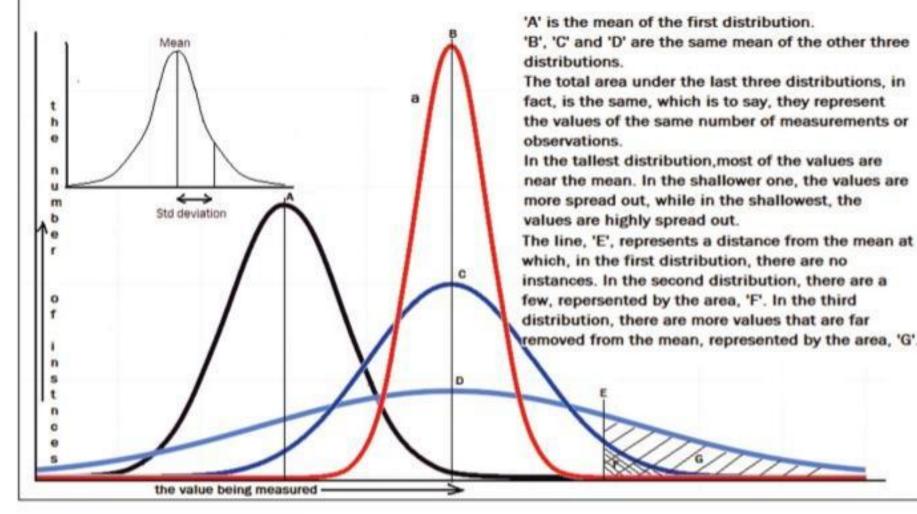


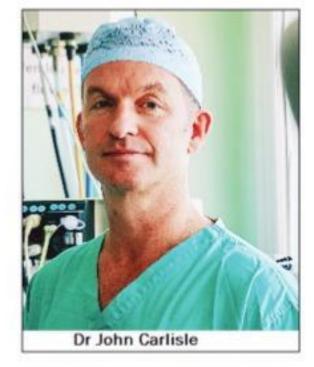
Numbers that reveal mappactice



S ANANTHANARAYANAN

a spring nor does one fine totle, fourth century BCE. A more recent English version is, swallow makes not a summer Support resarch claims "one swallow makes not a summer, nor one woodcock a winter". Both are adages to warn us against reliance on insufficient data. And so does the fable by Aesop, when it speaks of the man who sold his coat when he saw the first swallow and thought the winter was over. In the modern world, the reliability of data used for quality control, market forecast and for proving the effect of drugs has grown in importance. With automation of data collection and the power of computers, statistical analysis has become a routine activity and high levels of safety and economy have become possible. And so has competitiveness. But competition has pushed some researchers and the world of commerce to tweak the numbers and use statistics to make claims that are not justified. The journal, *Nature*, carried an article about the work on statistics by Dr John Carlisle, an anesthetist working in the general hospital of the seaside town of Torquay, South-west England. Dr Carlisle dabbles in statistical analysis and he uses his skills to look for features which help question the figures submitted by researchers as statistical support for their findings.

The properties of numbers ne swallow does not make can help detect instances of day," is attributed to Aris-tatle fourth contum PCF





effectiveness of drugs or the extent of their side effects.

Thus, where a random process should result in a certain distribution of measurements, researchers alter one or more conditions in the trial and see how the behaviour of the values



The methods that Dr Carlisle uses are based on the idea that "reallife data have natural patterns that artificial data struggle to replicate." These features were noticed in 1938 by Frank Benford, a US engineer-physicist, and have been used since then to check and verify statistical data.

A peculiar feature of the numbers that appear in all forms of data, known as Benford's law, is that the numbers are more likely to start with a smaller digit, like 1,2,3 than a larger digit like 7,8,9. The reason this happens can be understood if we see how quickly the first digits are repeated as we go from small to larger numbers. From 1 to 9, all digits appear at the same rate, just once. From 10 to 19, however, the digit 1 comes after just 8 numbers, and now it appears another 10 times as a first digit. This happens with the number 2 as well, but this number has to wait till 17 numbers have passed, for the decade, 20 to 29. