| 24 | 0.685 | 1.179 | 1.318 | 1.487 | 1.711 | 2.064 | 2.492 | 2.797 | 3.745 |
| 25 | 0.684 | 1.198 | 1.316 | 1.485 | 1.708 | 2.060 | 2.485 | 2.787 | 3.725 |
| 26 | 0.684 | 1.177 | 1.315 | 1.483 | 1.706 | 2.056 | 2.479 | 2.779 | 3.707 |
| 27 | 0.684 | 1.176 | 1.314 | 1.482 | 1.703 | 2.052 | 2.473 | 2.771 | 3.690 |
| 28 | 0.683 | 1.175 | 1.313 | 1.480 | 1.701 | 2.048 | 2.467 | 2.763 | 3.674 |
| 29 | 0.683 | 1.174 | 1.311 | 1.479 | 1.699 | 2.045 | 2.462 | 2.756 | 3.659 |
| 30 | 0.683 | 1.173 | 1.310 | 1.477 | 1.697 | 2.042 | 2.457 | 2.750 | 3.646 |
| 35 | 0.682 | 1.170 | 1.306 | 1.472 | 1.690 | 2.030 | 2.438 | 2.724 | 3.591 |
| 40 | 0.681 | 1.167 | 1.303 | 1.468 | 1.684 | 2.021 | 2.423 | 2.704 | 3.551 |
| 45 | 0.680 | 1.165 | 1.301 | 1.465 | 1.679 | 2.014 | 2.412 | 2.690 | 3.520 |
| 50 | 0.679 | 1.164 | 1.299 | 1.462 | 1.676 | 2.009 | 2.403 | 2.678 | 3.496 |
| 60 | 0.679 | 1.162 | 1.296 | 1.458 | 1.671 | 2.000 | 2.390 | 2.660 | 3.460 |
| 70 | 0.678 | 1.160 | 1.294 | 1.456 | 1.667 | 1.994 | 2.381 | 2.648 | 3.435 |
| 80 | 0.678 | 1.159 | 1.292 | 1.453 | 1.664 | 1.990 | 2.374 | 2.639 | 3.416 |
| 100 | 0.677 | 1.157 | 1.290 | 1.451 | 1.660 | 1.984 | 2.364 | 2.626 | 3.390 |
| 500 | 0.675 | 1.152 | 1.283 | 1.442 | 1.648 | 1.965 | 2.334 | 2.586 | 3.310 |
| 1000 | 0.675 | 1.151 | 1.282 | 1.441 | 1.646 | 1.962 | 2.330 | 2.581 | 3.300 |
| 00 | 0.674 | 1.150 | 1.282 | 1.440 | 1.645 | 1.960 | 2.326 | 2.576 | 3.291 |

For degrees of freedom *d.f.* not in the table, use the closest *d.f.* that is *smaller.*

