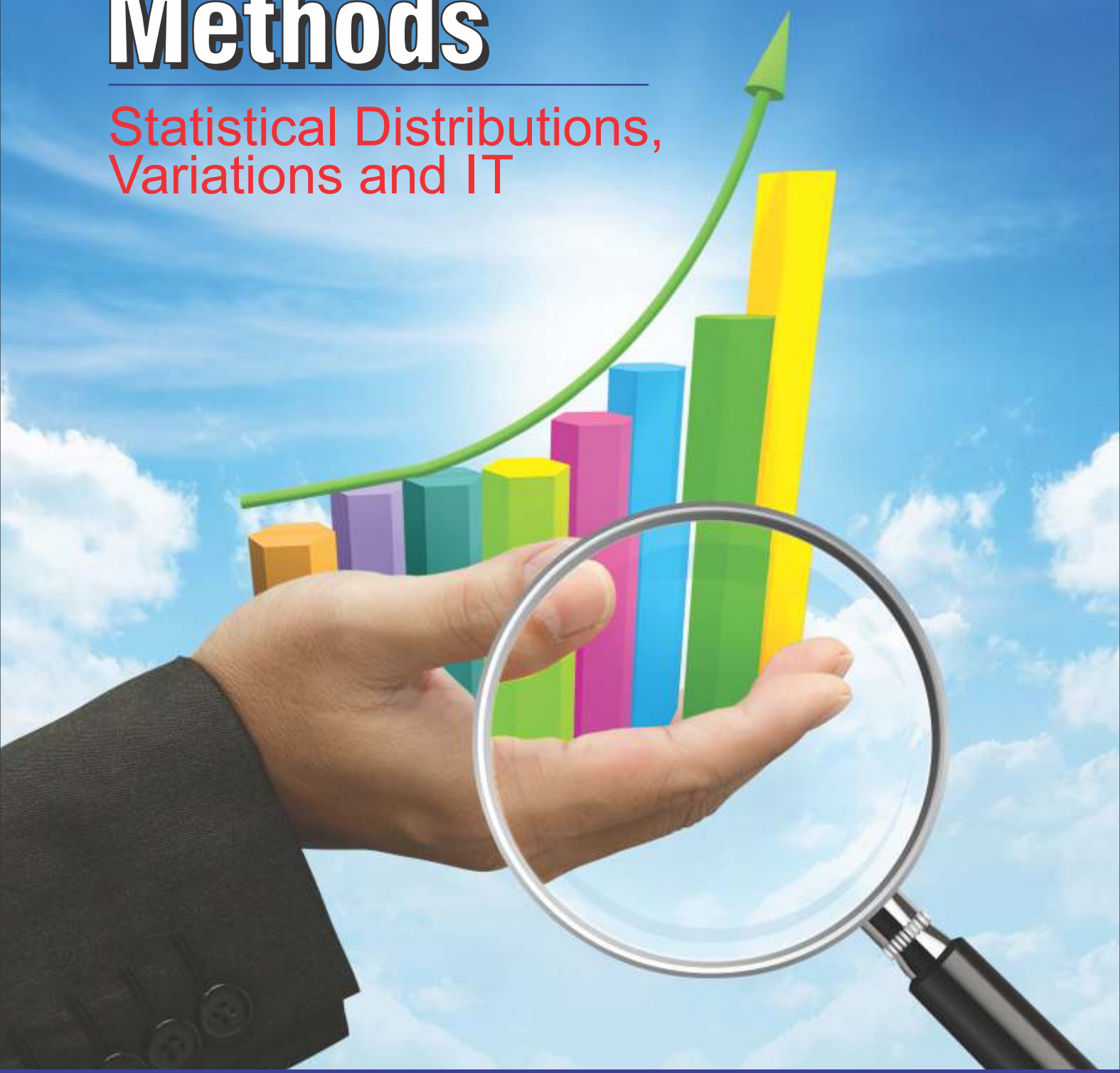
Block-IV

Quantitative Methods

Statistical Distributions, Variations and IT



Block
IV

STATISTICAL DISTRIBUTIONS, VARIATIONS AND IT

Ref. No. QM SLM 102021B4

For any clarification regarding this book, the students may please write to The ICFAI Foundation for Higher Education (IFHE), Hyderabad specifying the unit and page number.

While every possible care has been taken in type-setting and printing this book, The ICFAI Foundation for Higher Education (IFHE), Hyderabad welcomes suggestions from students for improvement in future editions.

Our E-mail id: cwfeedback@icfaiuniversity.in

Center for Distance and Online Education (CDOE)
The ICFAI Foundation for Higher Education
(Deemed-to-be-University Under Section 3 of UGC Act, 1956)
Donthanapally, Shankarapalli Road, Hyderabad- 501203.

BLOCK IV: STATISTICAL DISTRIBUTIONS, VARIATIONS AND IT

Decision making involves choosing an action from several available alternatives subject to different linear constraints. In a business, the idea is to choose the course of action which would in some sense optimize the results obtained. One way of decision making is applying intuitive or subjective methods based on 'hunches' or previous experiences or the knowledge of the person taking decisions; the other way is to apply quantitative or mathematical methods to the process of decision making. To apply quantitative or mathematical methods in a business environment, it is necessary to list out a statistical program. In statistics, there are a number of statistical programming techniques. IT can be used for statistical purposes. This block consists of Chi-Square Test and Analysis of Variance, Role of IT in Modern Business Enterprise and Statistical Software tools.

Unit-10 Chi-Square Test and Analysis of Variance helps to gain knowledge about statistical techniques called Chi-Square Test and Analysis of Variance. The first technique allows us to test whether the difference between the proportions representing more than two samples is significant or not, the second facilitates the testing of the same with regard to means i.e., to consider it they are equal to each other or not. The Chi-Square test is also employed to determine whether the two attributes according to which a population is categorized are independent of each other or not; and it also serves as a test for distributional goodness of fit.

Unit-11 Role of IT in Modern Business Enterprise introduces computers, classification of computers and discusses software concepts. It explains problem solving using computers. It discusses the design of computer based information systems. Enterprise Resource Planning (ERP) and Internet concepts are also explained in the unit.

Unit-12 Statistical Software Tools highlights the use of computers in finance. It introduces SPSS (Statistical Package for Social Sciences). SPSS is widely used for data analysis purpose in business, marketing, management, HR, psychology, and social sciences research. The data analysis using SPSS is discussed in the unit. The unit introduces the features and applications of SAS package. SAS is Statistical Analysis Systems.

Unit 10

Chi-Square Test and Analysis of Variance

Structure

10.1 Introduction

In the previous unit, we learnt various control charts. $\bar{X}$ charts, R charts and P chart are the four techniques that a quality control program uses to check the overall quality of its products. In this unit we will learn Chi-Square analysis and Analysis of Variance.

Chi-Square Test is used to determine whether the two attributes according to which a population is categorized are independent of each other or not and it also serves as a test for distributional goodness of fit. Analysis of Variance tests whether the difference between the proportions representing more than two samples is significant or not and it also tests whether the variability among them is significant or not.

10.2 Objectives

After going through the unit, you should be able to:

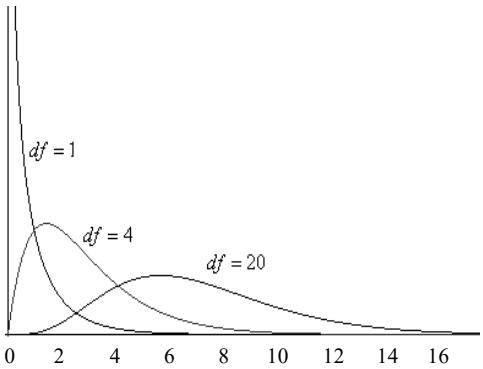
- Define chi-square test as a test of independence;
- Explain chi-square test as a distributional goodness of fit;
- Define analysis of variance; and
- Examine the population variance.

10.3 Chi-Square Distribution

The Chi-Square distribution, like the 't' distribution has a single parameter called degree of freedom. It's a distribution of random variable called χ^2, so it

assumes only positive value. The shape of chi-square distribution depends on the degrees of freedom. The total area under the curve for each of the degrees of freedom is one and the entire chi-square distribution lies on the right of the vertical axis. If the number of degrees of freedom is small, the curve will be skewed to the right. As the number of degrees of freedom increases, the curve also gradually tends to become symmetrical in nature (see Figure 10.1).

Figure 10.1: Chi-Square Distribution at Various Level of dof



Since different degrees of freedom have different sampling distributions, it is difficult to have a table that would give us the areas under the curve for all possible values. Chi-Square gives us the areas in the right tail at the significance level, which are most commonly used in tests.

10.4 Chi-Square Test as a Test of Independence

A test of independence is performed using contingency tables. As discussed in the previous unit, frequency tables of two variables presented simultaneously are called contingency tables. They are constructed by listing all the levels of one variable as rows in a table and the levels of the other variables as columns. The test of independence tests if the null hypothesis that the two characteristics of the element of a given population are independent against the alternative hypothesis. Let us explain the whole process with the help of an example.

A manufacturer of automobile wished to determine whether there are any differences in three media (Magazine, TV, Radio) in terms of recall of an advertisement and made a study. The results of the study were as follows:

Recall Ability	Media			
	Magazine	TV	Radio	Total
Purchase the brand	50	20	30	100
Do not purchase the brand	70	80	50	200
Total	120	100	80	300

The above table is called a Contingency table. Let the level of significance be α = 10%. If the proportion of the total population of consumers in each of the three media is denoted by p_m, p_t, and p_r, then the null and the alternative hypothesis will be set up as follows:

H_0: $p_m = p_t = p_r$ (Null Hypothesis: Proportion of consumers from each of the three media in terms of recall is equal).

H_1: $p_m \neq p_t \neq p_r$ (Alternative Hypothesis: Proportion of consumers from each of the three media in terms of recall is not equal).

In other words, we assume that the null hypothesis H_0 is a statement where the row and column variables are independent and alternative hypothesis H_1 is a statement where the row and column variables are dependent.

10.4.1 Test Statistics for a Test of Independence

Test statistics for a test of independence is given by

$$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

Where,

f_0 is the observed frequency

f_e is the expected frequency.

Now for every frequency in contingency table, we have to calculate the expected frequency expressed as:

$$\text{Expected Frequency} = \frac{(\text{Row Total})(\text{Column Total})}{\text{Grand Total}}$$

The observed and expected frequencies are shown below:

Recall Ability	Media			
	Magazine	TV	Radio	Total
Purchase the brand	$\frac{100}{300} \times 120 = 40$	$\frac{100}{300} \times 100 = 33.33$	$\frac{100}{300} \times 80 = 26.67$	100
Do not purchase the brand	$\frac{200}{300} \times 120 = 80$	$\frac{200}{300} \times 100 = 66.67$	$\frac{200}{300} \times 80 = 53.33$	200
Total	120	100	80	300

Now let us calculate the test statistic.

f_0	f_e	$f_0 - f_e$	$(f_0 - f_e)^2$	$(f_0 - f_e)^2/(f_e)$
50	40	10.00	100.00	2.50
20	33.33	–13.33	177.69	5.33
30	26.67	3.33	11.09	0.42
70	80	–10.00	100.00	1.25
80	66.67	13.33	177.69	2.67
50	53.33	–3.33	11.09	0.21
Total				12.38

The value of the Chi-Square Statistic is $\chi^2 = 12.38$.

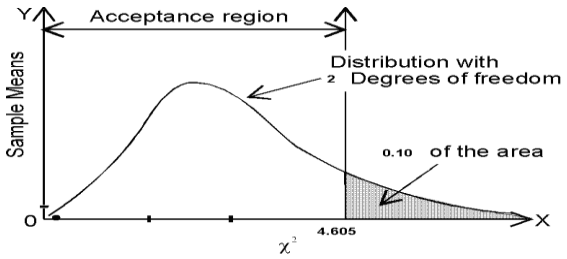
10.4.2 Determining Degrees of Freedom

Degrees of freedom for test of independence = (row – 1) (column – 1) as this reflects the fact that we can freely assign frequency to only (r – 1) row and (c – 1) column before the frequency of every cell is determined.

In this case, degrees of freedom = (2 – 1) (3 – 1) = 2.

This value can be used for looking up the value of the chi-square statistic. The value in the table is found at the intersection point of number of degrees of freedom (calculated from the contingency table) and the corresponding level of significance. This value denotes the area to the right end of the tail. According to the problem, at α = 10% and 2 degrees of freedom, the area under the right tail is 0.10 and the chi-square statistic is 4.605. Since the sample chi-square statistic (12.38) does not fall in the acceptance region, we reject the null hypothesis.

Figure 10.2: The Critical Region for Chi-square Value 12.38



10.5 Chi-Square Test as a Distributional Goodness of Fit

The unit 4 on 'Probability Distribution' discussed the binomial experiment. It is an experiment where the number of trials is fixed, probability of outcomes remain constant from success or failures and outcomes belong to one of the two categories. Multinomial experiments are similar in all except that the outcomes of each trial is classified into a different category. A method for testing a claim that in a multinomial experiments, frequency observed in the different category fits a particular distribution is discussed. As this test is to know-how well an observed frequency distribution conforms to some theoretical distribution, the method is called as goodness of fit test. Chi-Square distribution can be used to verify the appropriateness of the distribution and to conclude whether any significant difference exists between the observed (experimental) and the expected theoretical distribution employed.

Steps in Goodness of Fit Test

a. *Defining Null and Alternative Hypothesis:* Null and Alternative hypothesis for binomial experiment can be defined as follows:

   H_0: Experiment with p = $\bar{p}$ is best described by a Binomial Distribution (Null Hypothesis)

   H_1: The experiment cannot be described by the Binomial Distribution (Alternative Hypothesis).

b. *Finding the Value of the Chi-Square Statistic:*

The Chi-Square statistic is given by:

$$\chi^2 = \Sigma \frac{(f_0 - f_e)^2}{f_e}$$

Where,

f_0 is the observed frequency

f_e is the expected frequency.

c. *Determining the Degrees of Freedom for the Distribution:*

Degrees of freedom = k – 1

where 'k' refers to the number of points sought to compare while calculating observed and theoretical frequencies.

d. *Referring the Tables for the χ^2 Statistic at (k – 1) Number of Degrees and 'α' Level of Significance:*

In this step, the tables for the χ^2 statistic at (k – 1) degrees of freedom and required level of significance are referred.

e. *Accepting or Rejecting the Null Hypothesis:*

In this step, it is determined whether the computed statistic falls in the accepted region or not, and accordingly accepts or rejects the null hypothesis.

Example 1

A survey of 80 families with 4 children in Hyderabad revealed the following distribution:

No. of Male	0	3	1	2	4
No. of Female	4	1	3	2	0
No. of Families	10	15	18	24	13

Test the hypothesis that male and female births are equally probable. Use 1%.

The process can be described by a binomial distribution, and accordingly a probability of male birth p = 0.5 and probability of female birth q = 0.5 is assigned. This hypothesis should be tested at a level of significance of 1%.

To begin with, the hypothesis is set-up.

H_0: The process with p = 0.5 is best described by a Binomial Distribution (Null Hypothesis)

H_1: The process cannot be described by the Binomial Distribution (Alternative Hypothesis)

Significance level = 1%.

To test this hypothesis, seek whether the difference between actual values and the expected values is significant or not. The expected values refer to the values got if binomial distribution was the appropriate one to be used for describing this process. Expected values are computed by employing theoretical binomial probabilities or else by the formula shown below.

$$P(r) = {}^{n}C_r P^r (1-P)^{n-r}$$

$$= {}^{4}C_r\, 0.5^r (1-0.5)^{4-r} \quad = {}^{4}C_r\, 0.5^4.$$

Substituting r = 0,1,2,3,4, we have the probability of 1, 2, 3 and 4 male births.

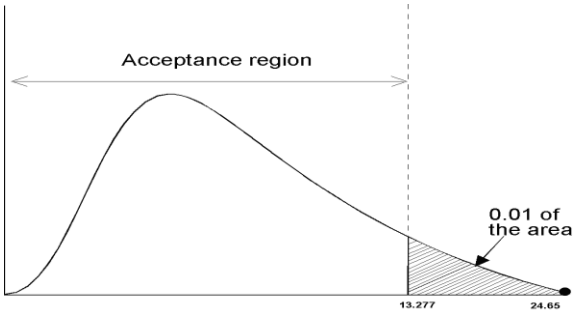
Now we can calculate the expected frequencies as:

No. of Male Birth (r)	Probability
0	${}^{4}C_0.0.5^4 = 1 \times 0.0625 = 0.0625$
1	${}^{4}C_1\ 0.5^4 = 4 \times 00625 = 0.25$
2	${}^{4}C_2\ 0.5^4 = 6 \times 0.0625 = 0.375$
3	${}^{4}C_3\ 0.5^4 = 4 \times 0.0625 = 0.25$
4	${}^{4}C_4\ 0.5^4 = 1 \times 0.0625 = 0.0625$

No. of Male Births	Total Family	Expected Probability (E)	Expected Frequency (E)	Observed Probability (O)	(O – E) (5) – (4)	$(O-E)^2$	$(O-E)^2/E$ (7) / (4)
(1)	(2)	(3)	(4) = (2) x (3)	(5)	(6)	(7)	(8)
0	80	0.0625	5	10	5	25	5
1	80	0.25	20	15	– 5	25	1.25
2	80	0.375	30	18	– 12	144	4.8
3	80	0.25	20	24	4	16	0.8
4	80	0.625	5	13	8	64	12.8
Total		1.00	80	80			$\chi^2 = 24.65$

So, the χ^2 for the test is 24.65. Degrees of freedom = 5 – 1 = 4.

