Numeracy for Everyone

1. Why Quantification?

Anil P Gore and S A Paranjpe

Introduction

India is a land blessed with beautiful and bountiful nature. We have the world's tallest mountains in the north and the heaviest rainfall in the east. There are dry deserts and long perennial rivers. Our forest wealth and wildlife are legendary. Western Ghats and north-eastern mountains are two of the world's megabiodiversity hotspots. No wonder so many of us love and cherish this treasure. It is to such nature lovers that we address the next few articles on numeracy, quantification and biostatistics.

We can anticipate the reaction of many. Some may say that as nature lovers they enjoy going out, trekking in national parks, watching birds and butterflies and photographing landscapes. Silence and serenity of the wild makes them feel recharged to face the humdrum of life once again. They would prefer not to spoil their pleasure by getting into dreary details of numbers and statistics. Yet another group may argue that they took up study of biology, at least partly because they did not care for mathematics. And how can any one expect them to go through formulas and equations?

To these nature enthusiasts, we will say, hold your horses. While we are not apologetic about equations, formulas or numbers, this series is not about them. It is about nature and its conservation. Every detail of biostatistics to be discussed herein should be judged on the basis of its relevance to these matters. It is our plan to spare the reader of almost all technical details and give only the concepts necessary to understand processes of degradation and restoration of nature. The only assumption we make is that the reader is a serious nature enthusiast. Every such

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.

Numeracy for Everyone: Series Introduction

Do you know about three R's in the traditional list of minimum skills to be acquired by every school-going child? They are reading, (w)riting and (a)rithmetic. Perhaps arithmetic was needed mainly to carry out common transactions such as buying grocery, vegetables or paying rent.

In today's world of exploding information, it is not enough to know these three R's. It is also necessary to learn techniques to consolidate and interpret the continuously bombarding information. Statistics is the science of identification and art of interpreting patterns in numbers. Its distinctiveness lies in trying to understand uncertain events. Death is certain but age of death is not. We can predict rain but cannot guarantee it.

The aim of this series of articles is to introduce the reader to the basic ideas of statistics applied in many spheres of life. Thus we begin by emphasizing the need for quantification and illustrating how to represent the data using graphs. Then the concepts of probability and association/correlation are discussed.

As statistical consultants, our main interaction during the last decade and a half has been with ecologists/biologists. This has influenced our writing and many illustrations naturally come from the field of ecology. Some statistical techniques specially developed for applications in ecology are also discussed.

Any scientific experiment needs planning. Statistical techniques called design of experiments and sample surveys help a lot in such planning. An outline of these techniques is given.

Areas of biological sciences other than ecology, where statistics is used extensively, include health, epidemiology, clinical trials etc. Social sciences have a large component of uncertainty. Statistics is a science that searches for laws applicable to large groups of individuals (populations) in the presence of uncertainties. Its applications in social science lead to interesting findings, reviewed briefly in subsequent articles.

In the present liberalised and globalised economy, quality of an industrial product decides its fate in the market. Maintaining high quality of production, adjusting the production process to minimise the loss due to rejection of product are topics where statistical expertise can make a difference. We conclude the series by overviewing this area.

If a bird species disappears altogether from an ecosystem, most birdwatchers will notice it promptly. But if there is a gradual decline in numbers, only regular counting will reveal the dangerous trend before it is too late.

person routinely encounters news, views and comments that contain so much numerical information. Here are a few illustrations.

- Bharatpur is a Mecca for birdwatchers, especially those interested in winter migrants including ducks and waders. The jewel of them all is the Siberian crane. For years experts have been warning of a decline in the numbers of Siberian cranes returning to Bharatpur. If a bird species disappears altogether from an ecosystem, most birdwatchers will notice it promptly. But if there is a gradual decline in numbers, only regular counting will reveal the dangerous trend before it is too late.
- In the eighties, this man-made wetland was closed to cattle grazing. Subsequently *Paspalum* grass and other weeds have increased encroachment upon the waterbody which is

getting choked. If the trend continues the wetland may be converted into a grassland.