## F Test Table Values (5% Level of Significance)

Percentage points of *Fisher*'s distribution

$f_{0.05,v_1,v_2}$

| $v_2$ \ $v_1$ | Degrees of freedom of the numerator ($v_1$) | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
| 1 | 161.4 | 199.5 | 215.7 | 224.6 | 230.2 | 234.0 | 236.8 | 238.9 | 240.5 | 241.9 | 243.9 | 245.9 | 248 | 249.1 | 250.1 | 251.1 | 252.2 | 253.3 | 254.3 |
| 2 | 18.51 | 19.00 | 19.16 | 19.25 | 19.30 | 19.33 | 19.35 | 19.37 | 19.38 | 19.40 | 19.41 | 19.43 | 19.45 | 19.45 | 19.46 | 19.47 | 19.48 | 19.49 | 19.50 |
| 3 | 10.13 | 9.55 | 9.28 | 9.12 | 9.01 | 8.94 | 8.89 | 8.85 | 8.81 | 8.79 | 8.74 | 8.70 | 8.66 | 8.64 | 8.62 | 8.59 | 8.57 | 8.55 | 8.53 |
| 4 | 7.71 | 6.94 | 6.59 | 6.39 | 6.26 | 6.16 | 6.09 | 6.04 | 6.00 | 5.96 | 5.91 | 5.86 | 5.80 | 5.77 | 5.75 | 5.72 | 5.69 | 5.66 | 5.63 |
| 5 | 6.61 | 5.79 | 5.41 | 5.19 | 5.05 | 4.95 | 4.88 | 4.82 | 4.77 | 4.74 | 4.68 | 4.62 | 4.56 | 4.53 | 4.50 | 4.46 | 4.43 | 4.40 | 4.36 |
| 6 | 5.99 | 5.14 | 4.76 | 4.53 | 4.39 | 4.28 | 4.21 | 4.15 | 4.10 | 4.06 | 4.00 | 3.94 | 3.87 | 3.84 | 3.81 | 3.77 | 3.74 | 3.70 | 3.67 |
| 7 | 5.59 | 4.74 | 4.35 | 4.12 | 3.97 | 3.87 | 3.79 | 3.73 | 3.68 | 3.64 | 3.57 | 3.51 | 3.44 | 3.41 | 3.38 | 3.34 | 3.30 | 3.27 | 3.23 |
| 8 | 5.32 | 4.46 | 4.07 | 3.84 | 3.69 | 3.58 | 3.50 | 3.44 | 3.39 | 3.35 | 3.28 | 3.22 | 3.15 | 3.12 | 3.08 | 3.04 | 3.01 | 2.97 | 2.93 |
| 9 | 5.12 | 4.26 | 3.86 | 3.63 | 3.48 | 3.37 | 3.29 | 3.23 | 3.18 | 3.14 | 3.07 | 3.01 | 2.94 | 2.90 | 2.86 | 2.83 | 2.79 | 2.75 | 2.71 |
| 10 | 4.96 | 4.10 | 3.71 | 3.48 | 3.33 | 3.22 | 3.14 | 3.07 | 3.02 | 2.98 | 2.91 | 2.85 | 2.77 | 2.74 | 2.70 | 2.66 | 2.62 | 2.58 | 2.54 |
| 11 | 4.84 | 3.98 | 3.59 | 3.36 | 3.20 | 3.09 | 3.01 | 2.95 | 2.90 | 2.85 | 2.79 | 2.72 | 2.65 | 2.61 | 2.57 | 2.53 | 2.49 | 2.45 | 2.40 |
| 12 | 4.75 | 3.89 | 3.49 | 3.26 | 3.11 | 3.00 | 2.91 | 2.85 | 2.80 | 2.75 | 2.69 | 2.62 | 2.54 | 2.51 | 2.47 | 2.43 | 2.38 | 2.34 | 2.30 |
| 13 | 4.67 | 3.81 | 3.41 | 3.18 | 3.03 | 2.92 | 2.83 | 2.77 | 2.71 | 2.67 | 2.60 | 2.53 | 2.46 | 2.42 | 2.38 | 2.34 | 2.30 | 2.25 | 2.21 |
| 14 | 4.60 | 3.74 | 3.34 | 3.11 | 2.96 | 2.85 | 2.76 | 2.70 | 2.65 | 2.60 | 2.53 | 2.46 | 2.39 | 2.35 | 2.31 | 2.27 | 2.22 | 2.18 | 2.13 |
| 15 | 4.54 | 3.68 | 3.29 | 3.06 | 2.90 | 2.79 | 2.71 | 2.64 | 2.59 | 2.54 | 2.48 | 2.40 | 2.33 | 2.29 | 2.25 | 2.20 | 2.16 | 2.11 | 2.07 |
| 16 | 4.49 | 3.63 | 3.24 | 3.01 | 2.85 | 2.74 | 2.66 | 2.59 | 2.54 | 2.49 | 2.42 | 2.35 | 2.28 | 2.24 | 2.19 | 2.15 | 2.11 | 2.06 | 2.01 |
| 17 | 4.45 | 3.59 | 3.20 | 2.96 | 2.81 | 2.70 | 2.61 | 2.55 | 2.49 | 2.45 | 2.38 | 2.31 | 2.23 | 2.19 | 2.15 | 2.10 | 2.06 | 2.01 | 1.96 |
| 18 | 4.41 | 3.55 | 3.16 | 2.93 | 2.77 | 2.66 | 2.58 | 2.51 | 2.46 | 2.41 | 2.34 | 2.27 | 2.19 | 2.15 | 2.11 | 2.06 | 2.02 | 1.97 | 1.92 |
| 19 | 4.38 | 3.52 | 3.13 | 2.90 | 2.64 | 2.63 | 2.54 | 2.48 | 2.42 | 2.38 | 2.31 | 2.23 | 2.16 | 2.11 | 2.07 | 2.03 | 1.98 | 1.93 | 1.88 |
| 20 | 4.35 | 3.49 | 3.10 | 2.87 | 2.71 | 2.60 | 2.51 | 2.45 | 2.39 | 2.35 | 2.28 | 2.20 | 2.12 | 2.08 | 2.04 | 1.99 | 1.95 | 1.90 | 1.84 |
| 21 | 4.32 | 3.47 | 3.07 | 2.84 | 2.68 | 2.57 | 2.49 | 2.42 | 2.37 | 2.32 | 2.25 | 2.18 | 2.10 | 2.05 | 2.01 | 1.96 | 1.92 | 1.87 | 1.81 |
| 22 | 4.30 | 3.44 | 3.05 | 2.82 | 2.66 | 2.55 | 2.46 | 2.40 | 2.34 | 2.30 | 2.23 | 2.15 | 2.07 | 2.03 | 1.98 | 1.94 | 1.89 | 1.84 | 1.78 |
| 23 | 4.28 | 3.42 | 3.03 | 2.80 | 2.64 | 2.53 | 2.44 | 2.37 | 2.32 | 2.27 | 2.20 | 2.31 | 2.05 | 2.01 | 1.96 | 1.91 | 1.86 | 1.81 | 1.76 |
| 24 | 4.26 | 3.40 | 3.01 | 2.78 | 2.62 | 2.51 | 2.42 | 2.36 | 2.30 | 2.25 | 2.18 | 2.11 | 2.03 | 1.98 | 1.94 | 1.89 | 1.84 | 1.79 | 1.73 |
| 25 | 4.24 | 3.39 | 2.99 | 2.76 | 2.60 | 2.49 | 2.40 | 2.34 | 2.28 | 2.24 | 2.16 | 2.09 | 2.01 | 1.96 | 1.92 | 1.87 | 1.82 | 1.77 | 1.71 |
| 26 | 4.23 | 3.37 | 2.98 | 2.74 | 2.59 | 2.47 | 2.39 | 2.32 | 2.27 | 2.22 | 2.15 | 2.07 | 1.99 | 1.95 | 1.90 | 1.85 | 1.80 | 1.75 | 1.69 |
| 27 | 4.21 | 3.35 | 2.96 | 2.73 | 2.57 | 2.46 | 2.37 | 2.31 | 2.25 | 2.20 | 2.13 | 2.06 | 1.97 | 1.93 | 1.88 | 1.84 | 1.79 | 1.73 | 1.67 |
| 28 | 4.20 | 3.34 | 2.95 | 2.71 | 2.56 | 2.45 | 2.36 | 2.29 | 2.24 | 2.19 | 2.12 | 2.04 | 1.96 | 1.91 | 1.87 | 1.82 | 1.77 | 1.71 | 1.65 |
| 29 | 4.18 | 3.33 | 2.93 | 2.70 | 2.55 | 2.43 | 2.35 | 2.28 | 2.22 | 2.18 | 2.10 | 2.03 | 1.94 | 1.90 | 1.85 | 1.81 | 1.75 | 1.70 | 1.64 |
| 30 | 4.17 | 3.32 | 2.92 | 2.69 | 2.53 | 2.42 | 2.33 | 2.27 | 2.21 | 2.16 | 2.09 | 2.01 | 1.93 | 1.89 | 1.84 | 1.79 | 1.74 | 1.68 | 1.62 |
| 40 | 4.08 | 3.23 | 2.84 | 2.61 | 2.45 | 2.34 | 2.25 | 2.18 | 2.12 | 2.08 | 2.00 | 1.92 | 1.84 | 1.79 | 1.74 | 1.69 | 1.64 | 1.58 | 1.51 |
| 60 | 4.00 | 3.15 | 2.76 | 2.53 | 2.37 | 2.25 | 2.17 | 2.10 | 2.04 | 1.99 | 1.92 | 1.84 | 1.75 | 1.70 | 1.65 | 1.59 | 1.53 | 1.47 | 1.39 |
| 120 | 3.92 | 3.07 | 2.68 | 2.45 | 2.29 | 2.17 | 2.09 | 2.02 | 1.96 | 1.91 | 1.83 | 1.75 | 1.66 | 1.61 | 1.55 | 1.55 | 1.43 | 1.35 | 1.25 |
| ∞ | 3.84 | 3.00 | 2.60 | 2.37 | 2.21 | 2.10 | 2.01 | 1.94 | 1.88 | 1.83 | 1.75 | 1.67 | 1.57 | 1.52 | 1.46 | 1.39 | 1.32 | 1.22 | 1.00 |

Degrees of freedom of the denominator ($v_2$)

# F Test Table Values (1% Level of Significance)

Percentage points of *Fisher*'s distribution

$f_{0.01,v_1,v_2}$

| $v_2$ \ $v_1$ | Degrees of freedom of the numerator ($v_1$) | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
| 1 | 4052 | 4999.5 | 5403 | 5625 | 5764 | 5859 | 5928 | 5982 | 6022 | 6056 | 6106 | 6157 | 6209 | 6235 | 6261 | 6287 | 6313 | 6339 | 6366 |
| 2 | 98.50 | 99.00 | 99.17 | 99.25 | 99.30 | 99.33 | 99.36 | 99.37 | 99.39 | 99.40 | 99.42 | 99.43 | 99.45 | 99.46 | 99.47 | 99.47 | 99.48 | 99.49 | 99.50 |
| 3 | 34.12 | 30.82 | 29.46 | 28.71 | 28.24 | 27.91 | 27.67 | 27.49 | 27.35 | 27.23 | 27.05 | 26.87 | 26.69 | 26.00 | 26.50 | 26.41 | 26.32 | 26.22 | 26.13 |
| 4 | 21.20 | 18.00 | 16.69 | 15.89 | 15.52 | 15.21 | 14.98 | 14.80 | 14.66 | 14.55 | 14.37 | 14.20 | 14.02 | 13.93 | 13.84 | 13.75 | 13.65 | 13.56 | 13.46 |
| 5 | 16.26 | 13.27 | 12.06 | 11.39 | 10.97 | 10.67 | 10.46 | 10.29 | 10.16 | 10.05 | 9.89 | 9.72 | 9.55 | 9.47 | 9.38 | 9.29 | 9.20 | 9.11 | 9.02 |
| 6 | 13.75 | 10.92 | 9.78 | 9.15 | 8.75 | 8.47 | 8.26 | 8.10 | 7.98 | 7.87 | 7.72 | 7.56 | 7.40 | 7.31 | 7.23 | 7.14 | 7.06 | 6.97 | 6.88 |
| 7 | 12.25 | 9.55 | 8.45 | 7.85 | 7.46 | 7.19 | 6.99 | 6.84 | 6.72 | 6.62 | 6.47 | 6.31 | 6.16 | 6.07 | 5.99 | 5.91 | 5.82 | 5.74 | 6.65 |
| 8 | 11.26 | 8.65 | 7.59 | 7.01 | 6.63 | 6.37 | 6.18 | 6.03 | 5.91 | 5.81 | 5.67 | 5.52 | 5.36 | 5.28 | 5.20 | 5.12 | 5.02 | 4.95 | 4.46 |
| 9 | 10.56 | 8.02 | 6.99 | 6.42 | 6.06 | 5.80 | 5.61 | 5.47 | 5.35 | 5.26 | 5.11 | 4.96 | 4.81 | 4.73 | 4.65 | 4.57 | 4.48 | 4.40 | 4.31 |
| 10 | 10.04 | 7.56 | 6.55 | 5.99 | 5.64 | 5.39 | 5.20 | 5.06 | 4.94 | 4.85 | 4.71 | 4.56 | 4.41 | 4.33 | 4.25 | 4.17 | 4.08 | 4.00 | 3.91 |
| 11 | 9.65 | 7.21 | 6.22 | 5.67 | 5.32 | 5.07 | 4.89 | 4.74 | 4.63 | 4.54 | 4.40 | 4.25 | 4.10 | 4.02 | 3.94 | 3.86 | 3.78 | 3.69 | 3.60 |
| 12 | 9.33 | 6.93 | 5.95 | 5.41 | 5.06 | 4.82 | 4.64 | 4.50 | 4.39 | 4.30 | 4.16 | 4.01 | 3.86 | 3.78 | 3.70 | 3.62 | 3.54 | 3.45 | 3.36 |
| 13 | 9.07 | 6.70 | 5.74 | 5.21 | 4.86 | 4.63 | 4.44 | 4.30 | 4.19 | 4.10 | 3.96 | 3.82 | 3.66 | 3.59 | 3.51 | 3.43 | 3.34 | 3.25 | 3.17 |
| 14 | 8.86 | 6.51 | 5.56 | 5.04 | 4.69 | 4.46 | 4.28 | 4.14 | 4.03 | 3.94 | 3.80 | 3.66 | 3.51 | 3.43 | 3.35 | 3.27 | 3.18 | 3.09 | 3.00 |
| 15 | 8.68 | 6.36 | 5.42 | 4.89 | 4.36 | 4.32 | 4.14 | 4.00 | 3.89 | 3.80 | 3.67 | 3.52 | 3.37 | 3.29 | 3.21 | 3.13 | 3.05 | 2.96 | 2.87 |
| 16 | 8.53 | 6.23 | 5.29 | 4.77 | 4.44 | 4.20 | 4.03 | 3.89 | 3.78 | 3.69 | 3.55 | 3.41 | 3.26 | 3.18 | 3.10 | 3.02 | 2.93 | 2.84 | 2.75 |