Table value of χ^2 for 4 degrees of freedom and significance level of 1%, is 13.277.



As the calculated value χ^2 falls under the rejection region, the null hypothesis may be rejected and can be concluded that binomial distribution is not the appropriate distribution to be used in this situation.

Example 2

Hyderabad Runners, the largest running community in Hyderabad, is organizing a marathon event. As part of the event, they provide Running T-Shirts to the participants. They claim the proportion of T-Shirt sizes is 60% Large (L), 30% Medium (M) and 10% Small (S). Suppose a random sample of 100 actual sizes has 50 Large, 45 Medium, and 5 Small sizes. Is this consistent with Hyderabad Runners 'claim? Use 0.05 as the significance level.

Solution

There are 4 steps required to solve this problem:

1) State the Hypotheses.
2) Frame analysis plan.
3) Analyze the sample data.
4) Interpret the Results.

Following the steps:

State the Hypotheses

Null and the alternate hypothesis

Null Hypothesis: H0 = The proportion of Large Size (30%), Medium Size (60%), and Small Size (10%) respectively.

Alternate Hypothesis: Ha = One or more the proportions in the null hypothesis is different.

Frame an analysis plan

Significance level is 0.05.

Using sample data, a Chi Square Goodness of Fit test would be used.

Analyze Sample Data

For chi-square goodness of fit test to sample data, we compute the degrees of freedom. The expected frequency counts, and the chi-square test statistic are ascertained. Based on the chi-square statistic, DF, the P-value is determined.

DF = k - 1 = 3 - 1 = 2 (Ei) = n * pi

(E1) = 100 * 0.30 = 30

(E2) = 100 * 0.60 = 60

(E3) = 100 * 0.10 = 10

X2 = Σ [(Oi - Ei)2 / Ei]

X2 = [(50 - 30)2 / 30] + [(45 - 60)2 / 60] + [(5 - 10)2 / 10]

X2 = (400 / 30) + (225 / 60) + (25 / 10) = 13.33 + 3.75 + 2.50 = 19.58

DF = degrees of freedom

k = number of levels of the categorical variable

n = is the number of observations in the sample

Ei is the expected frequency count for level i,

Oi is the observed frequency count for level i,

and X2 is the chi-square test statistic.

The P-value is the probability that a chi-square statistic having 2 degrees of freedom is more extreme than 19.58.

Using Chi-Square Distribution Calculator to find P(X2 > 19.58) = 0.0001.

Interpret the results

The P-value (0.0001) is less than the significance level (0.05).

So, we reject the null hypothesis.

It implies that, at least one of the proportions is different.

The approach is appropriate because the sampling method was simple random sampling. The variable under study was categorical. Each level of the categorical variable had an expected frequency count of at least 5.

Check Your Progress - 1

1. The psychology department of an institute has developed an IQ test. It has compiled the result of the test conducted on a number of students and faculty members of the institute. After analyzing the test results they have arrived at this conclusion that the distribution of results can be said to follow a binomial distribution with the probability of scoring above average IQ level to be 60%. They want to test whether this particular result really follows a binomial distribution. Which of the following statistical techniques should be used to verify these findings?
   a. Analysis of variance.
   b. Regression analysis.
   c. Correlation analysis.
   d. Chi-square test.
   e. Marginal analysis.
2. A contingency table prepared for chi-square test contains two rows and two columns. The degrees of freedom in this case will be
   a. 0
   b. 1
   c. 2
   d. 3
   e. 4.

3. Which of the following is the denominator while calculating expected frequencies in Chi-square analysis?

   a. Row Total
   b. Column Total
   c. Grand Total
   d. Row Total x Column Total
   e. Row Total x Grand Total

4. Which of the following is first step in Chi-square analysis?

   a. Stating Hypothesis
   b. Finding degrees of freedom
   c. Calculate chi-square
   d. Referring Standard Table
   e. Deciding on accepting null hypothesis

5. Which of the following is one of the uses of Chi-square analysis?

   a. Skewness
   b. Appropriateness of distribution
   c. Sampling
   d. Covariance
   e. Central tendency

10.6 Analysis of Variance

Analysis of Variance (ANOVA) is a technique of testing the equality of three or more population means by analyzing sample variance. Recall from the unit "statistical inference" – the main drawback of the method is the limitation of the number of population mean. Using Z or t-test we cannot test the difference of population of more than two means. The F-test can be employed for testing significance of proportions for more than two populations.

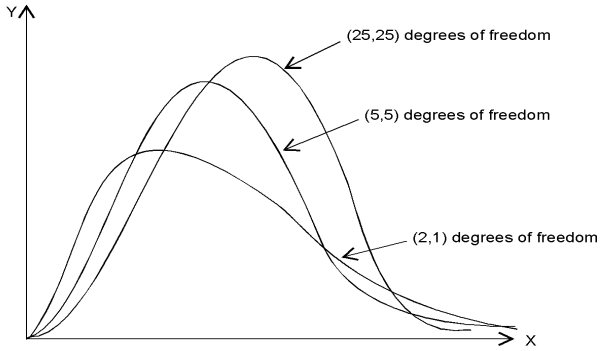
10.6.1 F Distribution

F-distribution, unlike the chi-square and t distribution uses two parameters, degrees of freedom of denominator and degrees of freedom of numerator. The shape of F-distribution depends on two degrees of freedom. The other properties of F-distribution are summarized below:

- F-distribution is not a probability distribution so that the area under the curve need not be 1.
- F-distribution is continuous and skewed to the right.
- The units of F-distribution are non-negative.
- The number of Degrees of Freedom (d.f.) is expressed as (N_1, N_2).

- The shape of F-distribution depends on the degrees of freedom.

Figure 10.3: Three F Distributions at Various Level of d.f.



10.6.2 F-Statistic

The F-statistic that arise from a one-way analysis of variance of the sample data is used to test the hypothesis about the population mean. F-statistic is defined as the ratio of variation among sample means and natural variation within groups. Mathematically, F-statistic can be expressed as:

$$\text{F ratio} = \frac{\text{Population variance obtained from the variance among the sample means (between column variance)}}{\text{Population variance obtained from the variance within the individual samples (within column variance)}}$$

$$\text{or } F = \frac{\text{Between column variance}}{\text{Within column variance}}$$

10.6.3 Assumptions made in ANOVA

i. The populations from which the samples are drawn should be normally distributed.

ii. The populations from which the samples are drawn should have same variance.

iii. The populations from which the samples are drawn should be random and independent.

iv. The analysis of variance is based on a comparison of two estimates of the population variance. One estimate is obtained from variance among the sample means and the second estimate is obtained from variation that exists within the samples.

10.6.4 Steps in ANOVA

The basic concept in one-way analysis is that the variation among the data values in the overall sample has to be separated in two parts:

i. Difference between the group means.

ii. Difference within the group means.

Total Variation = Variation between Group + Variation within Groups.

In comparing the means of population represented by the independent sample, null and alternative hypothesis can be written as

$H_0 = \mu_1 = \mu_2 = \mu_3 ..... = \mu_n$ i.e., sample means are equal

H_1 = The means are not equal.

Notice that alternative hypothesis does not require that all means must differ from each other.

10.6.5 Measuring Variation between Groups

The variation between group means is measured as the weighted sum of sample mean and the overall mean of all data.

$$S^2 = \frac{\Sigma\left(\bar{X} - \bar{\bar{X}}\right)^2}{k-1}$$

Where,

$\bar{X}$ denotes the sample mean,

$\bar{\bar{X}}$ the mean of all sample means, and

k denotes the number of sample means.

In ANOVA, the mean of all sample means is called as the grand mean ($\bar{\bar{X}}$).

Standard error of the mean is given by the equation.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Where,

$\sigma_{\bar{x}}$ is the standard error of mean (in this case, it is of sample means)

σ is the population standard deviation

n is the size of the sample.

By replacing the value of σ with $\sqrt{s^2}$.

An estimate of the population variance will be: $\hat{\sigma}^2 = \frac{\Sigma n(\bar{X} - \bar{\bar{X}})^2}{k-1}$

Since, 'n' and '$\bar{X}$' represent individual sample size and sample mean, its generalized form would be: $\hat{\sigma}^2 = \frac{\Sigma n_j(\bar{X} - \bar{\bar{X}})^2}{k-1}$.

10.6.6 Calculation of the Grand Mean

Grand mean can either be calculated as the arithmetic mean of given data set or weighted average of sample mean. The weights will be the respective sample sizes.

Method I

$$\text{Grand Mean} = \frac{X_1 + X_2 + X_3 + \ldots\ldots + X_n}{N}$$

Method II

$$\text{Grand Mean} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_2\bar{X}_3 + \ldots\ldots\ldots + n_n\bar{X}_n}{N}$$

Where,

n is the size of sample and N is the total number of data sets.

10.6.7 Measuring Variation within Groups

Population variance based on the variance within the samples can be calculated by the formula:

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

Since it is assumed that all the sample variances need to be equal, either of them obtained from the four samples can be used or else to get a better estimate the weighted average of the four sample variances can be used. Employ weighted average method of the sample variance, the resulting equation would be:

$$\hat{\sigma}^2 = \Sigma\left(\frac{n_j - 1}{n_T - k}\right)S_j^2$$

Where,

- $\hat{\sigma}^2$ is the second estimate of the population variance
- n_j is the size of the sample
- n_T is the total number of elements present in all the samples
- k is the number of samples
- S_j is the sample variance of the sample j.

10.6.8 Measuring Degrees of Freedom

The degrees of freedom (d.f.) for the numerator are given by a general rule as shown below: (k – 1)

Where,

k stands for the number of samples.

d.f. for the denominator is given by: $\sum_{k=1}^{n} (n_j - 1)$

Where,

- n_j denotes the number of elements in each of the sample
- Σ Symbol indicating that the sum of number of d.f. of each sample has to be taken.

The same is obtained from $(n_T - k)$

Where,

n_T denotes the number of total elements

k denotes the number of samples.

10.6.9 Value of F-Statistic from Table

The value of F-statistic for a given degrees of freedom for denominator and numerator can also be found.

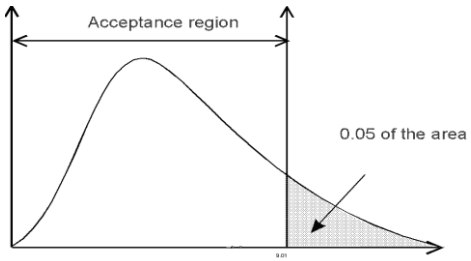
For example, to find the value of F-statistic for 5 degrees of freedom for numerator and 3 degrees of freedom of denominator and at 0.05 level of significance, the value of F-statistic will be 9.01.

10.6.10 Taking Decision about Null Hypothesis

While deciding whether the null hypothesis is accepted or rejected, the calculated

F-statistic with the value of the F-statistic from the tables can be compared. Then depending on whether the computed F value falls in the acceptance region or not, accept or reject the null hypothesis. For F-statistic of 9.01, the acceptance and rejection region will be as follows:

Figure 10.4: Critical Region for F Distribution



Example 3

Ocimum Labs is a renowned company manufacturing biochemicals based products. It has produced clones of popular genes such as DNA, RNA and Oops. The number of pairs sold for each of the genes on different days is as follows:

DNA	RNA	Oops
8	11	9
4	10	7
2	4	5
10	10	9

A test of ANOVA is to be performed for testing the equality of the mean number of pairs sold of the three genes.

$\bar{X}_1$ (Mean sale of DNA); $\bar{X}_2$ (Mean sales of RNA); $\bar{X}_3$ (Mean sales of Oops)

The means of the samples are:

$$\bar{X}_1 = \frac{24}{4} = 6; \quad \bar{X}_2 = \frac{35}{4} = 8.75; \quad \bar{X}_3 = \frac{30}{4} = 7.5$$

We calculate variance as below:

$$\hat{\sigma}^2 = \sum\left(\frac{n_j - 1}{n_T - k}\right) \times s_j^2$$

$$s_1^2 = \frac{(8-6)^2 + (4-6)^2 + (2-6)^2 + (10-6)^2}{4-1} = \frac{40}{3}$$

$$s_2^2 = \frac{(11-8.75)^2 + (10-8.75)^2 + (4-8.75)^2 + (10-8.75)^2}{4-1} = \frac{30.75}{3}$$

$$s_3^2 = \frac{(9-7.5)^2 + (7-7.5)^2 + (5-7.5)^2 + (9-7.5)^2}{4-1} = \frac{11}{3}$$

So,

$$\hat{\sigma}^2 = \left(\frac{4-1}{12-3}\right) \times \left(\frac{40}{3} + \frac{30.75}{3} + \frac{11}{3}\right) = 9.083$$

Example 4

A retailer ran a sales promotion campaign across 5 Indian cities viz. Mumbai, Bangalore, Hyderabad, Delhi, and Pune for a period of 30 days at the point of sale. It was classified as gift item, cash discounts, gift vouchers, and buy one get one free. The sales in the five different cities were listed:

	Cities				
Type of Promotion	Hyderabad	Delhi	Mumbai	Bangalore	Pune
Gift Vouchers	78	87	81	89	85
Gift Item	94	91	87	90	88
Cash Discount	73	78	69	83	76
Buy one get one free	79	83	78	69	81

a) Estimate the population variance using the between-column variance.

b) Estimate the population variance using the within-column variance computed from the variance within the samples.

c) Compute the mean unit sales for each type of promotion and then determine the grand mean.

e) Calculate the F-statistic. At the 0.01 level of significance, do the promotions produce different effects on sales?

Solution

a)

	Gift Vouchers	Gift Item	Cash Discount	Buy one get one free
	78	94	73	79
	87	91	78	83
	81	87	69	78
	89	90	83	69
	85	88	76	81
Total	**420**	**450**	**379**	**390**
n	5	5	5	5
x bar	84	90	75.8	78
Sigma x^2	35,360	40,530	28,839	30,536
s^2	20	7.5	27.7	29

Grand Mean= (420+450+379+390)/20 = 81.95

b) Population Variance Estimate= 5 $[(84-81.95)^2 + (90 - 81.95)^2 + (75.8-81.95)^2 + (78-81.95)^2]$ / (4-1)

= 612.15/3 = 204.05

c) Estimate = 4(20+7.5 +27.7 + 29) / (20 -4) = 336.8 /16 = 21.05

d) F = 204.05 / 21.05 = 9.69

With 3 DF in the numerator, 16 DF in the denominator, $\alpha = 0.01$, the critical value of F is 5.29, so reject H_0 , because 9.69 > 5.29.

This analyses to promotions has a significantly different effect on sales

10.7 Examining the Population Variance

Apart from population mean and population proportion, we need to estimate and control variance. This part will discuss the inference about population variance.

A cinema hall has a cool drinks fountain supplying Coco-Cola. When the machine is turned on, it fills 250 ml of the required drink. Note that machine will not put exactly 250 ml every time, sometimes it will provide more than 250 ml and sometimes it will provide less than 250 ml. If the variance is large it means there is large variation from the actual 250 ml. To keep the variance within the range, the machine will be adjusted from time to time. Before taking a decision to adjust the machine, the manager must conduct a test to find whether variance is within acceptance limit or not. For this he can select different samples. The probability distribution of variance of these samples is called sampling distribution of variance.

10.7.1 Inference about a Population Variance

Suppose the population from which sample is selected is normally distributed, then chi-square distribution with n – 1 degrees of freedom is given by the formula:

$$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$$

Here, the sample variance (s^2) is employed as an estimate for the population variance (σ^2).

Estimation of Population Variance

At $(1 - \alpha)$% of confidence interval, population variance is given by:

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} \quad \text{and} \quad \frac{(n-1)s^2}{\chi^2_{(1-\frac{\alpha}{2})}}$$

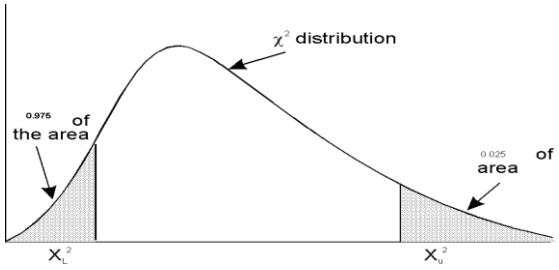
$\chi^2_{\frac{\alpha}{2}}$ and $\chi^2_{(1-\frac{\alpha}{2})}$ can be obtained from chi square distribution table for a given degrees of freedom.

Example 5

What is the value of $\chi^2_{\frac{\alpha}{2}}$ and $\chi^2_{(1-\frac{\alpha}{2})}$ for 10 degree of freedom and 95% confidence interval?

