

BBA(G)

104: Quantitative Techniques & Operational Research In Management

Unit I

Statistics: Definition, Importance & Limitation. Collection of data and formation of frequency distribution. Graphic presentation of frequency distribution – graphics, Bars, Histogram, Diagrammatic. Measures of central tendency – mean, median and mode, partition values – quartiles, deciles and percentiles. Measures of variation – range, IQR, quartile, deciles and percentiles. Measures of variation – range, IQR, quartile deviation and standard deviation and Lorenz Curve.

Unit II

Correlation Analysis: Correlation Coefficient; Assumptions of correlation analysis; coefficients of determination and correlation; measurement of correlation- Karl Person's Methods; Spearman's rank correlation; concurrent deviation the correlation coefficient; Pitfalls and limitations associated with regression and correlation analysis; real world application using IT tools..

Unit III

Linear Programming: Concept a assumptions usage in business decision making linear programming problem: formulation, methods of solving: graphical and simplex, problems with mixed constraints: duality; concept, significance, usage & application in business decision making.

Unit IV

Transportation, assignment problems & Game Theory: General structure of transportation problem, solution procedure for transportation problem, methods for finding initial solution, test for optimality. Maximization transportation on problem, transportation problem. Assignment problem approach of the assignment model, solution methods of assignment problem, maximization in an assignment, unbalanced assignment problem, restriction on assignment.



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

UNIT 1

Statistics

Statistics is the study of the collection, organization, analysis, interpretation and presentation of data. It deals with all aspects of data, including the planning of data collection in terms of the design of surveys and experiments.

Importance of Statistics in Different Fields

Statistics plays a vital role in every fields of human activity. Statistics has important role in determining the existing position of per capita income, unemployment, population growth rate, housing, schooling medical facilities etc...in a country. Now statistics holds a central position in almost every field like Industry, Commerce, Trade, Physics, Chemistry, Economics, Mathematics, Biology, Botany, Psychology, Astronomy etc..., so application of statistics is very wide. Now we discuss some important fields in which statistics is commonly applied

(1) **Business:**

Statistics play an important role in business. A successful businessman must be very quick and accurate in decision making. He knows that what his customers wants, he should therefore, know what to produce and sell and in what quantities. Statistics helps businessman to plan production according to the taste of the costumers, the quality of the products can also be checked more efficiently by using statistical methods. So all the activities of the businessman based on statistical information. He can make correct decision about the location of business, marketing of the products, financial resources etc...

(2) In **Economics:**

Statistics play an important role in economics. Economics largely depends upon statistics. National income accounts are multipurpose indicators for the economists and administrators. Statistical methods are used for preparation of these accounts. In economics research statistical methods are used for collecting and analysis the data and testing hypothesis. The relationship between supply and demands is studies by statistical methods, the imports and exports, the inflation rate, the per capita income are the problems which require good knowledge of statistics.

(3) In **Mathematics:**

Statistical plays a central role in almost all natural and social sciences. The methods of natural sciences are most reliable but conclusions draw from them are only probable, because they are based on incomplete evidence. Statistical helps in describing these measurements more precisely. Statistics is branch of applied mathematics. The large number of statistical methods like probability averages, dispersions, estimation etc... is used in mathematics and different techniques of pure mathematics like integration, differentiation and algebra are used in statistics.

(4) In Banking:

Statistics play an important role in banking. The banks make use of statistics for a number of purposes. The banks work on the principle that all the people who deposit their money with the banks do not withdraw it at the same time. The bank earns profits out of these deposits by lending to others on interest. The bankers use statistical approaches based on probability to estimate the numbers of depositors and their claims for a certain day.

(5) In State Management (Administration):

Statistics is essential for a country. Different policies of the government are based on statistics. Statistical data are now widely used in taking all administrative decisions. Suppose if the government wants to revise the pay scales of employees in view of an increase in the living cost, statistical methods will be used to determine the rise in the cost of living. Preparation of federal and provincial government budgets mainly depends upon statistics because it helps in estimating the expected expenditures and revenue from different sources. So statistics are the eyes of administration of the state.

(6) In Accounting and Auditing:

Accounting is impossible without exactness. But for decision making purpose, so much precision is not essential the decision may be taken on the basis of approximation, know as statistics. The correction of the values of current asserts is made on the basis of the purchasing power of money or the current value of it.

In auditing sampling techniques are commonly used. An auditor determines the sample size of the book to be audited on the basis of error.

(7) In Natural and Social Sciences: Statistics plays a vital role in almost all the natural and social sciences.

Statistical methods are commonly used for analyzing the experiments results, testing their significance in Biology, Physics, Chemistry, Mathematics,

Meteorology, Research chambers of commerce, Sociology, Business, Public Administration, Communication and Information Technology etc...

(8) In

Astronomy:

Astronomy is one of the oldest branch of statistical study, it deals with the measurement of distance, sizes, masses and densities of heavenly bodies by means of observations. During these measurements errors are unavoidable so most probable measurements are founded by using statistical methods.

Limitations of statistics.

Statistics is indispensable to almost all sciences - social, physical and natural. It is very often used in most of the spheres of human activity. In spite of the wide scope of the subject it has certain limitations. Some important limitations of statistics are the following:

1. Statistics does not study qualitative phenomena:

Statistics deals with facts and figures. So the quality aspect of a variable or the subjective phenomenon falls out of the scope of statistics. For example, qualities like beauty, honesty, intelligence etc. cannot be numerically expressed. So these characteristics cannot be examined statistically. This limits the scope of the subject.

2. Statistical laws are not exact:

Statistical laws are not exact as in case of natural sciences. These laws are true only on average. They hold good under certain conditions. They cannot be universally applied. So statistics has less practical utility.

3. Statistics does not study individuals:

Statistics deals with aggregate of facts. Single or isolated figures are not statistics. This is considered to be a major handicap of statistics.

4. Statistics can be misused:

Statistics is mostly a tool of analysis. Statistical techniques are used to analyze and interpret the collected information in an enquiry. As it is, statistics does not prove or disprove anything. It is just a means to an end. Statements supported by statistics are more appealing and are commonly believed. For this, statistics is often misused. Statistical methods rightly used are beneficial but if misused these become harmful. Statistical methods used by less expert hands will lead to inaccurate results. Here the fault does not lie with the subject of statistics but with the person who makes wrong use of it.

Data collection

Data collection usually takes place early on in an improvement project, and is often formalized through a plan that often contains the following activity.

1. Pre collection activity — agree on goals, target data, definitions, methods
2. Collection — data collections
3. Present Findings — usually involves some form of sorting analysis and/or presentation.

Prior to any data collection, pre-collection activity is one of the most crucial steps in the process. It is often discovered too late that the value of their interview information is discounted as a consequence of poor sampling of both questions and informants and poor elicitation techniques. After pre-collection activity is fully completed, data collection in the field, whether by interviewing or other methods, can be carried out in a structured, systematic and scientific way.

A formal data collection process is necessary as it ensures that data gathered are both defined and accurate and that subsequent decisions based on arguments embodied in the findings are valid. The process provides both a baseline from which to measure and in certain cases a target on what to improve.

Frequency and Frequency distribution

The frequency of a particular data value is the number of times the data value occurs. For example, if four students have a score of 80 in mathematics, and then the score of 80 is said to have a frequency of 4. The frequency of a data value is often represented by f .

A frequency table is constructed by arranging collected data values in ascending order of magnitude with their corresponding frequencies.

Example

The marks awarded for an assignment set for a Year 8 class of 20 students were as follows:

6 7 5 7 7 8 7 6 9 7
4 10 6 8 8 9 5 6 4 8

Present this information in a frequency table.

Solution:

To construct a frequency table, we proceed as follows:

Step 1:

Construct a table with three columns. The first column shows what is being arranged in ascending order (i.e. the marks). The lowest mark is 4. So, start from 4 in the first column as shown below.

Mark	Tally	Frequency
4		
5		
6		
7		
8		
9		
10		

Step 2:

Go through the list of marks. The first mark in the list is 6, so put a tally mark against 6 in the second column. The second mark in the list is 7, so put a tally mark against 7 in the second column. The third mark in the list is 5, so put a tally mark against 5 in the third column as shown below.

Mark	Tally	Frequency
4		
5		
6		
7		
8		
9		
10		

We continue this process until all marks in the list are tallied. Step 3:



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

Count the number of tally marks for each mark and write it in third column. The finished frequency table is as follows:

In general:

We use the following steps to construct a frequency table: Step

1:

Construct a table with three columns. Then in the first column, write down all of the data values in ascending order of magnitude.

Step 2:

To complete the second column, go through the list of data values and place one tally mark at the appropriate place in the second column for every data value. When the fifth tally is reached for a mark, draw a horizontal line through the first four tally marks as shown for 7 in the above frequency table. We continue this process until all data values in the list are tallied.

Step 3:

Count the number of tally marks for each data value and write it in the third column.

Class Intervals (or Groups)

When the set of data values are spread out, it is difficult to set up a frequency table for every data value as there will be too many rows in the table. So we group the data into class intervals (or groups) to help us organise, interpret and analyse the data.

Ideally, we should have between five and ten rows in a frequency table. Bear this in mind when deciding the size of the class interval (or group).

Each group starts at a data value that is a multiple of that group. For example, if the size of the group is 5, then the groups should start at 5, 10, 15, 20 etc. Likewise, if the size of the group is 10, then the groups should start at 10, 20, 30, 40 etc.

The frequency of a group (or class interval) is the number of data values that fall in the range specified by that group (or class interval).

Example

The number of calls from motorists per day for roadside service was recorded for the month of December 2003. The results were as follows:

28 22 217 130 120 86 80 90 120 140
70 40 145 137 113 90 68 174 192 170
100 75 102 97 75 123 100 82 109 120
81

Set up a frequency table for this set of data values.

Solution:

To construct a frequency table, we proceed as follows:

Smallest data value = 22

Highest data value = 217

Difference = Highest value - Smallest value

$$= 217 - 22$$

$$= 195$$

Let the width of the class interval be 40.

$$\therefore \text{Number of class intervals} = \frac{195}{40} = 4.7 = 5 \quad (\text{Round up to the next integer})$$

There are at least 5 class intervals. This is reasonable for the given data.

Step 1: Construct a table with three columns, and then write the data groups or class intervals in the first column. The size of each group is 40. So, the groups will start at 0, 40, 80, 120, 160 and 200 to include all of the data. Note that in fact we need 6 groups (1 more than we first thought).

Class interval	Tally	Frequency
0 - 39		
40 - 79		
80 - 119		
120 - 159		
160 - 199		
200 - 239		

Step 2: Go through the list of data values. For the first data value in the list, 28, place a tally mark against the group 0-39 in the second column. For the second data value in the list, 122, place a tally mark against the group 120-159 in the second column. For the third data value in the list, 217, place a tally mark against the group 200-239 in the second column.

Class interval	Tally	Frequency
0 - 39		
40 - 79		
80 - 119		
120 - 159		
160 - 199		
200 - 239		

We continue this process until all of the data values in the set are tallied.

Step 3: Count the number of tally marks for each group and write it in the third column. The finished frequency table is as follows:

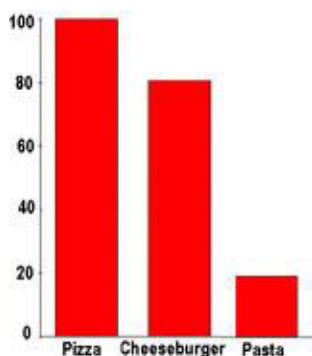
Class interval	Tally	Frequency
0 - 39		1
40 - 79		5
80 - 119		12
120 - 159		8
160 - 199		4
200 - 239		1
	Sum =	31

Graphic presentation of Fre quency distribution

Graphics

Graphics, also known as gra phical techniques, are information graph ics in the field of statistics used to visualiz e quantitative data.

Bar chart: A bar chart or bar graph is a chart with r ectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called a column bar chart.



A bar graph constructed from favorite foods mentioned by elementary school students.

Histogram

In statistics, a histogram is a graphical representation of the distribution of data. It is an estimate of the probability distribution of a continuous variable and was first introduced by Karl Pearson. A histogram is a representation of tabulated frequencies, shown as adjacent rectangles, erected over discrete intervals (bins), with an area equal to the frequency of the observations in the interval. The height of a rectangle is also equal to the frequency density of the interval, i.e., the frequency divided by the width of the interval. The total area of the histogram is equal to the number of data. A histogram may also be normalized displaying relative frequencies. It then shows the proportion of cases that fall into each of several categories, with the total area equaling 1. The categories are usually specified as consecutive, non-overlapping intervals of a variable. The categories (intervals) must be adjacent, and often are chosen to be of the same size. The rectangles of a histogram are drawn so that they touch each other to indicate that the original variable is continuous.

Measures of Central Tendency

Mean

In statistics, mean has two related meanings:

- the arithmetic mean (and is distinguished from the geometric mean or harmonic mean).
- the expected value of a random variable, which is also called the *population mean*.

There are other statistical measures that should not be confused with averages - including 'median' and 'mode'. Other simple statistical analyses use measures of spread, such as range, interquartile range, or standard deviation. For a real-valued random variable X , the mean is the expectation of X . Note that not every probability distribution has a defined mean (or variance); see the Cauchy distribution for an example.

For a data set, the mean is the sum of the values divided by the number of values. The mean of a set of numbers x_1, x_2, \dots, x_n is typically denoted by \bar{x} , pronounced "x bar". This mean is a type of arithmetic mean. If the data set were based on a series of observations obtained by sampling a statistical population, this mean is termed the "sample mean" (\bar{x}) to distinguish



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

it from the "population mean" (μ or μ_x). The mean is often quoted along with the standard deviation: the mean describes the central location of the data, and the standard deviation describes the spread.

An alternative measure of dispersion is the mean deviation, equivalent to the average absolute deviation from the mean. It is less sensitive to outliers, but less mathematically tractable.

Arithmetic mean (AM)

The *arithmetic mean* is the "standard" average, often simply called the "mean".

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

The mean may often be confused with the median, mode or range. The mean is the arithmetic average of a set of values, or distribution; however, for skewed distributions, the mean is not necessarily the same as the middle value (median), or the most likely (mode). For example, mean income is skewed upwards by a small number of people with very large incomes, so that the majority have an income lower than the mean. By contrast, the median income is the level at which half the population is below and half is above. The mode income is the most likely income, and favors the larger number of people with lower incomes. The median or mode are often more intuitive measures of such data.

For example, the arithmetic mean of six values: 5, 10, 13, 7, 25, 31 is

$$\frac{5 + 10 + 13 + 7 + 25 + 31}{6} = \frac{91}{6} \approx 15.$$

Weighted arithmetic mean

The weighted arithmetic mean is used, if one wants to combine average values from samples of the same population with different sample sizes:

$$\bar{x} = \frac{\sum_{i=1}^n w_i \cdot x_i}{\sum_{i=1}^n w_i}$$

The weights w_i represent the bounds of the partial sample. In other applications they



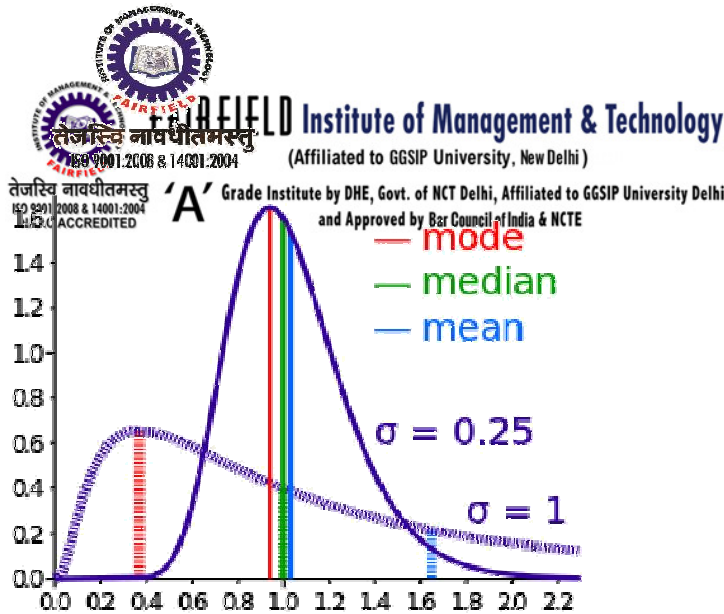
FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

represent a measure for the reliability of the influence upon the mean by respective values.

Problems



Comparison of mean, median and mode with different skewness.

The arithmetic mean may be mis interpreted as the median to imply that most values are higher or lower than is actually the case. If elements in the sample space increase arithmetically, when placed in some order, then the median and arithmetic average are equal. For example, consider the sample space {1,2,3,4}. The average is 2.5, as is the median. However, when we consider a sample space that cannot be arranged into an arithmetic progression, such as {1,2,4,8,16}, the median and arithmetic average can differ significantly. In this case the arithmetic average is 6.2 and the median is 4. When one looks at the arithmetic average of a sample space, one must note that the average value can vary significantly from most values in the sample space.

There are applications of this phenomenon in many fields. For example, since the 1980s in the United States median income has increased more slowly than the arithmetic average of income. Researchers dealing with frequency data must also be careful when reporting summary statistics such as means or median. Where a phenomenon is rare in general (for example, emergency room visits among the general population), but occurs frequently in some people (for example, dare devils), then the mean value may be much lower than the median.

Median.

The *median* of a finite list of numbers can be found by arranging all the observations from lowest value to highest value and picking the middle one. If there is an even number of



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

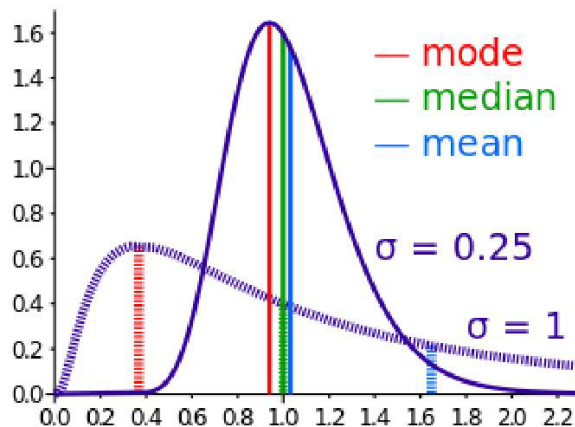
'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

observations, then there is no single middle value; the median is then usually defined to be the mean of the two middle values.

To find the median just line the numbers in order according to their value and cross off one number on each side. Example: 1 2 3 4 6 7 9 11 12 13, 14. In a sample of data, or a finite population, there may be no member of the sample whose value is identical to the median (in the case of an even sample size), and, if there is such a member, there may be more than one so that

the median may not uniquely identify a sample member. Nonetheless, the value of the median is uniquely determined with the usual definition

At most, half the population have values less than the *median*, and, at most, half have values greater than the median. If both groups contain less than half the population, then some of the population is exactly equal to the median. For example, if $a < b < c$, then the median of the list $\{a, b, c\}$ is b , and, if $a < b < c < d$, then the median of the list $\{a, b, c, d\}$ is the mean of b and c ; i.e., it is $(b + c)/2$.



Comparison of mean, median and mode with different skewness.

Medians in descriptive statistics

The median is used primarily for skewed distributions, which it summarizes differently than the arithmetic mean. Consider the multiset $\{1, 2, 2, 2, 3, 14\}$. The median is 2 in this case, as is the mode, and it might be seen as a better indication of central tendency than the arithmetic mean of 4.

Mode

In statistics, the mode is the value that occurs most frequently in a data set



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

Like the statistical mean and the median, the mode is a way of capturing important information about a random variable or a population in a single quantity. The mode is in general different from the mean and median, and may be very different for strongly skewed distributions.

The mode is not necessarily unique, since the same maximum frequency may be attained at different values. The most ambiguous case occurs in uniform distributions, where in all values are equally likely.

Comparison of mean, median and mode

Comparison of common averages of values { 1, 2, 2, 3, 4, 7, 9 }			
Type	Description	Example	Result
Arithmetic mean	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ Sum divided by number of values:	(1+2+2+3+4+7+9) / 7	4
Median	Middle value separating the greater and lesser halves of a data set	1, 2, 2, 3, 4, 7, 9	3
Mode	Most frequent value in a data set	1, 2, 2, 3, 4, 7, 9	2

Quartile

In descriptive statistics, the quartiles of a set of values are the three points that divide the data set into four equal groups, each representing a fourth of the population being sampled. A quartile is a type of quantile.

Definitions

- first quartile (designated Q_1) = lower quartile = splits lowest 25% of data = 25th percentile



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

- second quartile (designated Q_2) = median = cuts data set in half = 50th percentile
- third quartile (designated Q_3) = upper quartile = splits highest 25% of data, or lowest 75% = 75th percentile

The difference between the upper and lower quartiles is called the *interquartile range*.

Method

1. Use the median to divide the ordered data set into two halves. If the median is a datum (as opposed to being the average of the middle two data), include the median in both halves.
2. The lower quartile value is the median of the lower half of the data. The upper quartile value is the median of the upper half of the data.

Example 1

Data Set: 6, 47, 49, 15, 42, 41, 7, 39, 43, 40, 36 Ordered Data Set: 6, 7, 15, 36, 39, 40, 41, 42, 43, 47, 49

Method 1

$$\begin{cases} Q_1 = 25.5 \\ Q_2 = 40 \\ Q_3 = 42.5 \end{cases}$$

Example 2

Ordered Data Set: 7, 15, 36, 39, 40, 41

Method 1

$$\begin{cases} Q_1 = 15 \\ Q_2 = 37.5 \\ Q_3 = 40 \end{cases}$$

Deciles

In descriptive statistics, any of the nine values that divide the sorted data into ten equal parts, so that each part represents 1/10 of the sample or population

Percentile

In statistics, a percentile (or centile) is the value of a variable below which a certain percent of observations fall. For example, the 20th percentile is the value (or score) below which 20 percent of the observations may be found.

The 25th percentile is also known as the first quartile (Q_1), the 50th percentile as the median



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

or second quartile (Q_2), and the 75th percentile as the third quartile (Q_3).

Measures of variation

Range

The range is the simplest measure of variation to find. It is simply the highest value minus the lowest value.

$$\text{RANGE} = \text{MAXIMUM} - \text{MINIMUM}$$

Since the range only uses the largest and smallest values, it is greatly affected by extreme values, that is - it is not resistant to change.

The Inter-quartile Range

The inter-quartile range is a measure that indicates the extent to which the central 50 % of values within the dataset are dispersed. It is based upon, and related to, the median.

In the same way that the median divides a dataset into two halves, it can be further divided into quarters by identifying the upper and lower quartiles. The lower quartile is found one quarter of the way along a dataset when the values have been arranged in order of magnitude; the upper quartile is found three quarters along the dataset. Therefore, the upper quartile lies half way between the median and the highest value in the dataset whilst the lower quartile lies halfway between the median and the lowest value in the dataset. The inter-quartile range is found by subtracting the lower quartile from the upper quartile.

For example, the examination marks for 20 students following a particular module are arranged in order of magnitude.

	Lower quartile					Median		Upper quartile												
Student	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Mark	43	48	50	50	52	53	56	58	59	60	62	65	66	68	70	71	74	76	78	80

The median lies at the mid-point between the two central values (10th and 11th)



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

= half-way between 60 and 62 = 61

The lower quartile lies at the mid-point between the 5th and 6th values

= half-way between 52 and 53 = 52.5

The upper quartile lies at the mid-point between the 15th and 16th values

= half-way between 70 and 71 = 70.5

The inter-quartile range for this data set is therefore $70.5 - 52.5 = 18$ whereas the range is: $80 - 43 = 37$.

The inter-quartile range provides a clearer picture of the overall dataset by removing/ignoring the outlying values.

Like the range however, the inter-quartile range is a measure of dispersion that is based upon only two values from the dataset. Statistically, the standard deviation is a more powerful measure of dispersion because it takes into account every value in the dataset. The standard deviation is explored in the next section of this guide.

Variance

"Average Deviation"

The range only involves the smallest and largest numbers, and it would be desirable to have a statistic which involved all of the data values.

The first attempt one might make at this is something they might call the average deviation from the mean and define it as:

$$\text{Avg. Dev} = \frac{\sum (x - \mu)}{N}$$

The problem is that this summation is always zero. So, the average deviation will always be zero. That is why the average deviation is never used.

Population Variance

So, to keep it from being zero, the deviation from the mean is squared and called the "squared deviation from the mean". This "average squared deviation from the mean" is called the variance.

FIMT Campus, Kapashera, New Delhi-110037, Phones : 011-25063208/09/10/11, 25066256/ 57/58/59/60
 Fax : 011-250 63212 Mob. : 09312352942, 09811568155 E-mail : fimtoffice@gmail.com Website : www.fimt-ggsipu.org

$$\text{Population Variance} = \sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

Unbiased Estimate of the Population Variance

One would expect the sample variance to simply be the population variance with the population mean replaced by the sample mean. However, one of the major uses of statistics is to estimate the corresponding parameter. This formula has the problem that the estimated value isn't the same as the parameter. To counteract this, the sum of the squares of the deviations is divided by one less than the sample size.

$$\text{Sample Variance} = s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

Standard deviation

Standard deviation is a widely used measure of variability or diversity used in statistics. It shows how much variation or "dispersion" exists from the average (mean, or expected value). A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data points are spread out over a large range of values.

Consider a population consisting of the following eight values:

2, 4, 4, 4, 5, 5, 7, 9

These eight data points have the mean (average) of 5:

$$\frac{2 + 4 + 4 + 4 + 5 + 5 + 7 + 9}{8} = 5$$

To calculate the population standard deviation, first compute the difference of each data point from the mean, and square the result of each:

$$\begin{array}{ll} (2 - 5)^2 = (-3)^2 = 9 & (5 - 5)^2 = 0^2 = 0 \\ (4 - 5)^2 = (-1)^2 = 1 & (5 - 5)^2 = 0^2 = 0 \\ (4 - 5)^2 = (-1)^2 = 1 & (7 - 5)^2 = 2^2 = 4 \\ (4 - 5)^2 = (-1)^2 = 1 & (9 - 5)^2 = 4^2 = 16 \end{array}$$



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

Next compute the average of these values, and take the square root:

$$\sqrt{\frac{(9 + 1 + 1 + 1 + 0 + 0 + 4 + 16)}{8}} = 2$$

Unit 2

Correlation Coefficient (r)

The quantity r , called the *linear correlation coefficient*, measures the strength and the direction of a linear relationship between two variables. The linear correlation coefficient is sometimes referred to as the *Pearson product moment correlation coefficient*

in honor of its developer Karl Pearson.

The mathematical formula for computing r is:

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n} \right] \left[\sum y^2 - \frac{(\sum y)^2}{n} \right]}}$$

where n is the number of pairs of data.

✦ The value of r is such that $-1 < r < +1$. The + and – signs are used for positive linear correlations and negative linear correlations, respectively.

✦ **Positive correlation:** If x and y have a strong positive linear correlation, r is close to +1. An r value of exactly +1 indicates a perfect positive fit. Positive values indicate a relationship between x and y variables such that as values for x increase, values for y also increase.

✦ **Negative correlation:** If x and y have a strong negative linear correlation, r is close to -1. An r value of exactly -1 indicates a perfect negative fit. Negative values indicate a relationship between x and y such that as values for x increase, values for y decrease.

✦ **No correlation:** If there is no linear correlation or a weak linear correlation, r is close to 0. A value near zero means that there is a random, nonlinear relationship between the two variables.



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

✦ A *perfect* correlation of ± 1 occurs only when the data points all lie exactly on a straight line. If $r = +1$, the slope of this line is positive. If $r = -1$, the slope of this line is negative.

✦ A correlation greater than 0.8 is generally described as *strong*, whereas a correlation less than 0.5 is generally described as *weak*. These values can vary based upon the "type" of data being examined.

Assumption of correlation analysis

linearity, normality, and homoscedasticity

Calculations on correlation rely on a linear relationship. e.g. there is a correlation between experience and nervousness in driving a car but you may not get a **linear correlation** because the

nervousness factor first rises then falls with experience.

Normality means that neither set of data is independently skewed.

Homoscedasticity means that there needs to be a consistent scatter pattern over the whole range. Otherwise, you may get a positive correlation over a range of the data that is tainted by a unproven correlation in another area.

And then there is causality, which means that the two sets of data may have a third influencing factor e.g. I could say there is a positive correlation between crime and membership of the conservative party but that would be a bad correlation as both are dependent on population density.

Coefficient of Determination, r^2 or R^2 :

✦ The *coefficient of determination*, r^2 , is useful because it gives the proportion of the variance (fluctuation) of one variable that is predictable from the other variable. It is a measure that allows us to determine how certain one can be in making predictions from a certain model/graph.

✦ The *coefficient of determination* is the ratio of the explained variation to the total variation.

✦ The *coefficient of determination* is such that $0 < r^2 < 1$, and denotes the strength of the linear association between x and y .

✦ The *coefficient of determination* represents the percent of the data that is the closest to the line of best fit. For example, if $r = 0.922$, then $r^2 = 0.850$, which means that 85% of the total variation in y can be explained by the linear relationship between x and y (as described by the regression equation). The other 15% of the total variation in y remains unexplained.

✦ The *coefficient of determination* is a measure of how well the regression line represents the data. If the regression line passes exactly through every point on the scatter plot, it



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

would be able to explain all of the variation. The further the line is away from the points, the less it is able to explain.

KARL PEARSON'S

CORRELATION COEFFICIENT (r):---

In statistics, the Pearson product-moment correlation coefficient (r) is a common measure of the correlation between two variables X and Y. When measured in a population the Pearson Product Moment correlation is designated by the Greek letter rho (ρ). When computed in a sample, it is designated by the letter "r" and is sometimes called "Pearson's r." Pearson's correlation reflects the degree of linear relationship between two variables. It ranges from +1 to -1. A correlation of +1 means that there is a perfect positive linear relationship between variables. A correlation of -1 means that there is a perfect negative linear relationship between variables. A correlation of 0 means there is no linear relationship between the two variables. Correlations are rarely if ever 0, 1, or -1. If you get a certain outcome it could indicate whether correlations were negative or positive.

Mathematical Formula:--

The quantity r, called the linear correlation coefficient, measures the strength and the direction of a linear relationship between two variables. The linear correlation coefficient is sometimes referred to as the Pearson product moment correlation coefficient in honor of its developer Karl Pearson.

The mathematical formula for computing r is:

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum x)^2][N\sum y^2 - (\sum y)^2]}}$$

Where:

- N = number of pairs of scores
- $\sum xy$ = sum of the products of paired scores
- $\sum x$ = sum of x scores
- $\sum y$ = sum of y scores
- $\sum x^2$ = sum of squared x scores
- $\sum y^2$ = sum of squared y scores



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

Spearman's Rank Correlation Coefficient

The Spearman's Rank Correlation Coefficient is used to discover the strength of a link between two sets of data.

Spearman's Rank correlation coefficient

A correlation can easily be drawn as a scatter graph, but the most precise way to compare several **pairs of data** is to use a statistical test - this establishes whether the correlation is really significant or if it could have been the result of chance alone.

Spearman's Rank correlation coefficient is a technique which can be used to summarise the strength and direction (negative or positive) of a relationship between two variables.

The result will always be between 1 and minus 1.

Method - calculating the coefficient

- Create a table from your data.
- Rank the two data sets. Ranking is achieved by giving the ranking '1' to the biggest number in a column, '2' to the second biggest value and so on. The smallest value in the column will get the lowest ranking. This should be done for both sets of measurements.
- Tied scores are given the mean (average) rank. For example, the three tied scores of 1 euro in the example below are ranked fifth in order of price, but occupy three positions (fifth, sixth and seventh) in a ranking hierarchy of ten. The mean rank in this case is calculated as $(5+6+7) \div 3 = 6$.
- Find the difference in the ranks (d): This is the difference between the ranks of the two values on each row of the table. The rank of the second value (price) is subtracted from the rank of the first (distance from the museum).
- Square the differences (d^2) To remove negative values and then sum the $\sum (d^2)$.



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

तेजस्वि नावधीलमस्तु
ISO 9001:2008 & 14001:2004
NAAC ACCREDITED

'A'

Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

Convenience Store	Distance from CAM (m)	Rank distance	Price of 50cl bottle (€)	Rank price	Difference between ranks (d)	d ²
1	50	10	1.80	2	8	64

Convenience Store	Distance from CAM (m)	Rank distance	Price of 50cl bottle (€)	Rank price	Difference between ranks (d)	d ²
2	175	9	1.20	3.5	5.5	30.25
3	270	8	2.00	1	7	49
4	375	7	1.00	6	1	1
5	425	6	1.00	6	0	0
6	580	5	1.20	3.5	1.5	2.25
7	710	4	0.80	9	-5	25
8	790	3	0.60	10	-7	49
9	890	2	1.00	6	-4	16
10	980	1	0.85	8	-7	49
					$\Sigma d^2 = 285.5$	

Data Table: Spearman's Rank Correlation

- Calculate the coefficient (R) using the formula below. The answer will always be between 1.0 (a perfect positive correlation) and -1.0 (a perfect negative correlation).

When written in mathematical notation the Spearman Rank formula looks like this :

$$(R) = 1 - \frac{6 \sum d^2}{n^3 - n}$$

Now to put all these values into the formula.

- Find the value of all the d^2 values by adding up all the values in the Difference² column. In our example this is **285.5**. Multiplying this by **6** gives **1713**.

- Now for the bottom line of the equation. The value n is the number of sites at which you took measurements. This, in our example is **10**. Substituting these values into $n^3 - n$ we get **1000 - 10**

□ We now have the formula: $R = 1 - (1713/990)$ which gives a value for R :

$$1 - 1.73 = -0.73$$

Example

In this example, we will use the raw data in the table below to calculate the correlation between the IQ of a person with the number of hours spent in front of TV per week.

IQ, X_i Hours of TV per week, Y_i

106 7

86 0

100 27

101 50

99 28

103 29

97 20

113 12

112 6

110 17

First, we must find the value of the term $\frac{1}{n} \sum_{i=1}^n X_i Y_i$. To do so we use the following steps, reflected in the table below.



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

1. Sort the data by the first column (X_i). Create a new column x_i and assign it the ranked values $1,2,3,\dots,n$.
2. Next, sort the data by the second column (Y_i). Create a fourth column y_i and similarly assign it the ranked values $1,2,3,\dots,n$.

3. Create a fifth column d_i to hold the differences between the two rank columns (x_i and y_i).
4. Create one final column d_i^2 to hold the value of column d_i squared.

