L:24

STATISTICAL STUDY OF TELECOMMUNICATION TRAFFIC IN PAKISTAN.

A

THESIS

SUBMITTED

TO THE DEPARTMENT

OF ELECTRICAL ENGINEERING

EAST PAKISTAN UNIVERSITY OF ENGIN-

EERING & TECHNOLOGY, DACCA, IN PARTIAL

FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE(ENGG.)IN ELECTRICAL ENGINEERING.

BY

KAZI ABDUR ROUF

DEPARTMENT OF ELECTRICAL

ENGINEERING, EAST PAKISTAN UNI-

VERSITY OF ENGINEERING AND TECHNOLOGY.

D A C C A

1967.

ACCEPTED AS SATISFACTORY FOR PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF M.Sc. (ENGINEERING) IN ELECTRICAL ENGINEERING.

Examiners (i) .....F. Rahman
11. 2. 67.

(ii) ....
11.2.67

(iii) ....
11/2/67.

CERTIFICATE

THIS IS TO CERTIFY THAT THIS WORK WAS DONE BY ME AND IT HAS NOT BEEN SUBMITTED ELSEWHERE FOR THE AWARD OF ANY DEGREE OR DIPLOMA OR FOR PUBLICATION.

(signature of the Candidate)

Countersigned

(Superviser)

ACKNOWLEDGEMENT:

The Author expresses his heartfelt gratitude and indebtedness to Dr.A.M.Zahurul Haq, Professor of Electrical Engineering, East Pakistan University of Engineering and Technology, Dacca for his sympathetic and constant guidance throughout the progress of this work.

To Professor Dr.M.N.Azam, Head of the Department of Electrical Engineering and Dean of the faculty of Engineering. The author is grateful for his introduction to the statistical theory and the great interest he has shown in this work by extending his valuable suggestions continually.

The author wishes to express his appreciation to many friends and colleagues who have helped in the collection of data and making arrangement for typing of these pages.Particularly valuable assistance has been rendered by Mr. Shamsul Islam, Divisional Engineer Telegraphs and Mr. Habibur Rahman, A.D.E.T. of the T.& T. Department. The author is indebted to his wife for her assistance in the preparation of the manuscript.

The author gratefully acknowledges the help extended to him by Dr.V.G.Desa an Dr.Kamal Uddin of the Mechanical Engineering Department in working with the automatic calculator.

Finally the author is thankful to Mr.A.M.Patwari, Assistant Professor of Electrical Engineering, East Pakistan University of Engineering and Technology for his kind help in procuring a reference work from England.

CONTENTS:

CONTENTS:

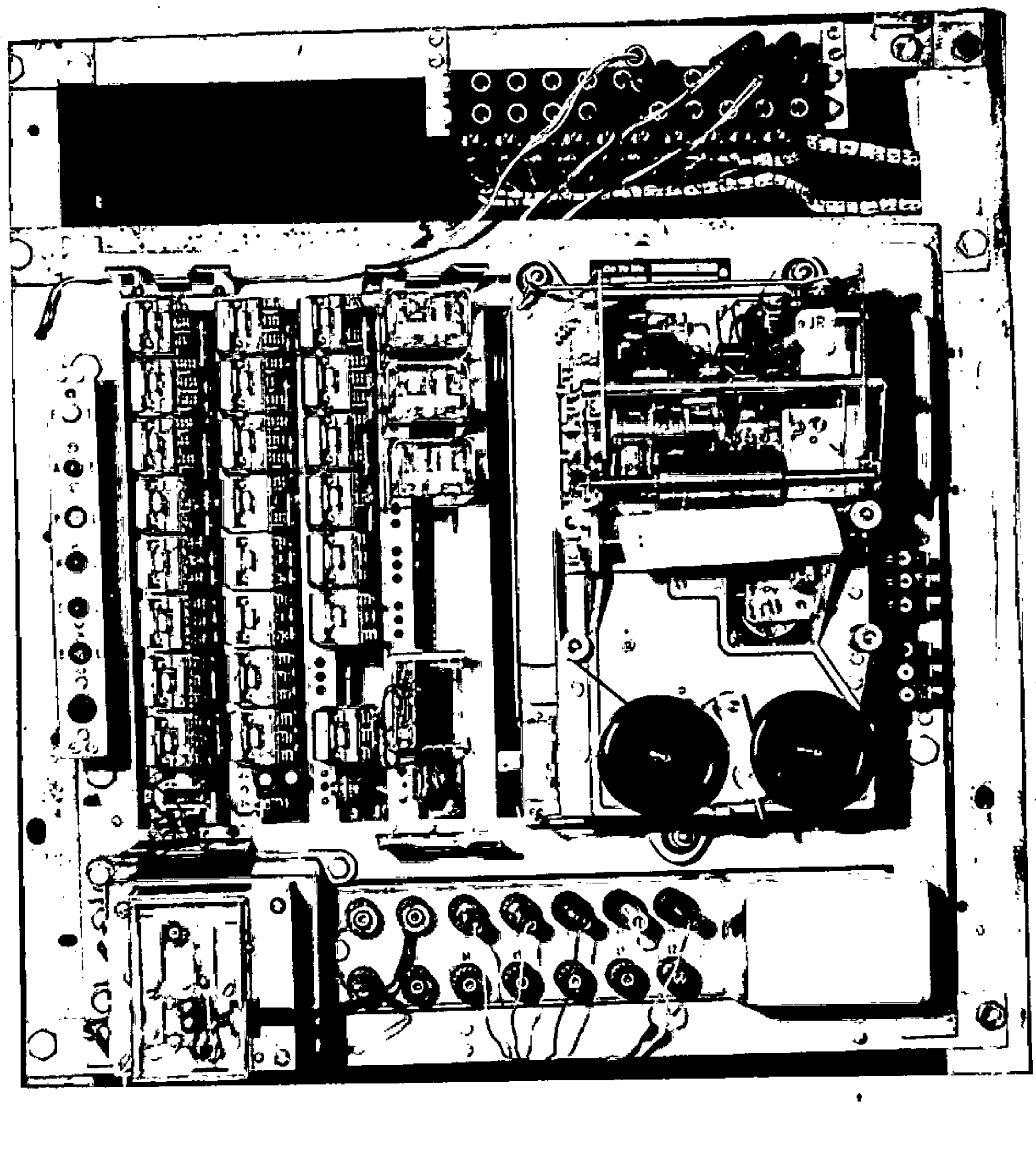
....

LIST OF FIGURES.

LIST OF FIGURES: (Contd) Page.2.

LIST OF TABLE:

LIST OF TABLE: (Contd) Page.2.



A B S T R A C T.

Some data regarding telecommunication traffic in Pakistan have been collected begining from 1954 upto September, 1966. Certain data pertaining to the local telephone subscribers at Dacca, interwing and Dacca-London telegraph traffic along with the technical losses, hourly booked and effective trunk calls on Dacca-Karachi circuit and subscribers trunk dialling on Dacca-Chittagong circuits have been studied and systematic statistical analysis has been attempted.

Significant correlation between Total interwing and Dacca-London effective telegraph traffic and technical loss are found to exist. Auto correlation function of the interwing and Dacca-London technical loss (hours) have been determined for various time intervals (numbers of months). The correlation of hourly booked calls and effective calls has been found and a high value has been obtained. The correlation between the effective minutes and hourly percentage of circuits time available has been determined with a view to examine the economic utilisation of this system. The statistical characteristic of the Dacca - Chittagong direct dialling system has been investigated in detail and certain significant conclusion regarding the effectiveness of the existing facilities have been arrived at.

CHAPTER - I.

INTRODUCTION.

Study of traffic flow is most important for efficient planning and maintenance of telecommunication systems. A complete idea of the characteristics of traffic is of very great help to any administration dealing in telecommunication. The telecommunication organisations are greatly interested in analysing the traffic pattern as the same helps them to predict the future pattern of traffic flow.

Full knowledge of the traffic pattern helps the administration to know whether the existing set up is adequate to deal with the demand. Sometimes it happens that the overall set up is adequate to deal with the existing traffic but the phasing is not proper. A proper analysis of the traffic pattern reveals such discrepancies which may be corrected appropriately and economically.

For quality control work as applied to telecommunication it is necessary to have complete study of the pattern of traffic.

In planning and equipment designing the knowledge of traffic is essential to calculate the quantities of different components.

It is not sufficient for the Engineer to say how a work can be done; he must also say how economically the work can be done.

Apart from this, a Telecommunication Organisation must be able to predict its requirements in men and material for a future period. Such prediction requires traffic analysis and its projected values in future.

A systematic study of such problems has not been done in Pakistan as far as is known to the author. In Europe and America such works have been done to a considerable extent. Problems in different countries differ because of the social and economic structure of the country. Problems faced by our country is not the same as it is in other countries due to obvious reasons.

One of the modern tendency of the present day is the steadily increasing use of statistical methods in a wide variety of scientific fields. Many Engineering firms have used statisfical methods with profit and it is expected that such methods of study in the field of Telecommunication Traffic would be equally profitable.

The fluctuations of traffic is very complex and frequently the under lying causes of fluctuations are not clearly understood. The process is known as a random process, although the random phenomenon may possess a statistical regularity. Because of the complexity and incomplete understanding of the phenomenon, no simple laws are available to describe their behaviour exactly. Hence an important characteristic of the random phenemenon is that in the process of its fluctuations the future can not often be predicted precisely.

Closely related to the notion of random phenomenon are the notions of random events and their probahilities. A random event is one, whose relative frequency of occurence, in a very long sequence of observations of randomly selected situation in which the event may occur, approaches a stable limit value as the number of observations is increased to infinity;

the limit value of the relative frequency is called the probability of the random event. A fundamental idea is that traffic should be considered statistically and described in accordance with the concepts of probability theory.

It is frequently argued that results of statistical investigation can be interpreted according to individual taste and are, therefore, unreliable. Of the fallacies about statistics, the most persistent is illustrated by the remarks of a judge that he was not influenced by statistics, he only relied on facts. In one sense, statistics are only accumulated facts and any distinction between statistics and facts is not understandable. If proper use is made of the data through statistical manipulation it will reveal facts inherent in the data, and in so far as this process makes evident relationship which could not otherwise have been detected, it does add to the total information. Telecommunication engineers are often discouraged from using statistical methods as they have no time to read a large number of theoritical volumes in order to cover completely their applied problems. The mathematical theory of probability is involved and is spread in a widely scattered literature.

We need not look far to find problems using statistical idea. The concept of average is used by every one in referring to a man of average measurement of ear to mouth distance. We use a single figure of an average to represent the whole group of men. We use the single quantity, the average, to describe the characteristic of the group. For example, the average distance between the lips and the centre of the ear of a group of men is 14 c.m.

We do not mean all men in the group have a measurement of 14 c.m. We use the average as number which describes the whole group of men, so that if we were to pick a man at random and measure the distance between his lips and centre of the ear we would expect it to be 14 C.M.

Another concept is that of dispersion or variability. An Engineer may say a group of condensers is more variable in quality than another group. For example the measurement between the lips and the centre of the ear varie's from 11 c.m. to 17 C.M. Dispersion measures how many data are away from the mean.

Correlation is another commonly used statistical description. Increase in one variable may be accompanied by an increase or decrease of another variable. The variables are then said to be correlated.

In the opening stages of statistical investigation we will need collection of a large amount of data from which to extract the quantities relevant to our purpose. Collection of data takes a long time and sometimes we need data of several years before we can deduce some useful results. Unfortunately we have practically no data available and whereever available, they are not complete. Certain observations have been made and data collected which in all cases may not be sufficient to reach any conclusion.

Many excellent statistical tools are available, some of which, although very simple, are powerful techniques in the method of solution of many important problems encountered in telecommunication traffic fields. This, in fact, will help our telecommunication organisation in evolving standard methods which may be utilised to tackle the ever expanding planning and maintenance problems.

CHAPTER II.
SURVEY OF PAST STUDIES

To the Author's knowledge, works on this field has not yet been done in Pakistan. In Pakistan, Telecommunication is entirely a Government enterprise. Recently a planning cell has been organised to do the perspective planning for the next 20 years.

A section has been organised to collect data and process it. It is in its initial stage and any works worth mentioning has yet to be done. In fact, statistical methods are yet to be introduced.

Works on this line are done in some telecommunication administrations in Europe and America. But very little is known about their works. A series of articles entitled "A simple introduction to the use of statistics in telecommunication Engineering" by J.F. DOUST and H.J.JOSEPHS, F.S.S. A.M.I.E.E. may be cited as an example.

These articles show how statistical methods are used in analysing samples which obey "Normal probability Law", the poisson probability law" etc.

They show how simple tests on specially prepared graph papers indicate whether or not statistical stability has been obtained. They also show how required probability characteristics are obtained from the graphs without going through laborious mathematics.

They have also shown how small samples can be dealth with "students' 't' method", using tables of 't' and probability chart for 't' test.

Bivariate distributions have been analysed clearly showing how the test of significance of correlation co-efficients are obtained from small samples.

In the first article of the above named series, a problem to determine the best length of hand micro-telephone has been considered. For this purpose measurements were taken at random from different individuals. At the first instant 1945 measurements were taken. It was processed in the statistical way and found by different tests that the data were not sufficient. Next, another set of 1945 data were taken and combined with the previous 1945. The whole was again processed and tested and it was concluded that sample was now big enough to represent a 'Population' and it was stable. The above sample followed normal probability law and it was shown how to use probability paper for such works.

Comparatively small sample has been dealth with and it was shown that data was sufficient as it responded to the tests prescribed.

In the second article it has been shown how to deal with small samples. When samples are small it may not be possible to draw exact conclusion as to the properties of the parent 'Universe' from which they were drawn, but the limits between which those properties must lie can be determined with a desired degree of certainty.

To check tne adjustment of a machine which produce electrical insulating washers a small sample of 10 washers were available. Students 't' test showed that deviations were not altogether due to chance, the machine adjustment might be wrong. This was significant. The truth of the assumption was doubtful. Table of 5% and 1% level of significance for different values of N was used to save laborious calculations.

In the third article the concept of discontinuous variables were introduced and Poisson probability law was utilised to calculate the number of lock outs occuring in 50 telephone connections of a certain 4-wire circuit. Theoritical and observed frequencies were tabulated and it was found to obey nearly 'Poisson Law'.

Probability charts have been introduced which are very helpful in calculation. Some other applications of Poisson probability paper have also been shown.

In the fourth article Force-displacement characteristics of carbon granules has been discussed. From the data obtained the co-efficient of correlation has been calculated.

By the significance test it was established that real carrelation exists.

Apart from tne above "Ki squarred test " (X^2 test) has been introduced to telecommunication.

The above articles are excellent demonstration of the use of statistical methods in telecommunication.

They can be used as guiding examples for further works on telecommunication traffic.

CHAPTER III.

CERTAIN STATISTICAL TECHNIQUES TO BE APPLIED.

It is necessary in the investigation of any phenomenon to observe and record some characteristics of the object under consideration. We must have observation of some form. The observation used as an illustration are the height measurement of a number of people.

3.1 Histogram

A Histogram is a picture of a number of observations. Suppose 5 measurements have been taken in nearest inches as follows:
68, 72, 66, 67, 68. We then represent them by equal units of area above a scale marked with those heights (figure 3.1).

If we take another few measurements we would get a picture from which we can form at sight an idea of how the heights are distributed at various values along a line in inches (Fig. 3.2).

It is not difficult to see from this histogram that the average height is around 69 or 70 inches and a small proportion of the heights are less than 66 inches or morethan 73 inches.

This picture or Histogram is particularly useful when a large number of measurements have been taken.

We have drawn here histogram assuming that height is a discrete variable. The above description of how to draw a histogram will thus apply to discrete measurements as they occur and will apply to continuous measurement if they are grouped into a number of categories. Even if the measurements had been made to a much greater precision, they could be represented by the above histogram by combining all measurements

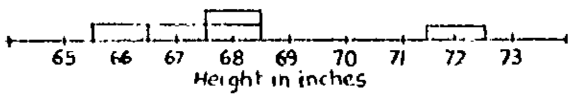


Fig. 3·1

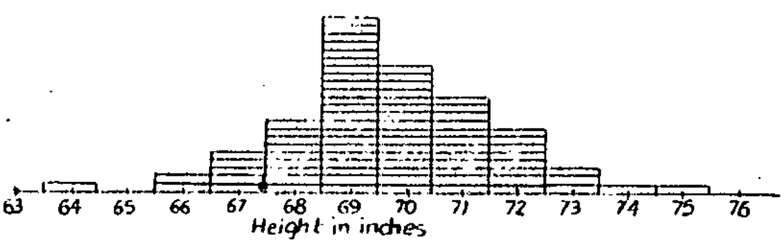


Fig. 3·2

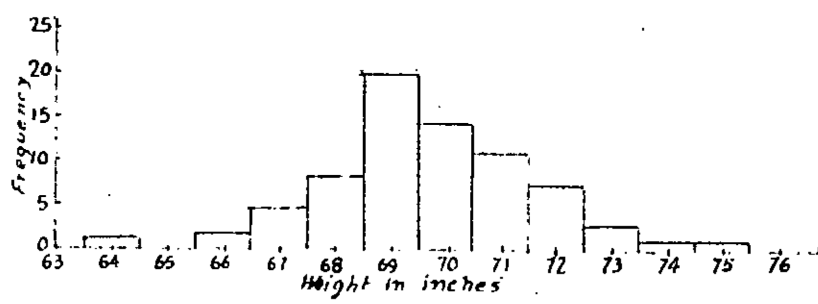


Fig. 3·3

between 68.5 to 69.5 into the group 69 inches etc. The measurements can of course be grouped into any length interval. If we had assumed that the measurements had been made to the nearest $\frac{1}{2}$ inch, we would have obtained a histogram similar in appearance to that above but the data would be grouped into about twice as many groups.

It is important to record observation to the required degree of precision. For most studies 8 to 20 groups would meet the usual requirements.

If more and more observations were taken and the histograms were constructed by grouping the measurements into intervals of less and less width, it is possible to imagine that the histograms would approach a smooth curve which would represent more truly the distribution of continuous measurements. Fig. 3.4 illustrate such a smooth curve.

The interval chosen for the classification is referred to as the group interval and the frequency in any particular group interval is the group frequency.

3.2 Frequency polygon.

Another graphical description with which each measurement occurs is given by a frequency polygon. This is constructed by joining midpoints at the top of each column in the Histogram. For the histogram in fig. 3.3 the frequency polygon would be as in fig. 3.5.

When these Histograms or frequency polygons, are drawn with proportion in place of actual frequencies, we call them "frequency distribution" of the measurements they represent.

Suppose that, instead of indicating the number or proportion of measurements at each length

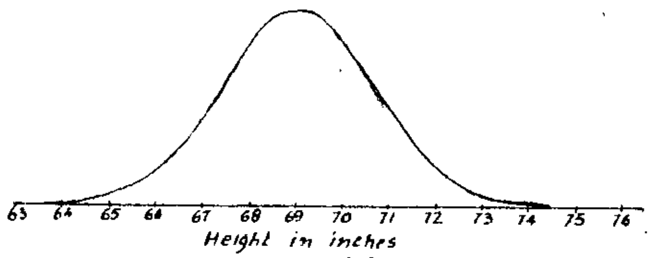


FIG. 3·4

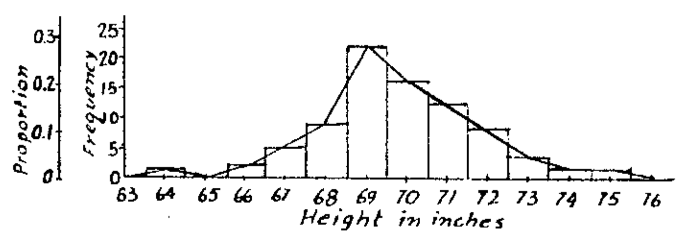


FIG. 3·5

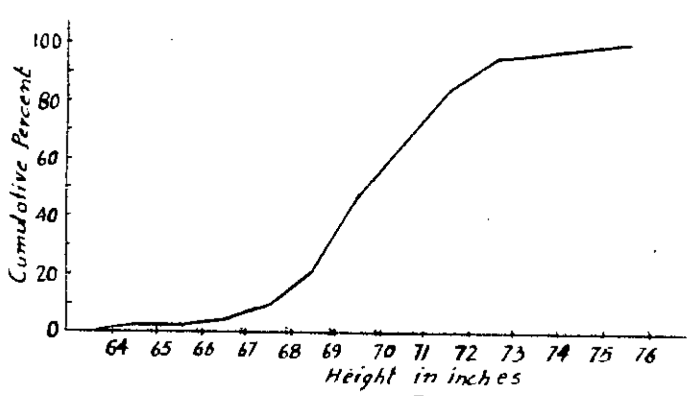


FIG. 3·6

as in the Histogram 3.3, we wished to draw a corresponding polygon which would indicate the percentage of measurements shorter than a given length. If we look at the Histogram, we can see that starting from the left and progressing towards right, when we have reached the point 63.5 inches 0 measurements or Zero percent is below that value; by the time we have reached the values of 64.5 inches, we know one measurement is shorter than 64.5 inches. If we continue in this manner and make the following columns, wherein the first column represent the height, the second column the number of individuals whose measurements are less than that height, the third column indicate the percentage of measurements which are less than that height and the fourth column indicate the proportion of measurements with height less than that height. If proportions are used in drawing the graph, the polygon is called a cumulative distribution polygon.