- India's forest lands are supposed to be shrinking i.e. the area under forest must be declining over the years. Further, even the areas nominally under forest are getting progressively degraded. This means that the number of trees per hectare must be going down.
- It was argued by ecologists a decade ago that if froglegs are exported on a large scale, the frog *Rana tigrina* may get endangered. What is the basis for this argument?
- Many wildlife enthusiasts have expressed reservations about the claims of forest departments that the number of tigers in Project Tiger areas is increasing continuously. People are very sceptical about the census methods adopted.
- Bird ringing programs of the Bombay Natural History Society (BNHS) have revealed that ducks migrate over great distances in very short times.

If you know the rationale behind all these statements, you probably will not see anything new in the sequel.

But many people are puzzled by these. When controversies and debates over environmental policies are heard, they are unable to make up their minds about which side is more reasonable. This is because they are unable to interpret quantitative information and arguments and to distinguish between valid evidence and junk. In that case you stand to gain something by continuing to read. We hope to discuss trends, harvesting strategies, diversity indices, census techniques, computer packages and a host of other things.

So are you willing to try a new adventure ?

Why Statistics: Ours is, what has often been called, an age of information explosion. Newspapers, magazines, videos, TV, cinema, radio and god knows what else, keep feeding us information, much of it numerical. It comes in the form of tables, graphs, charts, estimates, predictions etc. It is all mindboggling for the uninitiated. Why cannot things be stated simply?

The discipline of statistics helps us sort it all out, make it simple and comprehensible. Once that is done, we can go deeper and consider comparison of life span patterns.

Take the case of our life span. How long does man live? It is difficult to answer this question precisely. Salim Ali, the father of Indian Ornithology, died at the ripe old age of 92 while his wife Tehmina died young. In India nearly every tenth child dies in its first year of life. So what is the length of human life? Well, it is a variable. It can range from zero (still birth) to a hundred years or even more. This sounds like utter confusion. The discipline of statistics helps us sort it all out, make it simple and comprehensible. Once that is done, we can go deeper and consider comparison of life span patterns. Do tribals who are close to nature live longer than city dwellers? If not why not? Perhaps tribal life spans are shorter because of poverty, malnutrition and lack of medical facilities, or in other words ineffective efforts for tribal welfare. If you think tribal welfare has nothing to do with conservation ask yourself the following. Can forests be protected while forest dwellers are hungry, miserable and desperate?

In biology every individual has his own characteristics and yet the group as a whole follows a pattern (law). Statistics talks about these group properties or population laws.

Statistics as a science was born out of the study of biological problems. Peculiarity of biological phenomena is variation. The mango tree is the same and so is the caretaker, but different fruits are not identical in weight. In a eucalyptus plantation, all trees are of the same age, but their heights are different.

In physics if the law of gravitation holds for one apple, it is true for every apple. In chemistry if one drop of an acid shows a property, it is true for the whole lot. In biology things are not that simple. Every individual has his own characteristics and yet the group as a whole follows a pattern (law). Statistics talks about these group properties or population laws. When we say subabul trees are faster growing than tamarind, it is not meant that every subabul tree is taller than every tamarind tree of the same age. We mean the population of subabul trees as a whole is taller than the corresponding tamarind population. We will discuss later how this can be stated more precisely or convincingly.

Environmental impact assessment is a very popular phrase now-

a-days. Developmental projects of many kinds, a dam, a factory, a highway can have an adverse effect on the surroundings. Our society has become very conscious of these aspects and there is an endless debate about them. If one side puts up impressive figures in support of the project, the other side has to be equally effective in their arguments. Consider river pollution caused by effluents of sugar factories. If the discharge is very heavy relative to the flow of water, it will kill most of the natural biota. That is what you see in city wastewater disposal systems. Then the damage caused is obvious. Around many chemical factories, streams are brightly coloured or full of foam. Here environmental impact is clear. But can we measure the damage at moderate level of pollution? It will be in the form of reduced numbers of fishes of common types, disappearence of some species, increase in numbers of organisms typical of polluted waterbodies etc. All these aspects have to be estimated statistically.

We believe therefore that familiarity with methods of quantification of nature is necessary even to understand conservation debates intelligently.

Figuratively Speaking

Information in bulk has to be condensed and summarised for effective communication. It has to be presented in such a manner that space needed is not unreasonable and the reader understands it easily. One of the ways of doing this is use of graphs, charts and figures. Take any issue of a good magazine like *India Today* or *Frontline*. You are very likely to encounter one or more of bar charts, pie charts, histograms and scatter plots. It is useful to know what they are.

