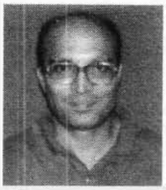
Numeracy for Everyone

7. Numeracy in Industry: Statistical Skills to Improve Quality and Maximize Profits

Anil P Gore and S A Paranjpe



A P Gore is currently Head, Department of Statistics, University of Pune. He initiated teaching and research in ecological statistics at Pune. He has been active for many years in promoting interaction between statisticians and biologists.

S A Paranjpe is Reader in the Department of Statistics, University of Pune. She has long experience of applying statistics in the fields of nutrition, forestry, ecology, agriculture and health.

If we ask a typical student of standard twelve about what her/his choice of higher education is, a very likely answer would be 'medicine' or 'engineering'. This is understandable in view of high employability, prestige and remuneration associated with these professions. There is a common belief that while students going to medicine need not be well versed in mathematics, those opting for engineering have to be. So it may come as a surprise to many that mathematics as practised in a typical engineering course does not prepare students in two crucial quantitative aspects of industrial environment viz. quality and profit. The purpose of this article is to highlight mathematical/statistical tools commonly used in quality improvement and optimization.

Quality Improvement

Have you heard of the terms, globalization and liberalization? Over the last decade or so, India has opened the gates of her economy and industrialists all over the world are now welcome to manufacture and sell their products here. From very sophisticated and essential (say IBM or Enron) to very unessential (say Pepsi or McDonalds), the whole spectrum of companies have set up shops here. Why do customers throng to (at least some) them? Partly it is image building and partly quality. People feel that Maruti cars are better than Premier (and worth the price). Hence the tilt. Quality has therefore become a key factor in industrial production in India.

For our limited purpose let us define quality as uniformity. This is exactly the opposite of the world of art in which uniqueness is at a premium. Uniform Ganesh idols can easily be made from a mould. However, those made one at a time cost more. Latter are regarded as of higher quality. So notion of quality is different. When it comes to industrial products, consumer expects high class performance. Whatever the specifications in terms of performance, each item (or batch or package) produced should fulfill the specifications and all should essentially be alike. This is quality. We take such uniformity in products for granted. When a spark plug of a scooter needs replacement, we buy one without a moment's thought as to whether it will fit 'our' vehicle. We assume that there is no individuality in scooters (of the same company and brand) or in spark plugs. In fact things are not quite 'the same'. There is variation. But it is within tolerable limits. Suppose the plug intended for a Bajaj Super scooter cannot be fitted to a vehicle because the OD (outer diameter) of the threaded portion of the plug is too large then we will say that it is beyond tolerance limits. Items which differ from specifications beyond tolerance limits are rejected by customers. They have to be reworked if possible or discarded as scrap. So a basic duality is that consumer wants assured good quality products while manufacturer wants to avoid rejections. Quality improvement protects interests of both sides. Understanding nature and causes of variability is the foundation for attaining high quality.

[1] Part 1. Why Quantification?, *Resonance*, Vol.4, No.5, 19-30, 1999.
Part 2. Dice of Life, *Resonance*, Vol.4, No.9,14–23, 1999.
Part 3. Just for Ecologists, *Resonance*, Vol.5, No.5, 15-25, 2000.
Part 4. Numeracy in Research Planning, *Resonance*, Vol.5, No.11, 14-25 , 2000.
Part 5. A Circular Path of Conjectures, Observations, Interpretation and New Conjectures about Diseases and the Causes there of, *Resonance*, Vol.6, No.3, 8-18, 2001.
Part 6. Numeracy in Social Sciences Art and Science of Quantifying Amorphous Concepts, Impressions and Unmeasurables, *Resonance*, Vol.6, No.6, 6-17, 2001.