IQ, X_i Hours of TV per week, Y_i rank x_i rank y_i d_i d_i^2

86	0	1	1	0	0
97	20	2	6	-4	16
99	28	3	8	-5	25
100	27	4	7	-3	9
101	50	5	10	-5	25
103	29	6	9	-3	9
106	7	7	3	4	16
110	17	8	5	3	9
112	6	9	2	7	49
113	12	10	4	6	36

found, we can add
 With d_i^2 them to find $\sum d_i^2 = 194$. The value of n is 10. So these values
 can now be substituted back into the equation,

$$\rho = 1 - \frac{6 \times 194}{10(10^2 - 1)}$$

which evaluates to $\rho = -0.175757575...$ with a P-value = 0.6864058 (using the t distribution)

This low value shows that the correlation between IQ and hours spent watching TV is very low. In the case of ties in the original values, this formula should not be used. Instead, the Pearson correlation coefficient should be calculated on the ranks (where ties are given ranks,



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

as described above).

Calculate Coefficient of Concurrent Deviation

Example:

Calculate Coefficient of Concurrent Deviation from the following data

X	:	1	2	3	4	5	6
Y	:	25	169	5	1	0	

Solution: Calculation of Coefficient of Concurrent-Deviations

X	Deviation from Previous value. dx	Y	Deviation from Previous value. dy	dx dy
1		25		
2	+	10		
3	+	9		
4	+	5		
5	+	1		
6	+	0		
n = 5		c = 0		

$$r_{c} = \pm \sqrt{\pm((2c-n)/n)} = \pm \sqrt{\pm((0-5)/5)} = \pm \sqrt{-(1)} = -\sqrt{1}$$



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

limitations to regression analysis

1. It is not linearity in variables, as we have seen
2. it is not linearity in parameters, although we have only covered the linear regression model here

3. Remember that by estimating the linear model we can use the results to estimate parameters that are non-linear functions of the estimated model
4. If the model is non-linear in the parameter from the outset, can use Non-Linear Least Squares to the best non-linear curve.
5. The real limitation to the regression model is when the regression function does not contain the parameter of interest

Limitations of Correlation

Although correlation is a powerful tool, there are some limitations in using it:

- Correlation does not completely tell us everything about the data. Means and standard deviations continue to be important.
- The data may be described by a curve more complicated than a straight line, but this will not show up in the calculation of r .
- Outliers strongly influence the correlation coefficient. If we see any outliers in our data, we should be careful about what conclusions we draw from the value of r .
- Just because two sets of data are correlated, it doesn't mean that one is the cause of the other.

UNIT 3

Linear Programming (LP) Problem

A **linear programming problem** is one in which we are to find the maximum or minimum value of a linear expression

$$ax + by + cz + \dots$$

(called the **objective function**), subject to a number of linear **constraints** of the form $Ax + By + Cz + \dots \leq N$



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

or

$$Ax + By + Cz + \dots \geq N.$$

The largest or smallest value of the objective function is called the **optimal value**, and a collection of values of x, y, z, \dots that gives the optimal value constitutes an **optimal solution**.

The variables x, y, z, \dots are called the **decision variables**.

Example

Here is an example of an LP problem:

Find the maximum value of

$$p = 3x - 2y + 4z$$

subject to

$$\begin{array}{rcccccc} 4x & + & 3y & - & z & \geq & 3 \\ x & + & 2y & + & z & \leq & 4 \\ x \geq 0, y \geq 0, z \geq 0 & & & & & & \end{array}$$

The objective function is $p = 3x - 2y + 4z$. The constraints are

$$\begin{array}{rcccccc} 4x & + & 3y & - & z & \geq & 3 \\ x & + & 2y & + & z & \leq & 4 \\ x \geq 0, y \geq 0, z \geq 0. & & & & & & \end{array}$$

Sketching the Solution Set of a Linear Inequality

To sketch the region represented by a linear inequality in two variables:

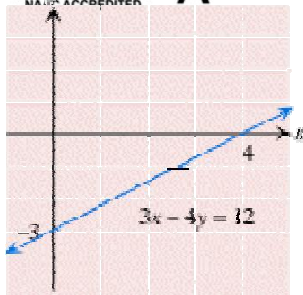
- Sketch the straight line obtained by replacing the inequality with an equality.
- Choose a test point not on the line ((0,0) is a good choice if the line does not pass through the origin, and if the line does pass through the origin a point on one of the axes would be a good choice).
- If the test point satisfies the inequality, then the set of solutions is the entire region on the same side of the line as the test point. Otherwise it is the region on the other side of the line. In either case, shade out the side that does not contain the solutions, leaving the solution region showing.

Example

To sketch the linear inequality

$$3x - 4y \leq 12,$$

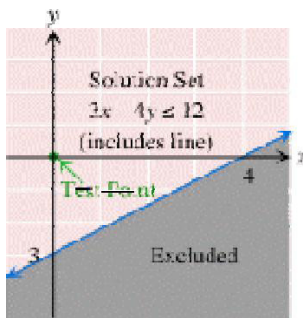
first sketch the line $3x - 4y = 12$.



Next, choose the origin (0, 0) as the test point (since it is not on the line). Substituting $x=0$, $y=0$ in the inequality gives

$$3(0) - 4(0) \leq 12.$$

Since this is a true statement, (0, 0) is in the solution set, so the solution set consists of all points on the same side as (0, 0). This region is left unshaded, while the (grey) shaded region is blocked out.



Feasible Region

The **feasible region** determined by a collection of linear inequalities is the collection of points that satisfy all of the inequalities.

To sketch the feasible region determined by a collection of linear inequalities in two variables: Sketch the regions represented by each inequality on the same graph, remembering to shade the parts of the plane that you do not want. What is unshaded when you are done is the feasible region.

Example

The feasible region for the following collection of inequalities is the unshaded region shown below (including its boundary).

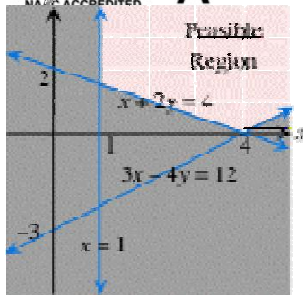


FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

$$\begin{array}{rclclcl} 3x & & - & & 4y & & \leq & & 12, \\ x & & + & & 2y & & \geq & & 4 \\ x & & & & \geq & & & & 1 \\ y \geq 0. & & & & & & & & \end{array}$$



Graphical Method

The **graphical method** for solving linear programming problems in two unknowns is as follows.

- Graph the feasible region.
- Compute the coordinates of the corner points.
- Substitute the coordinates of the corner points into the objective function to see which gives the optimal value.
- If the feasible region is not bounded, this method can be misleading: optimal solutions always exist when the feasible region is bounded, but may or may not exist when the feasible region is unbounded. The textbook shows a straightforward way for determining whether optimal solutions exist in the case of unbounded feasible regions.

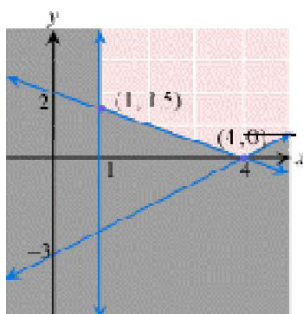
If you want to see a utility that automates the whole process, try our [Linear Programming Grapher](#). It does everything automatically!

Example


Minimize $C = 3x + 4y$ subject to the constraints

$$\begin{array}{rclcl}
 3x & - & 4y & \leq & 12, \\
 x & + & 2y & \geq & 4 \\
 x \geq 1, & & y \geq 0. & &
 \end{array}$$

The feasible region for this set of constraints was shown above. Here it is again with the corner points shown.



The following table shows the value of C at each corner point:

Point	$C = 3x + 4y$
(1, 1.5)	$3(1)+4(1.5) = 9$  mini mum
(4, 0)	$3(4)+4(0) = 12$

Therefore, the solution is $x = 1, y = 1.5$, giving the minimum value $C = 9$.

Standard Maximization Problem

A **standard maximization problem** in n unknowns is a linear programming problem in which we are required to maximize (not minimize) the objective function, subject to constraints of the form

$$x \geq 0, y \geq 0, z \geq 0, \dots,$$

and further constraints of the form

$$mAx + By + Cz + \dots \leq N,$$

where A, B, C, \dots and N are numbers with N nonnegative.

Note that the inequality here must be a " \leq ," and not " $=$ " or " \geq ."

Examples

The following is a standard LP problem:

Maximize $P = 2x - 3y + z$ subject to

$$4x - 3y + z \leq 3$$

$$x + y + z \leq 10$$

$$x \geq 0, y \geq 0, z \geq 0.$$

The following is *not* a standard LP problem:

Maximize $P = 2x - 3y + z$ subject to

$$4x - 3y + z \geq 3$$

$$x + y + z \leq 10$$

$$x \geq 0, y \geq 0, z \geq 0.$$



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

Simplex Method for Standard Maximization Problem

To solve a standard maximization problem using the **simplex method**, we take the following steps:

Step 1. Convert to a system of equations by introducing **slack variables** to turn the constraints into equations, and rewriting the objective function in standard form.

Step 2. Write down the initial **tableau**.

Step 3. Select the pivot column: Choose the negative number with the largest magnitude in the bottom row (excluding the rightmost entry). Its column is the pivot column. (If there are two candidates, choose either one.) If all the numbers in the bottom row are zero or positive (excluding the rightmost entry), then you are done: the basic solution maximizes the objective function (see below for the basic solution).