Here α = 5%, so $\frac{\alpha}{2}$ = 0.025 and $1 - \frac{\alpha}{2}$ = 0.975

For 10 degrees of freedom and $\frac{\alpha}{2}$ = 0.025, the χ^2 value is 20.483 and for $1 - \frac{\alpha}{2}$ = 0.975 the χ^2 value is 3.247.



Hypothesis Test

As discussed above, a complete hypothesis test about population variance consists of the following steps:

Step I : State the null and alternative hypothesis.

Step II : Select the distribution to use.

Step III : Determine the acceptance and rejection region.

Step IV : Calculate the value of test statistics.

Step V : Make the decision.

10.8 Inferences about Two Population Variance

The hypothesis about difference in two population means is discussed in unit on the statistical inference. In this section, an inference about two population variations is discussed. Inference about difference in population variations can be One-tailed test or Two-tailed tests.

10.8.1 One-Tailed Test

State the Null and Alternative Hypothesis

To test whether the differences between two populations are significant enough or not, set-up the hypothesis as follows:

H_0: $\sigma_1^2/\sigma_2^2 = 1$ (Null Hypothesis: Two variances are equal)

H_1: $\sigma_1^2/\sigma_2^2 > 1$ (Alternative Hypothesis: One of the variances is greater than the other).

Where,

p_1 and p_2 are two populations such that σ_1^2 and σ_2^2 are their respective variances, and the corresponding sample variances are given by s_1^2 and s_2^2.

Distribution to Use

F distribution is used to test the hypothesis about difference in population variance. For $n_1 - 1$ degrees of freedom in numerator and $n_2 - 1$ degrees of freedom in denominator, F distribution is defined as: $F = \frac{s_1^2}{s_2^2}$.

Determine the Acceptance and Rejection Region

A cautious approach should be taken while using the one-tailed test. The above F-statistic is used for alternative hypothesis $\sigma_1^2/\sigma_2^2 > 1$. But suppose alternative hypothesis is $\frac{\sigma_2^2}{\sigma_1^2} > 1$, the F ratio has to be adjusted as $\frac{s_2^2}{s_1^2}$ with $n_2 - 1$ degrees of freedom in numerator and $n_1 - 1$ degrees of freedom in denominator F.

10.8.2 Two-Tailed Test

The hypothesis will be:

H_0: $\sigma_1^2 = \sigma_2^2$ (Null Hypothesis: Populations have the same variance)

H_1: $\sigma_1^2 \neq \sigma_2^2$ (Alternative Hypothesis: Populations do not have the same variance)

Calculate the F-statistic, $F = \frac{s_1^2}{s_2^2}$.

Example 6

The organization believed that the performance of experienced employees is definitely better than that of the new employees. The performance is computed in terms of the rate of output. It means the rate of output is greater for the experienced employees. As per certain computations, the average unit output per hour for new employees for a particular type of work is 20 units per hour, and a variance of 56 units squared. For the group of 20 employees with 5 years' experience, the average output for the same type of work is 30 units per hour, and a sample variance of 28 units squared. Does the variability appear to differ at the two levels of experience? Test the hypothesis, using 0.05 level of significance.

Solution

We test the hypothesis at 0.05 level of significance

$H_{0:}\ \sigma^2 = 56$

$H_a: \sigma^2 \neq 56$

At $\alpha = 0.05$, the limits of the acceptance region are $\chi^2 = 8.907$ and $\chi^2 = 32.852$

The observed $\chi^2 = (n-1)\ s^2/\ \sigma^2 = 19\ (28)\ /56 = 9.5$.

Don't reject H_0 ; the variability is not significantly different.

Check Your Progress - 2

6. In a test involving ANOVA the following details are obtained:

   Estimated population variance based on the variance among the sample means = 20

   Estimated population variance based on the variances within the samples = 14.77

   What is the F-statistic for the data?

   a. 0.739.
   b. 1.354.
   c. 5.23.
   d. 34.77.
   e. 295.4.

7. The following details are available for a test involving ANOVA:

Number of samples	= 4
Size of the first sample	= 5
Size of the second sample	= 5
Size of the third sample	= 6
Size of the fourth sample	= 4

The number of degrees of freedom in the numerator and denominator of the F-ratio are

a. 4 and 6 respectively
b. 4 and 4 respectively
c. 5 and 4 respectively
d. 3 and 16 respectively
e. 4 and 20 respectively.

A retailing chain in order to improve sales has adopted four promotion policies at five of its stores. The data collected is shown below.

	Store 1	Store 2	Store 3	Store 4	Store 5
Free sample	78	87	81	89	85
One pack gift	94	91	87	90	88
Discount on cost price	73	78	69	83	76
Refund by mail if unsatisfied	79	83	78	69	81

8. What is the first estimation of population variance using between-column variance methods?

a. 204.05.
b. 203.25.
c. 201.25.
d. 205.
e. 206.63.

9. An analyst wants to have analysis and inference about two population variations. The analyst preferred to conduct the two-tailed test. What would be the null hypothesis?

a. $\sigma_1^2 = \sigma_2^2$.
b. $\sigma_1^2 \neq \sigma_2^2$
c. $\sigma_1^2 = 0$.
d. $\sigma_2^2 = 0$
e. $\sigma_2^2 \neq \sigma_1^2$.

10. A statistician wants to analyze the equality or otherwise of three population means by analyzing the sample variance. What is this statistical inference process referred to as by statisticians?

a. P test
b. Z test
c. Analysis of variance
d. Chi-square test
e. T test

10.9 Summary

- In this unit, the test of hypothesis about contingency table, called test of independence and test of hypothesis for experiment with more than two categories called goodness of fit-test have been discussed.
- Further, Test of hypothesis about the variance and standard deviation has been discussed. The Chi-Square distribution and F-distribution is used to perform this entire test. The Chi-Square distribution, like the t distribution has single parameter called degrees of freedom. It's a distribution of random variable called χ^2; so it assumes only positive value.
- The shape of chi-square distribution is dependant on the degrees of freedom. F-distribution unlike the chi-square and t-distribution uses two parameters – degrees of freedom of denominator and degrees of freedom of numerator. The shape of F-distribution is dependant on two degrees of freedom.

10.10 Glossary

Analysis of Variance (**ANOVA):** Analysis of variance is a technique of testing the equality of three or more population mean by analyzing sample variance.

Chi-Square Distribution: A distribution, like t-distribution has single parameter called degrees of freedom. The shape of Chi-square distribution is dependant on the degree of freedom, and the value of chi-square statistic is always positive.

Contingency Table: Tables constructed by listing all the levels of one variable as rows in a table and the levels of the other variables as column.

F-Distribution: A family of distributions differentiated by two parameters (d.f. numerator, d.f. denominator), used primarily to test hypotheses regarding variances.

10.11 Suggested Readings/Reference Material

1. Gupta, S. P. Statistical Methods. 46th Revised ed. New Delhi: Sultan Chand & Sons. 2021.
2. I. Levin Richard, H. Siddiqui Masood, S. Rubin David, Rastogi Sanjay. Statistics for Management. Pearson Education; Eighth edition, 2017.
3. Gerald Keller. Statistics for Management and Economics. Cengage, 2017.
4. Arora, P. N., and Arora, S. CA Foundation Course Statistics. 6th ed. S Chand Publishing, 2018.
5. Mario F Triola. Elementary Statistics. 13th ed., 2018.
6. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran. Statistics for Business and Economics. 13th Edition, Cengage Learning India Pvt. Ltd., 2019.

7. S D Sharma. Operations Research. KEDAR NATH RAM NATH, 2018.
8. Hamdy A. Taha. Operations Research: An Introduction. 10th ed., Pearson, 2016.
9. Malhotra, N. (2012), Marketing Research: An Applied Orientation, 7th ed., Pearson, 2019.
10. Cooper, D.R. and Schindler, P.S. and J. K. Sharma (2018), Business Research Methods, 12th edition, McGraw-Hill Education.

10.12 Self-Assessment Questions

1. Discuss the importance of χ^2-test.

2. Explain the Chi-Square Test as a Distributional Goodness of Fit.

3. What is Analysis of Variance? Explain clearly the technique of analysis of variance for data with one-way classification.

4. What assumption need to be made while using the technique of analysis of variance?

5. What is F-distribution? Explain the characteristics of F-distribution.

10.13 Answers to Check Your Progress Questions

1. (d) Chi-square test

2. (d) 3

3. (c) Grand Total

$$\text{Expected Frequency} = \frac{(\text{Row Total})(\text{Column Total})}{\text{Grand Total}}$$

4. (a) Stating hypothesis

The Chi-square analysis starts with defining null and alternate hypotheses.

5. (b) Appropriateness of distribution

Chi-Square distribution can be used to verify the appropriateness of the distribution and to conclude whether any significant difference exists between the observed (experimental) and the expected theoretical distribution employed.

6. (b) 1.354

(20/14.77)

7. (d) 3,16

(Samples-1, sum of sample sizes-4)

8. (d) 68.17

(Sum of 5 column variances*4/17)

9. (a) $\sigma_1^2 = \sigma_2^2$.

10. (c) Analysis of Variance.

Unit 11

Role of IT in Modern Business Enterprise

Structure

11.1 Introduction

In the previous unit, we learned Chi-square analysis and Analysis of Variance. Chi-Square Test is used to determine whether the two attributes according to which a population is categorized are independent of each other or not and it also serves as a test for distributional goodness of fit. Analysis of Variance tests whether the difference between the proportions representing more than two samples is significant or not and it also tests whether the variability among them is significant or not.

For any analysis, if the data points are huge, we need some computers to solve them. In this unit, we will gain knowledge on the role of computers and IT in a firm.

From Abacus to Pentium to Core i9 personal computers, we have come a long way. The IT wave has been changed to an extent which was not evident even during the industrial revolution. What more one can ask for as the modern research promises to take us beyond the unthinkable. What exactly will be the contours of future technology is difficult to say now. Irrespective of this, computers vis-à-vis the information technology have succeeded in creating an enviable position for themselves. In this scenario can one afford to ignore the rapidly changing environment around him? Obviously not, especially if one wants to conquer the competition. And it can be said that one can find a reliable ally in computers.

In this chapter, the first two topics 'Problem solving Using a Computer and Designing Computer Based Information Systems' will give you a brief overview of solving basic problems by using the concepts of flow charting and designing computer-based information systems in an organization. The third topic 'Enterprise Resource Planning' to which the whole world has been turning to make their organizations more competitive explains how to go about integrating the different functions that exist in an enterprise so that it functions as a single organism in a nutshell. The fourth topic deals with some of the basic terminology one gets to hear while talking about the current rage of the world: 'The Internet'.

11.2 Objectives

After going through the unit, you should be able to:

- Recall the steps required for solving problem using a computer;
- Recognize designing computer based information systems;
- Examine how Enterprise Resource Planning integrates different functions in an enterprise ; and
- Define basic terminology of Internet.

11.3 Problem Solving Using a Computer

The following steps are required for solving a problem using a computer system:

i. Problem Definition
ii. Program Planning (Preparation of a flow chart)
iii. Writing the Program
iv. Debugging
v. Documentation

11.3.1 Problem Definition

In this phase, a concise statement of the problem is made. All inputs available and outputs required are specified. A thorough understanding of the problem is essential, as otherwise the solutions aimed at may not be correct.

11.3.2 Program Planning (Preparation of a Flow Chart)

A flow chart is a graphic method for indicating a proposed or actual solution to the problem. A flow chart shows the steps a computer performs to solve a problem. A programer usually develops a program flow chart before coding. A flow chart uses special symbols for representing various operations.

Creating flow charts is an art, not a science. It is not a cut and dried process with firm rules. Rather it offers the analyst and the programer an opportunity to give expression to their creative urges and to apply their creative powers to problem solving.

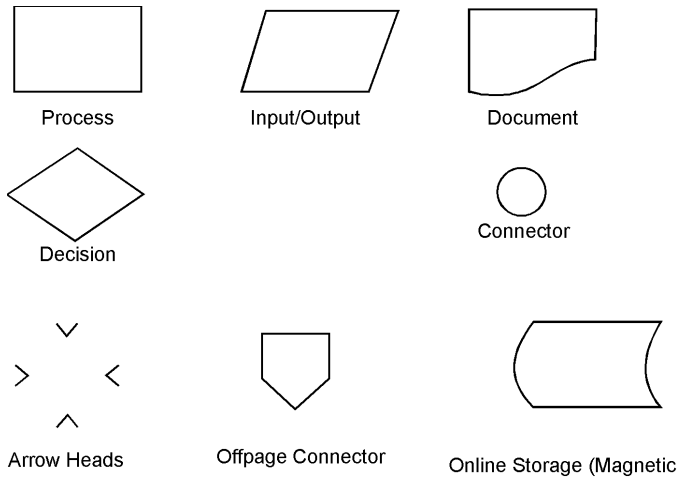
There is no unique 'Correct' flow chart for solving a given problem. Each programer can come up with his own flow chart provided some formal rules for drawing flow charts are observed. A set of commonly used flow chart symbols are given below.

Adhering to the following rules is considered to be a good flow charting technique.

a. The direction of flow should usually be from the top to the bottom of the page and from the left to the right.
b. Lines should never cross. Crossing lines can always be avoided by the use of a pair of connector symbols.
c. Arrow-heads should be used to show the direction of the flow, especially, if it is other than top to bottom or left to right.
d. The symbols may be of any size, only the shape is standard.
e. Every flow chart should begin with a 'Start' symbol and end with a 'Stop' symbol. The start and stop will appear only once in each flow chart.

   These merely indicate the physical starting and ending points of the flow chart.
f. The I/O symbol can appear at any point at which data is to be entered in the computer or any point at which data is being output from the computer.

Figure 11.1



g. The next step in completing the flow chart is to mentally analyze the problem in order to obtain a logical sequence of processes. The process symbol is used to indicate the steps involved in manipulating the data into the desired result.

h. The basic contributor to flow charting complexity is the fact that questions must be answered during the processing that is decisions must be made. In many instances, the computer must choose between two logical paths. Therefore, the program must reduce the required decision to a comparison of two values. The decision box must have at least two paths leading from it. Each path is labelled to indicate the condition which will cause the flow to follow a particular path.

i. In certain circumstances, a programmer may find it difficult to avoid crossing a line or to prevent drawing a long or jagged line between symbols. The connector symbol indicates a transfer of flow and thus, always must appear in pairs. An arrow-head leading into a connector indicates that control is to be transferred to the point at which that connector's counterpart appears. An arrow-head leading from a connector indicates the point at which control is to re-enter the flow chart.

j. The printer output symbol indicates that the results are to be output at the printer.

k. Program comment or annotation – a comment or a description inside the box can be used to clarify some point of the flow chart.

l. Off-page connector indicates reference to another point on a different page.

m. Magnetic-tape/Magnetic-disk/Magnetic-Drum indicates that input/output is from respective storage unit.

11.3.3 Principal Structures in Flow Charting

Structured programing calls for the usage of four principal structures to help simplify the program. They are:

a. Sequential control

b. IF-THEN-ELSE control

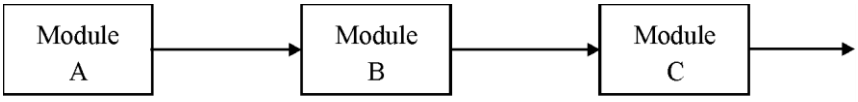
c. Replication PRE-TEST (DO WHILE) AND POST-TEST (DO UNTIL)

d. Case Structure PRE-TEST AND POST-TEST

These structures are illustrated in the 5 figures.

The first of the figures, is the conventional sequential flow in a given order. That is, Module A, followed by Module B and followed by Module C. For example: Data is read, data is processed and data is printed out.

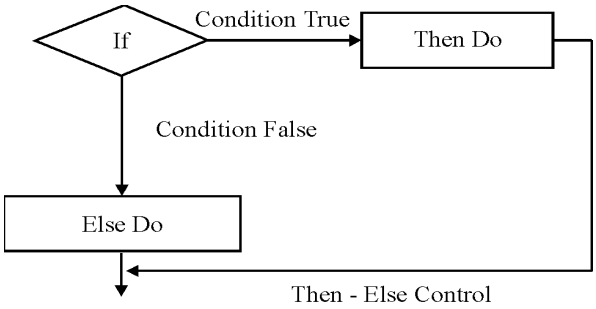
Figure 11.2



Sequential Structure

The second structure, is used for processing if a specific condition is true. For example: If the condition is true, then MODULE A is processed, ELSE the condition is false and MODULE B is processed.

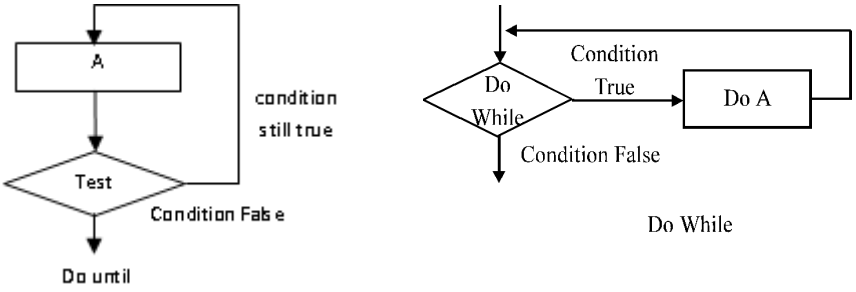
Figure 11.3



The third structure is used for repeated operations (replications). For example: Do the following sequence of instructions WHILE the condition is TRUE.