Each production process has its own intrinsic variability. Many factors contribute to it. Raw materials or components supplied by vendors may not be uniform. Performance of different operators or machines may be different. Things may vary from day shift to night shift and from one day to another. In a popular novel about automobile industry in USA, named '*Wheels*' by Irving Wallace, an insider recommends to a friend that one should buy a car produced on Wednesday. The reason for this is supposed to be that as the weekend approaches (Thursday/ Friday) workers become negligent. Reason for avoiding production on Monday/Tuesday is holiday hangover. The point is that some variation in products may remain in spite of all precautions. Attempts to ensure quality can be made at each of the three major stages – product design, controlling production

The aim is to be able to reject all bad lots and accept all good ones. It is not possible to avoid both errors but by choosing a sufficiently large sample size one can keep the probabilities of making these errors reasonably low.

process and final inspection. It turns out that statistical tools useful in these three stages were developed over time in the reverse order mainly because of the difficulty involved.

Acceptance Sampling

If you are not a producer but a (bulk) consumer you wish to ensure that supplies received are of requisite quality. It is not practical to check every item. Sometimes checking is destructive. (Suppose you wish to test if explosion of a fire cracker is loud enough to be used at 'Deepavali'. You cannot test one and have it too.) Hence a sample can be taken instead and the number of defective items in the sample counted. If this count is too high, the lot is rejected. Otherwise it is accepted. The key question is 'how many items should I inspect?' The aim is to be able to reject all bad lots and accept all good ones. It is not possible to avoid both errors but by choosing a sufficiently large sample size one can keep the probabilities of making these errors reasonably low. The actual calculations have to be based on the so called hypergeometric distribution. Let us consider a simple hypothetical example.

Suppose we have to buy a box of 200 lead pencils. We propose to check a random sample of 32 pencils. Our rule is to accept the lot if sample contains no more than 3 defective pencils. Suppose the lot contains 40 bad pencils and deserves to be rejected. Then what is the probability that our decision rule will fail to do so? This is the so called consumer's risk.

P(the lot will be accepted) $= P(\#$ of defectives in sample $\leq 3)$

$$= \sum_{i=0}^{3} \frac{\binom{160}{32-i} \cdot \binom{40}{i}}{\binom{200}{32}}$$

This turns out to be approximately 7.5%.

Now suppose the lot contains only 9 defective pencils and deserves acceptance. Under our rule now the probability of accepting the lot is,

$$P(\text{the lot will be accepted}) = \sum_{i=0}^{3} \frac{\binom{191}{32-i}\binom{9}{i}}{\binom{200}{32}}$$

This is approximately 96%. Here producer's risk is only 4% These calculations can be verified using EXCEL. In actual practice one has to select sample size and decision rule so as to ensure that consumer's and producer's risks are at pre-specified levels. Such a selection is called an 'acceptance sampling plan'. Such plans for commonly used risk levels are available in standard manuals.

In actual practice one has to select sample size and decision rule so as to ensure that consumer's and producer's risks are at pre-specified levels. Such a selection is called an 'acceptance sampling plan'.

Control Charts

Acceptance sampling plan reduces the risk of a consumer getting a substandard item. But rejected lots constitute a major loss to the producer. As in health, prevention is much better than cure. There is necessity to provide for midcourse correction. Opportunity for such correction is provided by control charts. They involve monitoring production flow, taking periodic samples, measuring relevant traits (e.g. OD of a spark plug) and plotting say the means on an appropriate graph. Each control chart has a centre line representing the desired value of the trait and upper/lower control limits (say at a distance of 3 sigma). If a sample value falls outside the control limits, it is treated as a warning of trouble in the production line. Let us consider an example from mechanical engineering. The data, collected by a student working on an industrial assignment in a factory in Pune, is about cylindrical bearings being made for an IC (internal combustion) engine. Wall thickness is the criterion of interest here. A sample of five pieces is selected every 15 minutes and average wall thickness is plotted. If $\bar{\bar{X}}$ is the grand average i.e. mean of sample means and S is the standard deviation of these sample means then the control limits are calculated