Step 4. Select the pivot in the pivot column: The pivot must always be a positive number. For each positive entry b in the pivot column, compute the ratio a/b , where a is the number in the Answer column in that row. Of these **test ratios**, choose the smallest one. The corresponding number b is the pivot.

Step 5. Use the pivot to clear the column in the normal manner (taking care to follow the exact prescription for formulating the row operations described in Chapter 2) and then relabel the pivot row with the label from the pivot column. The variable originally labeling the pivot row is the **departing** or **exiting variable** and the variable labeling the column is the **entering variable**.

Step 6. Go to Step 3.

Basic Solution

To get the **basic solution** corresponding to any tableau in the simplex method, set to zero all variables that do not appear as row labels (these are the **inactive** variables).

The value of a variable that does appear as a row label (an **active variable**) is the number in the rightmost column in that row divided by the number in that row in the column labeled by the same variable.

Example

In the following tableau

x y Z s t u p Ans



तेजस्वि नावधीतमस्तु
ISO 9001:2008 & 14001:2004
NAAC ACCREDITED

FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

-1	0	0	1	0	0	0	4
1	0	3	0	0	8	0	12
4	0	0	0	3	0	0	2
-5	2	0	0	0	6	0	4

6	0	0	0	0	0	5	-25
---	---	---	---	---	---	---	-----

the basic solution is

$x = 0, y = 2, z = 4, s = 4, t = 2/3, u = 0, p = -5,$
 and the active variables are $y, z, s, t,$ and $p.$

Nonstandard Constraints

To solve a linear programming problem with constraints of the form $Ax + By + \dots \geq N$ with N positive, subtract a **surplus variable** from the left-hand side, rather than adding a slack variable. The basic solution corresponding to the initial tableau will not be feasible since some of the active variables will be negative, and so the rules for the initial pivoting are different from those above.

Star all rows that give a negative value for the associated active variable (except for the objective variable, which is allowed to be negative). If there are starred rows, you will need to begin with

Phase I:

Phase I: Getting into the Feasible Region (Getting Rid of the Stars) In the first starred row, find the largest positive number. Use test ratios as in the preceding section to find the pivot in that column (exclude the bottom row), then pivot on that entry. If the lowest ratio occurs both in a starred row and an unstarred row, pivot in a starred row rather than the unstarred one. Repeat until no starred rows remain, then go on to **Phase II.**

Phase II: Use the Simplex Method for Standard Maximization Problems If there are any negative entries on the left side of the bottom row after Phase I, use the method described above for solving standard maximization problems.

Simplex Method for Minimization Problem

To solve a minimization problem using the simplex method, convert it into a maximization problem. If you need to minimize c , instead maximize $p = -c.$

Example

The minimization LP problem:

Minimize $C = 3x + 4y - 8z$ subject to the constraints

$$3x \quad \quad \quad - \quad \quad \quad 4y \quad \quad \quad \leq \quad \quad \quad 12,$$



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

तेजस्वि नावधीलमस्तु
ISO 9001:2008 & 14001:2004
NAAC ACCREDITED

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

$$\begin{array}{rcccccccc}
 x & & + & & 2y & & + & & z & & \geq & & 4 \\
 4x & & & & - & & 2y & & + & & 5z & & \leq & & 20 \\
 x \geq 0, & y \geq 0, & z \geq 0
 \end{array}$$

can be replaced by the following **maximization** problem:

Maximize $P = -3x - 4y + 8z$ subject to the constraints

$$\begin{array}{rcccccccc}
 3x & & & & - & & 4y & & & & \leq & & 12, \\
 x & & + & & & & 2y & & + & & z & & \geq & & 4
 \end{array}$$

$$4x - 2y + 5z \leq 20$$

$$x \geq 0, \quad y \geq 0, \quad z \geq 0.$$

Solving a Matrix Game Using the Simplex Method

A game may be solved using the simplex method as follows.

Before you start, check for saddle points. If you find one, you have already solved the game; the optimal strategies are the pure strategies passing through a saddle point. Otherwise, continue with the following steps.

1. Reduce the payoff matrix by dominance.
2. Add a fixed number k to each of the entries so that they all become non-negative.
3. Set up and solve the associated linear programming problem using the simplex method.
4. Obtain the optimal strategies and the expected value as follows:

Column Strategy

- a. Express the solution to the linear programming problem as a column vector.
- b. Normalize by dividing each entry of the solution vector by the sum of the entries, or by the value of the objective variable p .
- c. Insert zeros in positions corresponding to the columns deleted during reduction.

Row Strategy

- d. Write down the row vector whose entries are the numbers appearing in the bottom row of the final tableau underneath the slack variables.
- e. Normalize by dividing each entry of the solution vector by the sum of the entries.
- f. Insert zeros in positions corresponding to the rows deleted during reduction.

Value of the Game

$$e = 1/p - k.$$



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

Basic Duality

The study of duality is very important in LP. Knowledge of duality allows one to develop increased insight into LP solution interpretation. Also, when solving the dual of any problem, one simultaneously solves the primal. Thus, duality is an alternative way of solving LP problems.

However, given today's computer capabilities, this is an infrequently used aspect of duality.

Therefore, we concentrate on the study of duality as a means of gaining insight into the LP solution. We will also discuss the ways that primal decision variables place constraints upon the resource valuation information. The Primal problem can be written as:

Primal

$$\begin{aligned} \text{Max} \quad & \sum_j C_j X_j \\ \text{s.t.} \quad & \sum_j a_{ij} X_j \leq b_i \quad \text{for all } i \\ & X_j \geq 0 \quad \text{for all } j \end{aligned}$$

Associated with this primal problem is a dual resource valuation problem. The dual of the above problem is

Primal Dual Pair and Their Units

Primal

$$\begin{aligned} \text{Max} \quad & \sum_j C_j X_j \\ \text{s.t.} \quad & \sum_j a_{ij} X_j \leq b_i \quad \text{for all } i \\ & X_j \geq 0 \quad \text{for all } j \end{aligned}$$

where x is the variable and equals units

sold Max sum (per unit profits) * (units

sold)

s.t. Sum (per unit res. use)*(units sold) \leq res on hand

Dual

$$\begin{aligned} \text{Min} \quad & \sum_i U_i b_i \\ \text{s.t.} \quad & \sum_i U_i a_{ij} \geq c_j \text{ for all } j \\ & U_i \geq 0 \text{ for all } i \end{aligned}$$

U is the variable and equals per unit resource value

Min sum (per unit res value) * (res on hand)

s.t. sum (per unit res value) * (per unit res use)

≥ per unit profits

So resource values – shadow prices are set up so per unit profits are exhausted.

Primal Dual Pair and Example

$$\begin{aligned} \text{max} \quad & 2000X_{\text{fancy}} \quad \square \quad 1700X_{\text{fine}} \quad \square \quad 1200X_{\text{new}} \\ \text{s.t.} \quad & X_{\text{fancy}} \quad \square \quad X_{\text{fine}} \quad \square \quad X_{\text{new}} \leq 12 \\ & 25X_{\text{fancy}} \quad \square \quad 20X_{\text{fine}} \quad \square \quad 19X_{\text{new}} \leq 280 \\ & X_{\text{fancy}} \quad , \quad X_{\text{fine}} \quad , \quad X_{\text{new}} \geq 0 \end{aligned}$$

	(van cap)		(labor)		(profits)
min	$12U_1$	\square	$280U_2$		(Resource Payments)
s.t.	U_1	\square	$25U_2$	\geq	2000 (X_{fancy})
	U_1	\square	$20U_2$	\geq	1700 (X_{fine})
	U_1	\square	$19U_2$	\geq	1200 (X_{new})
	U_1	,	U_2	\geq	(nonnegativity)

Primal rows become dual columns

Degeneracy and Duality

The above interpretations for the dual variables depend upon whether the basis still exists after the change occurs.

When a basic primal variable equals zero, dual has alternative optimal solutions. The cause of this situation is generally primal constraints are redundant at the solution point and the range of right hand sides is zero.

$$\begin{aligned}
 & \text{Max } 3X_1 \quad \square \quad 2X_2 \\
 & X_1 \quad \square \quad X_2 \leq 100 \\
 & X_1 \quad \quad \quad \leq 50 \\
 & \quad \quad \quad X_2 \leq 50 \\
 & X_1 \quad , \quad X_2 \geq 0
 \end{aligned}$$

At the optimal solution, $X_1 = 50$, $X_2 = 50$, constraints are redundant.

If first slack variable is basic then $X_1 = 50$, $X_2 = 50$, $S_1 = 0$ while S_1 is basic. Shadow prices are 0, 3, and 2.



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

If S_3 basic $X_1 = 50$, $X_2 = 50$, $S_3 = 0$ with shadow prices 2, 1, 0. Same objective value -- multiple solutions.

Degeneracy and Duality

$$\begin{aligned} \text{Max } 3X_1 \quad & \square \quad 2X_2 \\ X_1 \quad & \square \quad X_2 \leq 100 \\ X_1 \quad & \leq 50 \\ & X_2 \leq 50 \\ x_1 \quad , \quad & x_2 \geq 0 \end{aligned}$$

$$u_1 \ u_2 \ u_3 = [0 \ 3 \ 2] \text{ or } [2 \ 1 \ 0]$$

The main difficulty with degeneracy is in interpreting the shadow prices as they take on a direction.

If one were to increase the first right hand side from 100 to 101 this would lead to a zero change in the objective function and X_1 and X_2 would remain at 50.

Decrease first constrain rhs from 100 to 99 then objective function which is two units smaller because X_2 would need to be reduced from 50 to 49.

This shows that the two alternative shadow prices for the first constraint (i.e., 0 and 2) each hold in a direction.

Similarly if bound on X_1 51, obj increases by 1, whereas, if moved downward to 49, it would cost 3.

Meanwhile, reducing X_2 bound costs 2 and increasing by 0. this explains all shadow prices

usage & application in business decision making

This is a simplified example will illustrates the way in which a problem relating to business is formulated. A large sub-contractor machines special parts to order. For one order he has two machines, X and Y available. Unfortunately, the machines perform more effectively on some jobs than others. Machine Y can do 10 units per hour on contracts from Alpha Ltd., 12 per hour on components from Beater & Co. and 26 per hour on those from Chester Inc., while Machine X can produce 16, 9 and 10 units per hour respectively.



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

However, this is complicated by certain constraints which are:

Maximum number of units per month of units from Alpha, Beater and Chester are 6,500, 4,440 and 800 per month respectively the machines X and Y are restricted to working a maximum of 260 and 350 hours per month respectively.

UNIT 4

TRANSPORTATION PROBLEMS

INTRODUCTION

A scooter production company produces scooters at the units situated at various places (called origins) and supplies them to the places where the depot (called destination) are situated.

Here the availability as well as requirements of the various depots are finite and constitute the limited resources. This type of problem is known as distribution or transportation problem in which the key idea is to minimize the cost or the time of transportation. In previous lessons we have considered a number of specific linear programming problems. Transportation problems are also linear programming problems and can be solved by simplex method but because of practical significance the transportation problems are of special interest and it is tedious to solve them through simplex method.

OBJECTIVES

After completion of this lesson you will be able to:

- solve the transportation problems by
 - a. North-West corner rule;
 - b. Lowest cost entry method; and
 - c. Vogel's approximation method.
- test the optimality of the solution.

A typical transportation problem is shown in Fig.. It deals with sources where a supply of some commodity is available and destinations where the commodity is demanded. The classic statement of the transportation problem uses a matrix with the rows representing sources and columns representing destinations. The algorithms for solving the problem are based on this matrix representation. The costs of shipping from sources to destinations are indicated by the entries in the matrix. If shipment is impossible between a given source and destination, a large cost of M is entered. This discourages the solution from using such cells. Supplies and demands are shown along the margins of the matrix. As in the example, the classic transportation problem has total supply equal to total demand.



Figure. Matrix model of a transportation problem.

The network model of the transportation problem is shown in Fig.. Sources are identified as the nodes on the left and destinations on the right. Allowable shipping links are shown as arcs, while disallowed links are not included .

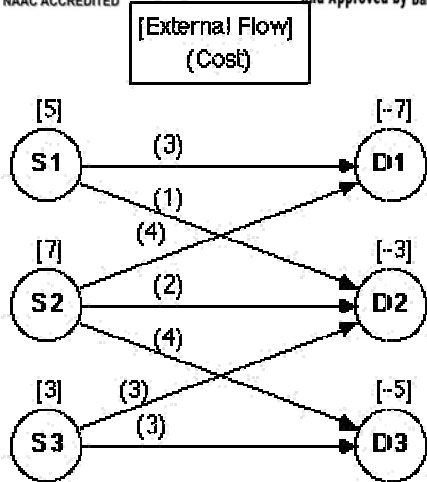


Figure. Network flow model of the transportation problem.

Only arc costs are shown in the network model, as these are the only relevant parameters. All other parameters are set to the default values. The network has a special form important in graph theory; it is called a bipartite network since the nodes can be divided into two parts with all arcs going from one part to the other.

On each supply node the positive external flow indicates supply flow entering the network. On each destination node a demand is a negative fixed external flow indicating that this amount must leave the network. The optimum solution for the example is shown in Fig.



Figure. Optimum solution, $z = 46$.

Variations of the classical transportation problem are easily handled by modifications of the network model. If links have finite capacity, the arc upper bounds can be made finite. If supplies represent raw materials that are transformed into products at the sources and the demands are in units of product, the gain factors can be used to represent transformation

efficiency at each source. If some minimal flow is required in certain links, arc lower bounds can be set to nonzero values.

Some Definitions

The following terms are to be defined with reference to the transportation problems:

- Feasible Solution (F.S.) A set of non-negative allocations $x_{ij} \geq 0$ which satisfies the row and column restrictions is known as feasible solution.
- Basic Feasible Solution (B.F.S.) A feasible solution to a m-origin and n-destination problem is said to be basic feasible solution if the number of positive allocations are $(m+n-1)$.
 If the number of allocations in a basic feasible solutions are less than $(m+n-1)$, it is called degenerate basic feasible solution (DBFS) (otherwise non-degenerate).
- Optimal Solution: A feasible solution (not necessarily basic) is said to be optimal if it minimizes the total transportation cost.

SOLUTION OF THE TRANSPORTATION PROBLEM

Let us consider the numerical version of the problem stated in the introduction and the mathematical formulation of the same in the next section, as below in Table

Depot Unit	B_1	B_2	B_3	B_4	Stock
A_1	$c_{11}=2$	$c_{12}=3$	$c_{13}=5$	$c_{14}=1$	$a_1=8$
A_2	$c_{21}=7$	$c_{22}=3$	$c_{23}=4$	$c_{24}=6$	$a_2=10$
A_3	$c_{31}=4$	$c_{32}=1$	$c_{33}=7$	$c_{34}=2$	$a_3=20$
Requirement	$b_1=6$	$b_2=8$	$b_3=9$	$b_4=15$	= =38

In order to find the solution of this transportation problem we have to follow the steps given below.

- Initial basic feasible solution
- Test for optimization.

Let us consider these steps one by one.

Initial Basic Feasible Solution (TEST OF OPTIMALITY)

There are three different methods to obtain the initial basic feasible solution viz.

- (I) North-West corner rule
- (II) Lowest cost entry method
- (III) Vogel's approximation method

(I) North-West corner rule

In this method we distribute the available units in rows and column in such a way that the sum will remain the same. We have to follow the steps given below.

- a) Start allocations from north-west corner, i.e., from (1,1) position. Here $\min(a_1, b_1)$, i.e., $\min(8,6)=6$ units. Therefore, the maximum possible units that can be allocated to this position is 6, and write it as 6(2) in the (1,1) position of the table. This completes the allocation in the first column and cross the other positions, i.e., (2,1) and (3,1) in the column.

Depot \ Unit	B_1	B_2	B_3	B_4	Stock
A_1	6(2)				$8-6=2$
A_2	×				10
A_3	×				20
Requirement	$6-6=0$	8	9	15	32

- (b) After completion of step (a), come across the position (1,2). Here $\min(8-6, 8)=2$ units can be allocated to this position and write it as 2(3). This completes the allocations in the first row and cross the other positions, i.e., (1,3) and (1,4) in this row



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

FAIRFIELD

Institute of Management & Technology

Managed by 'The Fairfield Foundation'

(Affiliated to GGSIP University, New Delhi)

नेपालिय वाणिज्यमन्त्रालय
189/9001:2008 & 14001:2004
Grade Institute by DNE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

Unit	B_1	B_2	B_3	B_4	Stock
A_1	6(2)	2(3)	x	x	2-2=0
A_2	x				10
A_3	x				20
Requirement	0	8-2=6	9	15	30

Maximization of transportation problem

Transportation is a minimization model. First, we should balance the problem by taking dummy where necessary (i. e. either on Supply side or Demand side). After balancing the problem, we should convert the problem in Minimization by converting it into Regret matrix.

To find Regret matrix, find maximum profit value in the original matrix. From this maximum profit value, subtract all profit values in the matrix. The resulting matrix is Regret matrix.

Description of the assignment problem

In the classical assignment problem, the goal is to find an optimal assignment of agents to tasks without assigning an agent more than once and ensuring that all tasks are completed. The objective might be to minimize the total time to complete a set of tasks, or to maximize skill ratings, or to minimize the cost of the assignments.



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

The assignment problem is a particular class of transportation linear programming problems with the supplies and demands equal to integers (often 1). Since all supplies, demands, and bounds on variables are integral, the assignment problem relies on a nice property of transportation problems that the optimal solution will be entirely integral. As you will see, the goal of the

objective of the assignment problem (unlike transportation problems) may not have anything to do at all with moving anything!

Assignment Problem

A typical assignment problem, presented in the classic manner, is shown in Fig. 12. Here there are five machines to be assigned to five jobs. The numbers in the matrix indicate the cost of doing each job with each machine. Jobs with costs of M are disallowed assignments. The problem is to find the minimum cost matching of machines to jobs.



Restriction on the assignment problem.

The network model is in Fig.. It is very similar to the transportation model except the external flows are all +1 or -1. The only relevant parameter for the assignment model is arc cost (not shown in the figure for clarity) ; all other parameters should be set to default values. The assignment network also has the bipartite structure.

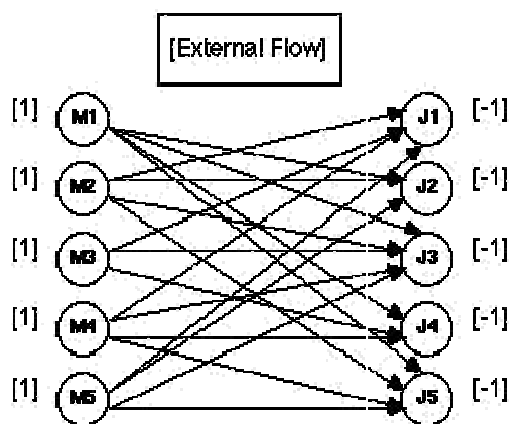


Figure. Network model of the assignment problem.

The solution to the assignment problem as shown in Fig has a total flow of 1 in every column and row, and is the assignment that minimizes total cost.



FAIRFIELD
Institute of Management & Technology
 Managed by 'The Fairfield Foundation'
 (Affiliated to GGSIP University, New Delhi)

Figure. Solution to the assignment Problem

Unbalanced assignment problem

If number of jobs is not equal to the number of operators then the assignment problem is known as unbalanced assignment problem. In words in a pay off matrix of assignment problem if number of rows and number of columns are not equal then it is unbalanced assignment problem.

To solve unbalanced assignment problem it is necessary to insert dummy row or dummy column such that the given pay off matrix become a square matrix. All values in dummy row or column are assume to be zero.

Scheduling problems

Job shop scheduling (or Job-shop problem) is an optimization problem in computer science and operations research in which ideal jobs are assigned to resources at particular times. The most basic version is as follows:

We are given n jobs J_1, J_2, \dots, J_n of varying sizes, which need to be scheduled on m identical machines, while trying to minimize the makespan. The makespan is the total length of the schedule (that is, when all the jobs have finished processing). Nowadays, the problem is presented as an online problem (dynamic scheduling), that is, each job is presented, and the online algorithm needs to make a decision about that job before the next job is presented.

Shop floor control

View all production jobs by work center
 Select workload by date range
 and shift Update operation com



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

pletions

Report completions by quantity or
percentage Review operation start and end
dates

Work center dispatch list

List of work by priority, order, line and type

Pay points automatically back-fill prior
operations Labor and machine load hours

Setup and run
times Completion
status

Outside processing
 status Next or prior
 work center

Labor and machine capacities with effectivity
 Efficiency factors, queue times and cost
 centers Work order shortages

Gantt chart

A Gantt chart is a type of bar chart, developed by Henry Gantt in the 1910s, that illustrates a project schedule. Gantt charts illustrate the start and finish dates of the terminal elements and summary elements of a project. Terminal elements and summary elements comprise the work breakdown structure of the project. Some Gantt charts also show the dependency (i.e. precedence network) relationships between activities. Gantt charts can be used to show current schedule status using percent-complete shadings and a vertical "TODAY" line as shown here.

Although now regarded as a common charting technique, Gantt charts were considered revolutionary when first introduced.^[1] In recognition of Henry Gantt's contributions, the Henry Laurence Gantt Medal is awarded for distinguished achievement in management and in community service. This chart is also used in information technology to represent data that has been collected.

Job shop and its principals

Job shops are typically small manufacturing systems that handle job production, that is, custom/bespoke or semi-custom/bespoke manufacturing processes such as small to medium-size customer orders or batch jobs. Job shops typically move on to different jobs (possibly with different customers) when each job is completed. In job shops machines are aggregated in shops by the nature of skills and technological processes involved, each shop therefore may contain different machines, which gives this production system processing flexibility, since jobs are not necessarily constrained to a single machine. In computer science the problem of job shop scheduling is considered strongly NP-hard.

A typical example would be a machine shop, which may make parts for local industrial



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

machinery, farm machinery and implements, boats and ships, or even batches of specialized components for the aircraft industry. Other types of common job shops are grinding, honing, jig-boring, gear manufacturing, and fabrication shops.

The opposite would be continuous flow manufactures such as textile, steel, food manufacturing and manual labor.

personnel scheduling

No matter what size company you own or manage, your Personnel department is the center of all employee-related information. From interviewing and hiring to exit interviews and dismissals, the HR department is responsible for gathering and keeping updated all personnel information and setting policies for personnel. The unique demands placed on your company's personnel department calls for a powerful management and scheduling tool designed to streamline the process of hiring, tracking and communicating with all of your company's staff.

ShiftPlanning provides custom scheduling software designed to address your company's Personnel demands including:

- Automated shift scheduling for every employee in every capacity.
- Streamlined employee rotations.
- Streamlined vacation scheduling.
- Incorporated overtime and regulatory rules.

Your company's Personnel department can take advantage of ShiftPlanning's soon to be integrated Payroll feature which, in addition to maintaining all of your company's personnel records and schedules, will automatically track and process employee work hours and overtime.

Department managers can now work with Personnel in the ShiftPlanning Scheduler to:

- Add and import unlimited employees through .csv, Address Book, Facebook and Google Apps.
- Add and import unlimited schedules via .csv and .xls.
- Prepare unlimited Schedules.
- Manage staff skills and find the right people for the job instantly.
- Allow employees to swap shifts electronically with no disruption to scheduling.
- View schedules online, 24/7.
- Create and generate multiple types of in-depth reports based on planned time and actual time.
- Synchronize staff calendars.

ShiftPlanning state-of-the-art scheduling software:

- Sets up a separate schedule for trainees.
- Easily creates separate schedules for every sector.



FAIRFIELD Institute of Management & Technology

(Affiliated to GGSIP University, New Delhi)

'A' Grade Institute by DHE, Govt. of NCT Delhi, Affiliated to GGSIP University Delhi
and Approved by Bar Council of India & NCTE

- Offers individual schedule management to team leaders.
- Allows sector managers to oversee and control shift trades, vacations, open shifts, etc.
- Provides a SMS notification system to make scheduling fast and easy.

Benefits:

- Improve customer service by keeping adequate staff on hand at all times.
- End unnecessary overtime costs.
- Reduce employee turnover by scheduling around each staff member's individual shift preferences.

- Reduce unauthorized absences.
- Comply with Health & Safety laws, union bylaws and any industry-specific guidelines or regulations that apply to your staff members.

ShiftPlanning's totally customizable software has powerful tools that integrate union regulations and labor laws into scheduling, reducing or eliminating any margin of error. ShiftPlanning is engineered to adapt to and implement updates and changes in regulatory rules as they arise without replacing software or disrupting business as usual. ShiftPlanning makes shift scheduling, managing, and now payroll easy, streamlined and so cost effective you can be reigning in out-of-control expenses within your first pay cycle.

Principles of job shop scheduling

Job shops are typically small manufacturing systems that handle job production, that is, custom/bespoke or semi-custom/bespoke manufacturing processes such as small to medium-size customer orders or batch jobs. Job shops typically move on to different jobs (possibly with different customers) when each job is completed. In job shops machines are aggregated in shops by the nature of skills and technological processes involved, each shop therefore may contain different machines, which gives this production system processing flexibility, since jobs are not necessarily constrained to a single machine. In computer science the problem of job shop scheduling is considered strongly NP-hard.

A typical example would be a machine shop, which may make parts for local industrial machinery, farm machinery and implements, boats and ships, or even batches of specialized components for the aircraft industry. Other types of common job shops are grinding, honing, jig-boring, gear manufacturing, and fabrication shops.

The opposite would be continuous flow manufactures such as textile, steel, food manufacturing and manual labor.

REFERENCE

- 1.Gupta, S.P. & Gupta, P.K., “ *Quantitative Techniques and Operations Research*”, Sultan Chand & Sons, 2005.
- 2.Vohra, N.D., “ *Quantitative Techniques in Management*”, Tata McGraw Hill, 2004.
- 3.Gupta, S.P., “ *Statistical Methods*,” Sultan Chand & Sons, 2004.
- 4.Srivastava, U.K., Shenoy, G.V. and Sharma, S.C., “ *Quantitative Techniques for managerial Decisions*”, New Age International Pvt. Ltd., 2nd edition, 2002.