The cumulative distribution indicates by its height, above a particular point on the scale the proportion of the area in the corresponding Histogram to the left of that point (fig.3.6).

We have shown how a distribution is represented by Histogram and frequency Polygon, but it is not always possible or convenient to do so.

3.3 Measurement of Central values.

Central value refers to the location of centre of the distribution. One of the intuitive notions of a centre of a distribution is the arithmatic mean or average. It is defined as the sum of all observations divided by the number of observations.

If there are small number of observations, we shall add them directly and divide the total by the number of observations. If we have four observations, X_1, X_2, X_3 and X_4 than 'm' mean =

$$\frac{X_1 + X_2 + X_3 + X_4}{4}$$

Many occasions will arise when we need adding very large number of observations. We then use symbol Σ for summation, for example $\sum_{i=1}^{4} X_i =$ $X_1 + X_2 + X_3 + X_4$ and $\bar{X} = \sum_{i=1}^{4} X_i / 4$

Similarly $\sum_{i=1}^{20} X_i$ is the sum of 20 observations.

The arithmatic mean is equal to the sum of the N observations divided by the number of observations.

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{N}$$

Let fi be the frequency of observation of value X. Then $f_1 X_1 + f_2 X_2 + \text{-----} f_K X_K = \sum_{i=1}^{K} f_i X_i$

Total number of observations is

$$f_1 + f_2 + \text{---------} f_K = \sum_{i=1}^{K} f_i = N$$

thus we have

$$\bar{X} = \sum_{i=1}^{K} f_i X_i \Big/ \sum_{i=1}^{K} f_i$$

3.4 <u>Other measures of Central value</u> -

Two other commonly used measure of a distribution are the mode and the median. The mode is the value which occur most frequently. In Histogram it is the mid point of the interval with the greatest frequency. The median is the point on the scale of observations which has an equal area under the Histogram on either side. Then the median is the middle observation if there are odd number of cases and is the mean of the two central observations if there are an even number of cases.

3.5 Measure of dispersion. The variance.

An important concept in statistics is that an average does not in itself give a clear picture of a distribution. The graphs of fig: 3.7 show a group of distribution all having the same arithmatic mean yet obviously differing in general appearance.

Another type of measure which helps to clarify the shape of the distribution is a measurement which indicate how the observations are spread out from the average, such a measure could be called a measure of dispersion, spread or variability. It is usually desirable to have a measure which will be large if observations are distant from the mean and small if they are close to the mean.

At first glance the sum of the deviations of the observations from the mean.

$$\sum_{i=1}^{N} (x_i - \bar{x})$$

may seem to be a good measure for thispurpose, but on further examination its value is seen to be always equal to zero.

This difficulty may be overcome by squaring the deviations before they are added. The variance is defined as the sum of squares of the deviations of the observations from $\hat{\bar{X}}$ divided by the total number of observations. In symbols

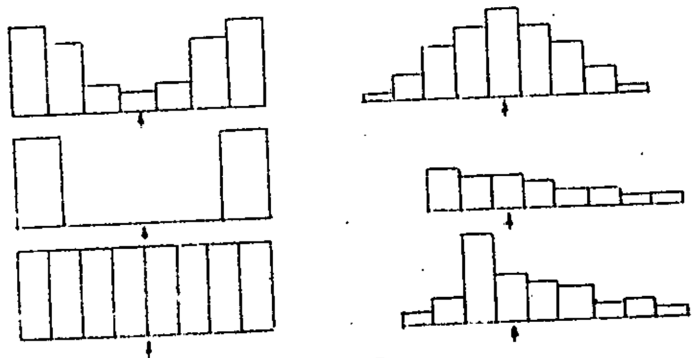


FIG. 3·7

$X_1 - \overline{X}$ is the deviation of the 1st observation from the mean, $X_2 - \overline{X}$ is the deviation of the 2nd observation from the mean, and so on. Thus the variance which we shall represent symbolically by S^2, is

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \text{-----} (X_n - \bar{X})^2}{N} = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^2}{N}$$

3.6. Standard deviation

Since all the deviations were squared at the start it is convenient to take the square root of the average sum of squares, so that the final measure has the same dimension as the original observations from which it was obtained. This is known as the standard deviation of the distribution.

Mean deviation - The calculation of this is best illustrated by an example.

Let us consider the set of cricket score of a cricketer A.

The mean is found as 50. Next subgtracting the mean score from each individual score we have.

$(X - \bar{X})$ values as

-50, 62, 47, - 46, 26, - 49, 38, 52, - 36, -44

Now since the sum of all these deviations is automatically zero, a measure of spread can not be based on them as they stand. If the magnitude of the deviation is used, irrespective of their sign, a measure is obtained which is a guide to the spread of the observations. This is called the mean deviation and is obtained by summing the deviations regardless of their sign and dividing by the number of observations involved.

3.7 Universe and Sample

Any set of individuals (or objects) having some common characteristics constitute a 'population' or 'Universe'. Any sub set of a population is a 'sample' from that population.

The term population may refer either to the individuals measured or to the measurements themselves. There is then a 'distribution of the measurements of a sample' which we actually observe and study and a 'distribution of the measurements of the universe' which may exist but usually not in observed or recorded form. One of the most important problems in statistics is to decide what information about the distribution of the population can be inferred from the study of the sample.

A third type of distribution consists of the distribution of measurement made on each of all possible samples of fixed size which could be taken from a universe. For example, if we took all possible samples of 10 students from a given school and computed the mean height of each sample, we would have a large number of means. These means form a distribution, which we call the sampling distribution of the mean of samples of size 10.

3.7.1 Population or universe

A universe can be finite or it can be infinite. In many cases a universe will be finite but so large that we treat it as though it were infinite.

Many populations are so scattered as to be inaccessible as a whole. For example, in studying some characteristics of all the telephone instruments produced in a factory it might be impractical to take measurements on all telephones but quite responsible to measure a sample.

The various descriptive words and measurements which we have previously mentioned are applied to the distribution of the universe, as well as to the other distributions. We shall be specially interested to the standard deviation and the mean. We shall denote the arithmatic mean of the universe by 'm' and the standard deviation of the universe by σ

Any measurable characteristics of an universe is called a para-meter. m and σ are examples of parameters.

3.7.2 Sample

The size of a sample usually represented by the letter N, is the number of individuals in the sample. A sample may be of any size from N = 1 to the number of items in the universe.

The measurements on the individuals in the sample will form a distribution which will have a mean, denoted by $\bar{X}$, and a standard deviation, denoted by 'S'. Presumably $\bar{X}$ and 'S', which we can measure, should give us some information about m and σ , whose values are usually not known. $\bar{X}$ and S are different from sample to sample, while m and σ are constant, i.e, have particular value for a particular universe.

A value computed entirely from a sample is called a statistic. $\bar{X}$, S, the madian, the mode of the sample are examples of statistics.

The method of choosing a sample is an important factor in determining what use can be made of the sample. If some individuals in the universe are more likely to be chosen than others the sample is said to be biased. If has been found that subjective methods of picking individuals from a sample often lead to biased samples, apparently due to subconscious or conscious preferences of the person making the selections. To prevent this bias and to avoid criticism of bias even if the sample is not biased, some non-subjective method of choosing a sample should be employed.

3.8 Random sampling

When every individual in the population has and equal and independent chance of being chosen for a sample, the sample is called a random sample. Technically every individual chosen should be measured and returned to the population before another selection is made. This means, of course, that an individual can be chosen twice in the same sample. If the population is large compared with the sample size, very little error will result from the procedure of not returning each individual to the population. In practice the individual is rarely returned, and in fact it is frequently impossible to do so, as in a case where the individual is changed or destroyed by the examination. In many cases care is actually taken to prevent the same individual from appearing twice in the sample. Where there is a small universe we return each individual before drawing another.

3.9 Sampling distribution

Any statistic (measurement on a sample) has a sampling distribution. Suppose, we have from a universe, all possible samples. On each sample we perform the necessary computations to find $\bar{X}$ and S. We should then have a universe of means and a universe of S. By examination of the sampling distribution of the mean we can tell how frequently the mean $\bar{X}$ will fall in any interval we wish to discuss. The values which are obtained for the mean are governed by probability in exactly the same way as any other quantity which is subject to accidental fluctuations. It has a certain distribution law with its own standard deviation; the same is true for any statistics that can be calculated. The question arises, therefore, as to how widely and according to what law a statistic, such as a mean or a standard deviation may be expected to vary. These two important statistics obey laws of distribution which are nearly "normal". They are exactly normal when derived from a data which satisfy the "normal law", and approximately "normal" in which the data satisfy skew (or unsymmetrical) distribution laws. Since the frequency distribution of a statistic follows a "normal law" it is apparent that the determination of its standard deviation (or standard error as it is called in this case) makes it possible to form an estimate of the confidence that can be placed in the observed values of that statistic. The standard error of the mean $\bar{x}$ and the standard deviation S, of a normal sample are given approximately by the simple expression $\sigma/\sqrt{N}$ and $\sigma/\sqrt{2N}$ respectively.

For a large sample the distribution tend to be 'normal'. There are two other principal distributions namely the Poisson distribution and the Binomial distribution. It is important to have a general knowledge of these distributions and the mathematical formula by which they are represented.

3.9.1 The normal distribution

A variate is said to be normally distributed when it takes all values from $-\infty$ to $+\infty$ with frequencies given by a definite mathematical law, namely, that the logarithm of the frequency at any distance d from the centre of the distribution is less than the logarithm of the frequency at the centre by a quantity proportional to d^2 . The distribution is symmetrical with the greatest frequency at the centre. Although the variation is unlimited, the frequency falls off to exceedingly small values at any considerable distance from the centre. Fig. 3.8 represents a normal curve of distribution. The equation of the normal curve is

$$y = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

where m = mean of the distribution

σ = standard deviation of the distribution

X = Abscissa, measurement marked on horizental axis,

Y = Ordinate, height of curve corresponding to an assigned value of X.

N = Number of observations.

The total area between the curve and the X - axis is one square unit.

The first step in analysing the data obtained from a large sample is to see whether or not the relative frequency points can be fitted by the bell shaped normal charactertistic shown in fig.3.9. In fig 3.9 A is the deviation from the arithmatic mean and the solid line is the curve joining the relative frequency points.

The probability that the deviation does not exceed A is given by the ratio of the area enclosed between the bell shaped curve and the ordinates erected at A, to the total area under the curve. It is shown in works of theory of statistics that it is simple to express the percentage of the total area in terms of standard deviation. The standard deviation measure the extent to which the individual values are scattered.

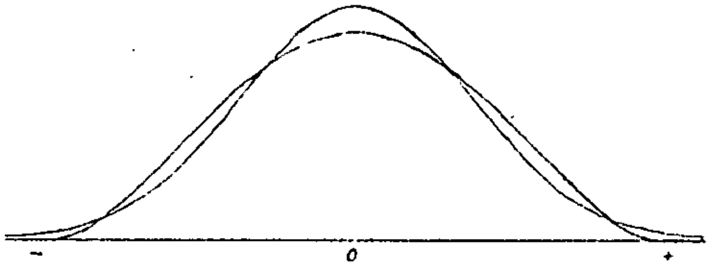


FIG. 3·8
Normal Distribution.

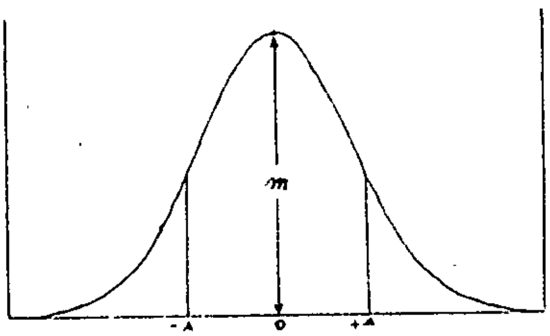


FIG. 3·9
Normal Distribution.

It can be shown that if the frequency distribution follows normal probability law, then

50%	of the area lies	within	$m \pm 0.6745\sigma$
68.27%	" "	"	$m \pm \sigma$
95.95%	" "	"	$m \pm 2\sigma$
99.73%	" "	"	$m \pm 3\sigma$

3.9.2 Probability Paper

From the data obtained it is possible to obtain the arithmatic mean and the standard deviation. The next phase is to find whether the data obtained from a random sample does or does not depart significantly from the normal probability law. Since the goodness of fit of the data with the bell-shaped curve is a matter of personal openion the difficulty of attaching the proper significance to deviations from the ideal bell-shaped curve is apparent.

The problem of attaching proper significance to these sampling deviations was attacked by Allen Hazen, an American Civil Engineer. Hazen found that by plotting the cumulative frequency per cent figures instead of actual frequency figures he obtained curves which were straight from 30% to 70%. This fact is illustrated in figure 3.10 which has been obtained by plotting cumulative frequency figures instead of relative frequencies. Hazen found that by changing the vertical scale he could readily transform the cumulative frequency curve into a straight line. It can be seen from fig. 3.10 that the scale needs little alteration for values from 20 to 80 percent. Beyond these limits the greater the difference from the middle 50% values the greater must be the alteration to the scale.

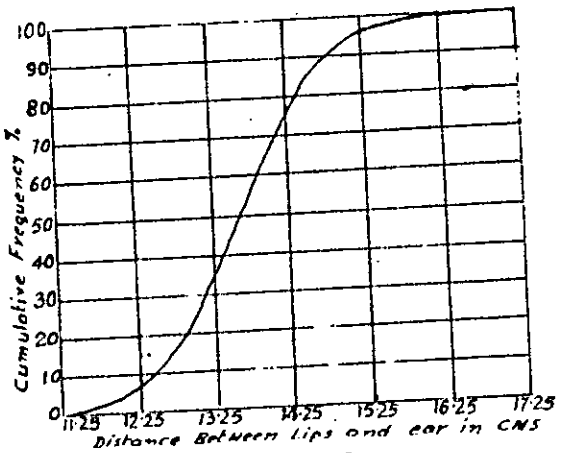


FIG. 3·10

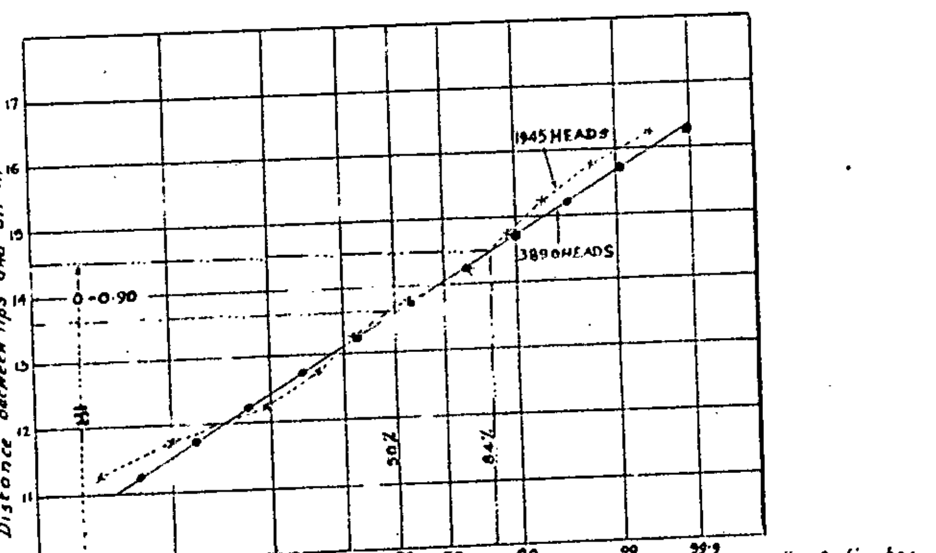


FIG. 3·11

Now a cumulative frequency curve is in effect an intergral curve, and this led Hazen to produce a scale such that the intergal of the normal bell-shaped probability curve would plot as a straight line. Consequently any series of observations which varied in accordance with the normal probability intigral would also plot as a straight line if such a scale was used on co-ordinate paper. In constructing this probability scale Hazen marked the centre of the scale 0.5. The other points on the scale are laid off in either direction from the centre point so that the pairs of values 0.4 and 0.6, 0.3 and 0.7, 0.1 and 0.9 etc. are respectively equidistant from the centre. The distances from the centre are made proportional to the corresponding values of area under the normal probability curve. This type of paper is known as "Arithmatic probability paper".

It can be seen that the dotted line drawn on such a probability paper from certain data as in fig 3.11 does not appear to be a good straight line. Hence it can be said that the sample does not appear to be quite stable. Now, if we draw a straight line joining some other data on the same or similar probability paper then the sample can be said to be stable. It can now be concluded that the value of m and S obtained from this sample closely agree with the real values that exists ~~on~~ in the parent universe from which the sample was drawn. On probability paper the standard deviation is approximately equal to the difference between the values of the ordinates at 84% line and the median value at 50% line. Thus, if the data plots as a straight line on the arithmatic probability paper the most probable value of the mean and the standard deviation can be approximately determined by direct inspection.

3.9.3 Small Samples

So far we have assumed that the number of observations are very large. The above is true even when the sample size is not very large but for very small samples, say of size 2 to ten, the above methods may not be very satisfactory.

The knowledge as to the nature of the distribution law obeyed by the mean and other statistics that are commonly calculated from measured data has been used in an elegant way in the costruction of control charts the purpose of which is to reveal at a glance whether or not the population from which a sample is drawn is of normal type. Provided the value of the parameter σ is known, the central chart can be made into an useful tool.

When samples are small it is not possible to draw exact conclusions as to the properties of the parent population from which they were drawn, but limits between which these properties must lie can be determined with desired degree of certainty. The determination of such limits is often of great value.

Out line of "Students" theory.

It has been stated that if samples of N items are drawn from a normal population, it will be found that the means of the samples are distributed around the mean of the population according to the normal probability law with a standard error of $\sigma/\sqrt{N}$ where σ is the standard deviation of the sampled population. To test any assumption concerning the mean of a random sample, as for example, that the mean $\bar{X}$ of the sample deviates by more than a specified amount from the mean m of the population, it is necessary to calculate,

$$t = \frac{\bar{X} - m}{\sigma/\sqrt{N}} \quad \text{-----------} (1)$$

which is the ratio of the deviation of the mean of the sample to the true standard error of the mean. Then the equation.

$$f = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} t} \quad \text{---------} (2)$$

with define the frequency distribution of the statistic, 't' if it is distributed normally. By calculating the area under this curve between -X and the ordinate at the value of 't' obtained from the sample, it would be possible to state the probability of obtaining a greater value of 't' and hence a greater deviation of $\bar{X}$ from m, if another sample were tested.

It is important to note, however, that the value of the standard deviation σ is usually unknown, and consequently it is necessary to estimate it from the N items in the sample. Use might be made of the standard deviation of the sample, namely,

$$S = \sqrt{(x - \bar{x})^2/N} \quad \text{--------} (3)$$

where the summation extends over all the observed $\bar{X}$ value in the sample. Prof. R.A.Fisher, however, has shown by the method of maximum likelihood, that the most probable estimate of σ is given by

$$S \times \sqrt{N/(N-1)} \quad \text{--------} (4)$$

$$S = \sqrt{\frac{\Sigma (x - \bar{x})^2}{N}} \quad \text{----------} (3)$$

Replacing σ by its estimated value, equation for 't' becomes

$$t = \frac{\bar{x} - m}{S/\sqrt{N-1}} \quad \text{-----------} (5)$$

where $\frac{S}{\sqrt{N-1}}$ is the estimated standard error. It will be observed that the quantity t is now the ratio of the deviation of the mean of a small sample to the estimated standard error of the mean. It was found by 'Student' that for small values of N, the ratio or statistic 't' given by (5) is not distributed normally like equation (1) but takes the form,

$$f = \frac{[(n-1)/2]!}{[(n-2)/2]!\sqrt{n\pi}} \cdot \left(1-\frac{t^2}{2}\right)^{-(n+1)/2} \quad \text{---} (6)$$

where $n = N-1$

Equation (6) is called the Student's distribution and although it is normal the probability of obtaining a greater or less value of t than that given by a particular sample can be obtained as before by integrating the curve between $-\infty$ and the ordinate at the value of t under consideration.

As N becomes larger, Student's distribution approaches a normal distribution. For $N>30$ it is permissible to use the tables of normal probability integrals calculated by Sheppard from equation (2) by replacing the independent variable by $t[(n-2)/n]^{\frac{1}{2}}$ and for $N > 100$ the quantity 't' itself may be used as the independent variable.

3.9.4 <u>Probability Chart.</u>

By integrating the frequency distribution expressed by the equation (6) between the limits $-\infty$ to t, R.A. Fisher and Student in 1925 tabulated the probability that the ratio (5) will be less than a specified value t for the various values of N. These tables, which are based on expansion of students' distribution, are bulky. It can be shown, however, that for Engineering application, it is possible to construct a nomograph from Fishers' solution and thereby dispose with the necessity for obtaining the original table. This probability chart is shown in fig. 3.12 and covers most of the practical values of t, N and P, likely to be met in Engineering practice. As a simple example of the use of this nomograph consider a small sample of 6 observations and suppose it is found by calculation that $t = 2.2$; then putting a straight edge on the chart between $N = 6$ and $t = 2.2$ the probability value of 0.96 can be read from the t scale. Thus the probability that random samples will give values of t less than 2.2 is 0.96 approximately. The probability is therefore (1-0.96) that random - sample of this size will have value of t greater than 2.2. and 2(1-0.96) that this value lies outside the range - 2.2 to + 2.2. In mathematical notation this may be written as

$$P (|t| > 2.4) = 0.08$$

which means that if a large number of samples of 6 were taken, it would be found that, on the average, about 8 samples per hundred had t values outside the range - 2.2 to + 2.2. Consequently it may be inferred that the deviations observed in the sample are not altogether due to chance and the assumption may be wrong. These considerations lead to the discussion of what are to be regarded as "significant probabilities".