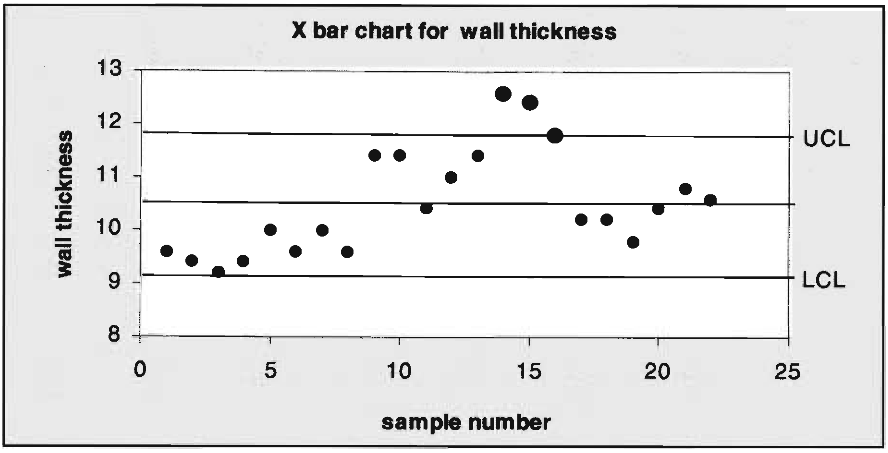


Figure 1. A control chart.

as $\overline{\overline{X}} \pm 3S$. *Figure* 1 shows the center line and upper and lower control limits (UCL,LCL). Each dot represents mean of one sample. Notice that dots corresponding to sample numbers 14 and 15 are above UCL while that for 16 is on the UCL. This suggests that the process had gone out of control. In such a situation the thing to do is to search for an assignable cause responsible for the process going out of control and to eliminate that cause. In this case inquiry showed that a cutting tool used had worn out and had to be replaced. Dots corresponding to samples 11,12 and 13 indicate a rising trend giving early warning, which was not heeded. Of course one may ask why three rising points all within control limits should be regarded as a sign of trouble. The objection is valid unless there is circumstantial evidence to suggest that trouble may be brewing. A thumb rule followed by some is to wait for occurrence of 7 successive points each higher than the previous one, before raising an alarm. The basis for this thumb rule is that probability of a point being higher than its predecessor is 1/2. Two successive points each being higher than previous one is an event with probability 0.25. Proceeding thus we note that at 7 the probability falls below 0.01 (a commonly accepted threshold) for the first time.

A thumb rule followed by some is to wait for occurrence of 7 successive points each higher than the previous one, before raising an alarm.

One may use charts not only for means but also for ranges, proportions, etc.

ISO 9000

Control charts are a technique for anticipating and detecting a failure at a particular stage in production. To ensure quality a manufacturing unit needs to follow appropriate techniques in all activities. So international norms e.g. ISO 9000, have been laid down as guiding principles to be implemented by a company. These include statistical methods as well. There are organizations which act as quality auditors. Many countries now a days import goods only from companies which conform to such procedural standards. This forces exporters to follow the norms seriously. You may have seen advertisements of companies regarding ISO 9000 certificate earned by them.

Taguchi Approach

In the last couple of decades the name of Taguchi has become synonymous with consideration of quality and productivity. This Japanese engineer introduced a novel approach to problems of quality in manufacturing. He argued that some variation in inputs is unavoidable and it is better to have a robust product design. In other words, design should be such that quality is invariant (and good) even when inputs are not uniform. Madhav Phadake (1989) gives the example of a power supply circuit to illustrate the underlying concept. We can use nonlinearity of a relationship to reduce effect of variation in input on product. Output voltage y is an increasing function of x say transistor gain. If we seek output voltage 110 (used in USA), the appropriate x value is A. But here small change in x causes big change in y. On the other hand output voltage $y=125$ can be gained at $x=B$ and here change in x does not cause major change in y. (See *Figure* 2). Hence a robust design will involve $x=B$. But now the stable value of y available is wrong. It has to be shifted to a desired level by introducing some resistors. Of course this is a rather simple example. In practice many factors are involved and identifying a robust combination can be rather difficult.