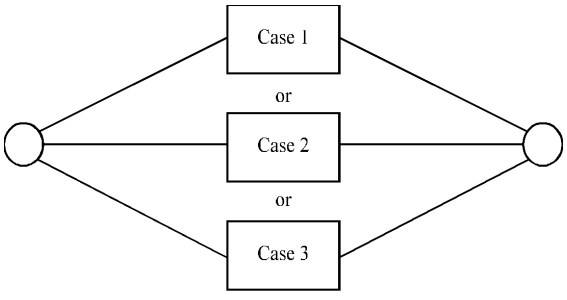
The DO AND TEST (or DO UNTIL) is similar to the DO WHILE except for an important difference. In the DO WHILE the condition for executing MODULE A is first checked and if not true, MODULE A is not executed. In the DO AND TEST, MODULE A is first processed, then the condition tested.

Figure 11.4



The following figure describes the case structure. This is applicable when there are multiple alternatives available and depending on the testing of a condition, exactly one case has to be processed.

Figure 11.5



Case Structure

Any program can be designed by combining these various structures.

11.3.4 Writing the Programs

Once the algorithm for solving the problem is specified using a flow chart, it has to be converted into a high-level language program for execution on the computer system. Specific high-level language is selected and using the rules of the language, program is written. Writing a program is often called "Coding".

Examples

a. Draw a flow chart to find the sum of fifty numbers. The flow chart is shown below.

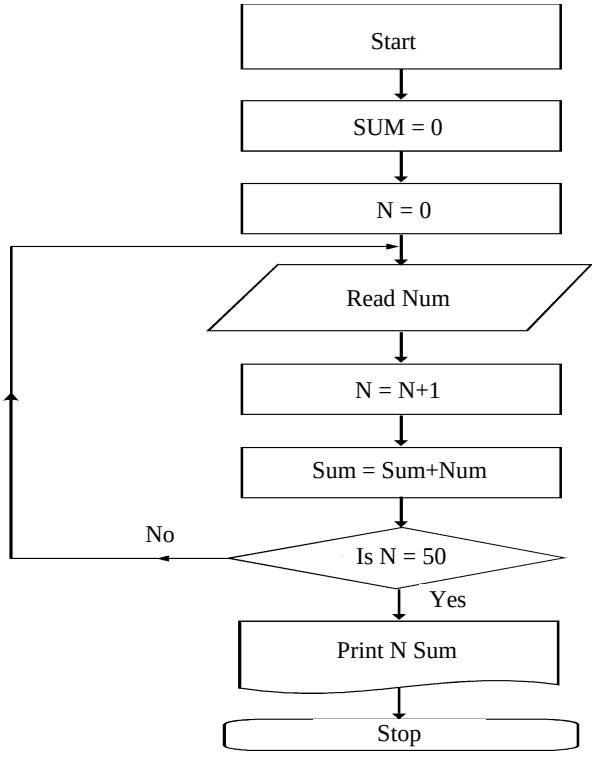
   After starting it reads the data, i.e. transmits information to the computer regarding the number. An accumulator is a variable where you can store totals. After the first number has been read the result will be stored in the accumulator. After the 2nd number is read its contents will be added to the accumulator (SUM). We keep on adding a number to the accumulator until we have read all the 50 numbers.

   'N' acts as a counter, it is incremented by one every time a number is read.

   In the decision box we test a condition and perform one of the two alternatives. In the example, if the condition N = 50 is true, then we execute "Yes" branch and if it is not true, then we execute the "No" branch.

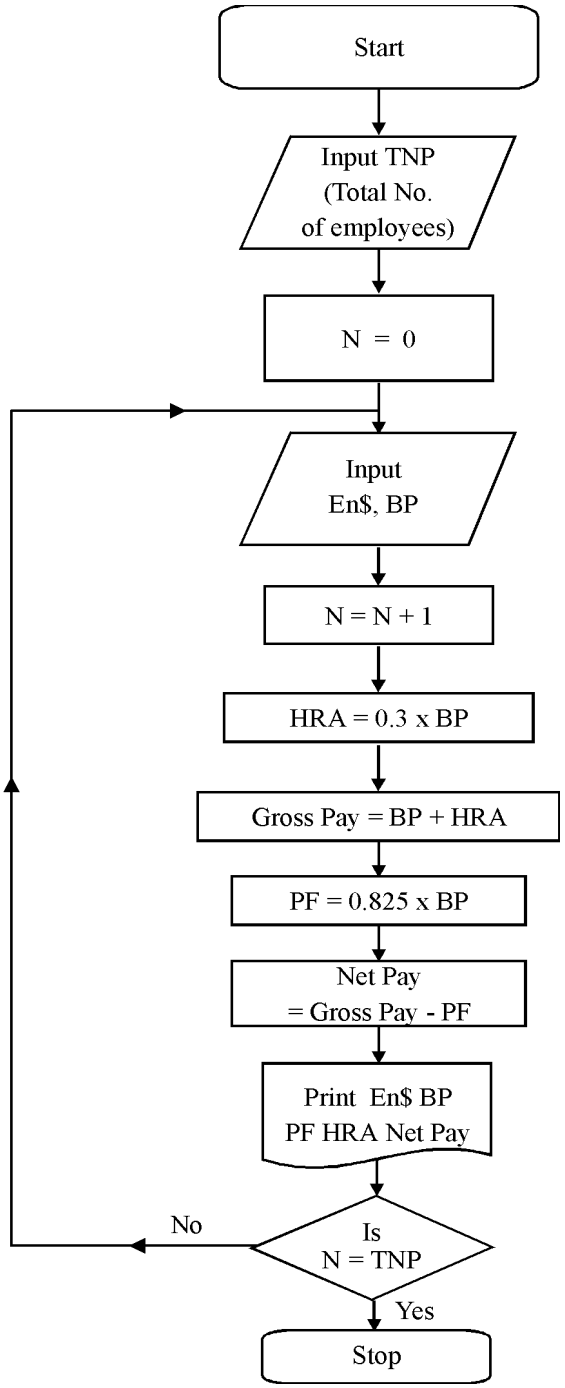
   A loop specifies repeated execution of program steps. Every loop should be properly terminated, lest we have an infinite loop, which can come about by logical error. In the example, the loop terminates when we have read 50 numbers.

Figure 11.6



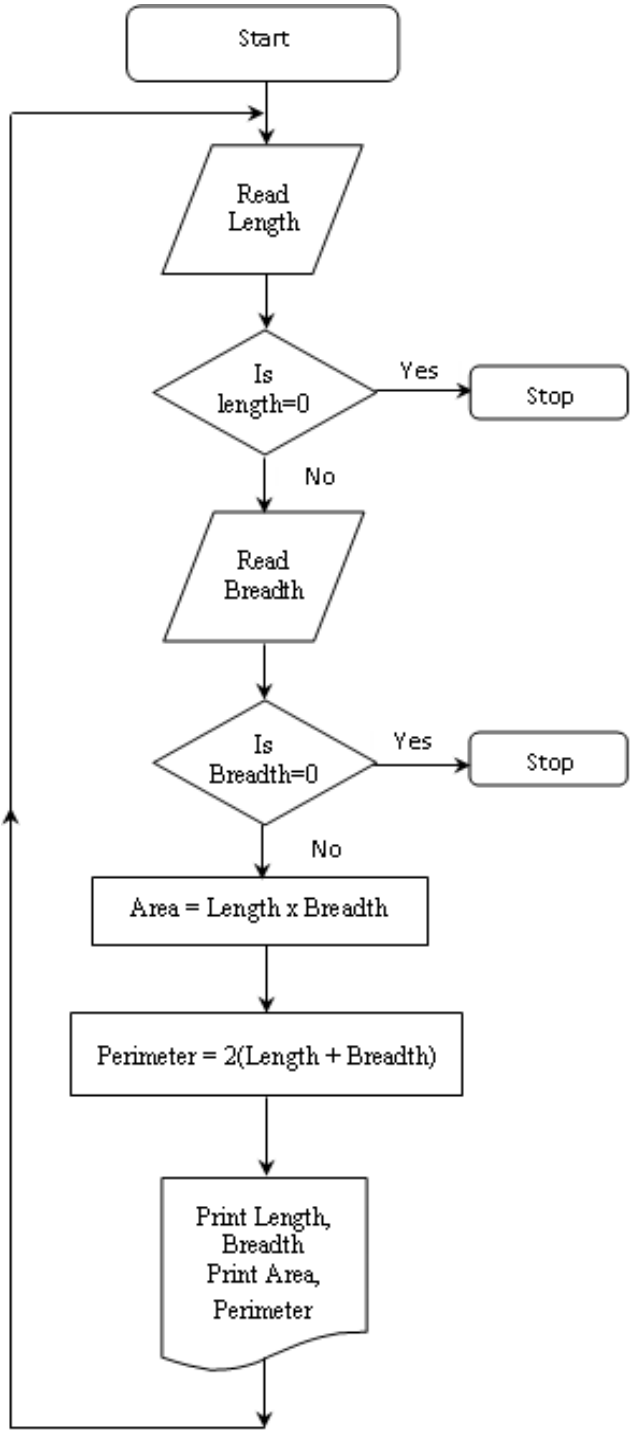
b. Draw a flow chart which will compute the net pay for several employees of an organization. The gross pay of an individual is made up of the basic pay and house rent allowance (which is 30% of the basic pay). The net pay is computed from the gross, by deducting the Provident Fund (which is 8.25% of the basic pay).

Figure 11.7



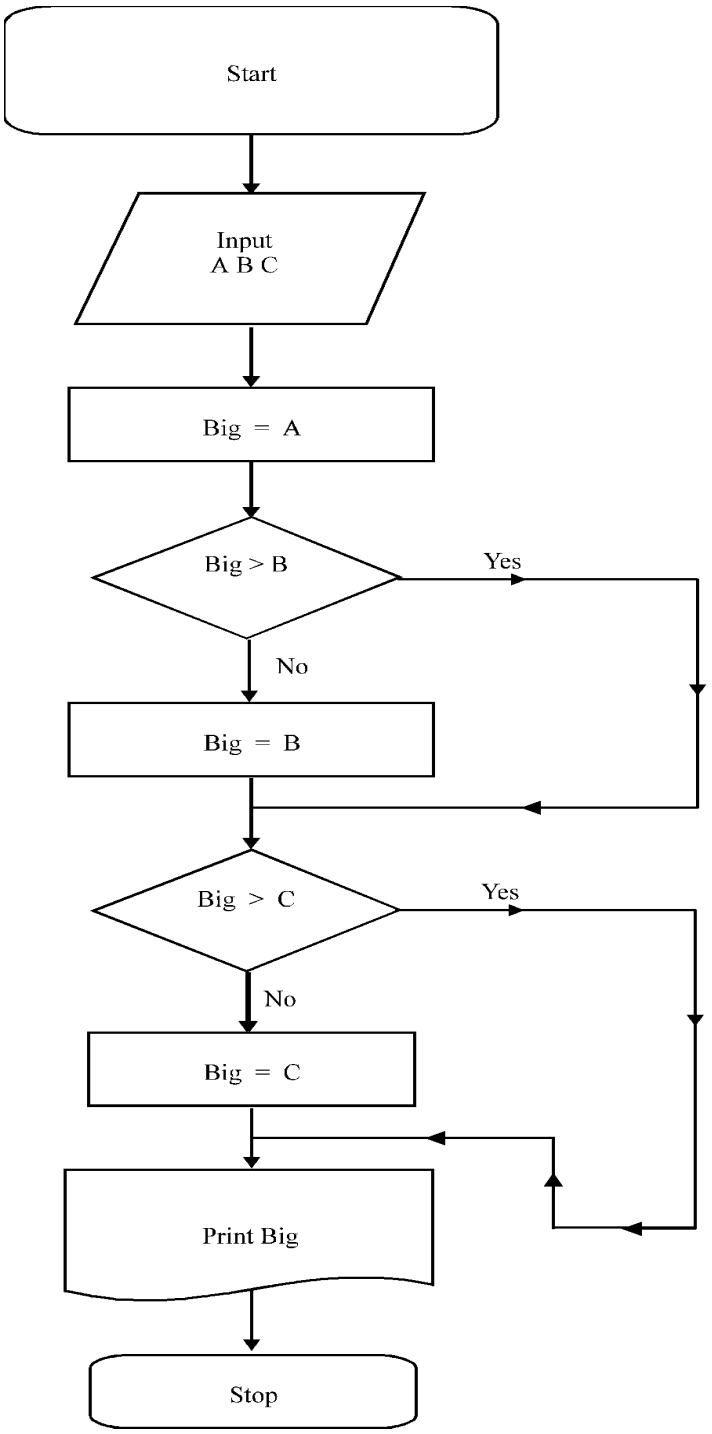
c. Draw a flow chart to input the length and breadth of a rectangle through a keyboard. Calculate the area and perimeter of the rectangle and print it. Keep on printing the area of various rectangles till either the length or the breadth is equal to zero.

Figure 11.8



d. Draw a flow chart to find the largest of three given numbers A, B, and C.

Figure 11.9



11.3.5 Debugging

Once a program has been written, it has to be tested first to ensure that it is free of errors. It is very normal that any program written will contain mistakes. When the source program is submitted to the compiler for conversion to object

program, compiler identifies errors so that the same can be corrected. This process of checking and correcting errors is known as "Debugging". The following steps are involved:

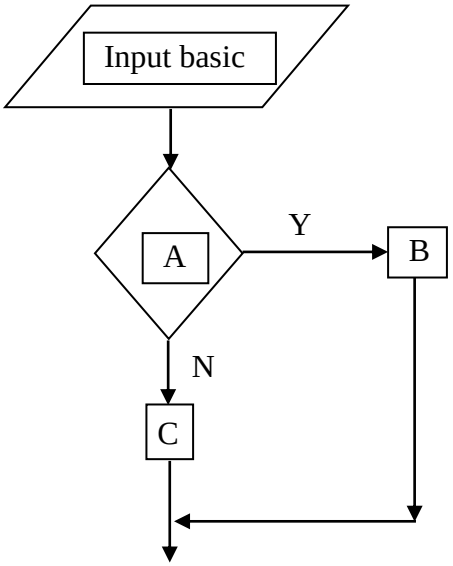
a. Desk checking
b. Compilation
c. Testing with test data
d. Running the program with actual data.

11.3.6 Documentation

In this phase, once the program has been tested and corrected, a complete documentation is made for future reference.

Check Your Progress - 1

1. The programmer from HItechsoft is making a flow chart for a payroll package. The rule says that, the HRA is 30% of basic for all those having 30000 or less salary, otherwise it is 45%. What shall be the equations at A, B, C in this chart?



   a. A: basic≤ 30000, B: hra=basic*0.45, C:hra=0.3*basic
   b. A: basic ≥ 30000, B: hra=basic*0.3, C:hra=0.45*basic
   c. A: basic = 30000, B: hra=basic*0.3, C:hra=0.45*basic
   d. A: basic≤ 30000, B: hra=basic*0.3, C:hra=0.45*basic

2. Mr Kunal, the software developer wrote the program code in a higher level language and during process found that, there are some errors.

He needs to correct them and resubmit for the program to work correctly. What is this process referred as?

a. Compilation
b. Debugging
c. Correcting
d. Writing Code

3. Which of the following are not the steps followed in Debugging?

a. Desk checking
b. Compilation
c. Testing with test data
d. Documentation.

4. Which of the following is true in the symbols of Flow charts?

a. Any Size
b. Uniform size
c. One shape for all
d. Cross marks are used.

5. In which of the following steps flow chart is prepared while solving a problem using a computer system:

a. Problem Definition
b. Program Planning
c. Debugging
d. Documentation

11.4 Designing Computer Based Information Systems

11.4.1 Background

The role of Management Information Systems (MIS) in an organization can be considered as providing management with required information in right time to help in their decision processes. Information systems are not restricted to any particular functional areas like production, finance, etc. or managerial level. It covers all segments of the company covering all areas. It is also known that additional role of providing required information to external agencies has to be fulfilled.

Information Systems does not necessarily imply usage of computers. However, in today's context, where computers have become very common, generally information systems and computers go hand in hand.

All organizations have some form of MIS. The data originating daily is captured in some form, manually or using computers, stored and processed to fulfill

necessary requirements. The main issue remains whether or not the existing MIS is able to fulfill the information needs within the required time frame and provide analyzed information to managers to arrive at a better decision.

Designing of an information system is not an easy task. Firstly, specific areas where computer application may help need to be identified. For each specific area, thorough understanding of the specific system is required, followed by information needs of managers. Only then any improvements can be conceptualized. This process is popularly known as "Systems Analysis and Design".