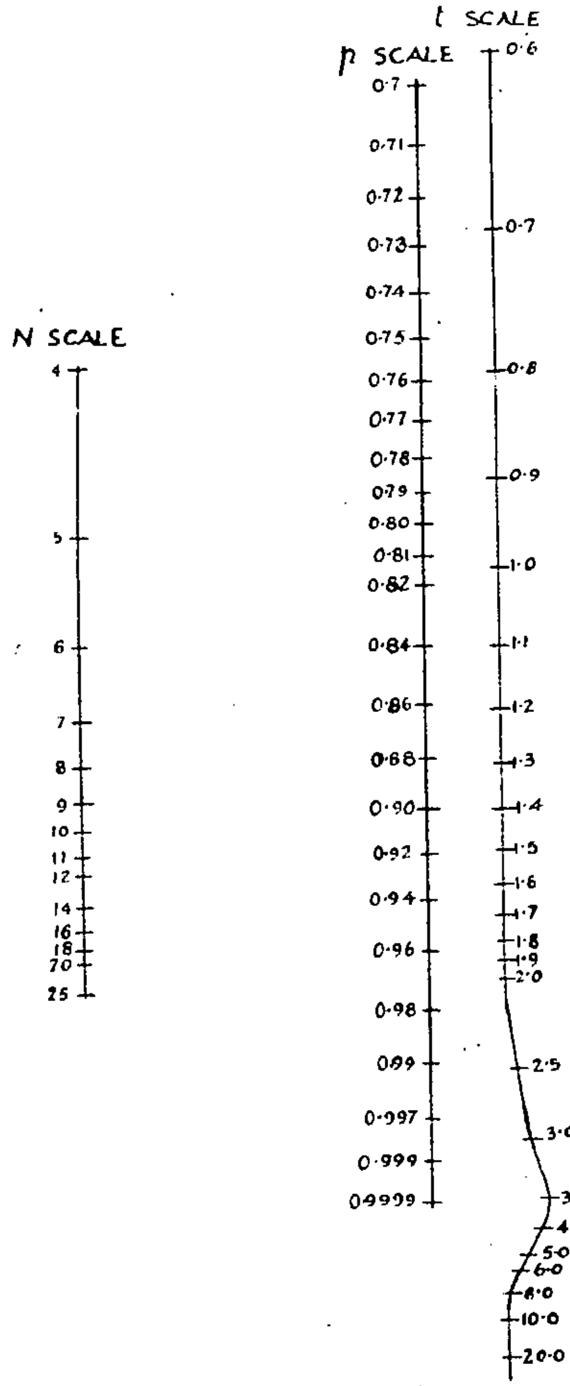


FIG. 3·12 - Probability Chart for 't' Test

3.9.5 <u>Significant probability Table</u>.

At this stage it will be interesting to see how 't' behaves in multiple sampling. As an example consider the distribution of t values obtained from 500 samples of 10 items. The frequency distribution of the t values obtained from these samples are shown by dots in fig.3.13 and the solid curve to fit these dots may be calculated from 'students' distribution. It will be found that the departure of the t distribution from the normal distribution is much too small to be noticed. Of the area under the curve only 5% lies beyond the ordinates errected at -a and + a and only 1 percent lies beyond the ordinates erected at +b and -b. For samples of 10 items the t value corresponding to $\pm$ a is $\pm$ 2.26, the t value corresponding to $\pm$ b is $\pm$ 3.25. It will be seen that only a few values of t are larger than 2.26 or smaller than -2.26, and very few values indeed are found to exceed the limit of $\pm$ 3.25. Nevertheless, these extreme values do occur at times in sampling, but the probability of their occurrence (usually written P($|t| >$ t) is very small. Thus, the significance of the mean $\bar{x}$, is determined by the value of P ($|t| >$ t), which is the probability that, if another sample of the same size be taken, the value of $|t|$ calculated from it will exceed the value calculated from the original sample. If this probability is greater than 5%, than the probability is regarded as 'not significant'. In other words, the deviations observed in the sample are such as

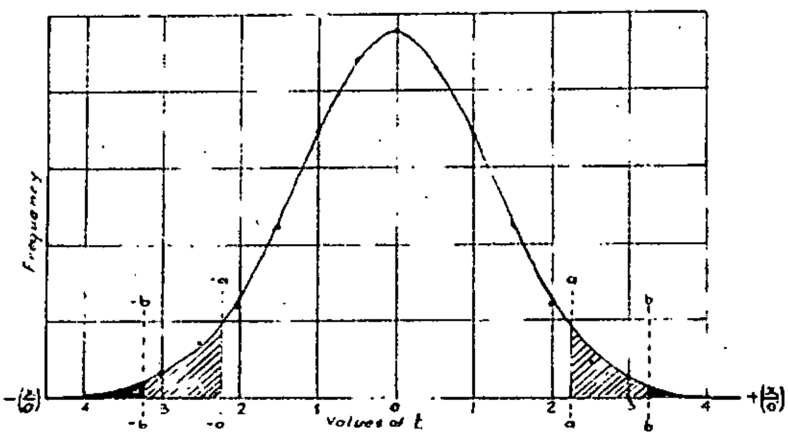


FIG. 3·13

Theoretical Frequency Curve Drawn to Fit Frequencies Obtained From 500 Values of "t" in Samples of 10.

may be ascribed to pure chance. If this probability lies between 1 percent and 5%, then the probability is regarded as 'significant'. In other words, the deviations observed may not be due to pure chance, and then the probability $P(|t| > t)$ is referred to as 'significant'. If the probability is much less than 1% the deviations observed are almost certainly not due to chance and than the probability $P(|t| > t)$ is referred as 'highly significant'. Hence the actual probability limits below which samples are regarded as significant are:

(i)	$P(\|t\| > t)$	≥ 0.05	not significant.
(2)	"	$\geq .01 < .05$	significant.
(3)	"	$< .01$	highly significant.

In fig:3.13 only one percent of the total area under the curve lies beyond theordinate drawn at b and -b and this small area is shaded black in the figure. The t value corresponding to $\pm$ b is 3.25 and is a value that will be exceeded, on an average, only once per 100 samples. If a sample of 10 items gave a t value greater than 3.25 the assumption about the value of the mean M would be abandoned. It will be noticed that the distribution of t changes with the number N of items in the sample. To meet this difficulty the 5 percent and 1 percent probability levels of (t) have been calculated for various values of N and are given in the table 1.

3.9.6 <u>Confidence limits</u>.

In the previous case of a sample of 10 a mean of 100 was assumed and t test showed that

TABLE I.

N	5%	1%	N	5%	1%	N	5%	1%
2	12.706	63.657	13	2.179	3.055	24	2.069	2.807
3	4.303	9.925	14	2.160	3.012	25	2.064	2.797
4	3.182	5.841	15	2.145	2.977	30	2.045	2.756
5	2.776	4.604	16	2.131	2.947	35	2.030	2.724
6	2.571	4.032	17	2.120	2.921	40	2.021	2.704
7	2.447	3.707	18	2.110	2.898	50	2.008	2.678
8	2.365	3.499	19	2.101	2.878	60	2.000	2.660
9	2.303	3.355	20	2.093	2.861	80	1.990	2.638
10	2.262	3.250	21	2.086	2.845	100	1.984	2.626
11	2.228	3.169	22	2.080	2.831	200	1.972	2.601
12	2.201	3.106	23	2.074	2.819	1000	1.962	2.581

the truth of the assumption was doubtful. Instead of making guesses about the true value of mean m and applying t test to each guess, it is better to determine, from equation (1) the limits of m required to satisfy given probability conditions. Suppose it is required to be 95% sure that the mean m of the population will lie within certain limits. For the 10 items in the sample, the t value corresponding to a probability of (100-95) per cent or 5 per cent is, from table above, found to be 2.262.

Suppose the $\bar{X}$ and S value of the sample are 95.2 and 6. If, then, m is taken to satisfy either of the relations,

$$\frac{95.2 - m}{\frac{6}{\sqrt{9}}} = 2.262 \text{ or } \frac{m - 95.2}{\frac{6}{\sqrt{9}}} = 2.262$$

Then the value of m would satisfy 95% condition of certainty. These limiting values are 90.676 and 99.724 respectively and consequently the inequality $90.68 \leq m \leq 99.72$ may be written. If m were taken as less than 90.676 or greater than 99.724, then 't', found from the sample values of $\bar{X}$ and S, would exceed 2.262 and the sample would be regarded as inconsistent at 5% level of significance. But whatever the true value of m may be, in repeated samples, simultaneous values of $\bar{X}$ and S would be obtained such that $|t| \leq 2.262$ in 95% of the cases. The values of 90.68 and 99.72 may be called the 95% confidence limits of m corresponding to 10 items in the sample and it may be concluded that,

m = (95.2 ± 4.5) with 95 percent confidence.

It is obvious, of course, that the 1% level of t may be used to give confidence limits at that level, and it is convenient to notice that the two equations giving the confidence limits can be written in the general form,

$$\frac{\bar{x} - m}{S/\sqrt{N-1}} = \pm t \quad \text{or} \quad m = \bar{x} \pm \frac{S}{\sqrt{N-1}} \cdot t$$

where $S/\sqrt{N-1}$ is the estimated standard error.

The two mean t test.

R.A.Fisher has shown that students' distribution can be applied to a variety of problems. Of these problems, one that appears quite frequently is the comparison of two samples obtained from different sources. The means may be different and it is then necessary to determine the probability that the difference is due to the samples having been drawn from different populations or the probability that the samples have been drawn from the same population and that the difference is due to chance. These probabilities may be determined as follows:-

Let $\bar{X}_1$ and $\bar{X}_2$ be the means of two samples of N1 and N2 items respectively and let S1 and S2 be their standard diviations. Then to test for significance, we calculate,

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\left(\frac{N_1 + N_2}{N_1 + N_2 - 2}\right)^{\frac{1}{2}} \left(\frac{N_1 S_1^2 + N_2 S_2^2}{N_1 N_2}\right)^{\frac{1}{2}}}$$

and $N = N_1 + N_2 - 1$

By means of the chart the probability that two other similar samples would give a value of 't' exceeding that given by the samples under consideration can be determined. The significance or otherwise of this probability, P ($|t| > t$), can be found from table 1

The following sample will illustrate the method.

Samples of rubber insulating material were received from two sources and to determine the amount of mineral matter in each a number of ash content determinations were made. The results were as shown in table below.

Percent ash content.

No.	Source A	Source B
1	24.3	18.2
2	20.8	16.9
3	23.7	20.2
4	21.3	16.7
5	17.4	-
Mean	21.5	18

Applying the above test,

for source A, $N_1 = 5$, $X_1 = 21.5$ and

$$S_1^2 = 6.004$$

for source B, $N_2 = 4$, $X_2 = 18$ and

$$S_2^2 = 1.945$$

Substituting these value in the expression for 't',

t

$$t = \frac{21.5 - 18.0}{\sqrt{\left(\frac{9}{7}\right)\frac{(30.02 + 7.78)}{5 \times 4}}} = 2.245$$

and $N = 5 + 4 - 1 = 8$
For $t = 2.245$ and $N = 8$. The probability chart gives P= 0.97 and thus the probability $P(|t| > 2.245)$ is given by $2(1-0.97) = 0.06$ thus there are about 6 chances in 100 of observing a greater difference in percentage ash content, and at the 5 percent level of significance we find from table 4. that for $N = 8$

$$\underline{2.245 < 2.365}$$

and therefore, no significant difference has been detected between the two sources on this count.

3.9.7 Discontinuous distributions.

Frequently a variable is not able to take all possible values, but is confined to a particular series of values such as the whole numbers. The normal distribution is the most important of the continuous distributions; but among discontinuous distribution the poisson series is of the 1st importance. If a variate can take the values 0, 1, 2, 3 ....X ...... and the relative frequencies with which the values occur are given by the series

$$e^{-m}\left(1, m, \frac{m^2}{2!}, \cdots \frac{m^x}{x!} \cdots\right)$$

then number is distributed in the poisson series. The total frequency is unity, since

$$e^{m} = 1 + m + \frac{m^2}{2!} + \frac{m^3}{3!} + \cdots$$

Whereas the normal curve has two unknown parameters, m and σ, the Poisson series has only one. This value may be estimated from a series of observations, by taking their mean, the mean being a statistic as appropriate to the poisson series as it is to the normal curve.. It may be shown theoritically that if the probability of an event is exceedingly small, but a sufficiently large number of independent cases are taken to obtain a number of occurrences, then this number will be distributed in a poisson series. For example the chance of a man being killed by horse kick on any one day is exceedingly small, but if an army corps of men are exposed to this risk for a year, often one or more will be killed in this way. The following data were obtained fromthe records of 10 army corps for 20 years, supplying 200 such observations.

Death	frequency observed	Expected.
0	109	108.67
1	65	66.29
2	22	20.22
3	3	4.11
4.	1	0.63
5	..	0.08
6	..	0.01

The average, $\bar{x}$, is 0.61 and taking this as an estimate of m the numbers calculated agree excellently with those observed.

The poisson distribution can be derived as an approximation to the binomial distribution for a certain range of values for N and P. This distribution can be shown to arise from a different axiomatic approach.

We consider a fixed unit of time, T (Perhaps a second, day or week), in which certain events may occur. We shall assume that event occur independently and that for short periods of time, Δt. The probability of one event occurring is proportional to the

length of time Δt , that is, is equal to C $\frac{\Delta t}{T}$,
where C is constant during the time period T. We shall assume that the probability of two or more events occuring in Δt is so small that it may be neglected. With these assuptions it can be shown, using tools of calculars, that

$$P_r = \frac{m^r}{r!} e^{-m} \qquad\qquad m = c\,\frac{t}{T}$$

If we are observing real events and conditions under which they are being observed seem to correspond to the above assumption i.e., the events are reasonably independent and the probability that an event occurs is proportional to the length of time interval, then the frequency of an observed event may follow the poisson distribution.

The poisson distribution for several values of mean m are shown in fig 3.14 .Smooth continuous curves have been drawn through the discrete points which represent the distribution merely to show more clearly the form of the distribution. It can be seen from figure 3.14 that the skewness of the poisson distribution is very large for distributions which have small values of m and is less in those with larger value of m. The poisson distribution tends to approach that of the normal law as m increases and would eventually become identical with the normal law if m become infinite. Since, however, m must alaways remain finite, the distribution can never reach exact equivalence to the normal distribution, although for large values of m the difference between the two distributions is extremely small.

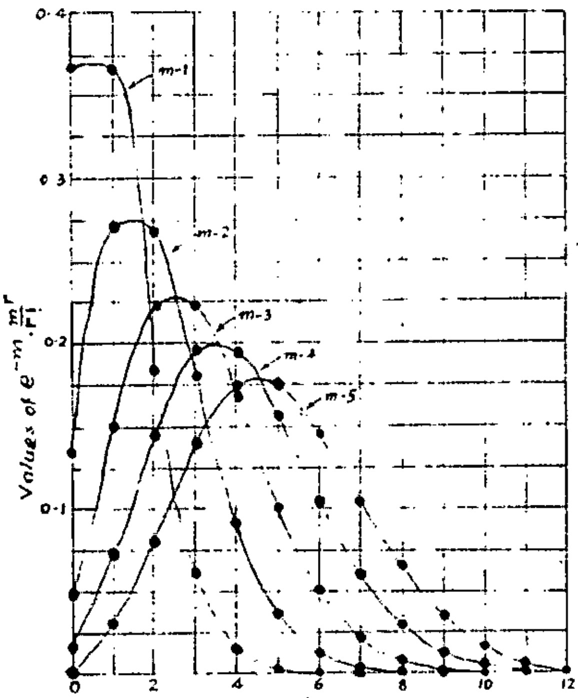


FIG. 3·14 Poisson Distribution For Various Values of the mean.

The characteristics of the distribution are simple in that, although skew, they are completely determined by one parameter alone, namely, the arithmatic mean m .

According to the poisson distribution the probabilities of occurrence of the values o, 1, 2 etc. are given by the series.

$$e^{-m}\left(1,\ m,\ \frac{m^2}{2!} \text{----} \frac{m^r}{r!}\right)$$

where m is the arithmatic mean. Thus according to the series the chance that there are exactly 4 occurrences of an event in a large group of trials in given by $m^4 e^{-4}/4!$. The chance that there will be at least 4 occurrences in the group will be given by the sum of the series starting at 4, namely

$$e^{-m}\left(\frac{m^4}{4!} + \frac{m^{4+1}}{(4+1)!} + \text{-----} \frac{m^r}{r!}\right)$$

In general the chance or probability Pr that an event will happen at least r times in a large group of trials, may be written as

$$e^{-m}\left(\frac{m^r}{r!} + \frac{m^{r+1}}{(r+1)!} + \text{----}\ etc.\right)$$

This expression can be written as

$$Pr = e^{-m} \sum_{n=r}^{n=\infty} \frac{m^n}{n!}$$

This series is known as poiss~~y~~on's exponential summation and is one of the fundmental expressions in modern probability theory. For purposes of calculation, however, it is possible to transform Pr into a more convenient form.

Noting that $e^{-m} \sum_{n=0}^{n=\infty} \frac{m^n}{n!}$ the sum of all possible chances, is equal to unity. Pr may be written in the finite form

$$Pr = 1 - e^{-m} \sum_{n=0}^{n=r-1} \frac{m^n}{n!}$$

which is convenient for numerical calculations. In engineering application of the poisson's law the summation form is generally used instead of the individual term form, because usually the chance of exceeding a certain limit, rather than the chance of obtaining any one particular value, is required.

It has been found that the rapid application of the Poisson summation to practical problems is greatly facilitated by use of what may described as Poisson probability paper. Such a table is shown in figure 3.15 . The curves of figure 3.15 give the relationship between m, the mean number of occurrences of an event in a large group of trails and the probability Pr that the actual number of occurrences in any such group of trials will equal or exceed any given number r. A linear scale is used for m, the curves indicate particular values of r starting at r = 1. To cover on a large scale a smaller range of m and r values, the probability chart shown in fig 3.15 has been produced which give the individual r curves in unit steps from m = o to m = 15. This probability chart and the chart at figure 3.16 were prepared by Dr.G.A.Campbell. These probability charts enable the Comparison of actual data with the corresponding poisson summation to be easily made graphically, thereby avoiding tedious calculations. The charts can be used as co-ordinate papers on which to plot any distribution of data, theoritical or observed, provided

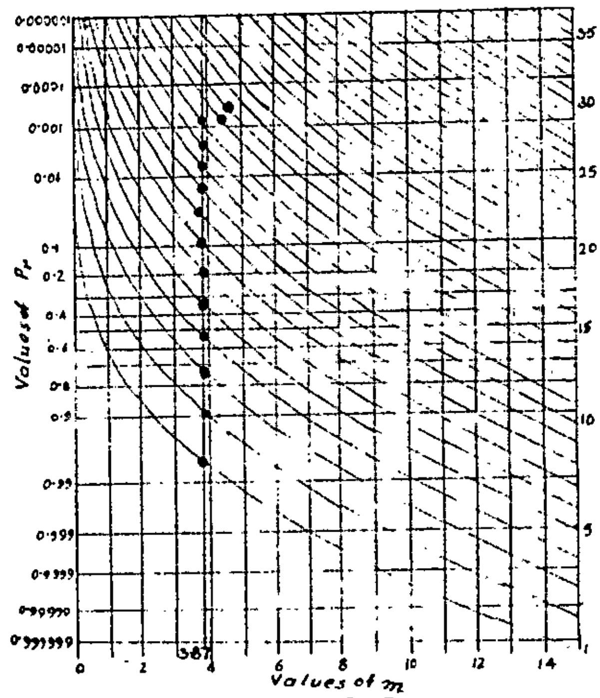


FIG. 3·15:

Comparison of an observed distribution of Relay Contact faults with its corresponding theoretical poisson distribution.