| 17 | 8.40 | 6.11 | 5.18 | 4.67 | 4.34 | 4.10 | 3.93 | 3.79 | 3.68 | 3.59 | 3.46 | 3.31 | 3.16 | 3.08 | 3.00 | 2.92 | 2.83 | 2.75 | 2.65 |
| 18 | 8.29 | 6.01 | 5.09 | 4.58 | 4.25 | 4.01 | 3.84 | 3.71 | 3.60 | 3.51 | 3.37 | 3.23 | 3.08 | 3.00 | 2.92 | 2.84 | 2.75 | 2.66 | 2.57 |
| 19 | 8.18 | 5.93 | 5.01 | 4.50 | 4.17 | 3.94 | 3.77 | 3.63 | 3.52 | 3.43 | 3.30 | 3.15 | 3.00 | 2.92 | 2.84 | 2.76 | 2.67 | 2.58 | 2.59 |
| 20 | 8.10 | 5.85 | 4.94 | 4.43 | 4.10 | 3.87 | 3.70 | 3.56 | 3.46 | 3.37 | 3.23 | 3.09 | 2.94 | 2.86 | 2.78 | 2.69 | 2.61 | 2.52 | 2.42 |
| 21 | 8.02 | 5.78 | 4.87 | 4.37 | 4.04 | 3.81 | 3.64 | 3.51 | 3.40 | 3.31 | 3.17 | 3.03 | 2.88 | 2.80 | 2.72 | 2.64 | 2.55 | 2.46 | 2.36 |
| 22 | 7.95 | 5.72 | 4.82 | 4.31 | 3.99 | 3.76 | 3.59 | 3.45 | 3.35 | 3.26 | 3.12 | 2.98 | 2.83 | 2.75 | 2.67 | 2.58 | 2.50 | 2.40 | 2.31 |
| 23 | 7.88 | 5.66 | 4.76 | 4.26 | 3.94 | 3.71 | 3.54 | 3.41 | 3.30 | 3.21 | 3.07 | 2.93 | 2.78 | 2.70 | 2.62 | 2.54 | 2.45 | 2.35 | 2.26 |
| 24 | 7.82 | 5.61 | 4.72 | 4.22 | 3.90 | 3.67 | 3.50 | 3.36 | 3.26 | 3.17 | 3.03 | 2.89 | 2.74 | 2.66 | 2.58 | 2.49 | 2.40 | 2.31 | 2.21 |
| 25 | 7.77 | 5.57 | 4.68 | 4.18 | 3.85 | 3.63 | 3.46 | 3.32 | 3.22 | 3.13 | 2.99 | 2.85 | 2.70 | 2.62 | 2.54 | 2.45 | 2.36 | 2.27 | 2.17 |
| 26 | 7.72 | 5.53 | 4.64 | 4.14 | 3.82 | 3.59 | 3.42 | 3.29 | 3.18 | 3.09 | 2.96 | 2.81 | 2.66 | 2.58 | 2.50 | 2.42 | 2.33 | 2.23 | 2.13 |
| 27 | 7.68 | 5.49 | 4.60 | 4.11 | 3.78 | 3.56 | 3.39 | 3.26 | 3.15 | 3.06 | 2.93 | 2.78 | 2.63 | 2.55 | 2.47 | 2.38 | 2.29 | 2.20 | 2.10 |
| 28 | 7.64 | 5.45 | 4.57 | 4.07 | 3.75 | 3.53 | 3.36 | 3.23 | 3.12 | 3.03 | 2.90 | 2.76 | 2.60 | 2.52 | 2.44 | 2.35 | 2.26 | 2.17 | 2.06 |
| 29 | 7.60 | 5.42 | 4.54 | 4.04 | 3.73 | 3.50 | 3.33 | 3.20 | 3.09 | 3.00 | 2.87 | 2.73 | 2.57 | 2.49 | 2.41 | 2.33 | 2.23 | 2.14 | 2.03 |
| 30 | 7.56 | 5.39 | 4.51 | 4.02 | 3.70 | 3.47 | 3.30 | 3.17 | 3.07 | 2.98 | 2.84 | 2.79 | 2.55 | 2.47 | 2.39 | 2.30 | 2.21 | 2.11 | 2.01 |
| 40 | 7.31 | 5.18 | 4.31 | 3.83 | 3.51 | 3.29 | 3.12 | 2.99 | 2.89 | 2.80 | 2.66 | 2.52 | 2.37 | 2.29 | 2.20 | 2.11 | 2.02 | 1.92 | 1.80 |
| 60 | 7.08 | 4.98 | 4.13 | 3.65 | 3.34 | 3.12 | 2.95 | 3.82 | 2.72 | 2.63 | 2.50 | 2.35 | 2.20 | 2.12 | 2.03 | 1.94 | 1.84 | 1.73 | 1.60 |
| 120 | 6.85 | 4.79 | 3.95 | 3.48 | 3.17 | 2.96 | 2.79 | 2.66 | 2.56 | 2.47 | 2.34 | 2.19 | 2.03 | 1.95 | 1.86 | 1.76 | 2.66 | 1.53 | 1.38 |
| ∞ | 6.63 | 4.61 | 3.78 | 3.32 | 3.02 | 2.80 | 2.64 | 2.51 | 2.41 | 2.32 | 2.18 | 2.04 | 1.88 | 1.79 | 1.70 | 1.59 | 1.47 | 1.32 | 1.00 |