Systems analysis is the process of evaluating and understanding a particular system and its environment. In a way, it will involve studying the inputs to a system, outputs required and the associated processing to convert inputs to desired outputs.

Understanding the existing system is often much more difficult than expected. It requires gathering a lot of information. This will include, interviews with key personnel, study of forms and reports, norms for evaluation and quite often direct observation of production/service processes. This phase involves dealing with human beings, who quite often resist any type of change and hence co-operation to the System Analyst may be rather low.

In the context of computers, the resistance is much more as there is always a fear in the mind of the concerned persons like fear of losing jobs, more control, etc.

Systems analysis is probably the most difficult part in the development of a computer-based system. A System Analyst is expected to have:

i. Knowledge of information technology

ii. Understanding of management process

iii. Ability of effective communication with different levels.

The Analyst is expected to make a compromise between available technology and current business needs.

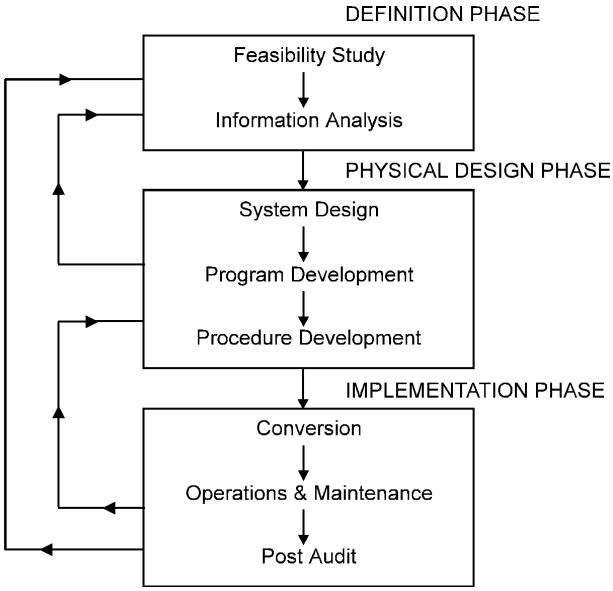
After the information has been collected, it is required to pinpoint the weaknesses of the system, if any. The information collected is quite large and it becomes essential to use systematic tools for documentation and analysis. Some of the commonly used tools are:

- System Flow Charts
- Decision Tables
- Grid Charts
- Data Flow Diagrams
- Warnier Diagrams.

11.4.2 Systems Life Cycle Approach

Systems Analysis and Design follows a life cycle for development, as shown below. Brief discussion on this approach is given here:

Figure 11.10: Systems Life Cycle Approach



Definition Phase

In this phase firstly, a feasibility study is conducted to check the viability of the project. This consists of checking of Economic feasibility, Technical feasibility and the Operational feasibility. Once a decision is taken to initiate the project with predefined objectives, then a thorough study of the existing system is made. This involves studying and documenting formal as well as informal flow of information. The documentation is done using standard systems analysis tools like System Flow Chart, Grid Chart, Decision Tables, Data Flow Diagrams, Warnier Diagrams, etc. Besides these, the Systems Analyst has to identify the information needs of various managers. This is generally done using personal interviews.

The comparison between the existing system and managerial requirements will provide deficiencies, weaknesses of the system, if any. Based on this, a conceptual design of the proposed system is generally arrived at, which is submitted to user managers and a feedback is taken.

Physical Design

The conceptual design accepted by user managers is the starting point of this phase. Here, a detailed design is created with I/O design, data base designs and processing module design. Coding in high-level language is done. At this stage, detailed control procedures, security guidelines and back up procedures are laid down in detail.

The documentation of the designed system is extremely important. The following manuals must be created:

- User Manual
- Systems Manual

After the completion of this phase, the system is ready to be implemented in practice.

Implementation

During this phase, actual changeover from the old system to the new system takes place. Either, the two systems are run in parallel for sometime and then the old system is dropped after the new one stabilizes. Alternatively, at a fixed instant the old system is dropped and the new system implemented. Generally, in Indian context, parallel conversion is considered to be the safer one.

Irrespective of the quality of the system implemented, there will always be continual changes in user requirements and also some error might crop up during processing. There is a need for on-going maintenance of the system. It has been found that over the total life cycle of a software package, more than 50% of budget is spent on maintenance.

The last item of system life cycle is Post Audit. In this module, efforts are made to compare the planned versus actuals from the view of gaining valuable insight in software development. All aspects of the system are reviewed and documented. As can be seen from the figure, there are feedback loops from one phase to another. The whole process is repeated till user satisfaction is achieved.

11.4.3 Data Processing

In any computer-based system, data storage and retrieval plays an important role. Data storage involves decision about the encoding of data, assignment of data to meaningful classes and the organization of data in the form of files. On the conceptual level, data is stored in related groupings of records which constitute the files of the system. In a data processing application, two types of files are used for structured data (There are various other approaches while working with unstructured and semi-structured data handling the big data analytics applications)

a. Master Files

b. Transaction Files

Master file contains information on a particular instant (historical information). This information has to be updated on the basis of day-to-day transactions. The updating can be done at fixed time intervals or instantaneously (on-line Realtime processing – Batch Processing).

The information generated between the updating cycle has to be stored in what is called as a transaction file. Let us consider an example of an inventory system:

Master file will contain the status of stock position as on a particular time. Each record may contain:

- Inventory Code
- Description
- Stock on hand
- Stock on order
- Price, etc.

The stock items will be issued and received regularly over the day. The issues and receipts will form the transactions file. It may contain the following data:

- Inventory Code
- Type of Transaction (Issue or Receipt)
- Quantity, etc.

As can be seen, the two files have inventory code as a common item. The master file can be updated using the transaction file and generate a new master file. In a real time system, each transaction is independently processed and the files are updated instantaneously.

11.4.4 File Organization

There are three types of file organization used in practice. Let us consider the example of a dictionary. The dictionary contains listing of English words sorted alphabetically. The top left and right hand corners provide the starting and the ending of each page. Let us assume that the information of the dictionary is stored on a disk. Now suppose, we want to find a word, say, "Computer". This can be done in the following ways:

a. **Sequential Access:** We start from the first word and check whether or not it is the desired word. We continue checking till the required word is found. This method implies searching the whole file from beginning to end and is known as sequential processing. Time required to locate a specific word will obviously depend on its location from the starting point.

b. **Indexed Sequential Access:** To locate "Computer", we can also search only the top corner words to first identify which particular page the word belongs to. After knowing the specific page, we can then search sequential till the required word is found. In this method, firstly an index has been searched followed by a complete sequential search. This method is known as indexed sequential method and is much faster compared to sequential method and is very commonly used in practice.

c. **Random Access:** Suppose each word is assigned a specific sequence number and "Computer" was 25th from the beginning. Then we could go directly to the 25th location on the disk without any searching. This type of processing is known as Random Access and the response in this method will be very fast. However, the concept of allocating a unique number, to each data item, becomes difficult as the complexity increases.

11.5 Enterprise Resource Planning

What has football got to do with it? asks management guru Michael Hammer in his latest best seller Beyond Reengineering. The answer is almost everything. Football is indeed a good analogy of the modern business environment. Companies have to play like teams of people with diverse skills for a common goal under some of the toughest competitive conditions.

For organizations to stay ahead and remain profitable, the entire company must be able to block, dribble, pass and run better than competition. In other words read markets quicker, manufacturer faster, deliver outstanding customer value at lower operating costs.

The company's challenge lies in understanding the markets and determining their goals. On this basis identify what and how every function can do towards this end, not as individually efficient units but as a coordinated team. In short, be a focused enterprise having an integrated operating structure.

Information is the key towards this. Information system couple the various activities of the firm leading to efficient operations and effective management. Easy access to timely accurate and relevant data leads to informed decision making, minimizing the response time and enabling better coordination in the firm. Computer-based information systems are the obvious solution to the vast information requirements of companies today. For professionals and entrepreneurs, a basic understanding of information system has become imperative. There is no escape from the fact that computers are here to stay and change the way companies compete. Making skillful use of the available host of information technologies today is no longer a matter of choice, it is a matter of competitive survival.

The latest tool, information technology has lent to business applications is ERP. ERP stands for enterprise resource planning. It is enterprise wide systems solution. It comprises of a single system which has organization wide applications built into it. The ERP solutions seek to streamline and integrate operational processes and information flows in the company, hence synergize the resources of an organization viz., men, money, material, machine through information.

11.5.1 Evolution

The origin of using computers for business, traces the following line of story. Originally, they were designed to support the repetitive and time consuming

functions. The applications started in the 1960s with Materials Resource Planning or MRP. It was the simple and conventional inventory management system which planned the production or procurement of components of the finished goods. These systems took the production schedules as input and gave the procurement schedules by calculating amount required, keeping lead time in view. Though MRP is successful in streamlining production, it deals with only a part of the system. If the input i.e., the production schedule or its basis (customer orders, sales forecast, inventory, bill of materials etc.) is flawed then MRP is a very efficient way of planning procurement of the wrong parts.

The next step was to automate the entire manufacturing process, this brought the next generation of computer application called Manufacturing Resource Planning or MRP II. This system sought to put the entire enterprise be it taking customer orders or ordering raw material automation, systems is that, it assumes a static view of the company and fits the system into the enterprise. It is thus a solution which automates the existing process based on existing work flows.

The approach so far, was organized around tasks. Systems solutions were developed to meet the needs of order entry group, inventory department or the marketing department, but not as an entire business process of delivering ordered goods.

These business applications for business are automation solutions which simply shift time-consuming, mundane tasks on to the powerful processing and memory of machines. Out of this evolve task-centric islands of information and efficiency. Say, order fulfillment is broken down into receipt of orders, entering them into computer, production scheduling, allocating inventory, packing, loading and delivering. A typical automation solution speeds up each task. This delivers efficiency but lacks synergy. The flaw here is that companies are trying to solve the problem by looking at parts and not the whole. Loading IT on inefficient process is only going to reinforce tasks that need not be done at all to achieve the desired results.

11.5.2 What is new?

ERP, on the other hand begins with a fresh relook at the existing business. In fact it is the result of a happy marriage between Business Process Reengineering (BPR) and Information Technology (IT). BPR is fundamental rethinking or radical redesign of business process to bring about dramatic improvement in performance. It is enabling tool through which a business can achieve it. ERP thus enables a business to reengineer by providing a single-system enterprise wide application.

It enables a company to run its business in an integrated manner with a company wide angle which leads to a better orientation towards the customer needs, the company’s goals and how each operating process supports them.

With this, companies are able to develop and manufacture products more quickly and profitably. Also, they can respond quickly to change as and when an external factor necessitates change.

11.5.3 What constitutes ERP?

ERP typically comprises a client server architecture, fourth generation language, RDBMS and graphical user interface in an open systems environment.

It operates in a client server environment forming a software layer having the application logic and business rules. A ERP package solution meets the needs of the business application through basic modules such as manufacturing, financials, sales & distribution, human resources. These modules are developed to cater to the wide functionality of every process and are benchmarked against the best business practices worldwide. These modules can be customized to meet the specific requirements of individual businesses.

ERP streamlines the organization from the back-office to the board room. It does so, by first giving the existing business a hard look. The working of the organization is split into a logical process not tasks. Often many practices arise which are unnecessary, complex or simply exist by habit or are developed without keeping the whole picture in mind. The core processes are identified. Once the wholistic business analysis is complete, integration with the modules begins. Each business process is interwoven into the ERP package, thus ironing out inefficiency. In short, the entire functional knowledge and working of the firm is simulated on the system keeping in view the right business practices.

The ERP packages enable operational efficiency, implement continuous change and lead an organization through management by fact. It achieves this by operating at all management levels. Since the basic business processes have been specified, the package system has supporting procedures and system developed. At the lowest level or the actual operational level, all transactions are captured through office automation and transaction processing systems. The data is captured only once, at the point of origin and routed appropriately through the work flow.

At the next level, the interface level, it enables seamless integration of information across operating groups by maintaining a common data warehouse from all activities be it, marketing, sales, finance, HR, manufacturing. Changes in any individual activity have a ripple effect throughout related areas through this interfacing.

At the top management levels is the tactical and strategic levels ERP can be used as a source of competitive advantage, as it leads to transparent departments where relevant information is available instantaneously. Informed decision-making may well serve to take correct decisions at the right time.

11.5.4 The ERP Market

ERP on the above mentioned accounts is rising rapidly to the top of corporate IT agenda. Worldwide the ERP market is estimated to be about $39.34 billion in 2019. It is expected to grow to $86.30 billion by 2027, growing at a CAGR of 9.8% from 2020 to 2027.[1]

In India, the ERP market is estimated around $6.69 billion for 2020 and is expected to grow to $7.42 billion in 2021. Most of the ERP implementations in India are in the MNCs. This is primarily because of their experience or familiarity with ERP benefits in their global operations. With global competition knocking at the door, most professionally managed firms are expected to go the ERP way.

Already several of the industry majors like ABB, HLL, Maruti Udyog Ltd., Ispat Group, ONGC, Samsung have made a beeline for ERP products. The Indian market is hotting up and the solution providers are stepping up their Indian operations.

First tier companies (those with a turnover greater than Rs.10 billion) implement ERP to increase internal efficiency and external competitiveness. Once ERP is established at this level, these large companies begin to desire similarly increased efficiency from their suppliers. Hence, second tier companies are pressured to implement ERP, and a trickle-down effect ensues. Indian companies quickly understood that in order to work at maximum efficiency, ERP must be implemented at all levels.

The solution providers are both MNCs and homegrown. Ramco and Mastek are in the latter category. The major players and their products are given below:

Of the players in the Indian market, SAP leads the market. SAP AG is a German Company. It has blazed a trail in developing applications and is rightly considered the ERP missionary. In India it has made deep inroads having a customer base of 40 companies, having tied up with TATA group to provide solutions for all TATA companies. Oracle and Baan are competing for the second place. Oracle India has been appointing partners and riding piggy back on its success in the RDBMS market. It has a customer base comprising Sony, MUL, AP Rayon, Kellogs India, etc.

BAAN IC is developed by Netherlands based BAAN Company. It has backed customers like TISL, Clayton, Rallies, Godrej. It has set up its own technology park in Hyderabad as a hub for Software development. It has also tied up with Informix bounding Informix data base with Baan. Another major player is home grown RAMCO systems. The latest version MARSHAL 3.0 was launched by Bill Gates during his recent visit to India. It has sold 60 copies over 100

[1] https://www.alliedmarketresearch.com/ERP-market

locations of which 38 are in India. Its clients include HDFC Bank, Standard Chartered, Amtrex Appliances and Dr. Reddy Laboratories.

An implementation of a particular package may take place because the parent company is already using it. In this case implementation is rendered easy as the learning curve has already been traversed. However, most companies are implementing ERP either because the competition has moved towards it or are the first ones going in for it seeing the potential benefits to their line of business.

11.5.5 The Right Choice

The selection criteria of a company shopping for an ERP package differ on various accounts. ERP packages involve not only a technological improvement but demand business redesign as well. To add to this, costs are high as the hardware, software, networking, implementation and training costs amount to high sums. Also, reversing the implementation once it is underway can be disastrous both fiscally and functionally. Thus the selection of the right ERP package is of vital importance to a company.

The process of selection is often constrained by factors like finding concrete data about the packages. Usually, companies have to rely on the solution provider's marketing for information about the package. Another stumbling block is the lack of a proper framework or procedure for selection. Companies end up focusing on a narrow and focussed set of "what can the package do for me?" Hence, the selection process is prone to internal politics, package hype and intuitive feelings.

An objective and in-depth selection methodology would involve time and cost but the effort may be well worth it. The chief criteria would be to examine the required functionality. A key decision to be made here is whether to buy an integrated package or to buy a modular one. While integrated packages allow a consolidated system with high data integrity across modules, they also need an enterprise-wide effort. This involves more time, complex configuration. In view of the trade-off companies may prefer loosely coupled system for its functions, and implement modules which provide the requisite integration.

The other parameters for selection are the technological architecture which examines the hardware, the software, the development tools of the package, the cost involved and service provided by the vendor. The company must also keep in mind the future moves of the vendor as release upgradation and orientation towards future technologies will serve to change the infrastructure of the firm accordingly.

The business giants in India, taking a cue from the MNCs, have realized that ERP solution could be a sine qua non in tackling the problems unleashed in the aftermath of globalization. However, the bottom line here is, it is not what you have got which matters, it is what you do with it.

A Corporate Wish List

Functional requirements

- Support all their business process (internal processes as well as three way relation between the customer, the company and the vendor)
- Online integration of all business process and data
- Provide easy.
- Support rapid modifications and development of the system
- Support integration of new existing application systems
- Lastly and most importantly it should be easy to use
- Low cost
- The system should be portable and scalable
- Open system independent of hardware environment
- Short response time
- High throughput of the background processing
- Ease of system administration
- Short down time during release change

List of Companies & their ERP Products

Company	Product
SAP	R/3
BAAN	BAAN
ORACLE	ORACLE Manufacturing, Financials HR
Scala India	Scala ERP package
SSA India	BPCS C/S
QAD	MFG/PRO
RAMCO	MARSHAL
MASTEK	MAMIS
PEOPLESOFT Inc.	Peoplesoft 6

11.6 Internet

Internet: The global network of computers linked by exclusive and regular phone lines and microwave and satellite signals.