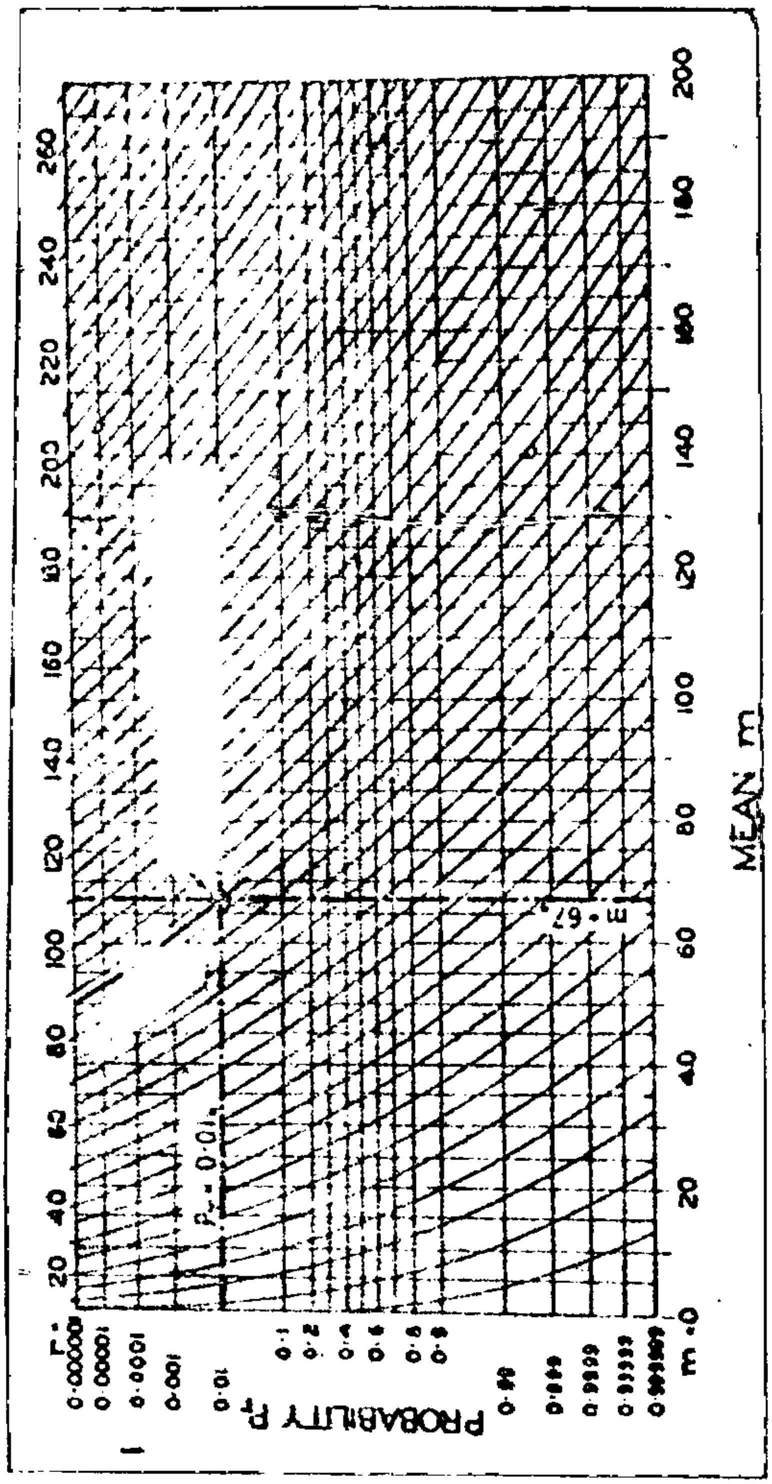


Fig 3·16

calculated frequencies fit the observed frequencies resonably closely.

For the purpose of testing the goodness of fit it is desirable to plot both the theoritical and observed frequencies on poisson probability paper. To use this chart as back ground for plotting the observed frequency distribution we begin by erecting a vertical line at point m = 3.87. The corresponding theoritical poisson distribution is represented by the point in which the vertical line cuts the r curves, or for convenience, simply by the vertical line itself as shown in figure 3.15. The observed distribution may be plotted with r and Pr as independent variable and the horizental deviations of these points from the vertical line serve as a measure of the discripancy between the two distributions. Since a goodness of fit test is to be made between a theoritical and an observed frequency distribution, the values of the values of the variable are limited to positive integers and zero.

The use of Poisson Probability paper.

In the vast majority of practical cases the Engineer will know whether or not he is dealing with a discontinuous variable. If observations are made of the occurrence of a discrete event which either definitely happens or fails to happen, it is likely that the frequency distribution will be of the poisson type. As an example consider the occurrence of faults on a particular type of relay contact in automatic exchanges as shown in the table below:

TABLE.

No. of contact faults per Exchange per week	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Observed frequency.	57	203	383	525	532	408	273	139	45	27	10	4	0	1	1 1
Calculated frequency	54	210	407	525	508	394	254	140	68	29	11	4	1	0	0

From the nature of the event under observation it may be assumed that the observed frequency distribution will follow a Poisson probability law. To test this assumption the theoritical frequencies are calculated and compared with the observed frequencies. The mean of the observations is 3.87 and using this value the theoritical frequencies can be calculated. From the table it can be seen that the of the probability Pr are the values of the observed relative frequencies and these x can be calculated from the observed data as shown in table 2.

It can be seen from figure 3.15 that there is a good fit between the observed dots and the vertical line and consequently it may be concluded that the sample examined indicates that the number of contact faults follows a poisson probability law.

A Poisson Significance test.

It is necessary to bear in mind several practical conditions which must work against any perfect agreement between an observed distribution and the corresponding poisson distribution.

In the first place, the sample considered will necessarily consist of a finite number N of items instead of an infinite number as assumed in the mathematical theory, and in the second place, the items considered may not be completely independent as assumed in the theory.

In a stable poisson sample the horizontal deviations from the vertical line erected at its mean value m, should be very small, whereas if the number of items n is too small to ensure stability, the plotted points will fall about a vertical line to the left or right of m, the distance being reduced

T A B L E II.

ARRANGEMENT OF DATA FOR A POISSON TEST:

Number of contact faults per exchange per week.	Observed frequency of occurrence.	Number with at least the values of 0,1,2,...etc.	Relative frequency (Pr)
0	57	2608	1.0000
1	203	2551	.978
2	383	2348	.900
3	525	1955	.753
4	532	1440	.552
5	408	908	.348
6	273	500	.192
7	139	227	.087
8	45	88	.034
9	27	43	.0165
10	10	16	.0061
11	4	6	.0023
12	0	2	.00077
13	1	2	.00077
14	1	1	.00038
Sum of frequencies	=2608		

as the number of items in the sample is increased. If, however, the mean of the population is unstable or varying in any manner, the plotted points will slope to the right as shown in fig 3.17 Curve 3. If a random sample has been drawn from a stable population, but not a poisson population, its distribution of dots will be away from the poisson vertical line and sloping to the left.

3.9.8 The Binominal probability law.

If the probability of occurrence of an event is p, andthe probability of failure is q, then if a random sample of n trials are taken, the frequency with which the event occurs, o, 1,2........ n times are given by the expansion of the binominal-$(q+p)^n$

The binominal distribution law is one of the fundamentals of probability theory established by Bernoulli in the 17th century.

When P=q and n≥10 the binomial distribution reduces to normal distribution. If $n > 1000$ and $p \neq q$, but neither is very small, the binomial distribution again reduces to the normal form. A special case arises when $p \rightarrow 0$ and $n \rightarrow \infty$ in such a way that the product np remains finite and equal to the mean m, then binomial distribution reduces to poisson exponential series.

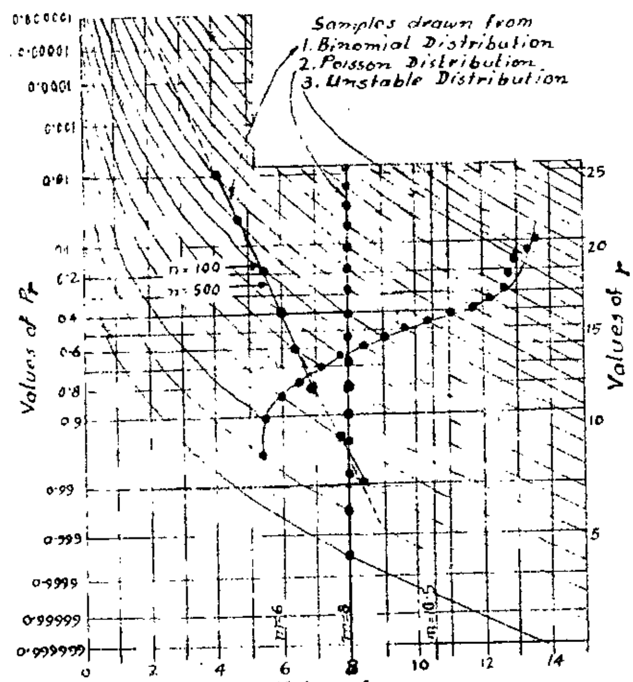


FIG. 3·17
A Poisson test for stability of mean

3.9.9 The X^2 Test

By the poisson test of stability we have shown whether certain observed data obey poisson distribution or other distribution.

There is a powerful method, dependent upon the statistic known as X^2 (Ki squared) distribution which can be used as a means of testing the agreement between observation and hypothesis.

The element common to these tests is the Comparison of the numbers actually observed to fall into any number of classes with the numbers which upon some hypothesis are expected. If the entire range of variables is divided into r classes and fo is the observed frequency in a particular class, and fE the frequency which could be expected to fall in that class on some hypothesis, then X^2 may be found by dividing the square of the difference between fo and fE by fE and summing these quotients for all of the r classes into which the variable has been divided. This in symbols,

$$X^2 = \sum \frac{(fo-fE)^2}{fE}$$

When X^2 has been calculated consultation of tables will indicate the probability P of the calculated value being exceeded as a result of the random sampling. If this probability is less than 0.05, then X^2 may be regarded as showing that the observed data depart significantly from the hypothesis which is being examined. Table 3 is an extract of the X^2 table for P = 0.05.

TABLE III.

n	X^2	n	X^2	n	X^2
1	3.841	11	19.675	21	32.671
2	5.991	12	21.026	22	33.924
3	7.815	13	22.362	23	35.172
4	9.488	14	23.685	24	36.415
5	11.070	15	24.996	25	37.652
6	12.592	16	26.296	26	38.885
7	14.067	17	27.587	27	40.113
8	15.507	18	28.869	28	41.337
9	16.919	19	30.144	29	42.557
10	18.307	20	31.410	30	43.773

T24

In order to utilise the table it is necessary to know also the value of n with which the table is to be entered. The rule for finding n is that n is equal to the number degrees of freedom in which the observed series may differ from the hypothetical; in other words, it is equal to the number of classes the frequencies in which may be filled up arbitrarily, without altering the expectation. As an example an analysis of the occurence of contract faults of a particular type of relay will be made. It is required to examine the hypothesis that the observed frequency distribution of faults form a poisson series. Table 4 contains the data.

To determine the hypothetical or expected frequencies fE on the assumption that the distribution of faults form a poisson series, we calculate the mean m and make the number N agree with the observed. The mean is 3.892; and since N, the total number of observations, is 398, it is, a simple matter to calculate the poisson frequencies recorded in the 3rd row of the table. Before X^2 can be calculated it will be necessary to divide the range of the variable into groups. Inspection of the table shows that 10 groups can be conveniently taken as 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 and over, the last six frequencies being combined into one group because it is an empirical rule that a class having a theoritical frequency less than 5 should never be used. The theoritical frequency for the last group is 7.28, this being the sum of the six frequencies classed together. Details of the arithmatic is shown in the

TABLE IV.

No.of contact faults per Exchange per week.	Observed frequency fo.	Hypthetical frequency (Calculated) fe.	Fo-fe	$(Fo-fe)^2/fe$
0	14	8.12	5.88	4.26
1	37	31.61	5.39	0.92
2	76	61.51	14.49	3.41
3	70	79.80	-9.80	1.20
4	64	77.64	-13.64	2.40
5	53	60.43	-7.43	0.91
6	31	39.20	-8.20	1.72
7	19	21.80	-2.80	0.36
8	14	10.60	3.40	1.09
9	9	4.59		
10	5	1.78		
11	5	0.63	12.72	22.23
12	1	0.20		
13	0	0.06		
14	0	0.02		

table. The sum of the 5th row is 38.50, which is the value of X^2. To get the required probability, table of X^2 is consulted; but before this is done it must be carefully noted that although 10 groups have been used to calculate X^2, the number of degrees of freedom n is only 8, since both n and m are fixed in the formation of the poisson series. From the table it is found for 8 degrees of freedom that X^2 is 15.507. The calculated value is much greater than this; hence the theoritical poisson series is a bad fit to the absorved fault distribution. From more detailed table it can be found that for this value of X^2 the probability P is less than 0.001; consequently it may be concluded that the observed fault distribution certainly does not follow poisson law.

If P is between 0.1 and 0.9 there is no reason to suspect the hypothesis tested. If it is below 0.02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. There will be a very small chance of error if values of P lower that 0.05 be taken to indicate a real discrepancy. There is, however, a large chance of error if very large values of P ($P > 0.95$) are obtained. This rather surprising result arises because a large value of P normally correspond to small value of X^2 that is to say a very close agreement between theory and fact. Now such agreements are rare; almost as rare as great divergencies; very close agreement between theory and fact is too good to be true. X^2 test can be applied to a number of problems in telecommunication.

3.10 Bivariate distributions:

In applying arithmatic probability paper to certain sets of observations it may be found that the points do not fall upon straight lines, but in smooth cures. Attempts have, therefore, been made to modify the paper by changing the linear scale in such a way that these special cases would plot as a straight line. In statistical theory the logarithmic transformation is of unique importance.

In many practical applications of logarithmic probability paper two distinct variables, each of which is subject to sampling flunctuations, are concerned.

It is relatively simple matter to fit straight lines to observations plotted on logarithmic probability paper, but this gives no direct information about the stability of the relationship between the variables or the magnitude of the sampling fluctuations, and has no practical value until the stability of the relationship between the two variables has been established. If, however, it can be shown by probability tests that the variables are calculated and that the plotted data are statistically stable, then the slope of the straight lines on logarithmic probability paper can be used to form a valueable statistical index for the forecasting of future trends.

3.11 The Coefficient of Correlation.

There are many instances in which an increase in one variable is accompanied by an increase in another. In other words, large values of one variable tend to be associated with large values of the second, whereas small value of the first tend to be associated with small values of the second, without the existence of a strict mathematical relationship between the two. The variables are then said to be correlated. Of course an increase in one variable may be accompanied by a decrease in the other, that is, large values of each variable may tend to be associated with small values of the other. This situation is known as inverse or negative correlations.

The most widely used measure of correlation for N pairs of values of the variables x and y is defined by the equation.

$$r = \sum \frac{(x-\bar{x})(y-\bar{y})}{N\,\sigma_x\,\sigma_y} \quad \text{---------- (1)}$$

In this equation $\bar{X}$ and $\bar{Y}$ are the means of the variable and σx and σy represent their standard deviations. Since the various values of the variables will always be considered in pairs, an X and a Y, therefore will always be the same number N of X's and Y's.

If there is a complete positive correlation between X and Y, the coefficient of correlation r has unity value; if it has complete negative correlation it has - 1 value. Complete correlation is, however, very rare; most practical cases give incomplete correlation with value of r between +1 and -1. If there is no correlation at all between the variables, r is zero.

For dealing with small samples equation (1) may be transformed into the form

$$r = \frac{\sum \frac{xy}{N} - (\bar{x}-\bar{y})}{\sigma_x\,\sigma_y} \quad \text{---------- (2)}$$

which is a more convenient form for dealing with small samples.

3.12 Test for significance of Correlation Coefficient.

The value of correlation coefficient will very rarely be found to be unity or Zero. It is, therefore, necessary to apply a test to decide whether the correlation coefficient differ significantly from zero or in other words, whether correlation actually exists. It must always be remembered that Correlation coefficients, like other statistics, are subject to fluctuations of the sampling. It can be shown that, when N is large, i.e., $N > 50$, $\sigma(r)$, the standard error of r takes the form

$$\sigma(r) = \frac{1-r^2}{\sqrt{N}} \quad \text{- - - - - - - - -} (3)$$

If $r > 2\ \sigma(r)$ the correlation coefficient is regarded as differing significantly from zero. If however, N is less than 50 equation (3) should not be used to calculate (r); the significance of a correlation coefficient could then be assessed by the t method. This method involves the calculation of t, where

$$t = \frac{r\sqrt{N-2}}{\sqrt{N-r^2}} \quad \text{- - - - - - - - -} (4)$$

The simplest way is to use Table 5 which has been prepared to enable the significance of correlation coefficients obtained by mere inspection. Table 5 is used by noting that if a calculated value of r is as big or bigger than the value given in the table for the appropriate value of n, then the correlation differs significantly from zero; in other words it indicates a real degree of correlation between the two variables.

TABLE V.

N	r	N	r
3	0.997	16	0.497
4	0.950	17	0.482
5	0.878	18	0.468
6	0.811	19	0.456
7	0.755	20	0.444
8	0.707	21	0.433
9	0.666	22	0.423
10	0.632	27	0.381
11	0.602	32	0.349
12	0.576	37	0.325
13	0.553	42	0.304
14	0.532	47	0.288
15	0.514	52	0.273

3.13 Regression line.

At this stage we may consider a bivariate distribution occurring in telecommunication. Suppose that 50 measurements were taken on busy hours for number of operators present X, effective calls passed Y. These measurements provide 50 paired numbers (x , y) which may be plotted as points in a plane. The resulting collection of points is called a scatter or dot diagram, an example of which is shown in fig ~~below~~. 3.18

There is often a tendency for values of X to conform in some way to the corresponding values of Y; and such tendency towards a functional relationship, although obscured by random fluctuations, will manifest itself in the scatter diagram by the greater density of dots along a certain line. This line is not sharply defined but its estimation is important to engineers, for it is a smudged image of a line which may be fine and clear cutin the parent population of which the observations are only a random sample. This latent curve or mathematical relationship Y=F(x) is called a regression or the regression of Y on X; likewise, the inverse form X=F (y) is called the regression of X on Y.

Owing to errors of random sampling, the observed regression line which arise are usually irregular, but it is usually possible to fit straight lines to them and to show mathematically that the observed regression lines do not depart significantly from the

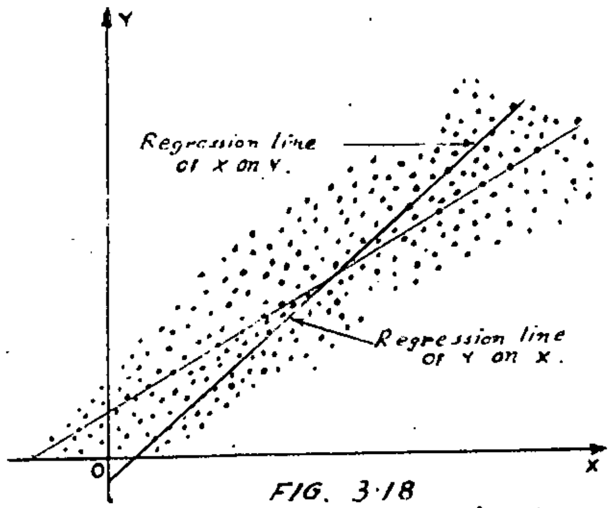


FIG. 3·18
Example of a "Scatter" or "Dot" Diagram.

fitted regression lines. It can be shown that these straight lines can be represented by simple equations

$$(y-\bar{y}) = r\frac{\sigma_y}{\sigma_x}(x-\bar{x}) \quad \text{- - - - - - - - - (5)}$$

$$(x-\bar{x}) = r\frac{\sigma_x}{\sigma_y}(y-\bar{y}) \quad \text{- - - - - - - (6)}$$

where the letters have same meaning as before. Equation (5) is the regression straight line of Y on X and the equation (6) is the regression straight line of X on Y. As before r is the coefficient of correlation between the variables x and y.

If r=1, the two regression lines are identical and the regression straight lines coincide. If r=0, the two lines are horizontal and vertical respectively and cut at right angles. The lines cross for any intermediate value of r, so that the larger the value of r, the smaller is the acute angle between them. The two lines always cross at points $\bar{X}$, $\bar{Y}$ on the graph. The angles these lines make to the horizontal and vertical axes respectively are measured by the expressions.

$$r\frac{\sigma_y}{\sigma_x} \quad \text{and} \quad r\frac{\sigma_x}{\sigma_y}$$

and these are called the coefficients of regression.

These equations can be used in practice to express one variable in terms of the other. The variables considered may be, for example, the power factor of electrolytic condensers and temperature, or the mutual conductance and anode current under given conditions in thermionic valves, or any other pairs of correlated variables arising in telecommunication. In all these examples it is necessary to calculate the means $\bar{X}$ and $\bar{Y}$ and their standard deviations σx and σy. Then the coefficient of correlation r can be calculated, and if this differs significantly from zero, equation (5) and (6) can be used. As an example, suppose that in an electrical investigation a large number of paired values of two variables x and y were measured and it was found that

$$\bar{X} = -0.21, \quad \bar{Y} = 0.81$$
$$\sigma x = 1.818; \quad \sigma y = 2.161$$

and σr = 0.666 was found to be more than twice its standard error (i.e., significantly ~~differently~~ different from zero).

Then, substituting these values in equation (5) and (6) there results -

$$Y - 0.81 = 0.7990 \quad (X+0.21)$$
$$\text{and} \quad X + 0.21 = 0.5551 \quad (Y-0.81)$$

which can be put into the form

$$Y = 0.7990x + 0.9778$$
$$\text{and } X = 0.5551Y - 0.6596$$

This simple example shows that it is only a matter of straight forward mathematics to estimate one variable in terms of the other and to plot the regression lines across the scatter diagram.

CHAPTER IV.

Collection of Data.