Degrees of freedom of the denominator ($v_2$)

Chi-Square ($\chi^2$) Distribution

| Degrees of Freedom | Area to the Right of Critical Value | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.995 | 0.99 | 0.975 | 0.95 | 0.90 | 0.10 | 0.05 | 0.025 | 0.01 | 0.005 |
| 1 | — | — | 0.001 | 0.004 | 0.016 | 2.706 | 3.841 | 5.024 | 6.635 | 7.879 |
| 2 | 0.010 | 0.020 | 0.051 | 0.103 | 0.211 | 4.605 | 5.991 | 7.378 | 9.210 | 10.597 |
| 3 | 0.072 | 0.115 | 0.216 | 0.352 | 0.584 | 6.251 | 7.815 | 9.348 | 11.345 | 12.838 |
| 4 | 0.207 | 0.297 | 0.484 | 0.711 | 1.064 | 7.779 | 9.488 | 11.143 | 13.277 | 14.860 |
| 5 | 0.412 | 0.554 | 0.831 | 1.145 | 1.610 | 9.236 | 11.071 | 12.833 | 15.086 | 16.750 |
| 6 | 0.676 | 0.872 | 1.237 | 1.635 | 2.204 | 10.645 | 12.592 | 14.449 | 16.812 | 18.548 |
| 7 | 0.989 | 1.239 | 1.690 | 2.167 | 2.833 | 12.017 | 14.067 | 16.013 | 18.475 | 20.278 |
| 8 | 1.344 | 1.646 | 2.180 | 2.733 | 3.490 | 13.362 | 15.507 | 17.535 | 20.090 | 21.955 |
| 9 | 1.735 | 2.088 | 2.700 | 3.325 | 4.168 | 14.684 | 16.919 | 19.023 | 21.666 | 23.589 |
| 10 | 2.156 | 2.558 | 3.247 | 3.940 | 4.865 | 15.987 | 18.307 | 20.483 | 23.209 | 25.188 |
| 11 | 2.603 | 3.053 | 3.816 | 4.575 | 5.578 | 17.275 | 19.675 | 21.920 | 24.725 | 26.757 |
| 12 | 3.074 | 3.571 | 4.404 | 5.226 | 6.304 | 18.549 | 21.026 | 23.337 | 26.217 | 28.299 |
| 13 | 3.565 | 4.107 | 5.009 | 5.892 | 7.042 | 19.812 | 22.362 | 24.736 | 27.688 | 29.819 |
| 14 | 4.075 | 4.660 | 5.629 | 6.571 | 7.790 | 21.064 | 23.685 | 26.119 | 29.141 | 31.319 |
| 15 | 4.601 | 5.229 | 6.262 | 7.261 | 8.547 | 22.307 | 24.996 | 27.488 | 30.578 | 32.801 |
| 16 | 5.142 | 5.812 | 6.908 | 7.962 | 9.312 | 23.542 | 26.296 | 28.845 | 32.000 | 34.267 |
| 17 | 5.697 | 6.408 | 7.564 | 8.672 | 10.085 | 24.769 | 27.587 | 30.191 | 33.409 | 35.718 |
| 18 | 6.265 | 7.015 | 8.231 | 9.390 | 10.865 | 25.989 | 28.869 | 31.526 | 34.805 | 37.156 |
| 19 | 6.844 | 7.633 | 8.907 | 10.117 | 11.651 | 27.204 | 30.144 | 32.852 | 36.191 | 38.582 |
| 20 | 7.434 | 8.260 | 9.591 | 10.851 | 12.443 | 28.412 | 31.410 | 34.170 | 37.566 | 39.997 |
| 21 | 8.034 | 8.897 | 10.283 | 11.591 | 13.240 | 29.615 | 32.671 | 35.479 | 38.932 | 41.401 |
| 22 | 8.643 | 9.542 | 10.982 | 12.338 | 14.042 | 30.813 | 33.924 | 36.781 | 40.289 | 42.796 |
| 23 | 9.260 | 10.196 | 11.689 | 13.091 | 14.848 | 32.007 | 35.172 | 38.076 | 41.638 | 44.181 |
| 24 | 9.886 | 10.856 | 12.401 | 13.848 | 15.659 | 33.196 | 36.415 | 39.364 | 42.980 | 45.559 |
| 25 | 10.520 | 11.524 | 13.120 | 14.611 | 16.473 | 34.382 | 37.652 | 40.646 | 44.314 | 46.928 |
| 26 | 11.160 | 12.198 | 13.844 | 15.379 | 17.292 | 35.563 | 38.885 | 41.923 | 45.642 | 48.290 |
| 27 | 11.808 | 12.879 | 14.573 | 16.151 | 18.114 | 36.741 | 40.113 | 43.194 | 46.963 | 49.645 |
| 28 | 12.461 | 13.565 | 15.308 | 16.928 | 18.939 | 37.916 | 41.337 | 44.461 | 48.278 | 50.993 |
| 29 | 13.121 | 14.257 | 16.047 | 17.708 | 19.768 | 39.087 | 42.557 | 45.722 | 49.588 | 52.336 |
| 30 | 13.787 | 14.954 | 16.791 | 18.493 | 20.599 | 40.256 | 43.773 | 46.979 | 50.892 | 53.672 |
| 40 | 20.707 | 22.164 | 24.433 | 26.509 | 29.051 | 51.805 | 55.758 | 59.342 | 63.691 | 66.766 |
| 50 | 27.991 | 29.707 | 32.357 | 34.764 | 37.689 | 63.167 | 67.505 | 71.420 | 76.154 | 79.490 |
| 60 | 35.534 | 37.485 | 40.482 | 43.188 | 46.459 | 74.397 | 79.082 | 83.298 | 88.379 | 91.952 |
| 70 | 43.275 | 45.442 | 48.758 | 51.739 | 55.329 | 85.527 | 90.531 | 95.023 | 100.425 | 104.215 |
| 80 | 51.172 | 53.540 | 57.153 | 60.391 | 64.278 | 96.578 | 101.879 | 106.629 | 112.329 | 116.321 |
| 90 | 59.196 | 61.754 | 65.647 | 69.126 | 73.291 | 107.565 | 113.145 | 118.136 | 124.116 | 128.299 |
| 100 | 67.328 | 70.065 | 74.222 | 77.929 | 82.358 | 118.498 | 124.342 | 129.561 | 135.807 | 140.169 |

**Chi-Square Table Values**