Extranet: An extranet is the part of a corporate internet that allows companies to communicate with the internets of their customers and suppliers facilitating electronic transactions.

World Wide Web (WWW): A big part of the network of networks – the internet – is called the WWW. It is an interlinked collection of the hypermedia system residing on web servers, that lets you browse through lots of information.

Browse: The client programs designed to work with Web servers are called the browsers. A server on the Web contains many sites, which in turn are composed of Pages. Web Sites offer text, pictures, graphics, sound and movies. The Web servers use the internet expressways to deliver data through a process called the packet switching. Each file is broken into small packages, tagged with its origin and destination and sent through routers, the computers which function like mail handlers, sending each package it receives through various routes. The client program on the other end reconstitutes the pieces.

Hypermedia and Multimedia: A hypermedia has a broader domain and apart from the text it can also contain pictures, sound and video. A multimedia is a combination of these features. You can see an animated film with sound or any other such combination.

Archie, Jughead, Gopher, and Veronica: Archie is a system that helps find files located anywhere on the internet. Jughead is a program that helps you find information in the gopher, a system where you find information using menus. And Veronica is friend of Archie and she helps you find things in the gopherspace.

Java: The language can run on any computer architecture and operating system and can be downloaded straight from internet. On the internet the programs written on Java are called Applets.

Protocol: A protocol is a set of rules governing the formatting of data transmitted between computers and terminals. A protocol is actually how computers will talk to each other. Protocol definitions range from how bits are placed on a wire to the format of an electronic mail message. Standard protocols allow computers from different manufacturers with different operating systems to communicate.

Bandwidth: It is the capacity of a cable or a phone line, measured in bits per second.

Bits Per Second: It is the speed at which bits are transmitted over a communication medium.

Baud: The rate at which the medium (usually a modem) can transfer groups of data. Multiple bits are transferred in each group, so the rate of data transfer may exceed the actual data rate.

FTP: It is File Transfer Protocol; or any application moving files using the File Transfer Protocol.

HTML: It is Hypertext Markup Language; the language in which World Wide Web documents are written.

HTTP: It is a protocol used on the web to transfer hypertext documents. A protocol is a set of rules that the computer use to communicate logically.

TCP/IP: Transmission Control Protocol/Internet Protocol, which allows all computers to speak the language of the internet and other networks.

Packet: It is a bundle of data. Packet sizes can vary from roughly 40 to 32,000 bytes, depending on network hardware and media.

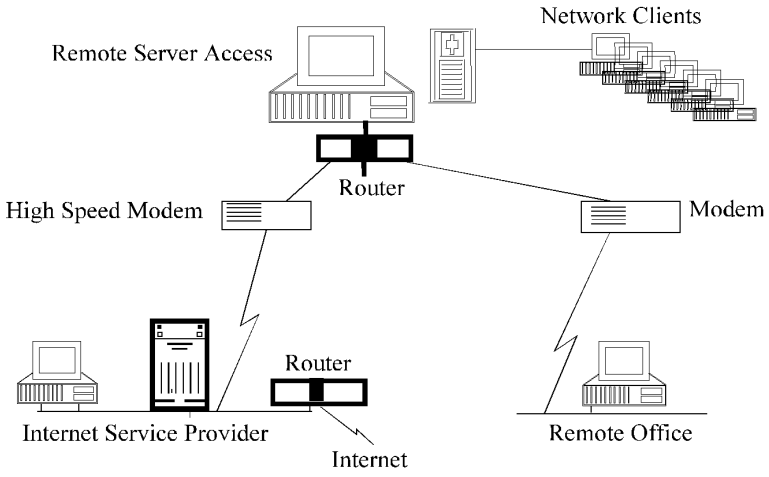
Hypertext: Hypertext is a system of Documents that contain links to other documents; selecting a link automatically calls up the other document.

ISDN: Integrated Services Digital Network; a digital telephone service. If your local phone company supports it, if you have the appropriate hardware and software, and if your local central office provides ISDN service (lots of ifs), ISDN allows high-speed home access to the internet (56 kilobytes per second).

Modem: A piece of equipment that connects a computer to a data transmission line. Today's fastest modems transmit data can down load data upto 1000 Megabits per second.

Electronic Data Interchange (EDI): It is the transfer of electronic messages from one company to another using a network. Companies use EDI to facilitate business-to-business transactions like purchasing orders, purchase conformation, invoicing and payments. These messages can be exchanged using VAN (vast area network) or the internet.

Figure 11.11: A corporate LAN



Uniform Resource Locator (URL): It is a web addressing scheme that spells out the exact location on an internal resource. When the user clicks on a link, the browser roads the link to the document. The information about the link is provided to the browser by the URL.

Address: Like any postal address, to reach a person via internet/e-mail, there is a unique mixture of letters, numerals, and punctuation marks that identify a mail box on a computer network. For example, take our address

AW@ICFAI.WIPROBT.EMS.VSNL.NET.IN

Now let us try to read it. The @ (at) symbol separates the person's mail box identifier (AW, to the left) from the name of the organization (ICFAI, to the right). In this case, WIPROBT a joint venture between Wipro and British Telecom has leased lines from VSNL, the internet service provider. EMS denotes that it is an electronic mail service and the term NET stands for network organization. The last term IN denotes India, which is the country code. Note that there are no spaces at all between any letters or characters and a full stop separates at times. This address will reach us from anywhere in the world.

Hacker: A computer expert who enjoys exploring and developing computer system and tinkers for fun.

Infobahn: Synonym for information superhighway.

11.6.1 Getting Ready for Internauting

The Sun Microsystems' motto says it all: "Network is the computer." There is no exaggeration in the statement. Networks comprise the backbone of any organization before it is hooked to the internet. But what are they really? Each group of the thousands of sites that go to make up the internet are networks by themselves. They may be small Local Area Networks (LANs) connecting computers within an office premises or Wide Area Networks (WANs) – independent networks belonging to the same enterprise connected through telephone/leased lines or Intranets.

Intranets, the latest corporate networks, interact with each other through 'browsers' on the internet. By applying the Web's technology for sharing information in their own corporate networks, companies are creating internal webs. The internet technology on which these webs are based makes it possible for any computer be it PC, Mac or workstation to get connected.

Because of the diversity of the networks connected to the internet, in terms of geographic location, operating systems, computer platforms and hardware, certain standards are required to be maintained by each of the networks that is connected to the internet, in order to ensure communication. Any organization which plans to get hooked on to intranet needs to plan its Local Area Network keeping this in perspective.

The internet uses a standard transmission protocol known as Transmission Control Protocol/Internet Protocol (TCP/IP). The networks that comprise the internet are connected to each other by something known as routers. Routers are devices that connect different types of LAN by converting formats and interpreting protocols. Routers also have a computing power for network management, fault identification and isolation and traffic flow management to avoid congestion.

The Internet Protocol (IP) makes sure that the routers know where to send the data by addressing it in small data envelopes. These envelopes are tiny, and can easily get damaged or lost in transit. The TCP breaks them down and effectively places them inside a secure packet. What it all means, is with the TCP/IP, you will be able to connect to the internet and participate fully in its applications no matter what machine you use.

The basic hardware required to connect all users on a corporate network are:

a. A LAN

b. An internet connection

c. A remote access server which acts as a gateway to the internet.

d. Telephone lines, modems.

e. Software for each user called client software which would allow him to:

- Connect over dial up links (TCP/IP)
- Configure and perform dialing through modems
- Have on-line access to the internet (Mosaic, Gopher, etc.)

Choosing a suitable networking technology is important in building an enterprise network. Among the ones available are Ethernet, Fiber Distributed Data Interface (FDDI), Token Ring and Asynchronous Transfer Mode (ATM).

ATM is a switched, connection-oriented local and wide-area networking technology. It offers a number of capabilities like integrating LANs and WANs, high performance, scalability to allow for growth, preservation of existing networking investments, support for multimedia applications and integration into WANs, etc. It also allows a theoretically unlimited number of users to have dedicated high-speed connections.

What can a corporate network connected to the internet offer?

- All networked users on the LAN can have on-line access to the internet.
- Remote office networks can be connected to the corporate LAN and thereon to the internet.
- Mobile users can access internet through the corporate LAN.

11.6.2 The Widening Web

The idea, the whole world is agog now had accidental origins. Here is how it all began. In early 1969, the U.S. Defence department financed a network for doing civilian research called Arpanet. Experts figured out that it is safe to avoid a concentration of network between researchers for apparent reason: What happens if a nuclear attack wipes out the entire research data? The apprehension led to the move of linking different networks of computers so that libraries of information remain at diverse locations. Eventually, the idea was enlarged to

encourage instant exchange of information and sharing of data between different institutions – research, defense and educational.

The primitive form of internet was earmarked by electronic mail (e-mail). It allows people to type messages from their PCs and have them electronically transmitted to recipients who can read, reply, delete, print, forward, or file them. The sender of e-mail can keep a copy in a computer file.

Communication became possible between these networks on a point-to-point basis. The idea gained so much currency that Arpanet became too big to sustain a network of military networks, educational networks, and scientific networks. The result: Milnet, NSNFNet and other 'sites' came up catering to the burgeoning information requirements of the elites (government) and intelligentsia (academia). The internet as we know it today had such obscure origins which received a fillip only in 1990.

The first international internet connection established in the U.K. and Norway in 1973 soon spread its web in France, Holland, Russia, Germany and most European countries. Even South-East Asian countries like Japan, Indonesia, Thailand, Taiwan, Singapore, South Korea are evincing keen interest to join the new communication craze. But it sure should surprise us that China and most of Africa are on the web.

As more and more countries join the bandwagon of infobahn, the internet world will become divided into the I-net haves and have-nots. And the sooner the countries latch on to this new communication engine, the smoother will be the road ahead to prosperity. Despite the widening web of the internet, about half of all registered internet "hosts" are located in North America (Canada included). That just shows that the protagonist, the U.S. still holds the leading edge in the information superhighway.

True, the nitty-gritties of internet still baffle many countries let alone convincing them. Take the Chinese for instance. The entire world knows that the Chinese always were paranoid about foreigners and tight-lipped about giving away information. Since information is the life-blood of an internet society, any draconian measure of the dragon-economy to restrict, censor or infiltrate the conduits of info-exchange dampens the purpose behind internet's rise.

Then come other concerns – pornography and terrorist activities. Reams of paper have already been written over the ability of internet to generate germs of a creepy kind. It is now a fact that less than one percent of the web sites being visited are anything about sex and violence. While the case of pornographic proliferation has been remarkably tackled by passing strict enforcements and laws against the offenders and creating passwords which prevent children and impressionable youth from surfing by the wayside, a lot is being debated about the way governments should tackle the spread of terrorist activities. The

downside of any move to curb terrorism is that internet may be prone to intervention, regulation and censorship by the government – anathemas to the laissez-faire of internet.

Then come the security concerns of the internet's new club of users – businessmen. Internets often find themselves subject to interception by hackers and cyberthieves who steal the credit card numbers and valued personal data for profit. Of course there are solutions, encryption or the art of camouflaging packets of data and information-in-transit.

Little wonder, the idea of internet like any great idea in the history of mankind has been so innocuous but the proliferation and germination of its derivative users has spawned a lot of doubts, delusions and distortions about its malefic nature. But then, if a phenomenon does not have a flip side, as it is often said, it ceases to be a phenomenon. At least, Bill Gates is convinced about the omnipotence of internet: "The surging popularity of the internet is the most important single development in the computer industry since the IBM PC was introduced in 1981... Like the PC, the internet is a tidal wave. It will wash over the computer industry and many others, drowning those who do not learn to swim in its waves."

And to take the point further, internet has already been recognized as a great equalizer between the world's richest and the poorest countries. If you are not on net, you are out of the race. Says Gates: "It is going to change societies and the relations between societies. Developing nations including China and India stand to gain a great deal, because they will be able to draw on the world's talent and knowledge while making the intellectual resources of their own people available to markets around the globe."

The global statistics are heartwarming: the electronic billboards ("web sites") are doubling every 53 days, 50 million PC users are linked by internet and around 1.5 million PCs are being planted every month. With such whopping changes transforming the way everything is being done, will India up the ante on internet?

Check Your Progress - 2

6. Ms Sahiti, the systems analyst had collected necessary information about the new application software to be built and needed to document these using systems analysis tools. Which one of the following is not a systems analysis tool?

    a. Data flow diagrams
    b. Warnier diagrams
    c. Grid charts
    d. Decision analysis

7. Ms Ananya, the systems analyst was following the life cycle for development of required artefacts and was at physical design phase. What are the activities during this phase?
   a. Information analysis, system design, procedure development
   b. System design, program development, conversion
   c. Feasibility study, System design, program development
   d. System design, program development, procedure development
8. Mr Basu, the documentation and purchase analyst, has to facilitate business-to-business transactions like purchasing orders, purchase conformation, invoicing and payments among their company, vendors, and customers. What is this process referred as?
   a. Electronic download information
   b. Electronic data interchange
   c. Electronic data exchange
   d. Electronic data information
9. Mr Kushal, the document specialist at Finservices group had to handle various types of data files, and need to retrieve them based on need for communicating with various stake holders. There are three types of file organization on the computer. What are they?
   a. Direct access, random access and indexed sequential access
   b. Sequential access, random access and direct indexed access
   c. Sequential access, direct access and indexed sequential access
   d. Sequential access, random access and indexed sequential access
10. Buildstrong cements had recently purchased the Enterprise Resource Planning (ERP) solutions for all of their activities, and ensured streamline and integration of all of their operational processes and information flows in the company to synergize their resources. Which are all the resources ERP synergizes?
    a. Men, money, material, marketing through information
    b. Men, marketing , manufacturing , machine through information
    c. Men, money, material, machine through information
    d. Men, money, material, manufacturing through information
11. The project manager was updating the management about the software application developed for their customer and its status. He informed them that the project is in operations and maintenance. Which phase of the system life cycle was the project at that moment?
    a. Definition phase
    b. Physical design phase
    c. Implementation phase
    d. Maintenance phase

11.7 Summary

- The following steps are required for solving a problem using a computer system: i. Problem Definition; ii. Program Planning (Preparation of a flow chart); iii. Writing the Program; iv. Debugging; v. Documentation
- Systems analysis is the process of evaluating and understanding a particular system and its environment. In a way, it will involve studying the inputs to a system, outputs required and the associated processing to convert inputs to desired outputs. It is the most difficult part in designing a system.
- ERP stands for enterprise resource planning. It is enterprise wide systems solution. It comprises of a single system which has organization wide applications built into it. The ERP solutions seek to streamline and integrate operational processes and information flows in the company, hence synergize the resources of an organization viz., men, money, material, machine through information.

11.8 Glossary

Electronic Data Interchange (EDI): It is the transfer of electronic messages from one company to another using a network.

Flow Chart: A flow chart is a graphic method for indicating a proposed or actual solution to the problem.

Internet: The global network of computers linked by exclusive and regular phone lines and microwave and satellite signals.

Systems Analysis: It is the process of evaluating and understanding a particular system and its environment.

Debugging: The process of checking and correcting errors is known as "Debugging".

11.9 Suggested Readings/Reference Material

1. Gupta, S. P. Statistical Methods. 46th Revised ed. New Delhi: Sultan Chand & Sons. 2021.
2. I. Levin Richard, H. Siddiqui Masood, S. Rubin David, Rastogi Sanjay. Statistics for Management. Pearson Education; Eighth edition, 2017.
3. Gerald Keller. Statistics for Management and Economics. Cengage, 2017.
4. Arora, P. N., and Arora, S. CA Foundation Course Statistics. 6th ed. S Chand Publishing, 2018.
5. Mario F Triola. Elementary Statistics. 13th ed., 2018.
6. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran. Statistics for Business and Economics. 13th Edition, Cengage Learning India Pvt. Ltd., 2019.

7. S D Sharma. Operations Research. Kedar Nath Ram Nath, 2018.
8. Hamdy A. Taha. Operations Research: An Introduction. 10th ed., Pearson, 2016.
9. Malhotra, N. (2012), Marketing Research: An Applied Orientation, 7th ed., Pearson, 2019.
10. Cooper, D.R. and Schindler, P.S. and J. K. Sharma (2018), Business Research Methods, 12th edition, McGraw-Hill Education.

11.10 Self-Assessment Questions

1. Define Flow Chart. Discuss the importance of it.
2. How ERP is beneficial to an organization? Justify.
3. What is systems analysis design? Explain.
4. Outline the importance of internet in the growth of an organization.

11.11 Answers to Check Your Progress Questions

1. (d) A: basic≤ 30000, B: hra=basic*0.3, C:hra=0.45*basic

2. (b) Debugging

3. (d) Documentation

4. (a) Any Size.

The symbols in flow charts may be of any size, only the shape is standard.

5. (b) Program Planning

Preparation of a flow chart is done in program planning step.