In the beginning of any Statistical work it is required to collect a large amount of raw material or data from which the relevant informations are extracted. Thus the telecommunication traffic investigator may have to count the busy hour traffic for several days or even several years. The method or care given to the collection of this raw material is important. The strength of a chain lies in the strength of the weakest link and it is useless to reach intricate conclusion from insufficient or inacurate data. Before any form of elaborate analysis it is essential to know the limitations and accuracy of statistical material and to be aware of the kind of errors that can arise.

4.1 : Most common sources of data are:

1) Questionaires - The data are obtained by forms designed by the statistician and completed by General public.
2) Observations - The data are collected by the investigator himself recording ~~to~~ the results of a series of observations but not necessarily relying on the general public for his information.

4.1.1 Questionaires and their completions is not a common practice in this Country as it is in Europe, America, Japan etc. These forms, if completed properly, can provide raw materials for a large number of studies.

Because of literacy factor and because of no standard practice in the country correct compilation of the data, which involve general public, is not possible. However, if such questionaire are confined to a community such as the telephone users in a city, we may expect correct result and be sure of its reliability for most practical purposes.

Clearly the efficiency of such collection of material depends largely on how these questions are framed, for bad questions can produce wrong informations. This applies to all enquiries made by questionnaire, and in the design of any statistical form there are certain rules which must be followed if resonable results are to be obtained. Briefly these rules are :-

(a) The form should be as concise as possible and there should be the minimum number of questions necessary to obtain the required information.

(b) The questions should be simple, and unambiguous in their possible interpretation.

(c) Questions which are likely to arouse strong feelings and hence attract inaccurate answers should be avoided. For example, asking a man whether he has any physical deformity is quite likely to produce an incorrect answer.

(d) The form should be made as attractive as possible to the eye by means of a suitable layout and clear type.

(e) When asking confidential information in a voluntary enquiry the persons' name should not be put on the form unless it is essential. This precaution is likely to produce accurate replies to the confidential questions.

From the above it is clear that the designing of any form entails a great deal of thoughts and that it is essential to imagine before hand the kind of answer that each question will provoke and whether that answer is the one that is required. It is, therefore, a good plan to try out a questionnaire on a few people before having a large number printed. A question that may seem simple and clear cut to the authors can quite easily prove a stumbling - block to users of the form. If a trained enumerator is to aid people to fill up the form then the questions asked can be a little more involved than if the survey is to be carried out by post, but it must be remembered that the public is less likely to disclose confidential personal information if the survey is carried out through an enumerator than through an impersonal medium.

4.1.2 A large amount of statistical data is obtained not by questionnaire but by the investigator himself going out into the field and counting or collecting items. This is the source described earlier as observations. Whatever the type of data it is essential to adopt a system of collection that is both logical and tidy, so that the data will still be understood at some later date. Results should, therefore, always be recorded in a note book since odd scraps of paper are easily lost. The heading of each page must give details of the raw material contained on that pages together with the date and place of collection. It is very tempting to say at the time of collection that this information is so obvious that it need not be recorded, but if the material is not used for some time such an omission may be regretted, especially if a large amount of other material has been used or collected meanwhile. Record should be made with a sharp pencil, as ink is very liable to smudge and become unreadable, especially in outdoor field work. If the data are required for future reference it should be converted to ink later.

The recording of the data in a note-book is quite straight forward if a systematic method is followed. Let us suppose a series of Barometer readings is being made every quarter of an hour at three different levels in a tall telecommunication building. Three columns are needed. The first for the time, the second for the level, and the third for the barometric height. Any one looking at the sheet at a future data would be able to understand what had been done and the results would leave no room for ambiguities. No set of raw materials should be accepted unless it contains all the relevant information.

When the nature of the survey has been agreed the observers can be briefed and then sent out for the actual collection of the data. At frequent intervals the results obtained to date is collected and inspected to make sure that they conform to the requirement and no unforeseen snags have arisen. At the end of certain period say one week all the data can be collected ready for analysis. At this stage it will consist of a large number of pages in observer's note books, and the next stage in the investigation is to reduce this large volume of figures to a few managable ones to bring out the particular points of interest.

4.2 From the mass of data it is usually difficult for any one to pick out the salient features of the data without missing important points. It is therefore necessary to condense the raw figures into some more managable form which will enable the investigator to pick out at once those features that are intended to be pursued further.

Tables are made in an orderly manner. A glance at the table gives a clear picture. Some times all the figures in the table are reduced to percentage of the total. It is much easier and conveys a more vivid impression to say 52.3% rather than to say 412 out of 788. In such cases the total number of observation on which the percentage is based must be given.

The information collected by the observer or available to him has now been reduced from long series of figures in a note book into a few well chosen and appropriate tables. Sometimes the tables themselves are sufficient to prove the point under discussion and no further analysis is necessary. But in many cases the tabulation has merely put the raw data in a form suitable for further comparisons.

Frequently the tables produced contain so much information that it is still impossible to disentangle the required bits and pieces of information. In these circumstances a further series of smaller tables are needed. Alternatively it may be possible to prepare some form of diagram, which many people will find easier to understand than a table of figures, to present in a clear and forceful manner, the important aspects of the information.

The second method i.e. the method of observation was adopted in collecting the raw materials. In the case of obtaining telephone and Telegraph traffic figures from old records a lot of difficulty was encountered. Most of the informations were recorded after sorting out relevant files from old records. No systematic records are available and it took several days only to find the relevant bits of papers from which to extract the informations.

In case of recording the information about the local calls originated by telephone subscribers a recording machine was used. This machine recorded the number of attempted calls, the number of calls matured and the number of ineffective calls. Maximum holding time, average holding time were obtained from the readings of the same machine (Frontis piece)

DA-Ctg. direct dialling readings were obtained from meters.

C H A P T E R - V.

R E S U L T S :

5.1 TABLE VI.

CHARACTERISTIC OF CERTAIN RANDOMLY SELECTED TELEPHONE SUBSCRIBERS AT DACCA.

Sl. No.	Date	Subscriber's category.	Subscriber.	No. of Calls attempted.	No. of calls ineffective.	Max. holding time.	Average holding time.
1	2	3	4	5	6	7	8
1.	11.7.65	Business	A	50	12		
	12.7.65	"		30	16		
	13.7.65	"		18	8	18^m-55^s	1^m-58^s
	14.7.65	"		86	46		
	15.7.65	"		46	22		
2.	16.8.65	"	B	4	-		
	17.8.65	"		49	30		
	18.8.65	"		22	8		
	19.8.65	"		12	8		
	20.8.65	"		18	5	7^m-25^s	1^m-56^s
	21.8.65	"		17	7		
	22.8.65	"		11	1		
	23.8.65	"		13	5		
3.	17.5.65	"	C	43	20		
	18.5.65	"		83	49		
	19.5.65	"		49	29	11^m-5^s	2^m-22^s
	20.5.65	"		79	50		
	21.5.65	"		91	58		
4.	24.6.65	"	D	1	-		
	25.6.65	"		5	4		
	26.6.65	"		2	1	2^m-45^s	1^m-16^s
	27.6.65	"		6	2		
	28.6.65	"		3	2		
5.	24.5.65	"	E	15	9		
	25.5.65	"		28	10	3^m-35^s	0^m-46^s
	26.5.65	"		15	10		
	27.5.65	"		23	11		
1.	21.9.64	Doctor	A	13	5		
	22.9.64	"		41	26		
	23.9.64	"		51	34	18^m-50^s	2^m-53^s
	24.9.64	"		23	9		
	25.9.64	"		14	7		
2.	28.11.64	"	B	3	1		
	29.9.64	"		29	23	4^m-15^s	1^m-13^s
	30.11.64	"		23	14		

T A B L E VI. (Contd). ~~Page.2~~.

Sl. No.	Date	Subscriber's category.	Subscriber.	No. of calls attempted.	No. of calls ineffective.	Max. holding time.	Average holding time.
1	2	3	4	5	6	7	8
3.	17.12.64	Doctor	C	11	5		
	18.12.64	"		6	1	3^m-50^s	1^m-19^s
	19.12.64	"		15	13		
4.	11.12.64	"	D	10	7		
	12.12.64	"		20	15	3^m-30^s	1^m-42^s
	13.12.64	"		26	9		
5.	30.11.64	"	E	14	13		
	1.12.64	"		15	11	11^m-35^s	3^m-13^s
	3.12.64	"		7	5		
6.	7.9.64	"	F	4	1		
	8.9.64	"		13	8		
	9.9.64	"		23	15		
	10.9.64	"		18	10		
	11.9.64	"		32	21	58^m-20^s	4^m-31^s
	12.9.64	"		10	4		
	13.9.64	"		10	5		
	14.9.64	"		28	17		
	15.9.64	"		33	24		
	16.9.64	"		16	-		
	17.9.64	"		20	7		
	18.9.64	"		14	6		
	19.9.64	"		22	20		
	20.9.64	"		9	3		
1.	1.9.65	Advocates		14	11		
	3.9.65	"		2	1		
	4.9.65	"		8	4		
	5.9.65	"		2	2		
2.	23.9.65	"		5	2		
	24.9.65	"		28	22		
	25.9.65	"		5	1		
	26.9.65	"		8	2	13^m-30^s	3^m-8^s
	27.9.65	"		14	9		
	28.9.65	"		11	5		
	29.9.65	"		8	4		
	30.9.65	"		18	12		
1.	24.1.65	Profession	D	20	6		
	25.1.65	"		51	28		
	26.1.65	"		45	21	14^m-35^s	1^m-5^s
	27.1.65	"		42	30		
	28.1.65	"		17	13		
	29.1.65	"		12	10		

T A B L E VI (Contd).

Sl. No.	Date	Subscriber's category.	Subscriber.	No. of calls attempted.	No. of calls ineffective.	Max. holding time.	Average holding time.
1.	29.9.64	Judge.	A	3	1		
	30.9.64	"		8	4		
	1.10.64	"		14	6	9^m-0^s	1^m-58^s
	2.10.64	"		18	5		
	3.10.64	"		19	12		
	4.10.64	"		6	-		
	5.10.64	"		1	-		
1.	30.7.64		C	10	5		
	31.7.64			43	34		
	1.8.64			34	26		
	2.8.64			20	23		
	3.8.64			24	11		
	4.8.64			13	8		
	5.8.64			8	3		
	6.8.64			19	10		
	7.8.64			36	22		
	8.8.64			3	2		
	9.8.64			16	11		
1.	16.4.64		B	6	2		
	17.4.64			5	-	9^m-15^s	2^m-2^s
	18.4.64			3	-		
	19.4.64			10	6		
	20.4.64			25	16		
2.	15.5.64	PBX Line		26	23	3^m-20^s	1^m-5^s
	16.5.64	"		6	1		
	17.5.64	"		13	2		

T A B L E VII

MONTHLY INTERWING AND DACCA-LONDON TELEGRAPH TRAFFIC & TECHNICAL LOSS.

Month	INTERWING		DACCA-LONDON	
	Total Traffic.	Tech. Loss.	Total Traffic.	Tech. Loss.
April/63	48,512	250.35	13,492	140.30
May/63	48,310	86.55	13,426	61.10
June/63	51,888	138.10	13,150	39.25
July/63	49,853	201.52	13,727	60.15
Augt/63	51,251	188.39	14,522	125.36
Sept/63	48,971	302.35	13,863	191.55
Oct/63	50,612	214.10	16,988	298.00
Nov/63	52,482	409.20	17,687	372.00
Dec./63	54,746	422.40	16,706	372.10
Jany/64	49,721	355.30	16,935	509.20
Feby/64	51,760	409.15	16,638	311.35
March/64	57,562	354.45	17,132	179.55
April/64	54,769	402.55	16,885	293-35
May/64	55,250	451.18	16,243	226.20

T A B L E VII. (Contd) ~~Page :2~~

Month	INTERWING		DACCA-LONDON	
	Total Traffic.	Tech. Loss.	Total Traffic.	Tech. Loss.
June/64	59,553	189.20	16,536	252.30
July/64	53,244	334.50	17,470	212.25
Aug/64	55,475	438.00	17,021	177.00
Sept/64	54,688	452.40	16,959	249.55
Oct/64	54,221	358.00	17,263	268.10
Nov/64	52,282	396.52	16,064	357.10
Dec./64	57,679	333.30	16,863	277.30
Jany/65	54,011	280.50	16,212	420.50
Feby/65	50,241	282.27	15,165	357.10
March/65	66,058	474.40	18,784	448.30
April/65	67,696	253.30	16,625	226.50
May/65	73,341	477.35	16,213	313.55
June/65	69,607	334.10	15,451	194.45

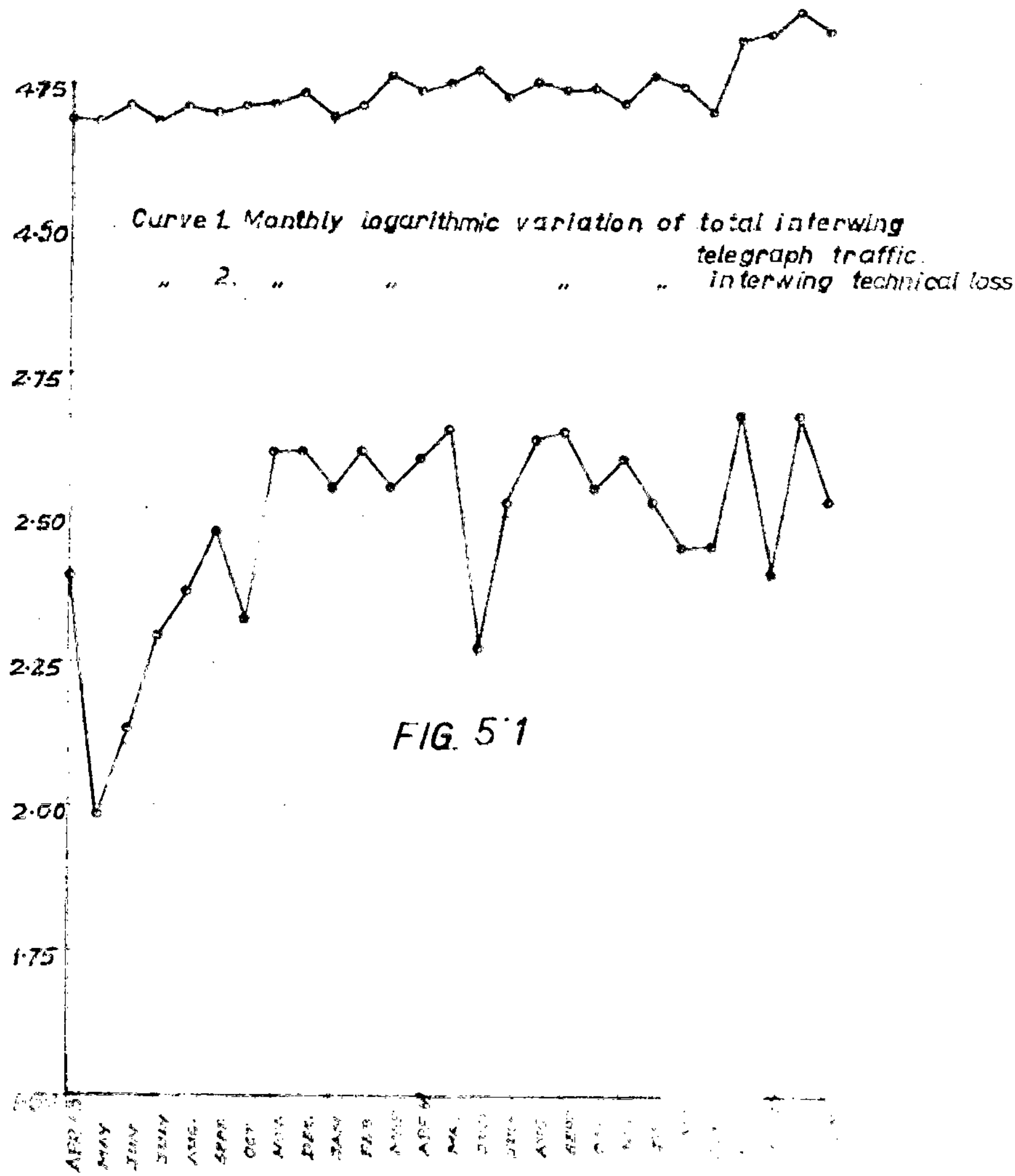

4·75
4·50
2·75
2·50
2·25
2·00
1·75
Curve 1. Monthly logarithmic variation of total interwing telegraph traffic.
„ 2. „ „ „ „ interwing technical loss.
MAY
JUN
JULY
AUG.
SEPT
OCT
NOV
DEC
JAN
FEB

FIG. 5·1

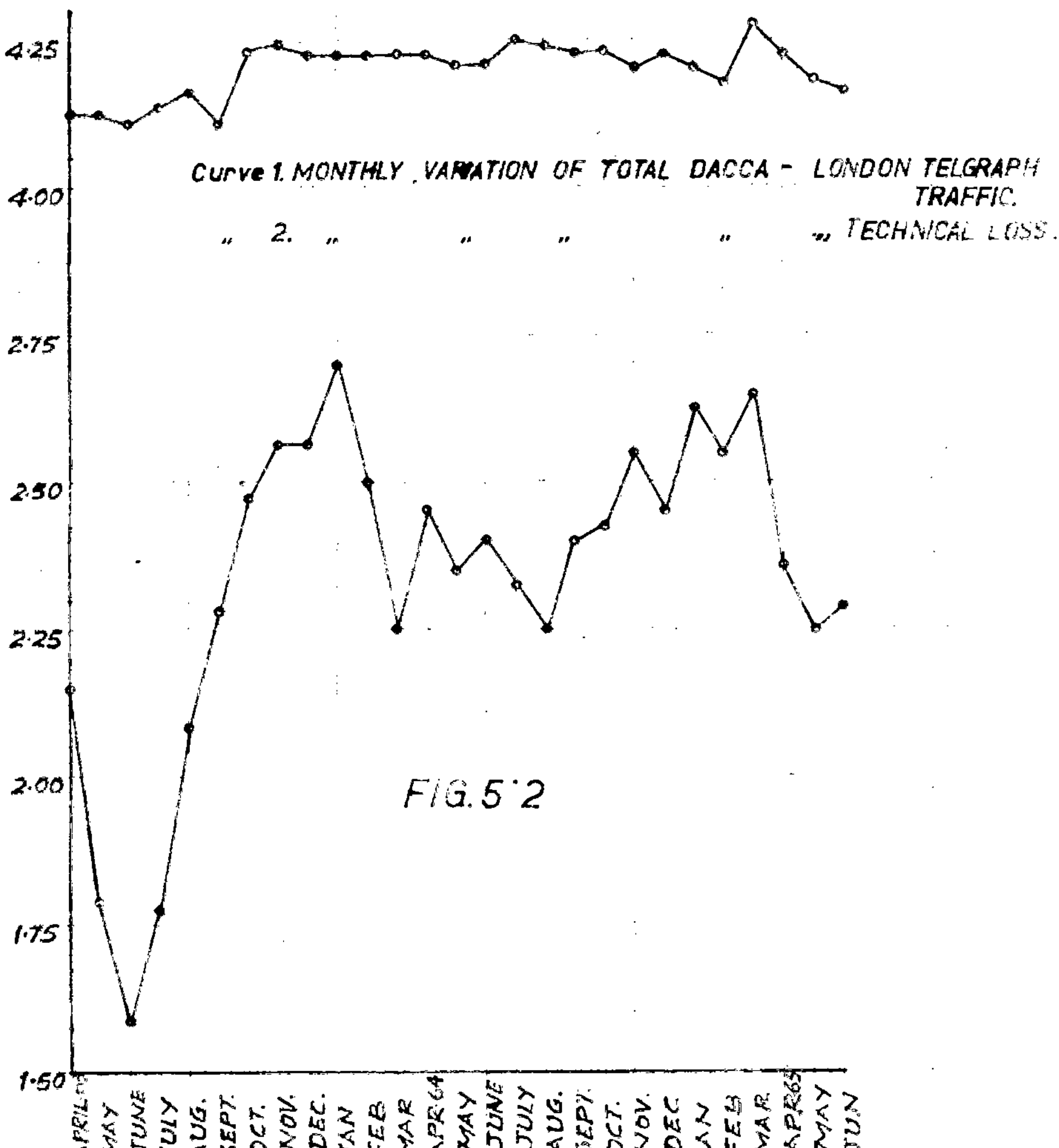

4·25
4·00
2·75
2·50
2·25
2·00
1·75
1·50
Curve 1. MONTHLY VARIATION OF TOTAL DACCA - LONDON TELGRAPH TRAFFIC.
" 2. " " " " " TECHNICAL LOSS.
APRIL
MAY
JUNE
JULY
AUG.
SEPT.
OCT.
NOV.
DEC.
JAN
FEB
MAR
APR64
MAY
JUNE
JULY
AUG.
SEPT.
OCT.
NOV.
DEC
JAN
FEB
MAR
APR65
MAY
JUN

FIG. 5·2

T A B L E VIII.

HOURLY VARIATION OF BOOKED CALLS AND EFFECTIVE CALLS ON DACCA-KARACHI CIRCUIT(PERIOD OF OBSERVATION 22.11.65 TO 27.11.65).

Sl. No.	Hours.	Calls booked.	Calls effective.
1	0-6	278	151
2	6-7	16	8
3	7-8	72	31
4	8-9	134	83
5	9-10	193	102
6	10-11	186	110
7	11-12	214	151
8	12-13	145	121
9	13-14	171	116
10	14-15	106	86
11	15-16	119	78
12	16-17	105	86
13	17-18	117	84
14	18-19	114	62
15	19-20	112	46
16	20-21	99	54
17	21-22	111	40
18	22-23	56	9
19	23-24	23	23

T A B L E IX-A.