Bar chart: Pollination is a crucial stage in regeneration of flowering plants. *Figure* 1 describes the relative importance of different pollinators of *Martynia annua*, devil's claw [1]. Notice that on the x-axis we have the climatic conditions. The corresponding bars give the percentage of visitors of three types of pollinators. Clearly digger bees are important on a cloudy day

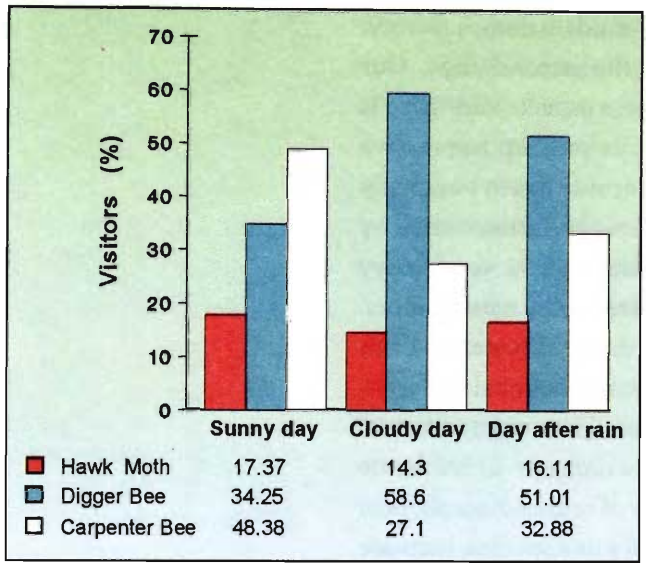
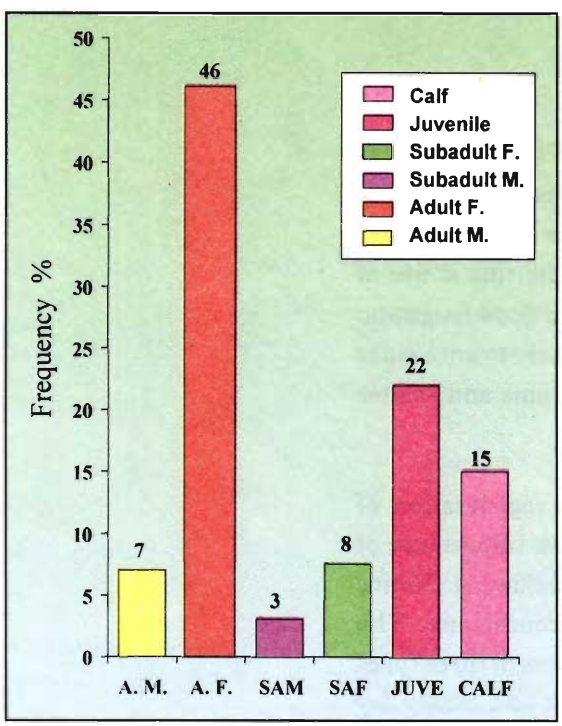


Figure 1. Pollinators of devil's claw.

Figure 2. Age–sex distribution of elephant population in Parambikulam.



or day after rain whereas on a sunny day carpenter bee plays an important role.

In *Figure* 2 we have age-sex composition of elephant herds in Parambikulam wildlife sanctuary [2]. On x- axis we see the words male, female, sub-adult etc. The corresponding bar shows the percentage of males (or females) of each type. Clearly adult males are few. This may be because they are loners and missed in count. Or they are poached for ivory.

Can you imagine what a bar chart giving total annual rainfall in 1998 at Jaipur, Bombay and Cherapunji would look like?

Pie charts: Pie is a western baked food item that is circular in shape. It is cut into portions for serving just like a cake. *Figure* 3 shows the approximate composition of species of flora and fauna in India, described in scientific literature [3]. Clearly insect species are the largest in number. About 5000 molluscs are known. The number of bird species is close to 1200. Mammals, reptiles and amphibians together are about 900 species. Did you know that the insect species are so overwhelming in number? The famous British/Indian biologist J B S Haldane was a non-believer. He is supposed to have said once, "*I don't know if God exists. But if he does, he must love beetles. He created more kinds of beetles than any other animal !*"

A pie chart you will invariably see after the announcement of the government budget

shows how a government earns income. The break-up gives percentage earning from income tax, excise and customs etc. Similar pie chart shows how a rupee is spent, i.e. in salaries, subsidies, development projects etc.