6. (d) decision analysis

7. (d) System design, program development, procedure development

8. (b) Electronic data interchange

9. (d) Sequential access, random access and indexed sequential access

10. (c) Men, money, material, machine through information

11. (c) Implementation phase

Unit 12

Statistical Software Tools

Structure

12.1 Introduction

In the previous unit we gained knowledge on the role of computers and IT in a business enterprise. In this unit, we get an overview of software to do data analysis. This unit describes about SPSS and SAS in detail and emphasizes on terminologies and benefits.

Statistical software is specialized computer programs for analysis in statistics and econometrics, finance, retail etc. Statistical package for Social Sciences (SPSS) is an integrated family of products that help to address the entire analytical process, also helps in generating reporting and deployment. The features of SPSS Statistics are accessible via pull-down menus or can be programmed with a command syntax language. The pull-down menu interface helps in displaying the output.

Statistical Analysis Systems (SAS) is a software suite that can mine, alter, manage and retrieve data from a variety of sources and perform statistical analysis on it. SAS provides a graphical point-and-click user interface for non-technical users and more advanced options through the SAS language. In order to use Statistical Analysis System, Data should be in a spreadsheet table format or SAS format. As programs have a DATA step, which retrieves and manipulates data, usually creating a SAS data set, and a procedure step, which analyzes the data.

12.2 Objectives

After going through the unit, you should be able to:

- Explain the use of computers in Finance;
- Identify basics of SPSS;
- Explain data analysis using SPSS; and
- Demonstrate SAS in data analysis.

12.3 Use of Computers in Finance

The discovery of computers has transformed the financial industry and how business deals are transacted. Computers influence both business and personal financial management. Computers enhance financial communication within members of an organization through a network system connecting various departments. With the use of computers, key financial decision makers are able to send financial reports and strategies instead of holding meetings. This system allows finance people to get updates on world stock exchange, and price changes affecting their business.

Data Storage:

Computerization makes data storage manageable and less bulky. Computers enable organizations to store large amounts of files in a small space, allowing us to have a large track of historical transactions, while avoiding consumption of space that would otherwise be consumed by piles of files in cabinets. Keeping financial records is critical for organizations, as well as individuals, as it allows tracking of payment records, debts owed, purchases done and bank transactions made. Having detailed financial records helps an organization to continually audit and analyze business performance.

Financial Information:

Computers have enhanced access to wide financial information through Internet access. Business operators or individuals can now access information on investment prospects, and conduct a detailed research on its profitability. There is also wide financial information available on the Internet, such as world prices, tax changes, inflation rate and currency exchange rates making it easier for institutions and individuals to make informed financial decisions. The Internet provides a wide source of information and access to specific financial reports of companies in trade

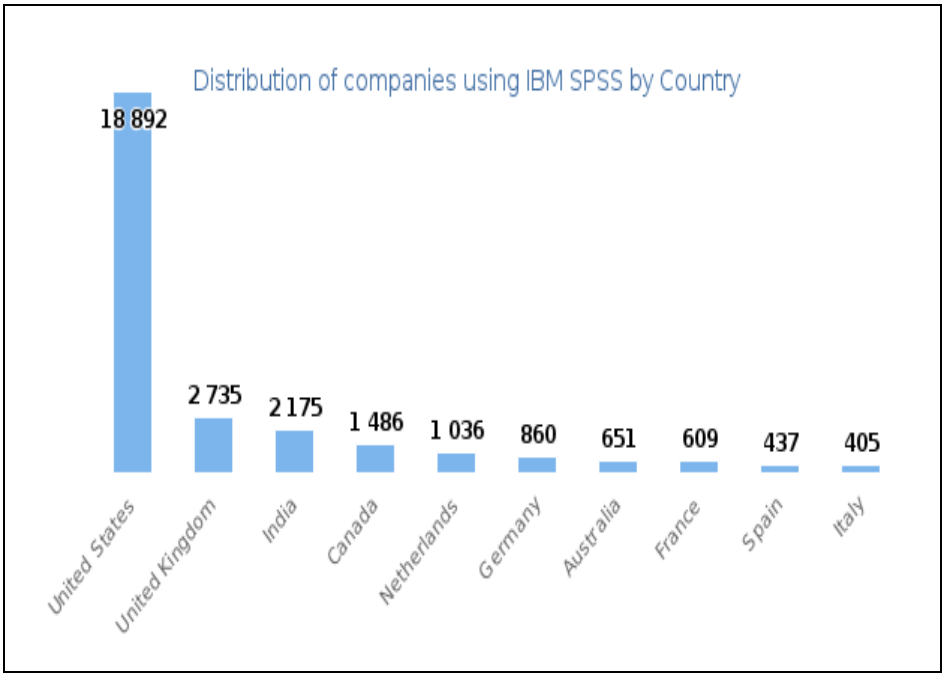
12.4 Introduction to SPSS

SPSS is the acronym of Statistical Package for the Social Sciences. This package can perform highly complex data analysis using statistical and mathematical functions by means of simple instructions. SPSS for Windows

provides a powerful statistical analysis and data management system in a graphical environment, using descriptive menus and simple dialog boxes. Statistical analyses carried out through SPSS can range from simple descriptive statistics like means and deviations to advanced analyses like regression and factor analysis.

Exhibit 12.1 gives few statistics about SPSS Usage. 51% of IBM SPSS customers are in United States, 7% are in United Kingdom and 6% are in India.

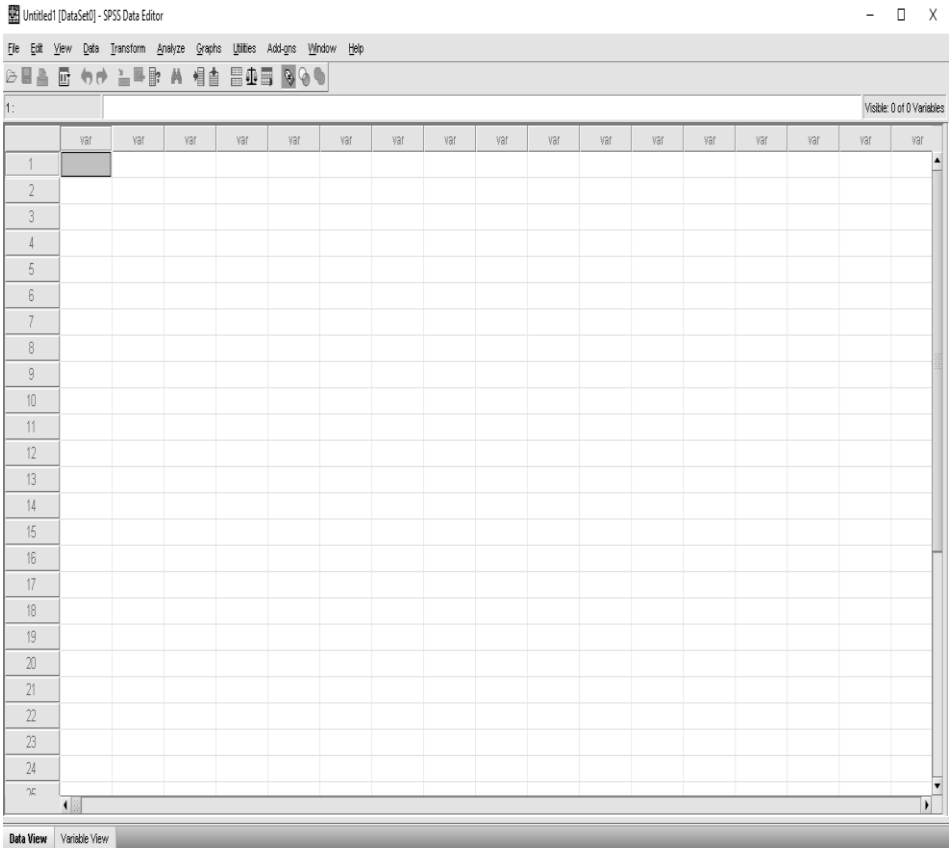
Exhibit 12.1: SPSS Usage Statistics



Adapted from https://enlyft.com/tech/products/ibm-spss

SPSS is widely used for researchers in social science and statistics as well as business research. The package enables summarization and display of data with tables and graphs.The package comprises several modules with the SPSS Base System being the basic module. The package supports Macintosh, Windows as well as UNIX operating systems. To begin with the SPSS application we have to select the following Menu options from the Desktop: Start > All Programs > SPSS for Windows > SPSS 13.0 for Windows

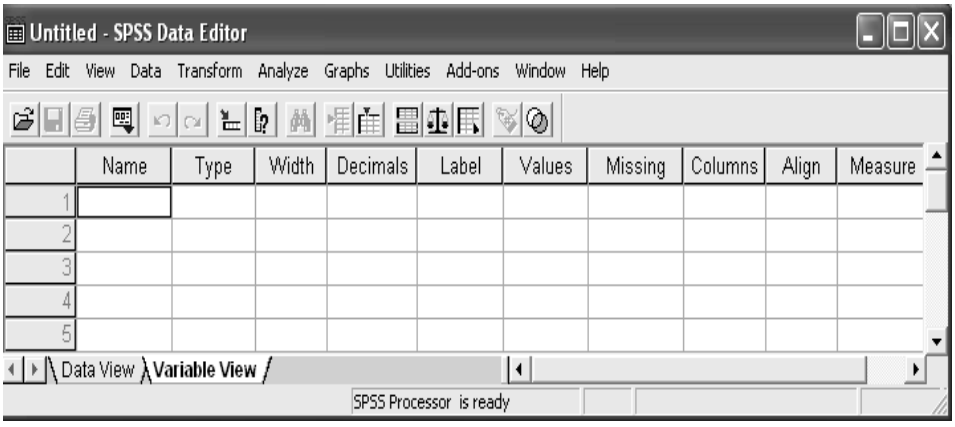
SPSS will start up with an open Data Editor window and a dialog box with the query "What would you like to do?" The choices are to run a tutorial, type in new data, run an existing database query, create a new query, or open an existing dataset. Click on the radio button for “Type in data” and click on OK. When you begin with SPSS for Windows, you will get a screen that is similar to an excel sheet.



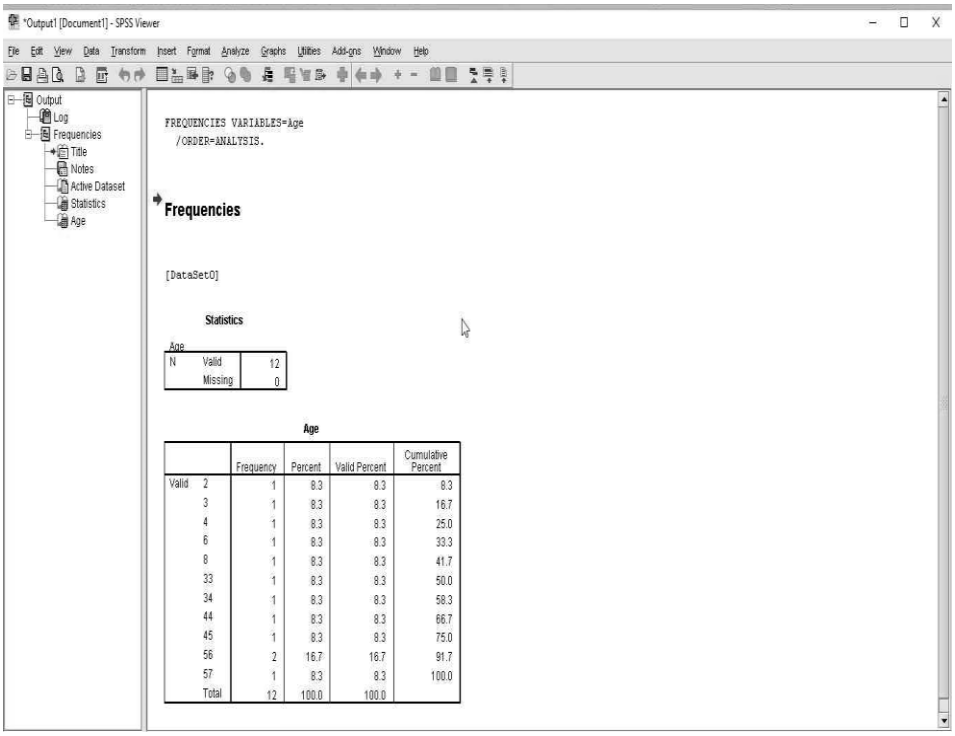
There are some basic windows in SPSS that act as the user interface namely the data editor view comprising the data view and the data variable view, the output view, the draft output view, and the script view and the syntax editor view.

The Data View: The data view displays the actual data that you have entered and any new variables created.

The Variable View: The variable view window contains the definitions of each variable in the data set, including its name, type, label, size, alignment etc. To go to the variable view, click the variable view tab at the bottom of the data window.

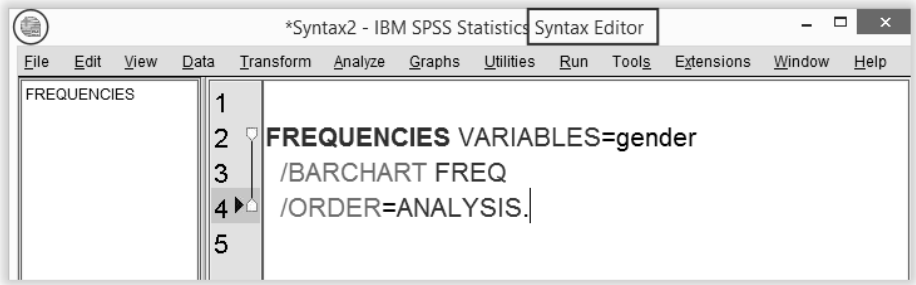


The Output View: The output window shows the results of queries like say frequency distributions, cross-tabs, statistical tests, and charts.



The Draft View: The draft view gives a view of the output as it is generated for printing.

The Syntax View: Syntax refers to the actual computer code that produces a specific output. Through this the exact steps or code used for a particular analysis can be saved and also refined to generate specific output.



SPSS Menus
File: To open an existing file, to read data (from a Text file/ Excel file) into data editor window, to save the data file, and to exit SPSS for Windows.
Edit: To perform functions like copy, cut or paste to the data editor window.
Data: To define variables, insert variables or cases, sort cases, merge files, split files, select cases and use a variable to weight cases.

Transform: To do some computations on variables, to create new variables from existing ones or recode old variables.
Analyze: To do statistical analysis, from descriptive statistics to testing of hypothesis.
Graphs: To obtain high resolution plots and graphs that can be edited in chart editor window.
Utilities: To run script and to display information on the contents of SPSS data files.
Window: To move to any open window or to see which window is active. The window with a check mark is the active one.
Help: To get help on topics in SPSS and to ask the statistics coach some questions.

It is important to be familiar with some basic terms that are used with reference to the SPSS software before we proceed with learning the basic functions:

Data Set: A specific set of data that is in a file format which resembles a spread sheet with rows and columns. The rows contain the information for each person or case; this is called a record. The columns contain the specific information for each person like age, gender, income, etc.

Variable Name: This is the specific name that you assign to a variable. For example, in the typical data set, age, gender, income, etc. would be some common variable names. These appear in the column of that variable in the data view window. It is restricted to eight characters and should begin with a letter (no special characters).

Variable Labels: Variable labels describe the variables closer to the convention of the English language, so that the output is understood by the reader as it need not be confined to eight characters. It also allows spaces to be used. It is generated from the variable view.

Values: Each variable comprises a value. If the variable is gender, then the values are male and female. In a data base, however, values are typically coded to take up less space. So gender might be coded as M or F, or as 0 or 1.

Value Labels: These labels are used to identify a particular code if data has been codified for use. For example, if we have represented gender with 0 or 1.The output will show the value labels of Male and Female rather than 0 or 1. It is generated from the variable view.

- SPSS allows a user to input data in its file on which the analysis has to be performed. It also allows data to be transferred from other files like ASCII files or an Excel sheet.

12.5 Data Analysis Using SPSS

SPSS enables comprehensive data analysis using various statistical methods. The section below describes the procedure to be adopted to analyze data using SPSS and also understand the output generated through the analysis.

Descriptive Statistics

To calculate Frequencies the following procedure is followed:

Step 1: From the menus choose: Analyze>Descriptive Statistics>Frequencies

Step 2: In the Frequencies dialog box, names of all the variables in the data set appear on the left side of the dialog box

Step 3: Choose the variable from the list. Click the arrow button. The selected variable appears in a box on the right. Click on OK. The output will appear on the output view.

The output shows the absolute frequency of observations, relative frequency as a percentage of all observations, the valid percent i.e. the adjusted frequency as a percentage of the number of respondents who provide a record answer rather than leaving it blank or providing a meaningless answer. In this case it is 100%. The table also shows the cumulative percentage.

To calculate the Descriptive for the four continuous variables in the data set:

Step1: From the menus choose: Analyze >Descriptive Statistics >Descriptives. A dialog box will appear. Names of all the numeric variables in the data set appear on the left side of the dialog box.

Step 2: Click the variable and click the arrow button to the right of the selected variable

Step 3: Do the same thing for the other variables. The selected variables will appear in the box on the right. Click OK. The output will be displayed.