Taguchi argued that product design must be such that quality is invariant even when inputs are non uniform.

Starting from inspection after production we have ended up

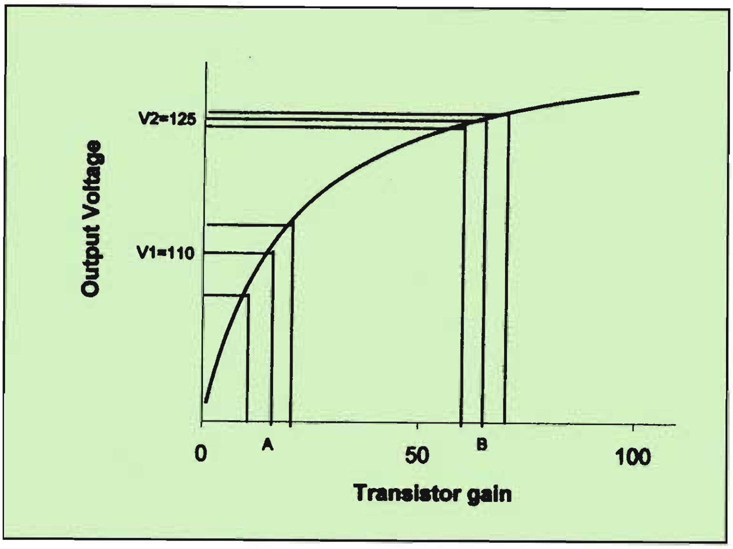


Figure 2. Nonlinear relationship and reduction in variability.

with the idea of robust product design for quality. This is the spectrum of statistical quality control ideas today. We may note that basic ideas in this field were developed mostly by statisticians in the west. Perhaps the most famous name among these is that of W E Deming. However, it was the Japanese industry that enthusiastically embraced the faith and harvested bumper profits. In the seventies this was recognized all over and western industries caught up with the Japanese. We in India have a long way to go in this respect.

Optimization

The second major concern in industry is profit maximization. This can be achieved by optimization in resource use. Numerical techniques can play an important role here too. Let us begin with a simple yet very real example.

The second major concern in industry is profit maximization. This can be achieved by optimization in resource use.

In a large automobile manufacturing plant, hundreds and thousands of gauges are used regularly. These must be accurate or else measurements go wrong. For this, they have to be checked periodically and recalibrated. How often should this be done? In a factory in India, the convention was to check each gauge every two weeks. There was a suspicion that this method did not work satisfactorily. Hence a simple study was launched. Gauges were classified by frequency of usage.

Frequency	Percentage of gauges with such use
100 times per day or more	20%
10 to 99 times per day	25%
upto 9 times a day	55%

Consultants suggested that calibration frequency should be commensurate with frequency of use. Where there is heavy use (i.e. category 1) there should be weekly check as error in these gauges affects many measurements. Moderately used gauges (category 2) can continue to be checked once a fortnight. The infrequently used ones (category 3) need be checked once a month only. It turns out that the above simple modification leads to a reduction in measurement errors and a small amount of saving in calibration effort.

Of course there are many methods in mathematics that come handy in solving a variety of optimization problems in industry. Operations Research (OR) is the name given to such a tool box. This discipline crystallized from military related research done during the Second World War. There are many aspects of OR – linear, nonlinear and dynamic programming, queues and inventories, modeling and simulation, etc. After the war, the methods came to be used by people at large. We will take a brief look at some of these techniques.

There are many aspects of Operations Research – linear, nonlinear and dynamic programming, queues and inventories, modeling and simulation, etc.