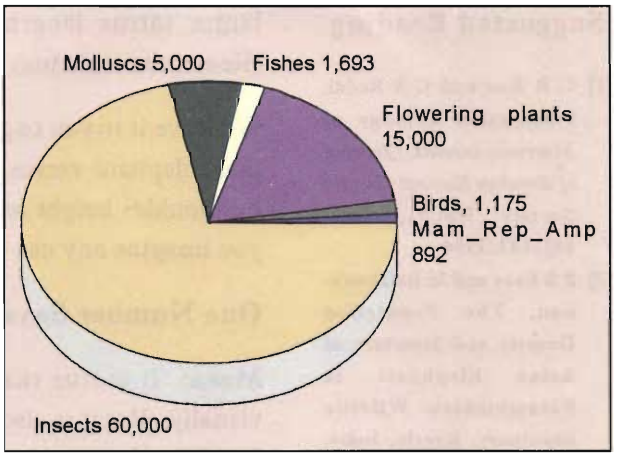
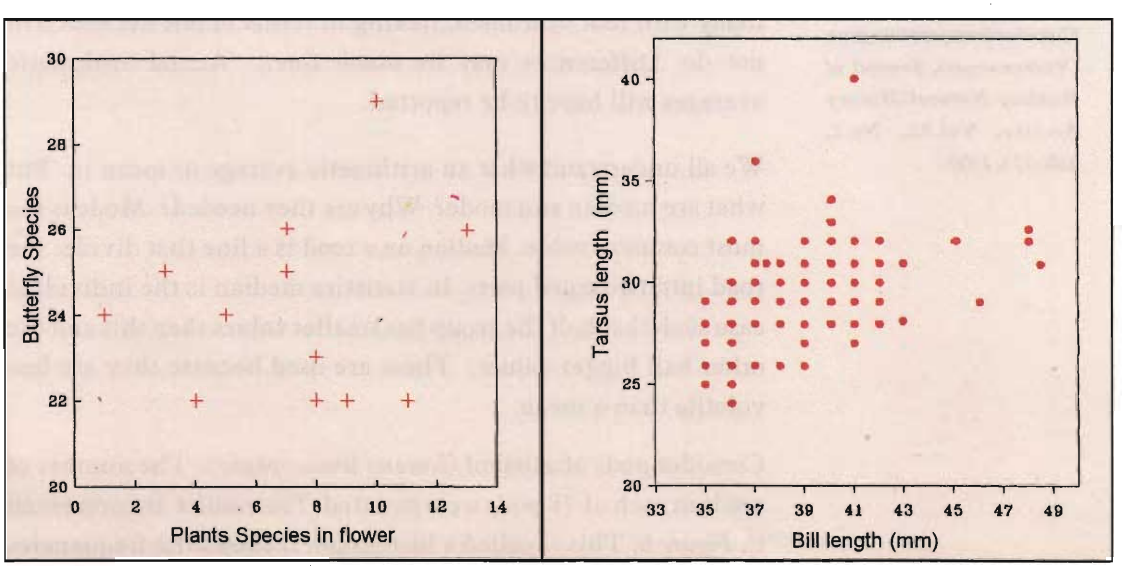


Figure 3. Described Indian species.

Scatter plot: Here is a puzzle : How are plants in flowering related to butterflies? That is trivial. Butterflies get nectar from flowers and plants get pollinated in return. Fine. Then how should number of plant species in flower be related to number of butterfly species present in an ecosystem? One guess is that greater the number of plant species in flower, greater should be the number of butterfly species present. *Figure* 4 gives the relevant data for Guindy National Park, Madras for 1991 [4]. One is struck by the fact that number of butterfly species occurring does not seem to be sensitive to number of flowering plant species. Perhaps the relation may hold for plant numbers and butterfly numbers, not their types. In that case a different scatter plot (i.e. two variables plotted against each other) may be needed. *Figure* 5 is an example of such a plot. It suggests that

Figure 4. (left) Number of butterfly species and plant species in flower in Guindy National Park, Madras.
Figure 5. (right) Scatter plot of body measurements of Garganey.