The mean, standard deviation, minimum, and maximum are displayed by default in the order in which you selected them.

Check Your Progress - 1

1. In SPSS, what is ‘Data View’?
    a. A table summarizing the frequencies of data for one variable
    b. A spreadsheet into which data can be entered
    c. A dialog box that allows you to choose a statistical test
    d. A screen in which variables can be defined and labeled
    e. A table where output can be saved.

2. In SPSS menus what is the role of Analyze?
    a. To run script for displaying information on the contents of SPSS.
    b. To get actual computer code which produces specific output.
    c. To do statistical analysis from descriptive statistics
    d. To find the definition of each variable in the data set.
    e. To describe the variables closer to defined fuction.
3. What is the role of value labels in SPSS?
    a. To make commands more understandable
    b. To identify a particular code if data has been codified for use
    c. To perform functions like copy, paste etc.
    d. To describe the variables closer to defined fuction.
    e. To input data into file.
4. The SPSS windows provides
    a. Powerful Statistical Analysis
    b. Data management system in a graphical environment
    c. Simple descriptive statistics like means and deviations to advanced analysis like regression and factor analysis
    d. Simple descriptive statistics like means and deviations to advanced analysis like regression and factor analysis, Data management system in a graphical environment
    e. Powerful Statistical Analysis, Simple descriptive statistics like means and deviations to advanced analysis like regression and factor analysis, Data management system in a graphical environment
5. Role of Utilities in SPSS menus:
    a. To define bigdata
    b. To run script and to display information on the contents on SPSS data files
    c. To describe statistics
    d. To describe financial data
    e. To remove patches

12.6 Introduction to SAS

What is SAS System?

The SAS system is an integrated system of software products. It provides tools to make the data useful and meaningful. Initially used to be known as Statistical Analysis Systems. Since then they have moved to diverse fields. Now, only known as SAS. Developed in the early 1970s at North Carolina State University.

Application of SAS:

Data entry, data retrieval, data management, statistical and mathematical applications, Operation Research, Project management, business planning and forecasting, customizing applications, Quality Improvement, data mining and Data warehousing for large scale business solutions, financial management and decision supporting, HRM, CRM, ERP etc. Exhibit 12.2 shows how SAS is used in industries.

Exhibit 12.2: SAS in Industries

Vayam Technologies, an IT solutions company in India, Supports government in making 'Digital India' a success. Used SAS platform to computerize the legacy data-base along with reporting services and analytical intelligence for analytics, alert generation, dash-boards and reporting.

Prominent manufacturing companies such as Ulbrich Stainless Steel & Lockheed Martin are using SAS analytics to drive Internet of Things (IoT) innovation in their facilities in order to be more productive and efficient.

Adapted from https://www.iot-now.com/2019/05/02/95365-lockheed-martin-ulbrich-stainless-steel-use-sas-artificial-intelligence-conquer-iot-challenges/ and https://www.business-standard.com/article/news-ani/vayam-technologies-to-support-government-in-making-digital-india-a-success-116101700607_1.html

SAS benefits in Finance:

- Reliable.
- Handel large datasets.
- Powerful procedures.
- Communicate and execute strategies more effectively.
- Create more frequent, accurate forecasts.
- Publish financial reports more quickly.

The SAS Language

Many software applications are either menu driven, or command driven (enter a command—see the result). SAS is neither. With SAS, you use statements to write a series of instructions called a SAS program. The program communicates what you want to do and is written using the SAS language.

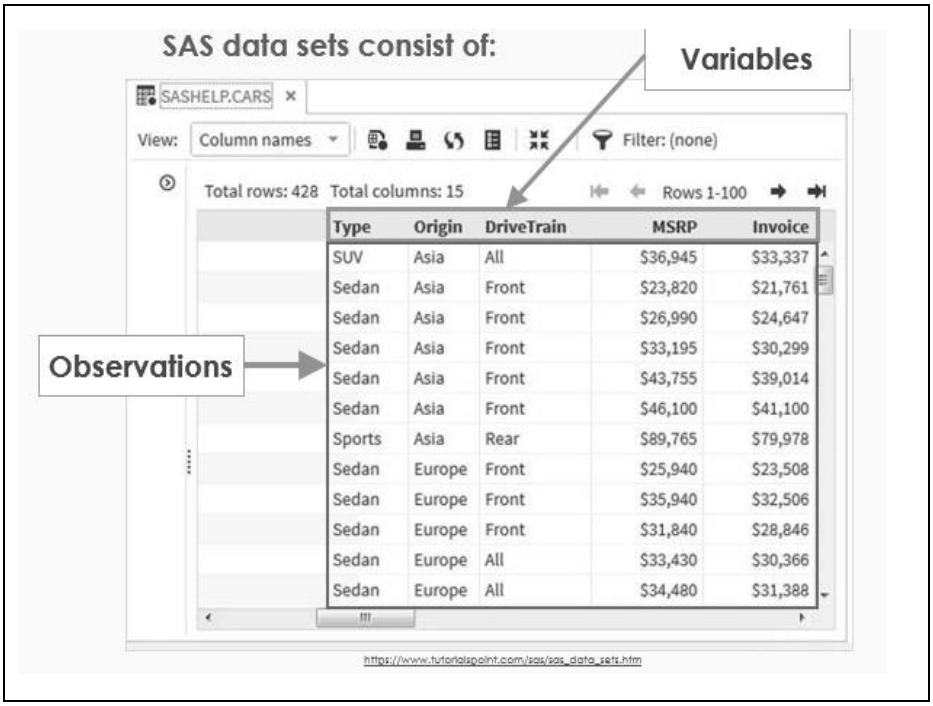
SAS statements: As with any language, there are a few rules to follow when writing SAS programs. Fortunately for us, the rules for writing SAS programs are much fewer and simpler than those for English.

The most important rule is every SAS statement ends with a semicolon.

Comments: To make your programs more understandable, you can insert comments into your programs. It doesn't matter what you put in your comments—SAS doesn't look at it.

There are two styles of comments you can use: one starts with an asterisk (*) and ends with a semicolon (;). The other style starts with a slash asterisk (/*) and ends with an asterisk slash (*/).

Variables and observations: Data, of course, are the primary constituent of any data set. In traditional SAS terminology the data consist of variables and observations. Adopting the terminology of relational databases, SAS data sets are also called tables, observations are also called rows, and variables are also called columns.



Data types: In SAS there are just two data types: numeric and character. Numeric fields are, well, numbers. They can be added and subtracted, can have any number of decimal places, and can be positive or negative. In addition to numerals, numeric fields can contain plus signs (+), minus signs (-), decimal points (.), or E for scientific notation. Character data are everything else. They may contain numerals, letters, or special characters (such as $ or!) and can be up to 32,767 characters long.

If a variable contains letters or special characters, it must be character data. However, if it contains only numbers, then it may be numeric or character. Any variable followed by a $ (dollar) sign is declared a character variable by the SAS system

Missing data: Sometimes despite your best efforts, your data may be incomplete. The value of a particular variable may be missing for some observations. In those cases, missing character data are represented by blanks, and missing numeric data are represented by a single period (.).

SAS procedures:

Once your data are accessible as a SAS data set, you can analyse the data and write reports using a set of utilities known as SAS procedures. Prewritten programs that analyse and process data sets and display results.

SAS procedures analyse data in a SAS data set for producing univariate descriptive statistics, frequency tables, cross tabulation tables, tabular reports consisting of descriptive statistics, charts, plots, and so on.

Examples:

- Proc APPEND
- Proc COMPARE
- Proc CONTENTS
- Proc MEANS
- Proc SUMMARY
- Proc PRINT
- Proc SORT
- Proc SQL
- Proc FREQ

Size of SAS data sets: Prior to SAS 9.1, SAS data sets could contain up to 32,767 variables. Beginning with SAS 9.1, the maximum number of variables in a SAS data set is limited by the resources available on your computer—but SAS data sets with more than 32,767 variables cannot be used with earlier versions of SAS. The number of observations, no matter which version of SAS you are using, is limited only by your computer's capacity to handle and store them.

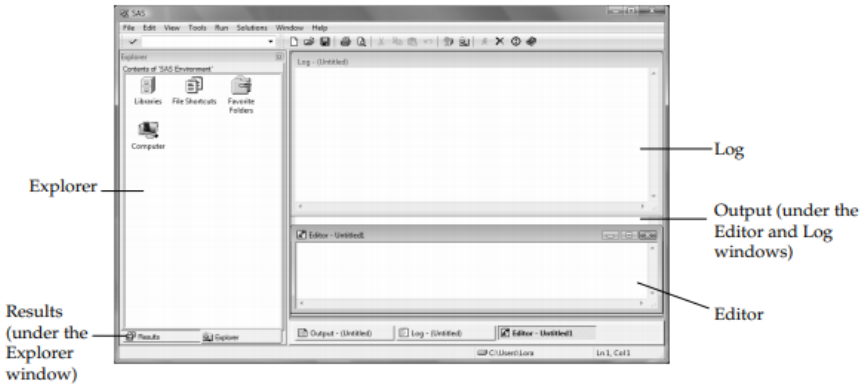
Rules for SAS names: You make up names for the variables in your data and for the data sets themselves. It is helpful to make up names that identify what the data represent, especially for variables. While the variable names A, B, and C might seem like perfectly fine, easy-to-type names when you write your program, the names Sex, Height, and Weight will probably be more helpful when you go back to look at the program six months later. Follow these simple rules when making up names for variables and data set members:

- Names must be 32 characters or fewer in length.
- Names must start with a letter or an underscore (_).
- Names can contain only letters, numerals, or underscores (_).
  No %$!*&#@,
- Names can contain upper- and lowercase letters.

The SAS Windows:

There are five basic SAS windows: the Results and Explorer windows, and three programming windows: Editor, Log, and Output. It is possible to bring up SAS without all these windows, and sometimes the windows are not immediately visible (for example, in the Windows operating environment, the Output window comes up behind the Editor and Log windows), but all these windows do exist in your SAS session. There are also many other SAS windows that you may use for tasks such as getting help, changing SAS system options, and customizing your SAS session.

SAS Windos



Source; https://www.sas.com/storefront/aux/en/splsb/61860_excerpt.pdf

Editor: This window is a text editor. You can use it to type in, edit, and submit SAS programs as well as edit other text files such as raw data files. In Windows operating environments, the default editor is the Enhanced Editor. The Enhanced Editor is syntax sensitive and color codes your programs making it easier to read and find mistakes. The Enhanced Editor also allows you to collapse and expand the various steps in your program. For other operating environments, the default editor is the Program Editor whose features vary with the version of SAS and operating environment.

Log: The Log window contains notes about your SAS session, and after you submit a SAS program, any notes, errors, or warnings associated with your program as well as the program statements themselves will appear in the Log window.

Output: If your program generates any printable results, then they will appear in the Output window.

Results: The Results window is like a table of contents for your Output window; the results tree lists each part of your results in an outline form.

Explorer: The Explorer window gives you easy access to your SAS files and libraries.

The SAS Commands:

There are SAS commands for performing a variety of tasks. Some tasks are probably familiar, such as opening and saving files, cutting and pasting text, and accessing Help. Other commands are specific to the SAS System, such as submitting a SAS program, or starting up a SAS application. You may have up to three ways to issue commands: menus, the toolbar, or the SAS command bar (or command line).

Menus: Most operating environments will have pull-down menus located either at the top of each window, or at the top of your screen. If your menus are at the top of your screen, then the menus will change when you activate the different windows (usually by clicking on them). You may also have, for each window, context-sensitive pop-up menus that appear when you press the right or center button of your mouse.

Toolbar: The toolbar, if you have one, gives you quick access to commands that are already accessible through the pull-down menus. Not all operating environments have a toolbar.

SAS command bar: The command bar is a place that you can type in SAS commands. In some operating environments the command bar is located with the toolbar; in other operating environments you may have a command line with each of the SAS windows (usually indicated by Command=>). Most of the commands that you can type in the command bar are also accessible through the pull-down menus or the toolbar.

The Output window: After submitting your program in the SAS windowing environment, your results will go to the Output window. If you have the SAS Explorer option turned on (some operating environments have this turned on by default, while others do not), then you will also see a listing of the different parts of your output in your Results window.

The Results window:

When you have a lot of output, the Results window can be very helpful. The Results window is like a table of contents for your output. It lists each procedure that produces output, and if you open, or expand, the procedure in the Results tree, you can see each part of the procedure output.

Check Your Progress - 2

6. Every SAS statement ends with a
   a. Semicolon
   b. Colon
   c. Underscore
   d. Slash
   e. Asterisk.

7. What is role of toolbar in SAS?
    a. To type SAS commands
    b. Give quick access to commands that are accessible.
    c. Give access to SAS files
    d. Contains inheritance function
    e. Missing data values
8. How many types of windows are there in SAS system?
    a. 1
    b. 2
    c. 3
    d. 4
    e. 5
9. What is the purpose of log window?
    a. It contains all important notes of your SAS session. It appears notes error and warning after submitting your SAS program
    b. It contains output
    c. It can easy access to your SAS files and libraries
    d. Importing and exporting data
    e. Writing the SAS program.
10. Which of the following is not a SAS procedure?
    a. Proc APPEND
    b. Proc DESCRIPTIVE
    c. Proc CONTENTS
    d. Proc MEANS
    e. Proc SUMMARY

12.7 Summary

Computers were initially used in business research to collect data and perform simple analysis. However, the introduction of specialized packages like Statistical Research like SAS, Stata and SPSS enabled automated data analysis and output generation. SPSS can perform complex data analysis using statistical and mathematical functions. The chapter discusses the detailed procedure to carry out data analyses ranging from simple descriptive statistics like means to advanced analyses like regression using SPSS for Windows.

12.8 Glossary

SPSS: Statistical Package for the Social Sciences.

SAS programs: A SAS program is a sequence of statements executed in order. A statement gives information or instructions to SAS and must be appropriately placed in the program.

12.9 Suggested Readings/Reference Material

1. Gupta, S. P. Statistical Methods. 46th Revised ed. New Delhi: Sultan Chand & Sons. 2021.
2. I. Levin Richard, H. Siddiqui Masood, S. Rubin David, Rastogi Sanjay. Statistics for Management. Pearson Education; Eighth edition, 2017.
3. Gerald Keller. Statistics for Management and Economics. Cengage, 2017.
4. Arora, P. N., and Arora, S. CA Foundation Course Statistics. 6th ed. S Chand Publishing, 2018.
5. Mario F Triola. Elementary Statistics. 13th ed., 2018.
6. David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, Jeffrey D. Camm, James J. Cochran. Statistics for Business and Economics. 13th Edition, Cengage Learning India Pvt. Ltd., 2019.
7. S D Sharma. Operations Research. Kedar Nath Ram Nath, 2018.
8. Hamdy A. Taha. Operations Research: An Introduction. 10th ed., Pearson, 2016.
9. Malhotra, N. (2012), Marketing Research: An Applied Orientation, 7th ed., Pearson, 2019.
10. Cooper, D.R. and Schindler, P.S. and J. K. Sharma (2018), Business Research Methods, 12th edition, McGraw-Hill Education.

12.10 Self-Assessment Questions

1. Discuss the menus in SPSS.
2. Explain about SAS windows.
3. Explain how do you do frequency analysis using SPSS and SAS.

12.11 Answers to Check Your Progress Questions

1. (d) A screen in which variables can be defined and labeled is known as data view.

2. (c) To do statistical analysis from descriptive statistics is known as Analyze.

3. (b) Role of value labels is to identify a particular code if data has been codified for use

4. (e) Powerful Statistical Analysis, Simple descriptive statistics like means and deviations to advanced analysis like regression and factor analysis, Data management system in a graphical environment.

5. (b) Role of utilities is to run script and to display information on the contents on SPSS data files

6. (a) Every SAS statement ends with a semicolon.

7. (b) The role of toolbar is to give quick access to commands that are accessible.

8. (e) 5 types of windows are there in a SAS system.

9. (a) The purpose of log window is it contains all important notes of your SAS session. It appears notes error and warning after submitting your SAS program

10. (b) Proc DESCRIPTIVE

It is not a SAS procedure.

Quantitative Methods

Course Structure

Block	Unit Nos.	Unit Title
I Introduction to Statistics and Probability		
	1.	Arranging Data
	2.	Central Tendency and Dispersion
	3.	Probability
	4.	Probability Distribution and Decision Theory
II Statistical Relations and Hypothesis Testing		
	5.	Statistical Inference and Hypothesis Testing
	6.	Correlation and Linear Regression
III Statistical Regression and Quality Control		
	7.	Multiple Regression
	8.	Time Series Analysis
	9.	Quality Control
IV Statistical Distributions, Variations and IT		
	10.	Chi-Square Test and Analysis of Variance
	11.	Role of IT in Modern Business Enterprise
	12.	Statistical Software Tools
V Advanced Statistics		
	13.	Index Numbers
	14.	Simulation
	15.	Linear Programming
VI Business Research		
	16.	Introduction to Business Research Methods
	17.	Questionnaire Design
	18.	Report Writing