HOURLY VARIATION OF EFFECTIVE MINUTES AND THE PERCENTAGE OF THE CIRCUIT TIME AVAILABLE BETWEEN DACCA AND KARACHI.

Hours.	Date. 22.11.65.	Effective minutes.	Percentage of the circuits time available.
0-6	Monday	27	16.66
6-7	"	3	4.166
7-8	"	15	45.83
8-9	"	63	58.33
9-10	"	60	83.33
10-11	"	! 87	80.55
11-12	"	93	90.277
12-13	"	108	97.22
13-14	"	72	51.388
14-15	"	72	65.27
15-16	"	66	79.166
16-17	"	51	70.83
17-18	"	63	66.66
18-19	"	87	87.5
19-20	"	60	64.58
20-21	"	39	29.16
21-22	"	27	4.166
22-23	"	24	18.75
23-24	"	24	29.166

T A B L E IX-B.

HOURLY VARIATION OF EFFECTIVE MINUTES AND THE PERCENTAGE OF THE CIRCUIT TIME AVAILABLE BETWEEN DACCA AND KARACHI.

Hours	Date 24.11.65.	Effective minutes.	Percentage of the circuits time available.
0-6	Wednesday.	132	45.83
6-7	"	3	8.33
7-8	"	21	58.33
8-9	"	57	50.00
9-10	"	87	61.1
10-11	"	102	59.72
11-12	"	111	79.16
12-13	"	96	79.16
13-14	"	51	36.11
14-15	"	105	66.666
15-16	"	60	100.00
16-17	"	90	89.58
17-18	"	60	64.58
18-19	"	24	33.33
19-20	"	9	10.41
20-21	"	30	45.83
21-22	"	9	18.75
22 -23	"	0	0
23-24	"	27	37.5

T A B L E IX-C.

HOURLY VARIATION OF EFFECTIVE MINUTES AND THE PERCENTAGE OF THE CIRCUIT TIME AVAILABLE BETWEEN DACCA AND KARACHI.

Hours.	Date. 26.11.66	Effective minutes.	Percentage of the circuits time available.
0-6	Friday	186	41.66
6-7	"	0	0
7-8	"	12	29.16
8-9	"	39	45.83
9-10	"	117	68.05
10-11	"	111	80.55
11-12	"	120	97.22
12-13	"	75	86.11
13-14	"	69	76.38
14-15	"	27	70.83
15-16	"	51	95.83
16-17	"	72	85.416
17-18	"	72	50
18-19	"	45	58.33
19-20	"	33	64.58
20-21	"	81	72.916
21-22	"	15	8.333
22-23	"	0	0
23-24	"	30	41.66

5.2 TABLE X.

COEFFICIENT OF VARIATION OF CERTAIN RANDOMLY SELECTED TELEPHONE SUBSCRIBERS AT DACCA.

Sl. No.	Category.	Standard deviation of the total no. of calls.	Standard deviation of ineffective calls.	Coefficient of variation of total calls.	Coefficient of variation of ineffective calls.
1.	Business	23.05	15.42	0.50108	0.6451
2.	"	12.66	8.77	0.69369	1.0962
3.	"	19.27	14.27	0.27927	0.34466
4.	"	1.93	1.33	0.5676	0.7388
5.	"	5.54	.707	0.27358	0.0707
1.	Profession	6 .66	3.96	0.67566	0.99
2.	"	7.93	6.01	0.8045	1.252
3.	"	12.03	10.12	0.58554	0.76847
4.	"	15.24	9	0.48899	0.5
1.	Doctors.	3.56	3.41	0.2966	0.3541
2.	"	15.1	5.22	0.5317	0.322
3.	"	11.1	4.03	0.60556	0.7132
4.	"	3.68	4.99	0.3452	0.788
5.	"	6.59	3.4	0.3531	0.3291
6.	"	8.52	7.61	0.4733	0.7557
1.	Advocate.	4.97	3.94	0.7645	0.985
2.	"	7.3	6.54	0.60206	0.91789

T A B L E XI.

CORRELATION BETWEEN MONTHLY TOTAL INTERWING TELEGRAPH TRAFFIC AND TECHNICAL LOSS.

Sl. No.	Total Traffic.	Tech. Loss.	Average of X $\bar{X}$	Average of Y $\bar{Y}$	Corrrelation co-efficient.
1.	48.512×10^3	250.35			
2.	48.310 "	86.55			
3.	51.888 "	158.10			
4.	49.853 "	201.52			
5.	51.251 "	188.59			
6.	48.971 "	302.35			
7.	50.612 "	214.10			
8.	52.482 "	409.20			
9.	54.746 "	422.40			
10.	49.712 "	355.30			
11.	51.760 "	409.15			
12.	57.562 "	354.45			
13.	54.769 "	402.55			
14.	55.250 "	451.18			
15.	59.553 "	189.20			
16.	53.244 "	334.50			
17.	55.475 "	438.00			

T A B L E XI. (Contd). Page.2.

Sl. No.	Total Traffic.	Tech. Loss.	Average of x $\bar{X}$	Average of y $\bar{Y}$	corre-lation co-eff-icie-nt.
18.	54.688×10^3	452.40			
19.	54.221 "	358.00			
20.	52.282 "	396.52	55.32×10^3	325.53	.3644
21.	57.679 "	333.30			
22.	54.011 "	280.50			
23.	50.241 "	282.27			
24.	66.058 "	474.40			
25.	67.696 "	253.30			
26.	73.341 "	477.35			
27.	69.607 "	334.10			

T A B L E XII.

CORRELATION BETWEEN MONTHLY TOTAL DACCA-LONDON TELEGRAPH TRAFFIC AND TECHNICAL LOSS.

Sl. No.	Total Traffic.	Tech. Loss.	Average of x. $\bar{X}$	Average of y $\bar{Y}$	Corre-lation co-eff-icient.
1.	$13,492x10^3$	140.30			
2.	13.426x "	61.10			
3.	13,150x "	39.25			
4.	13,727 "	60.15			
5.	14.522 "	125.36			
6.	13.863 "	191.55			
7.	16.986 "	298.00			
8.	17.687 "	372.00			
9.	16.706 "	372.10			
10.	16.935 "	509.20			
11.	16.638 "	311.35			
12.	17.132 "	179.55			
13.	16.885 "	293.35			
14.	16.243 "	226.20			
15.	16.536 "	252.30			
16.	17.470 "	212.25			

TABLE XII.(Contd) Page.2

Sl. No.	Total Traffic.	Tech. Loss.	Average of x. $\bar{X}$	Average of y. $\bar{Y}$	Correlation co-efficient.
17.	17.021×10^3	177.00			
18.	16.859 "	249.55			
19.	17.263 "	268.10			
20.	16.064 "	357.10			
21.	16.863 "	277.30			
22.	16.212 "	420.50	16.075×10^3	111453.27	0.68887
23.	15.165 "	357.10			
24.	18.787 "	448.30			
25.	16.625 "	226.50			
26.	16.213 "	313.55			
27.	15.451 "	194.45			

T AB L E XIII.

AUTO CORRELATION OF (INTERWING) TECHNICAL LOSS.

Sl. No.	Delay.	Auto correlation.	% of the maximum value.
1	1	110817.10	0.826733
2	2	93414.69	0.69690
3	3	92345.85	0.688930
4	4	117284.94	0.87498
5	5	121430.49	0.905911
6	6	93698.57	0.6990
7	7	77781.22	0.58027
8	8	134042.28	1.00
9	9	180251.02	0.807588
10	10	119604.77	0.89229
11	11	94570.88	0.705529
12	12	101573.42	0.75777

Auto correlation of interwing technical loss.
FIG. 5·3

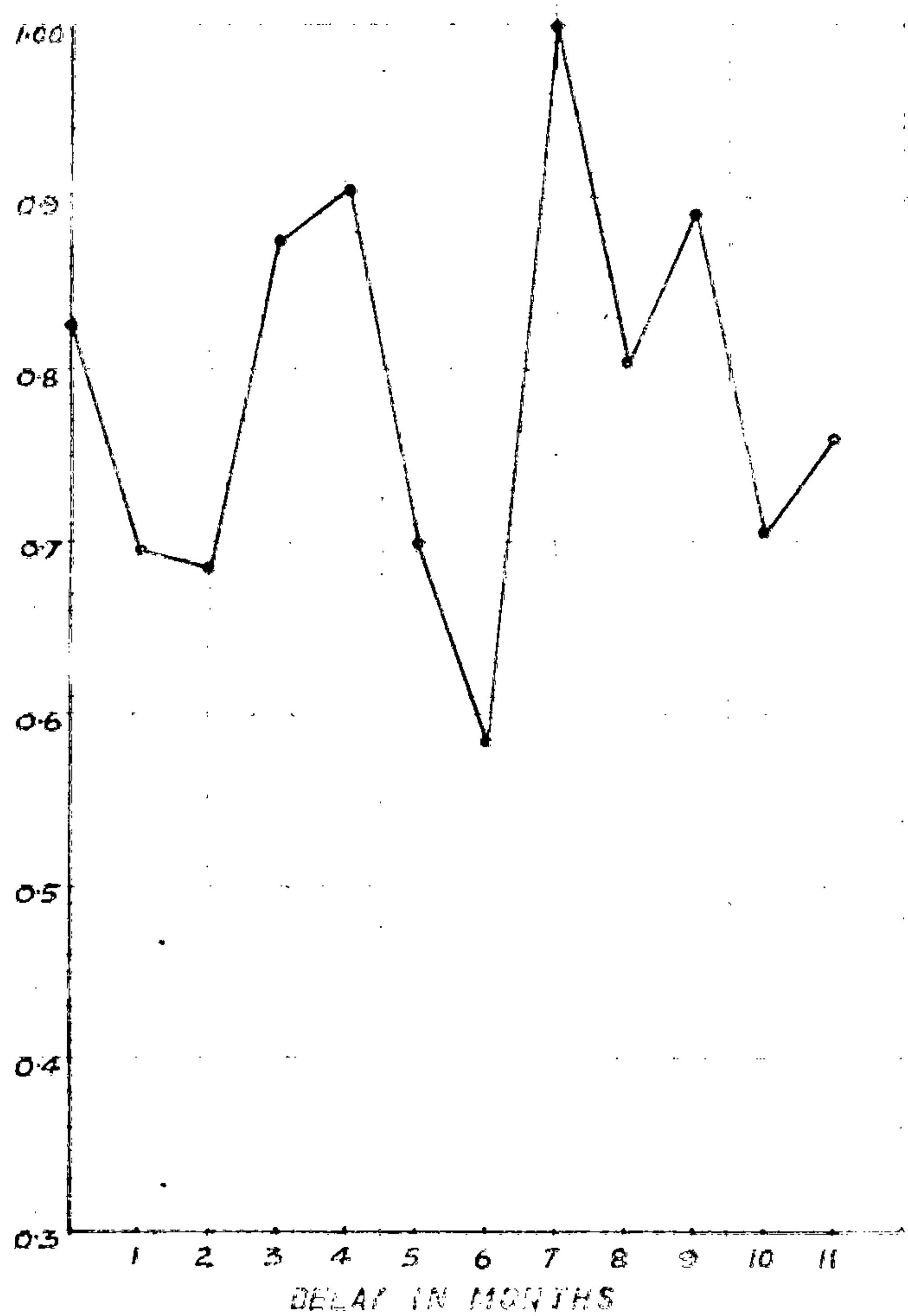


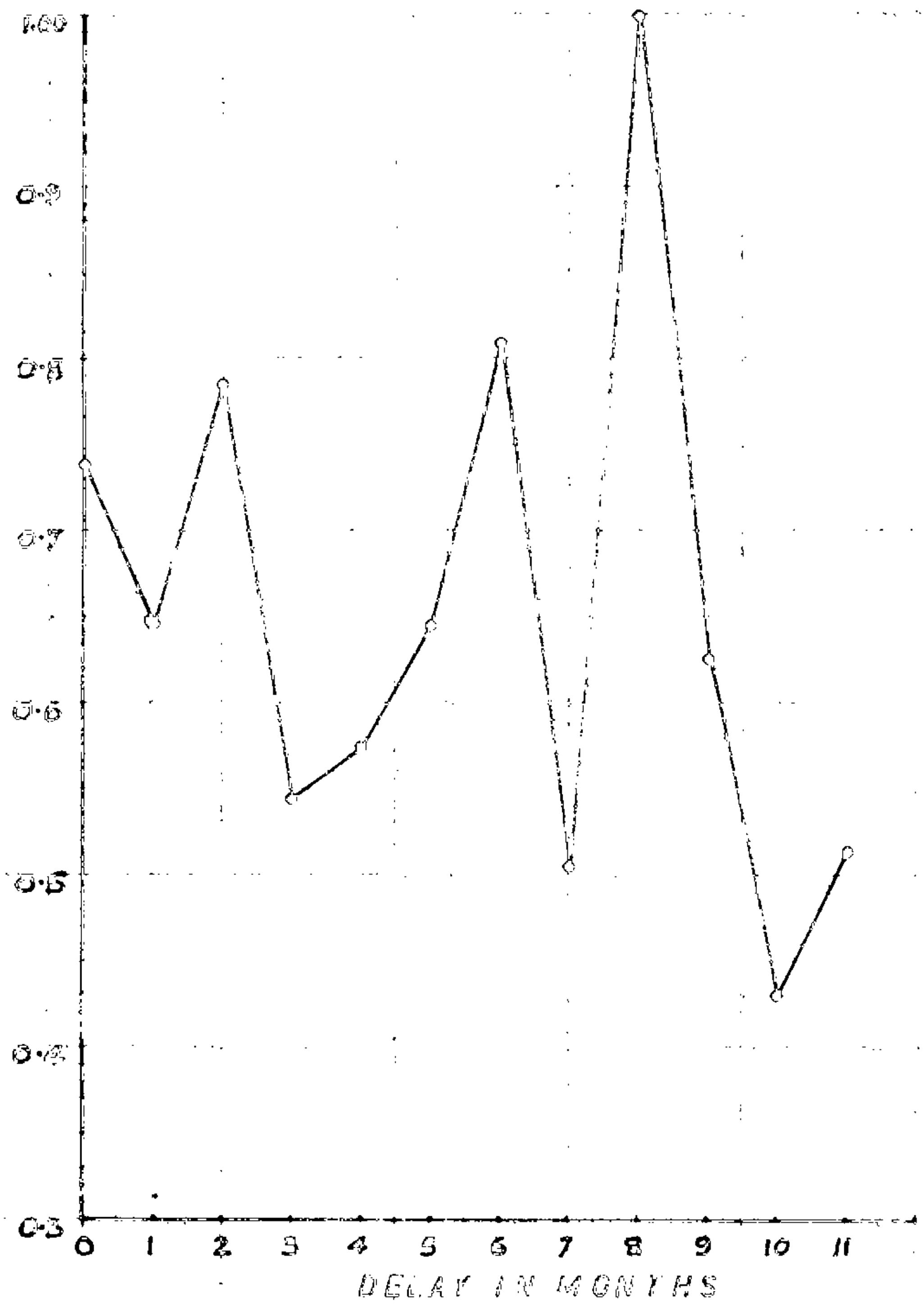
TABLE XIV.

AUTO CORRELATION OF DACCA-LONDON TECHNICAL LOSS.

Sl. No.	Delay.	Auto correlation.	% of the maximum value.
1	1	76832.88	0.738929
2	2	67144.83	0.648756
3	3	818282.76	0.786971
4	4	56200.58	0.54050
5	5	59680.38	0.573967
6	6	67149.87	0.645804
7	7	84046.45	0.808304
8	8	52719.27	0.50702
9	9	103978.64	1.00
10	10	65009.87	0.625223
11	11	44654.08	0.429454
12	12	53800.38	0.517417

AUTO CORRELATION OF DACCA LONDON TECHNICAL LOSS.

FIG. 5'4



T A B L E XV.

MONTHLY LOGARITHMIC VARIATION OF TOTAL INTERWING AND DACCA-LONDON TELEGRAPH TRAFFIC AND TECHNICAL LOSS.

Month.	Logof total traffic.	Log of Technical Loss.	Log of total traffic.	Log of Technical loss.
April/63.	4.68583	2.4002	4.1390	2.1470
May/63	4.6840	1.9373	4.1279	1.5938
June/63	4.6840	2,1402	4.1189	1.5938
July/63	4.6975	2.3021	4.1375	1.7793
Aug/63	4.7098	2.2750	4.1620	2.0881
Sept/63	4.6899	2.4805	4.1417	2.2822
Oct./63	4.7043	2.3306	4.2302	2.4742
Nov/63	4.7200	2.6118	4.2476	2.5705
Dec./63	4.7384	2.6257	4.2229	2.5706
Jany/64	4.6966	2.5506	4.2288	2.7070
Feby/64	4.7140	2.6118	4.2209	2.4933
March/64	4.7602	2.5495	4.2337	2.2545
April/64	4.7386	2,6047	4.2275	2.4674
May/64	4.7423	2.6544	4.2106	2.3545
June/64	4.7750	2.2770	4.2184	2.4019
July/64	4.7263	2.5243	4.2421	2.3268
Aug./64	4.7440	2.6415	4.2309	2.2480

TABLE XV. (Contd) Page.2.

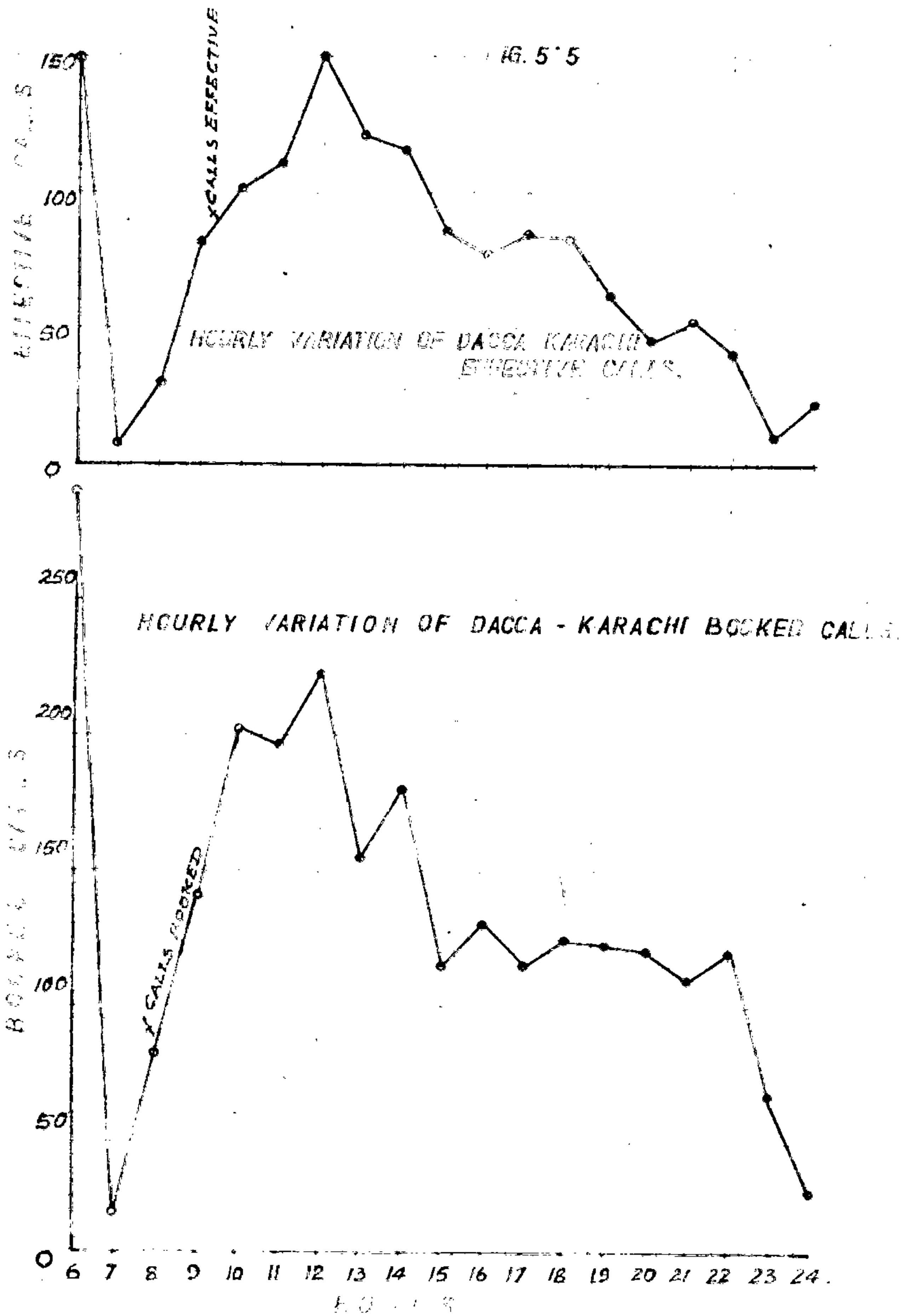
Month.	Log of total traffic.	Log of Technical loss.	Log of total traffic.	Log of technical loss.
Sept/64	4.7380	2.6555	4.2295	2.3972
Oct/64	4.7342	2.5539	4.2372	2.4283
Nov./64	4.7185	2.5982	4.2057	2.5528
Dec./64	4.7610	2.5228	4.2269	2.4430
Jany/65	4.7325	2.4480	4.2098	2.6237
Feby./65	4.7010	2.4505	4.1808	2.5528
March/65	4.8200	2.6762	4.2737	2.6517
April/65	4.8305	2.4036	4.2207	2.3551
May/65	4.8655	2.6789	4.2098	2.4936
June/65	4.8427	2.5238	4.1890	2.2888

T A B L E XVI.