Suggested Reading

[1] C B Rao and C S Reddi, Pollination Ecology of *Martynia Annua L.*, *Journal of Bombay Natural History Society*, Vol.91, No.2, 187–193, 1994.

[2] P S Easa and M Balakrishnan, The Population Density and Structure of Asian Elephants in Parambikulam Wildlife Sanctuary, Kerala, India, *Journal of Bombay Natural History Society*, Vol.92, No.2, 225–229, 1995.

[3] R Gadagkar, World's Biodiversity Needs to Be Preserved, *Down to Earth*, Vol.1, No.11, 43–44, 1992.

[4] B Rajasekhar, A Study on Butterfly Populations at Guindy National Park, Madras, *Journal of Bombay Natural History Society*, Vol.92, No.2, 275–278, 1995.

[5] T B Reddy and C S Reddi, Butterfly Pollination of *Clerodendrum infortunatum* (Verbenaceae), *Journal of Bombay Natural History Society*, Vol.92, No.2, 166–173, 1995.

larger tarsus length is associated with larger bill length. (Remember this may not be true across species.)

We leave it to you to guess what a scatter plot of tusk length of a male elephant versus age may look like. Alternatively it could be shoulder height or pugmark circumference versus age. Can you imagine any use of this scatter plot for a field biologist?

One Number Says it All

Mean: It is true that a graph or chart can convey a message visually. But it is also essential to summarise data numerically. Consider the question of how long an organism lives. An *E. coli* bacterium perhaps lives for a few minutes or hours before dividing. An adult mosquito lives for a few days. Life span of a rice or a wheat plant is a few months. A cow or a dog lives for a few years. Man lives for a few decades. Banyan trees perhaps live for a few centuries. The longest living organisms on earth are redwood trees in California that are supposed to live a couple of thousand years. Now remember not all mosquitoes live for the same length of time, nor do all men. But a typical, representative figure is still very useful here for comparative purposes. If the need is more stringent, we may have to be more precise than above. If we wanted to compare the life span of men in India today with that of women, talking in terms of few decades will not do. Differences may be much finer. Actual arithmetic averages will have to be reported.

We all understand what an arithmetic average or mean is. But what are median and mode? Why are they needed? Mode is the most common value. Median on a road is a line that divides the road into two equal parts. In statistics median is the individual case such that half the group has smaller values than this and the other half bigger values. These are used because they are less volatile than a mean.

Consider pods of subabul (*Lucena leucocephala*). The number of seeds in each of 78 pods were counted. The results are presented in *Figure* 6. This is called a histogram. It shows the frequencies

of pods with different numbers of seeds. Mean, median and mode calculated for the data set are shown on the histogram.

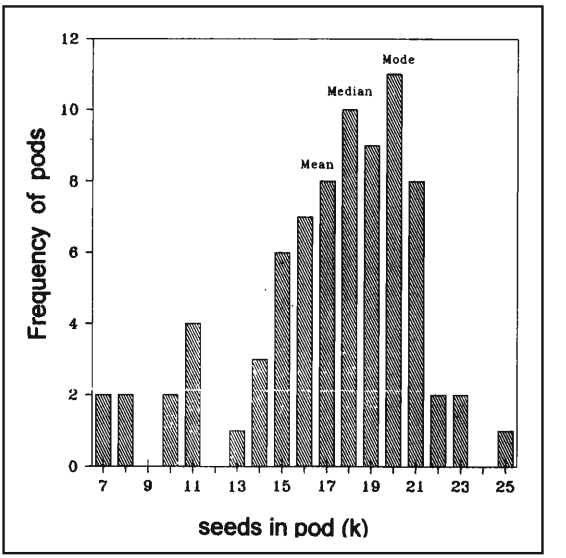


Figure 6. Frequency distribution of number of pods with k seeds/pod (Subabul).

Variation: One number that represents a whole set of values, like mean or median is called a measure of central tendency. But as noted earlier, variation is characteristic of biological data. Individual values are different from their average. In manufactured products, variation is undesirable. Variation shows lack of quality and lack of precision in manufacturing process. In nature, variation is the foundation of evolution. Because there is variation, some individuals are better adapted to environment. Hence natural selection favours them. If there were no variation, there would be no possibility of selecting better crop races/varieties. Variety is indeed the spice of life.