Linear Programming

Here the aim is to maximise a linear function of variables, which are subject to linear restrictions. Let us understand this technique through a simple example. A fictitious coal mining company, Bangalore Coal Fields produces two kinds of coal, type A and type B. Profit from selling one ton of type A is Rs. 400/- and a corresponding figure for type B is Rs. 300/-. Decision to be made is quantity x_1 of coal type A and x_2 of type B to be produced in a day so as to maximise the total profit $P = 400 * x_1 + 300 * x_2$.

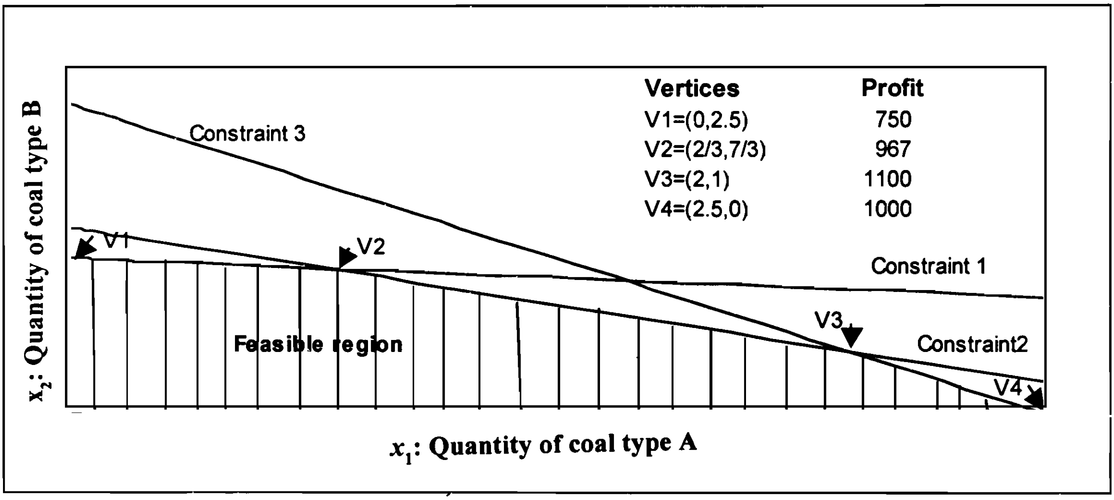
A mathematical theorem simplifies the choice by asserting that the best combination has to be at one of the vertices of the feasible region.

One cannot select very high values for both x_1 and x_2. This is because of restrictions on time, availability of cutting machine (10 hours in a day), screens (9 hours) and washing plant (10 hours). Time requirement for producing one ton of coal type A is 1, 3 and 4 hours of cutting machine, screens and washing plant, respectively. Corresponding values for coal type B are 4, 3 and 2. Hence mathematically the problem is to maximize total profit P subject to three restrictions viz.

$$x_1 + 4 * x_2 \leq 10$$
$$3 * x_1 + 3 * x_2 \leq 9$$
$$4 * x_1 + 2 * x_2 \leq 10$$
and $$x_1, x_2 \geq 0.$$

These constraints can be satisfied by several pairs of values of decision variables (x_1, x_2). This set is called feasible region because each of the choices is implementable. Among these we have to choose one which gives maximum profit. A mathematical theorem simplifies the choice by asserting that the best combination has to be at one of the vertices of the feasible region. *Figure* 3 shows the feasible region with four vertices indicated by arrows. Profit at each vertex is also shown. Clearly the optimal choice is at $x_1 = 2$ and $x_2 = 1$, which will give a profit of Rs. 1100/- per day.

Figure 3. Multiple constraints and feasible region.



As usual reality is more complex than a conveniently selected example like this. As the number of decision variables goes beyond 2, graphical presentation becomes difficult. For moderate number of decision variables and constraints, computer programs can be used to implement a procedure called simplex algorithm to arrive at the best choice. As the problem size goes beyond a point, even this approach begins to falter. For such super complex problems a new algorithm was devised by an Indian mathematician, Narendra Karmarkar who became an instant celebrity.