HOURLY CORRELATION BETWEEN BOOKED AND EFFECTIVE CALLS ON DACCA-KARACHI CIRCUIT.FOR A WEEK.

Hour.	Calls booked.	Calls effec -tive.	Average of x. $\bar{x}$	Average of y. $\bar{y}$	corre-lation co-eff -cient.
0-6	278	151			
6-7	16	8			
7-8	72	31			
8-9	134	83			
9-10	193	102			
10-11	186	110			
11-12	214	151			
12-13	145	121			
13-14	171	116			
14-15	106	86			
15-16	119	78	124.789	75.842	0.91361413
16-17	105	86			
17-18	117	84			
18-19	114-	62			
19-20	112	46			
20-21	99	54			
21-22	111	40			
22-23	56	9			
23-24	23	23			

FIG. 5·5



150
100
50
0
EFFECTIVE CALLS
X CALLS EFFECTIVE
HOURLY VARIATION OF DACCA KARACHI EFFECTIVE CALLS.
250
200
150
100
50
0
BOOKED CALLS
X CALLS BOOKED
HOURLY VARIATION OF DACCA - KARACHI BOOKED CALLS.
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
HOURS

TABLE XVII-A.

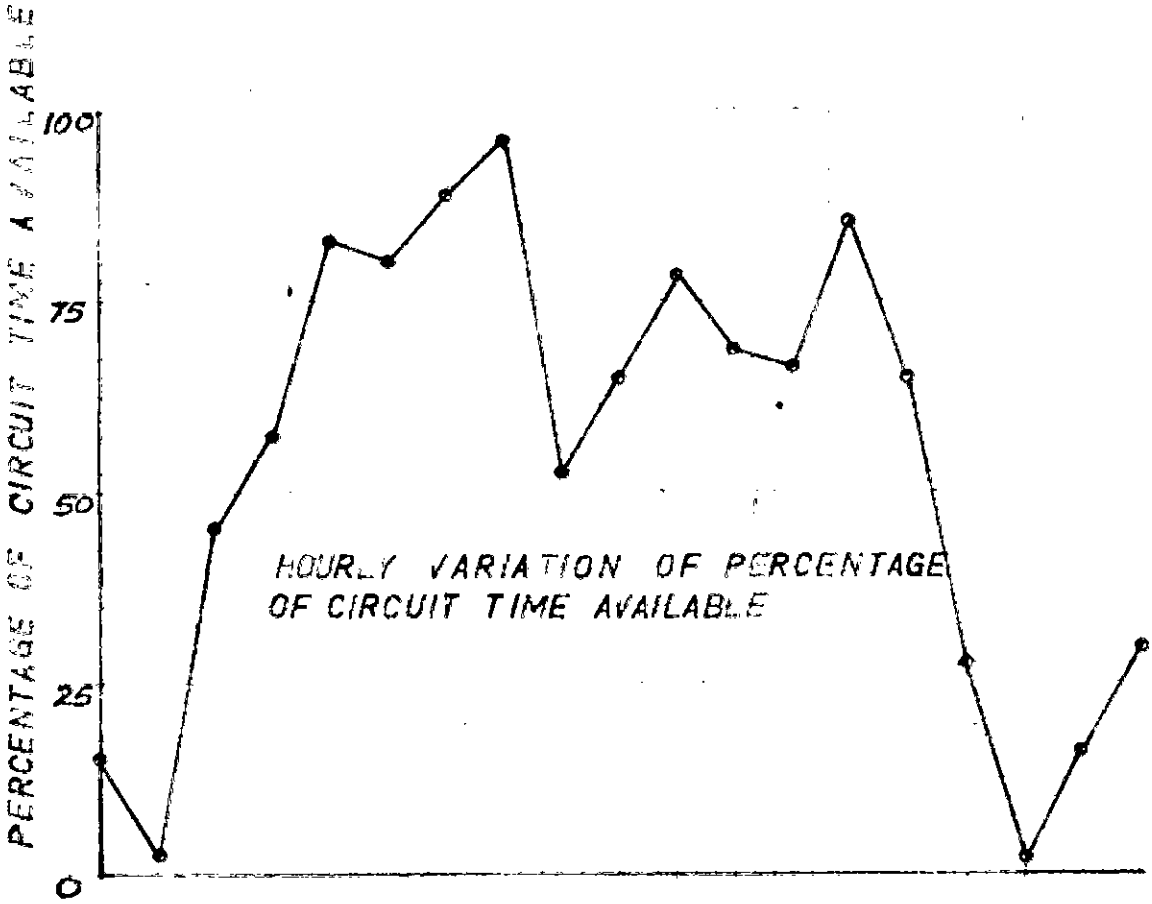
CORRELATION BETWEEN EFFECTIVE MINUTES AND PERCENTAGE OF CIRCUIT TIME AVAIL-ABLE ON DACCA-KARACHI CIRCUIT.

Monday:

Hour.	Calls booked.	calls effec-tive.	Average of x. $\bar{x}$	Average of y. $\bar{y}$	corre-lation co-effi-cient.
0-6	27	16.66			
6-7	3	4.166			
7-8	15	45.83			
8-9	63	58.33			
9-10	60	83-33			
10-11	87	80-55			
11-12	93	90-277			
12-13	108	97.22			
13-14	72	51.388	54.789	54.921	0.8933
14-15	72	65.27			
15-16	66	79.166			
16-17	51	70.83			
17-18	63	66.66			
18-19	87	87.50			
19-20	60	64.58			
20-21	39	29.16			
21-22	27	4.166			
22-23	24	18.75			
23-24	24	29.166			

MONDAY. - 22.11.65.

FIG. 5·6



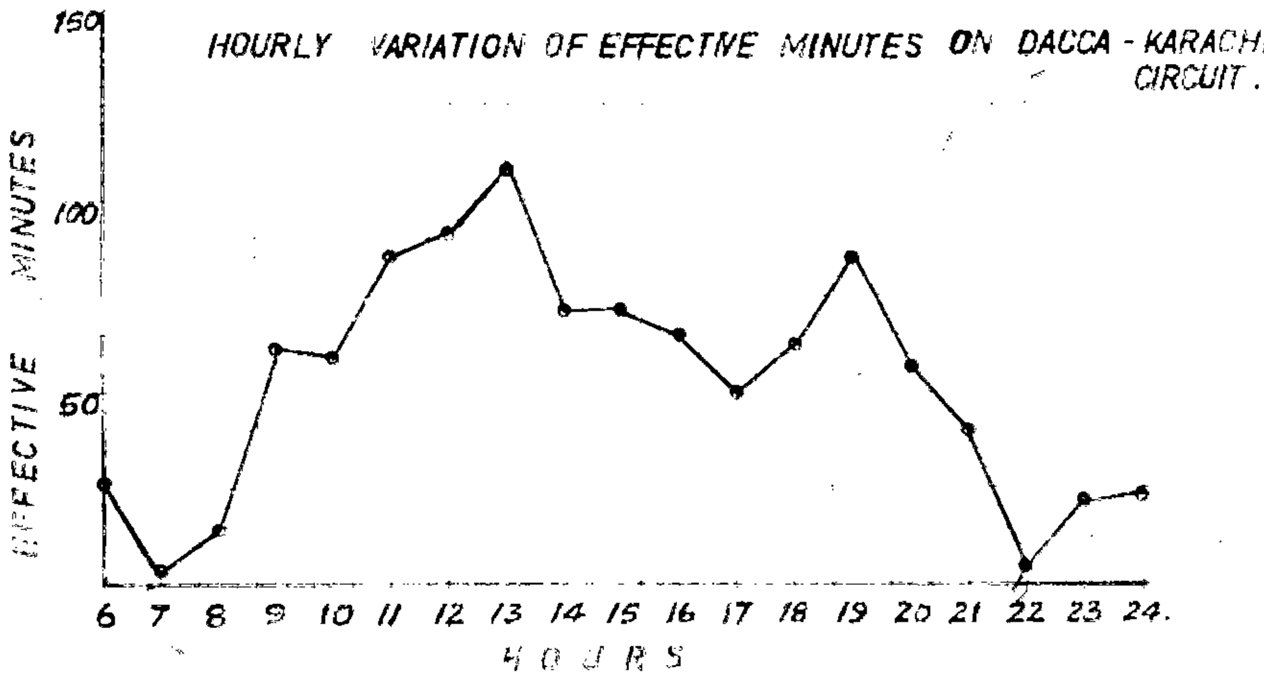


TABLE XVII-B.

CORRELATION BETWEEN EFFECTIVE MINUTES AND PERCENTAGE OF CIRCUIT TIME AVAILABLE ON DACCA-KARACHI CIRCUIT.

Wednesday:

Hour.	Calls booked.	Calls effective.	Average of x. $\bar{x}$	Average of y. $\bar{y}$	correlation co-efficient.
0-6	132	45.83			
6-7	3	8.33			
7-8	21	58.33			
8-9	57	50.00			
9-10	87	61.10			
10-11	102	59.72			
11-12	111	79.16			
12-13	96	79.16			
13-14	51	36.11			
14-15	105	66.66	56.5263	49.704	0.612716
15-16	60	100.00			
16-17	90	89.58			
17-18	60	64.58			
18-19	24	33.33			
19-20	9	10.41			
20-21	30	45.83			
21-22	9	18.75			
22-23	0	0.00			
23-24	27	37.50			

WEDNESDAY - 24.11.65.

FIG. 5·7

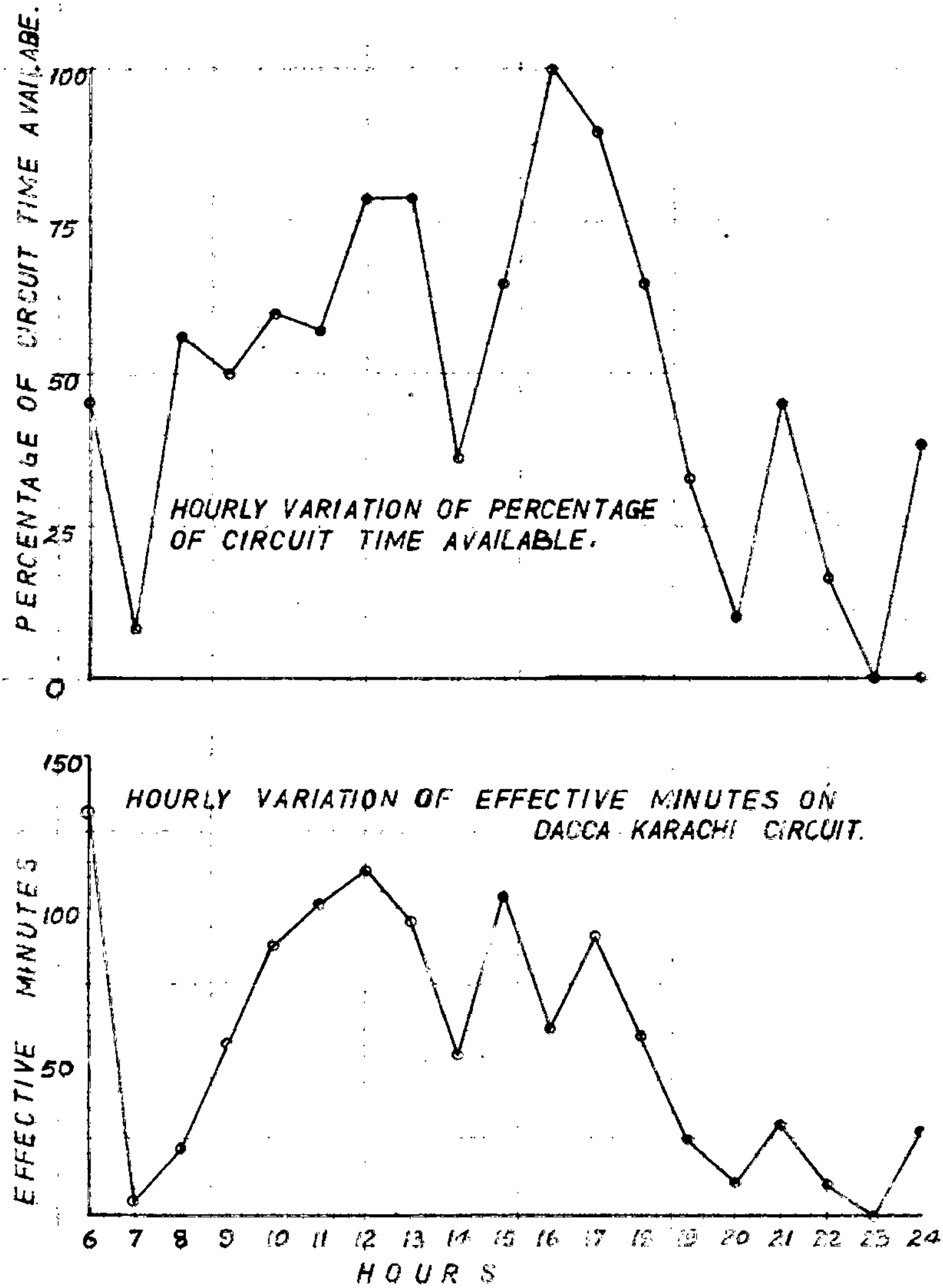


TABLE XVII-C.

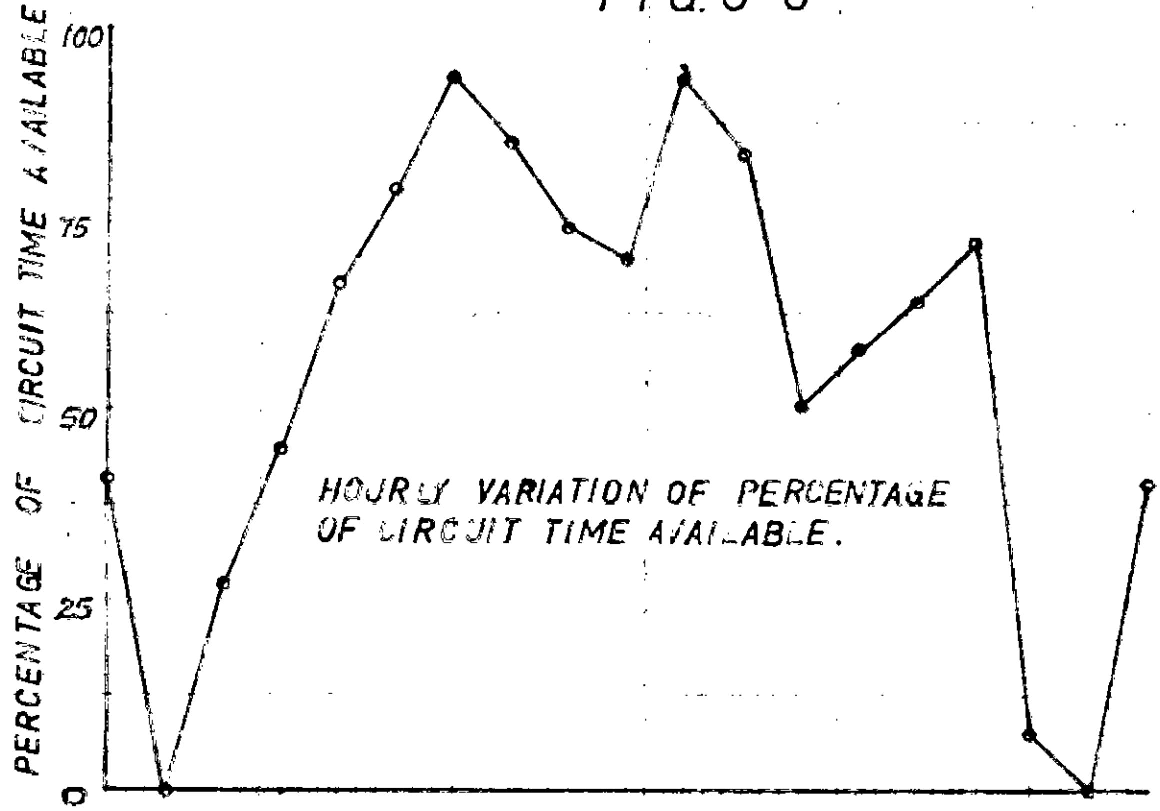
CORRELATION BETWEEN EFFECTIVE MINUTES AND PERCENTAGE OF CIRCUIT TIME AVAILABLE ON DACCA-KARACHI CIRCUIT.

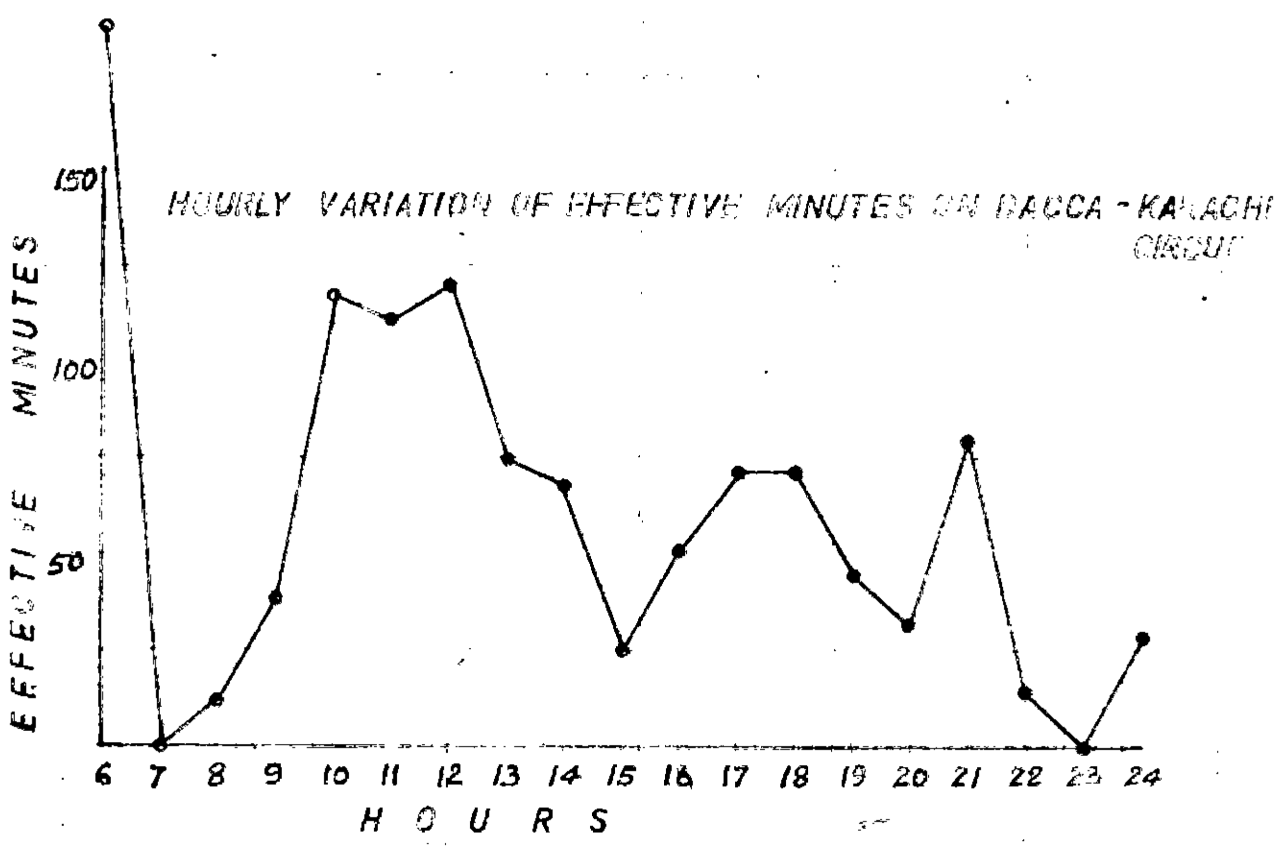
Friday:

Hour.	Calls booked.	Calls effective.	Average of x. $\bar{x}$	Average of y. $\bar{y}$	Correlation co-efficient.
0-6	186	41.66			
6-7	0	0.00			
7-8	12	29.16			
8-9	39	45.83			
9-10	117	68.05			
10-11	111	80.55			
11-12	120	97.22			
12-13	75	86.11			
13-14	69	76.38			
14-15	27	70.83			
15-16	51	95.83	60.7894	56.4606	0.517
16-17	72	85.41			
17-18	72	50.00			
18-19	45	58.33			
19-20	33	64.58			
20-21	81	72.91			
21-22	15	8.33			
22-23	0	0.00			
23-24	30	41 .66			

FRIDAY. - 26.11.65.

FIG. 5·8





5.3 A TELEPHONE TRUNKING PROBLEM:

Narayanganj Exchange has an equipped capacity of 2000 lines.