Poliomyelitus virus mutates very slowly and hence is stable. Vaccines have been developed successfully against it. Influenza virus mutates thousands of times faster. It has not been possible to develop a vaccine against it. Foot and mouth disease is a dreaded affliction of cattle. There is considerable variation in the genetic constitution of the pathogen. Hence different vaccines have to be injected for protection against different varieties.

So variation is important. Kelvin, the famous British scientist once remarked. '*If you can measure it then you understand it*', implying that if you cannot measure a thing, it is not understood well. How do we measure variation? One thing is range. Reddy and Reddi [5] measured proboscis length of butterflies foraging on the flowers of *Clerodendrum infortunatum*. The lengths in millimeter ranged from 8 to 38.

Range is useful as far as it goes. But it does not tell us the whole story. It indicates only the extremes. Sometimes that is enough. To build cages for rabbits it is enough to know the largest size

Length (mm)	No. of coots
≤190	111
190 – 200	1056
200 – 210	2926
210 – 220	1927
220 – 230	752
> 230	114
Total	6886

Mean = 208.7, Mode = 205.
Median = 208.9.

Table 1.

possible. For a domestic fish tank it is enough to know the range of lengths of fish species one wants to keep. On the other hand a fisherman is not satisfied to know only the range of sizes of a fish species, say Bombay duck, in his catch. He wants to know the number of fish in different size classes. The market price for each size class may be different. Range is the same if there is only one small fish and all others large or only one large fish and all others small. But the two cases are very different in terms of money they will fetch in the market. If we want to build a bridge over a river, we need to know the maximum water level. But in many other contexts extreme values are not enough. A store which sells shirts wants to stock various sizes. The manager needs to know how often each size will be asked for. In other words the entire distribution of shirt sizes has to be known.

In BNHS bird ringing project various body measurements were recorded for many birds. *Table* 1 shows the distribution of wing length of coots.

This frequency distribution is graphically shown in *Figure* 7. We notice that histogram is just like the bar chart encountered earlier except that here on the x-axis we have wing length values, instead of just some names. The histogram is sometimes called distribution (of wing lengths in this case). Above histogram shows that most birds have wing lengths near the average namely 209. As we go away from the average in either direction, the number of cases with that length declines.

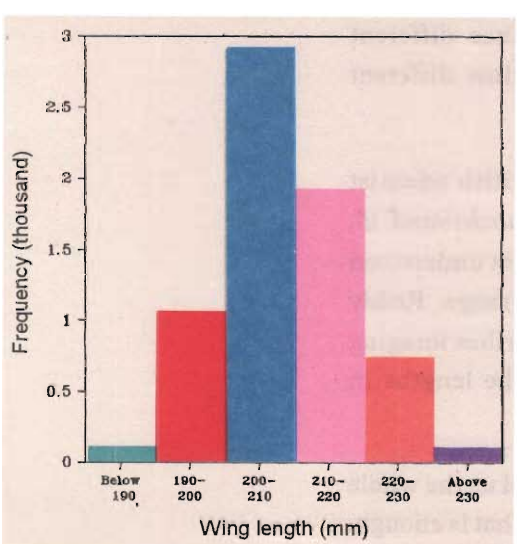


Figure 7. Frequency distribution of wing length of coots.

These features are conveniently captured in a measure called variance. The formula is not critical here and can be found in any standard textbook of statistics. This measure checks how far an individual value is removed from the average, squares the difference (to remove negative signs) and averages all such squares. Standard deviation (s.d.) is the square root of variance. It is very common in ecological

literature to report mean value of something together with s.d. e.g. for wing length data above s.d. is 7.9.

We will mention one more index encountered very frequently in ecological literature. That is the correlation coefficient.

Correlation Coefficient: Generally an older boy is heavier than a younger one. As ambient temperature grows warmer plants grow faster. These are examples where two measurements are related. The intensity of relation is measured by the correlation coefficient r. It takes values between -1 and $+1$. A value of zero indicates absence of relation. A +ve value suggests that as one quantity goes up so does the other, as in the above two examples. Can the reverse occur? That is to say when one thing goes up the other declines. An obvious example is prices and demand. When a fruit is expensive, not many of us buy it. As it gets cheaper we flock to the market. In the natural sphere, consider time budget of an animal. Since total time is fixed, if time spent on one activity is increased, that on another goes down. Thus a blackbuck male, during rutting season, must concentrate on protecting his status, access to females etc. and must be constantly ready to fight with challengers. He has then no time to feed.