For moderate number of decision variables and constraints, computer programs can be used to implement a procedure called simplex algorithm to arrive at the best choice.

Transportation Problem

Have you noticed that there are large godowns that store up industrial products, just outside the municipal limits of most cities? Cement, steel, paper, tires and what not. This is mainly to avoid unnecessary payment of octroi. But in any case every major manufacturer has to send shipments to widely distributed wholesale and retail vendors or customers. The problem here is to satisfy requirements of all customers while keeping the transport cost at a minimum. This is a special case of linear programming problem that requires special attention. The following example will illustrate the nature of this problem.

A company has three factories with production capacities 20,15 and 10, respectively. There are three godowns which can store 5, 20 and 20 units. The problem is that of shipping products from factories to godowns such that cost of transportation is minimized. Following is the matrix of transportation costs:

Cost of shipping one unit from factory to godown

From	To godown		
Factory	G1	G2	G3
F1	9	10	10
F2	10	14	8
F3	13	10	8

Queues are an inevitable feature of modern life.

A manager not trained in OR had adopted the following scheme: Send 5 units from factory F1 to godown G1 and 10 units to G2 and remaining 5 to G3. Send from F2 10 units to G2 and the remaining 5 to G3. Send everything from F3 to G3. Cost of this scheme is Rs. 455/-. Such a scheme which satisfies all the restrictions is called a feasible solution. However, the least cost solution may be different. To improve on a feasible solution, one identifies the most expensive element in the cost matrix and tries to reduce burden on that route. In our case such an element is F2 to G2 (cost Rs. 14/-). Hence we reduce the burden on this element from 10 to 5 and send these 5 units to G3. As a compensation we cancel shipment from F1 to G3 and send those 5 units to G2 instead. So the revised scheme is as follows:

Ship 5 units from F1 to G1
Ship 15 units from F1 to G2
Ship 5 units from F2 to G2
Ship 10 units from F2 to G3
Ship 10 units from F3 to G3

The cost of this solution is Rs. 425/-, which is an improvement over the earlier one. But even this is not the final answer.

The least cost solution turns out to be as follows.

Ship 5 units from F1 to G1
Ship 15 units from F1 to G2
Ship 15 units from F2 to G3
Ship 5 units from F3 to G2
Ship 5 units from F3 to G3

The cost of this solution is Rs. 405/-.

Let us remember that this is a simple problem and real problems are much more complex.

Queues

These are an inevitable feature of modern life. Queues are for a service. If you go to a barber shop on a Sunday you have to wait for long. Why aren't there more service points? You wonder. Even aircrafts have to form a queue for landing or take off. Why not build one more runway? The cost of adding an extra service point is often very substantial. Will the benefits more than outweigh costs? Experimentation is not possible. But we can do statistical simulation.

The cost of adding an extra service point is often very substantial. Will the benefits more than outweigh costs? Experimentation is not possible. But we can do statistical simulation.

This is the method of creating a mathematical reality, analogous to virtual reality created in computers. Arrivals of aircrafts are mimicked using random numbers from say an exponential distribution with a suitable arrival rate. This rate can be extracted from traffic data available. Time for which an arriving aircraft keeps airport facilities busy (service time) can be modelled using another suitable distribution.

If the model is good, it will generate waiting times and queue lengths comparable to the actual experience. This is validation of the model.

Now imagine that one extra runway is added. This will mean that service time distribution will have to be suitably changed. The revised model can be run for as many days as needed by using more and more random numbers to represent more and more arrivals. Changes in waiting time and queue length can then be estimated. They will form the basis of any rational judgement as to whether the cost of a new runway is commensurate with benefit. Similar exercises can be done with reservation counters, postal service counters, new berths in a port, etc.

For more examples of simulation see [1].