It is required to find out how many junction lines will be required to Dacca Exchange if with certain call duration

(a) One out of every thousand calls between two exchanges is allowed to fail (b) One out of every hundred calls between the two Exchanges is allowed to fail,(c) 5 out of every 100 calls are allowed to fail,(d)10 out of every 100 calls are allowed to fail (e) 15 out of every 100 calls are allowed to fail, (f) 20 out of every 100 calls are allowed to fail.

Supposed duration of the calls are $1\frac{1}{2}$ minutes, 2 minutes, 3 minutes, 4 minutes and 6 minutes. Also suppose that during the busy hour of the day each of the 2000 subscribers makes a call to Dacca Exchange.

Let us regard the problem as one involving independent Bernonlli trials. We suppose that for each telephone at Narayanganj, say the jth telephone, there is a probability that an outside line will be required (either as the result of an incoming call or an outgoing call) one could estimate by observing in the course of an hour how many minutes h_j an outside line is engaged, and estimating p_j by the ratio $h_j/60$. In order to have repeated Bernonlli trials, we assume $P_1=P_2 \cdots\cdots p_n=P$. We next assume independence of the events A_1, A_2, A_n in which A_j is the event that the jth telephone demands an outside line at the moment of time at which we are regarding the Narayanganj exchange. The probability that exactly r outside lines will be in demand at a given moment is, by the binomial law, given by --- $\binom{n}{r} p^r q^{n-r}$

Consequently, if we let r denote the number of outside lines connected to the Narayanganj switch room and make the assumption that they are all free at the moment at which we are regarding the Narayanganj Exchange, then the probability that a person desiring an outside line will find one available is the same as the probability that the number of outside lines demanded at that moment is less than or equal to

$$\sum_{r=0}^{r} \binom{n}{r} p^r (1-p)^{n-r} \doteq e^{-np} \sum_{r=0}^{r} \frac{(np)^r}{r!}$$

where the equality sign holds if the poisson approximation to the binomial applies.

For different values of holding time the value np will differ and hence the probabilities. To determine these probabilities the poisson probability paper (fig 3.16) has been used. This has saved the tedious calculations and lot of time (Chapter 3).

On the poisson probability paper the horizontal m scale determine the average number of simultaneous calls from the subscribers, the vertical Pr scale the probability that all the junction lines will be in use when a subscriber attempts to make a call; the number of junction lines is represented by the r curves. A vertical line is drawn at the appropriate value of m on the probability paper. The poisson distribution of the problem is then represented by the points in which the vertical line erected on the calculated m value cuts the r curves, or for convenience, simply by the vertical line itself.

To determine the number of junction lines required to meet say the 1 in 100 condition a horizontal line from Pr= 0.01 is drawn to cut the vertical line erected at the appropriate values of m. At the point of intersection of the two lines the value of r, the number of junction lines required to meet 1 in 100 condition is found.

For different values of probability of getting a free line number of outside lines required will be different. Those values will also differ for different values of holding time. Table XVIII gives the value of r for different holding times.

T A B L E XVIII.

THE VALUE OF r DETERMINED FROM FIG.3.16.

P	1/40	1/30	1/20	1/15	1/10
Poisson Approximation.					
.999 n=2000	74	95	132	171	245
.99 "	68	87	125	161	233
.95 "	62	81	117	153	223
. 90 "	60	78	113	149	218
..85 "	58	77	111	146	215
,80 "	57	95	108	143	211

5·4 DA-CTG. DIRECT DIALLING SYSTEM:

ASSUMPTIONS: Dacca-Karachi booked calls were examined for 6 days between 22.11.65 to 27.11.65 and it was found that the peak hour call is about 10.34% of the 24 hours traffic on an average, There were 1357 calls in the year 1963 between Dacca & Chittagong on an average when No. of local lines at Chittagong were 3000. At present there are 6000 lines at Chittagong and if we assume that the No. of outgoing calls and incoming calls are increased in that proportion, then there could be 2713 Trunk call at Chittagong (both incoming and outgoing).Roughly every subscriber will have half a call per day. Because of easy accessibility of lines between Chittagong and Dacca it will be reasonable to assume that about 20% of these calls will be originated during the busy hour instead of 10.34% as in the case of Karachi-Dacca.

Therefore, 20% of 2713 calls will be originated during the busy hour i.e. roughly 550 subscribers will make a call (either an incoming call or outgoing call) to Dacca. Because of delay all the subscribers did not use to make calls. We would expect another 550 subscribers to attempt a call during busy hour due to easy accessibility i.e.1100 subscriber will attempt a call during the busy hour. Table XIX gives the values of r for different holding times and different probabilities.

T A B L E XIX.

THE VALUE OF r DETERMINED FROM
FIG.3.15 & FIG.3.16.

P Poisson Approximation.		1/120	1/60	1/40	1/30	1/24
99.9%	n=1100	21	34	46	57	70
99 %	"	18	30	41	52	63
95 %	"	15	27	37	47	58
90 %	"	13	25	35	45	55
80 %	"	12	23	33	42	53
70 %	"	11	21	31	40	51
60 %	"	10	20	29	38	48
50 %	"	10	19	27	36	46
40 %	"	9	18	26	35	45
30 %	"	8	17	25	34	43
20 %	"	7	16	24	32	42
10 %	"	6	14	22	30	38
5 %	"	5	12	19	27	36
1 %	"	4	10	17	24	32
.1 %	"	2	7	13	20	28
	~~6/120~~	6/120	7/120	4/60	9/120	10/120
99.9%	"	80	91	103	113	124
99%	"	73	85	95	104	114
95%	"	67	79	88	97	107

T A B L E XIX. (Contd). ~~Page~~ 2.

—— P Poisson Approximation.	6/120	7/120	4/60	9/120	10/120
90% n=1100	65	75	85	95	104
80% "	62	71	81	90	100
70% "	60	68	78	87	97
60% "	57	66	76	85	94
50% "	56	64	74	83	92
40% "	54	62	72	81	90
30% "	52	60	70	77	87
20% "	49	57	68	75	85
10% "	46	54	63	72	81
05% "	43	50	58	68	75
01% "	40	47	56	62	70
00.1% "	35	41	49	57	64

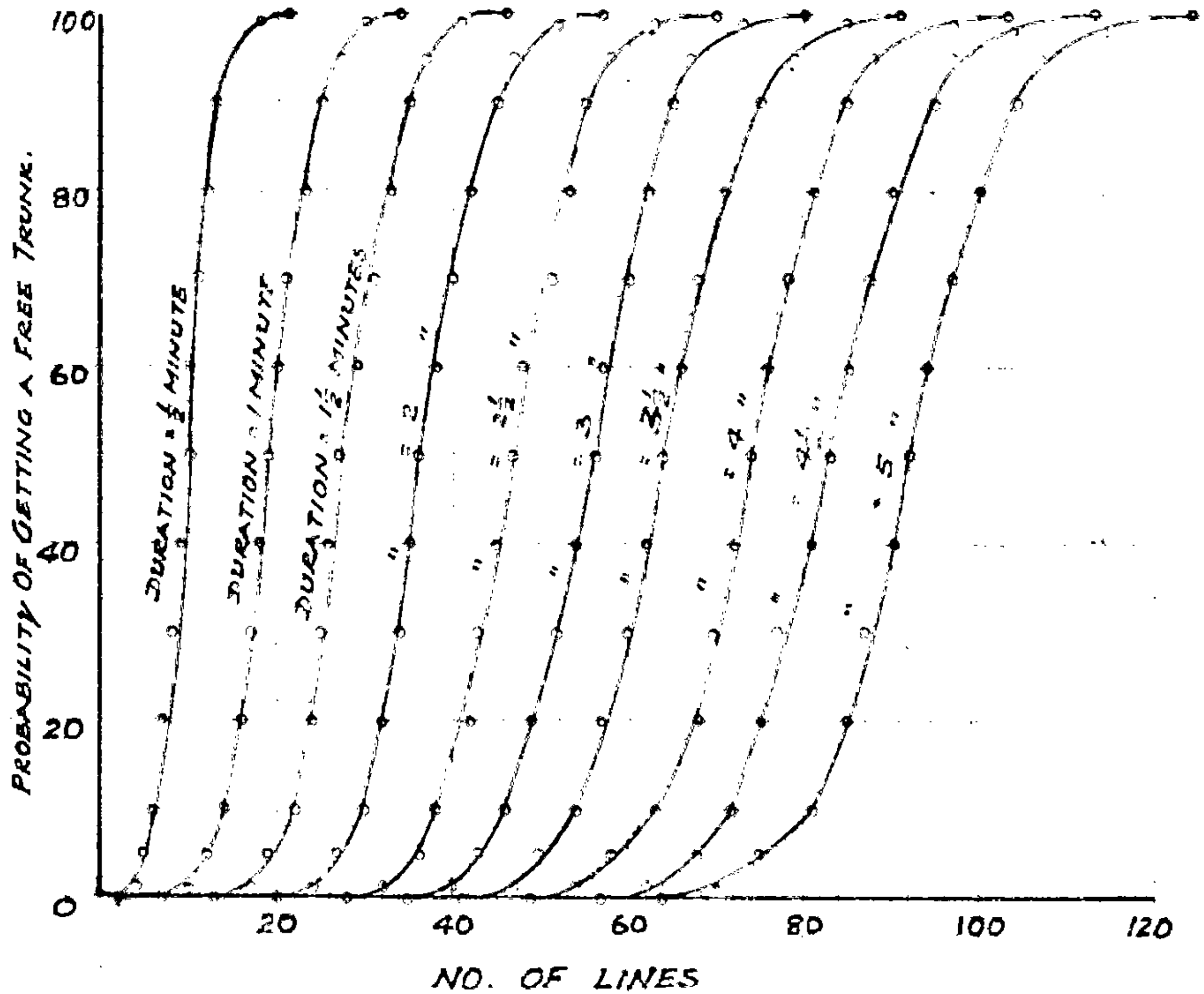

PROBABILITY OF GETTING A FREE TRUNK.
0
20
40
60
80
100
120
NO. OF LINES
DURATION = ½ MINUTE
DURATION = 1 MINUTE
DURATION = 1½ MINUTES
" 2 "
" 3 "
" 4 "
" 5 "

FIG 5.9

CHAPTER VI.

DISCUSSION OF RESULTS:

6.1 Characteristics of certain randomly selected telephone subscriber of Dacca.

Table VI is an example of how data for individual subscriber can be calculated and utilised. Average holding time and average number of calls originated by each type of subscriber can indicate the nature of the type of the subs. which can be utilised to calculate traffic pattern in future planning. For example a business Subs. A has an average holding time of 2 minutes and originates about 46 calls a day. Similarly relation between coefficient of variation and the ratio between average and maximum holding time could be investigated and possibly significant information about the characteristic of the Subs. of different types could be found with sufficient data. On the whole the data presented in this table can be treated as an example how statistical information regarding telephone subs. can be usefully treated.

6.2 Interwing Telegraph Traffic:

The existance of real corelation between total traffic and the technical loss, although of a low magnititude, may be taken as an indication that the ~~sufficient~~ causes of technical loss are not accounted with performance and operation of the equipment but due to other external causes, weather condition possibly being primarily responsible for the bulk of technical loss. However, the existence of the correlation itself can be interpreted as a fact that with larger volume of traffic more equipment would be in use and, therefore, more technical losses would originate due to more defects in the equipments themselves or in the lack of operating facilities.

6.3 Dacca-London Telegraph Traffic:

The same comments as above apply in genral to this case. The higher figure of corrrelation coefficient seems to indicate that in this case the maintenance of operating facilities of the equipment need more careful attention.

6.4 Auto correlation of technical losses, on Dacca-Karachi and Dacca-London Wireless Telegraph Traffic.

The auto-correlation of the technical losses with varying time gaps in months were studied with the aim to discover the effect of weather cycles, if any, on the technical losses. The irregular, but approximate weavy nature of the auto-correlation function curves obtained (figure 5.3 and 5.4) is perhaps a clear indication of the fact that the change in atmospheric condition due to "cycles of weather" is a significant factor in ~~reasoning~~ causing the technical losses. Due to altogether different types of climate existing at the same time in different Regions covered by the transmitted waves it would not be paying to try to make a detailed analysis of these cuves in terms of atmospheric conditions existing at the points of transmission and reception. However in each case maximum auto-correlations is found to exist with a time gap of 6-8 months which seems to be a likely phenemenon in as much as a period of 6-8 months covers a time length which may produce similar atmospheric conditions on either side of the peak of season.

6.5 Hourly variation of DA-KR booked calls & effective calls.

Hourly variation of Dacca-Karachi

booked calls and effective calls were studied with a view to find out whether the cct.available and the operation facilities available during any period of time were calculated in relation to the hourly variation of booked calls. Datas for six days between 22.11.65 to 27.11.65 were obtained and average hourly variations were found. Correlation coefficient of the hourly variation between effective calls and booked calls has been calculated and this is found to be .891. This is in fact very high and it can be said that the availability of ccts. and operating facilities were correctly phased.

6.6 <u>Hourly variation of paid minutes and percentage of circuit time available.</u>

To find out the economics of the system, correlation of the paid minutes per hour and percentage of cct. time available during that hour has been calculated. It is found that on Monday the 22.11.65 high correlation of .891 existed. Data for Wednesday the 24.11.65 and for 26.11.65 were obtained and analysed in a similar manner. There also correlation co-efficient of .63 and .51 respectively have been found which are higher than the significant correlation. They also show that real correlation between the percentage of cct. time available and the effective minutes on Dacca-Karachi cct. are in existance. Further more it could be said that on Monday the 22.11.65 the system was being far more economically used than on Wednesday and Friday. This may be an exceptional happening; however such a technic can be effectively used in assessing whether the facilities available in a telecommunication system are being most economically utilised.

6.7 Dacca-Narayanganj Trunking problem.

It is seen from table XVIII that 99.9% success in obtaining a free line between Narayanganj and Dacca, the number of junction lines required are 74, 95, 132, 171 and 245 for average call durations of 1.5 minutes, 2 minutes, 3 minutes, 4 minutes and 6 minutes respectively. The corresponding figure for 80% success are 57, 75, 108, 143 and 211 lines. It is estimated that the telephone communication system would be reasonably successful if the probability of obtaiing a free line during the peak hour ~~lies~~ is somewhere between 80% and 100%.

From study of the call durations in Dacca (refer Table VI) the average holding time is found to be about 2 minutes. Considering the facts that Narayanganj & Dacca are entirely two ~~xxxxx~~ separate towns the average duration of call between Dacca & Narayanganj would perhaps be slightly longer. With these facts in mind it seems reasonable to estimate that under the existing system of arrangement i.e. with 150 junctions between Dacca and Narayanganj there would be a reasonable degree of success in making a telephone call between the two localities at the peak hours.

It is to be remembered that the assumption that every subscriber at Narayanganj would attempt a call to Dacca at the peak hours is not based on any known findings, however, such an assumption seems to constitute a reasonable test on the efficiency of the telecom.system.

The fact that the method of analysis adopted in this particular type of situation with reasonable estimate of the average call duration found experimentally from a similar population of subscribers, gives as the required no. of junction lines figures in the range of actually existing number of junction lines, points to the usefulness of the application of statistical method of analysis.

6.8 <u>Dacca-Chittagong Direct dialling.</u>

6.8.1 Utilisation factor: - The durations within 24 hours for which each trunk line was in use were recorded. The total duration for both full rate and half rate calls were calculated. Figures for three typical normal days were 16017, 13822,12572. Now the total maximum minutes over the 24 hour period that the lines could have been in use = 48 x 24 x 60 =69120. From this the ratio between the actual minutes in use to the rated capacity in minutes (i.e.69120) come out to be $\frac{16017}{69120}, \frac{13822}{69120}, \frac{12572}{69120}$ i.e.,.2317, .200, .1820 respectively for the above mentioned 3 day. An average figure of .2012 be taken.

This average figure which is an indication of the extent to which the overall Dacca-Chittagong direct dialling system is in actual use, can be termed as the "utilisation factor" of the system.

<u>6.8.2 Statistical Analysis</u>:- A similar statistical method of analysis as in the case of Dacca Narayanganj telephone traffic is adopted (Refer 6.7) and the results presented in (Fig.5.9 and Table XIX).

The data corresponding to high probabilities (90% to 99.9%) of obtaining a free line in this system are expected to be of considerable practical use. For example it is found that for an average duration of call of two minutes and with 48 channels in service the probability of getting a free line at the peak hour is 95% (assuming that 1100 call originate in the peak hour) as assumed in section 5 (Table XIX). Addition of another 10 lines $\frac{10}{48} = 21\%$ to the system would increase the probability of getting a free line by only 4.9% to 99.9% and similarly for other durations. A similar type of conclusion could be drawn for other call duration.

The data for very low probability figures is of special interest in this connection. For example a very low probability figure (0.001) of getting a free line at the peak hour is a physical indication of the fact that all the 48 lines are in use for all the 60 minutes of the peak hour or in other words the calls are randomly spaced, over all the 48 channels successively in such a way that no further calls can be placed on the entire system at that hour which means that the system is being used at its rated capacity satistically.

The total duration that the system was in use for typical days (Section 6.8.1) are known. The total no.of calls both ways for those 3 days were also recorded and was found to be 8177,7348 and 8009 respectively. Average duration of calls is thus found to be about 2 minutes.

The maximum available minutes for 48 channels in the peak hour = 48 x 60 = 2880 channel minutes. With an average call duration of two minutes and all the lines being in use all the time throughout the hour i.e. Zero (.001) probability of getting free line, the total No. of Calls that can pass both ways = $\frac{2880}{2}$ = 1440

However, by refering to the probability calculation in fig.5.9 it is seen that ~~the~~ with 48 lines the condition for Zero probability of getting free line would require a call duration in the range of 3½ to 4 minutes; call duration of 3½ minutes at the peak hour would thus mean that only $\frac{2880}{2 x 3.5}$ = 823 calls would pass both ways at the peak hour. This figure is substantially different from 1440 obtained on the basis of 2 minutes average duration as indicated by the recorded figures.

The statistical assumption that 1100 persons are trying to originate a call at the peak hour seems, therefore, to have been on the lower side.

Using the observed call duration of 2 minutes with the given No.of channels = 48 and studying the case of Zero probability of getting a free line at the peak hour, reference is made to fig 3.16. The conditions imposed in this case are satisfied with a value of N= 73 x 30 = 2190 or in other words 2190 subscribers would probably be attempting a call at the peak hour instead of 1100 as assumed previously. As seen before with an average duration of two minutes, max.No.of calls that can be handled by the system =1440.

There would thus be an "OVERFLOW" of 730 calls at the peak hour. It seems therefore, reasonable to conclude that the system is still inadequate in handling the peak hour traffic even with all the channels in good condition.

From the peak hour call overflow of 730 the total overflow throughout 24 hours would be equal to 730x24 = 17520. However the actual over flow during 24 hours period would be lower than this.

From recordings of all overflows in minutes of the same 3 days figures 13869,13540, 5317 have been found.

6.8.3 System Effectiveness: - Two other figures indicative of the effectiveness of the existing system can also be derived from the above considerations :

A. System Effectiveness Factor:- This will be defined as the ratio;

$$\frac{\text{actual daily No.of materialized calls.}}{\text{Total daily No.of originating calls.}}$$

or

$$\frac{\text{Actual daily of No.of materialized calls}}{\text{Materialized calls +daily overflow.}}$$

In the system under consideration, system effectiveness factor for 3 typical working days are

$$\frac{8177}{8177+13869}, \quad \frac{7348}{7348+13540} \quad \frac{8009}{8009+5317}$$

or.362, .351, .605 i.e. an average of .442.

These figures are indications of the characteristic of the subscriber population of the two areas, as reflected on the communication system.

B. <u>Peak hour statistical effectiveness factor:</u>
This factor will be defined as the ratio;

$$\frac{\text{No.of maxm.possible calls of a given duration during the peak hr.}}{\text{Estimated statistical demand at peak hour.}}$$

In the DA-CTG direct dialling channels under investigation, for calls of average duration of 2 minutes,

Peak hour statistical Effectivness Factor

$$= \frac{1440}{2190}$$

$$= .66 \text{ or } 66\%$$

This means that at the peak hour, the number of channels provided is capable of handling about 66% of the actual demand on the system in the peak hour. This is obtained by statistically estimating the probable demand on the system. during the peak hour.

It is thus evident that adoption of a suitable statistical method of analysis of any telecommunication system is capable of providing much useful data, and practical guidance, in evaluating and improving existing systems, or in designing new telecommunication system.

B I B L I O G R A P H Y

1) Doust J.F. and JOSEPHS H.J.: A Simple introduction of statistics in telecommunication Engineering (Part I to IV)

2) Woodward P.M. : Probability and information theory with application to Radar (book)

3) Dixon J.W. and Massey F.J. Jr: Introduction to statistical analysis (book)

4) Fisher R.A. : Statistical methods for research workers (book)

5) Lee Y.W. : Statistical theory of Communication (book)

6) Bell D.A. : Statistical methods in Electrical Engineering (book)

7) Parzen E. : Modern Probability theory and its application (book)

8) Tolman R.C. : Principles of statistical mechanic (book)

9) Grant E.L. : Statistical quality Control (book)

10) Arkin and Colton : An outline of statistical methods (book)

11) Mood A.M. : Introduction to the theory of statistics (book)

12) Simon L.E. : Engineer's manual of statistical method (book)