Another example is time spent in brooding by Nilgiri laughing thrush. To help keep their chicks warm, Nilgiri laughing thrushes spend a lot of time on the nest (brooding). As the chicks grow, the time spent in brooding declines as seen in *Figure* 8. The correlation coefficient between age of chicks and percent time spent by parents in brooding is –0.98.

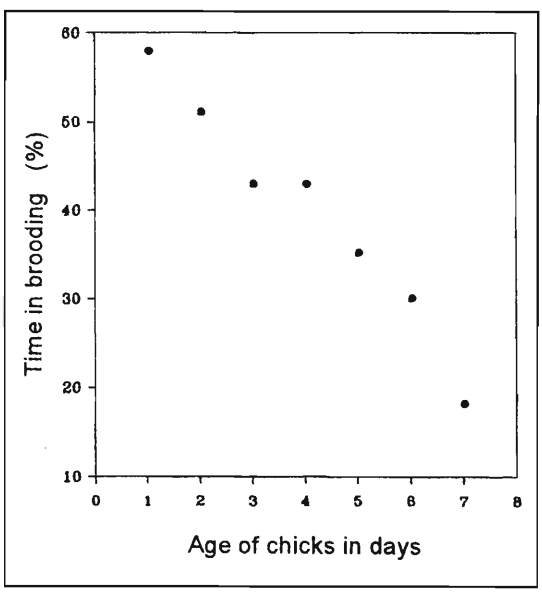


Figure 8. Time spent in brooding by G. cachinnans.

Exercises

In this year's budget of the Government of India, the receipts and expenditure are divided into various categories as shown in *Table* 2.

1. Prepare two pie charts using Excel.

Table 2. Percent contribution of various categories to receipts and expenditure.

Receipts		Expenditure	
Category	%	Category	%
Excise	20	Interest payment	27
Customs	15	Defence	14
Income tax	8	States share	14
Corporate tax	9	Central plan	13
Other taxes	2	Nonplan expenditure	13
Nondebt capital receipts	6	State and Union Territories plan assistance	10
Nontax revenue	15	Subsidies	7
Borrowing	25	Nonplan assistance	2
Total	100	Total	100

Suggested Reading for the series

[1] W B Fairley and F Mosteller, *Statistics and Public Policy*, Addison-Wesley, London, 1977.

[2] M O Finkelstein and B Levin, *Statistics for Lawyers*, Springer, NY, 1990.

[3] R Hooke, *How To Tell The Liars From The Statisticians*, Marcel Dekker, NY, 1983.

[4] A J Jaffe and H F Spirer, *Misused Statistics: Straight Talk For Twisted Numbers*, Marcel Dekker, NY, 1987.

[5] J M Tanur, F Mosteller, WH Kruskal, R F Link, R S Pieters and G R Rising (Eds.), "*Statistics, A Guide To The Unknown*, Holden-Day, San Francisco, 1972.

[6] H Zeisel, *Say It With Figures*, Harper, NY, 1957.

[7] H Zeisel and D Kaye, *Prove It With Figures*, Springer, NY, 1997.

Address for Correspondence
A P Gore & S A Paranjpe
Department of Statistics
University of Pune
Pune 411 007, India.

2. Prepare a pie chart of your personal activity time budget. Keep a diary of how you spend your time. Use a few simple categories like sleep, personal hygiene and nutrition, outdoor games, TV watching, interaction with other people outside place of work or education, reading etc. Draw pie charts for combined data as well as separate days of the week. Use Lotus/ Excel to draw the figure if you have an access to a PC.

3. Measure heights of all your classmates and draw a histogram of that data (separately for males, females and combined). Do you see any differences between sexes?

4. Number of words in a sentence is a simple quantitative feature of writing style. It is said that biblical style involves short and simple sentences. Long, complex sentences are characteristic of legal pronouncements. Obtain frequency distributions of sentence lengths in different kinds of writing such as (a) editorials in newspapers or this magazine, (b) sports columns, (c) tender and employment notices etc.

5. Students in a class differ from each other in very many ways. In an English medium school, the mother tongues of students may be different, or the district from which each one hails or one's favorite color etc. Select an attribute and prepare a bar chart (using Lotus).