Inventory Control

Mass manufacturing is an ongoing process. It requires continuous inputs of raw materials, bought out components, etc. You may have seen pictures of car assembly lines. As the car moves down the line, different parts are fitted to it. At each station there is a supply of a specific part, ready at hand. A worker at a station does his job and then the car moves on. A worker at a

A stores manager must take into account the time it takes to send an order and the delivery time before deciding the volume and frequency of order.

station does the same job repeatedly all day. In the famous film 'City Lights', Charlie Chaplin has humorously depicted psychological effects of such monotonous assembly line tasks on workers.

For smooth functioning of this set-up, there must be an uninterrupted supply of relevant parts. This means that parts must be stocked up. Such a stock is called an inventory. An important decision is the quantity to be stocked. Larger the stock, lower is the chance of interruption in production due to nonavailability. But larger stocks also imply that greater amount of capital remains locked up which involves a cost. A stores manager must take into account the time it takes to send an order and the delivery time before deciding the volume and frequency of order. If all systems are really fine tuned, it is possible to keep nearly zero stock level. The deliveries then have to be 'just in time'. This keeps inventory cost to a minimum.

One can see that there is lot of room for innovation in industrial management using these techniques and many more. In fact their use need not be restricted to industry and can easily be extended to all walks of life.

Exercises

1. In the study of bearings (see section on Control Charts), for each sample of five bearings maximum and minimum wall thickness was recorded. Data are given below.

Sample No.	Max.	Min.	Sample No.	Max.	Min.
1	11	8	12	13	10
2	11	8	13	13	10
3	11	8	14	13	12
4	10	9	15	13	11
5	11	8	16	13	10
6	11	8	17	11	9
7	11	9	18	12	9
8	11	9	19	11	9
9	12	11	20	11	10
10	13	10	21	11	10
11	11	10	22	12	9

The difference between maximum and minimum value in a sample is called sample range (R). A wide range indicates a poor quality. Calculate range for each sample and plot a control chart for range using the value R=2.23, LCL=0 and UCL=4.7. Check if any points fall outside the control limits.

2. Consider the problem of gauge calibration discussed in text. Verify that the saving in the revised calibration scheme is 7.5. Assume that cost of calibrating any gauge is the same. Try to work out the extent of reduction in measurement error.

3. A baker bakes two types of cakes each day, chocolate and a banana. He makes a profit of Rs.7.50 on one chocolate cake and Rs. 6.00 on a banana cake. A chocolate cake requires 4 units of flour and 2 units of margarine and a banana cake requires 6 units of flour and 1 unit of margarine. However only 96 units of flour and 24 units of margarine are available on each day. How many cakes of each type should the baker make on a day so as to maximize the profit?

Epilogue

We bring this series of articles to a conclusion now. Our attempt has been to describe briefly, applications of statistics in as many areas as possible. The intention was to paint with just a few bold strokes of the brush but almost totally devoid of details. We have referred to a large number of techniques, but only sketchily. We want to communicate to the readers our conviction that the core of statistics is not formulae but logical ideas. Given a sound idea, a suitable formula can always be devised to implement it. The vision of statistics essential for an intelligent layman is neither one of piled up numbers nor the third kind of lie in the famous Mark Twain quote but search for elegant patterns in a confusion of data. Perhaps readers will agree that this is also a good description of science itself.

Lastly, readers interested in dialogue about application of numeracy to any real life problems are welcome to write to us by e-mail or snail mail.

Suggested Reading

[1] S Kunte, Technique of Statistical Simulation, *Resonance*, Vol. 5, No. 4, pp.19-27, 2000.

[2] M S Phadake, *Quality Engineering Using Robust Design*, Prentice Hall, New Delhi, 1989.

[3] H A Taha, *Operations Research*, Macmillan, New Delhi, 1982.

[4] M L Crossley, *The Desk Reference of Statistical Quality Methods*, American Society for Quality, Quality Press, Milwaukee, Wisconsin, 2000.

Address for Correspondence
A P Gore and S A Paranjpe
Department of Statistics
University of Pune
Pune 411 007, India.
Email:
apgore@stats.unipune.ernet.in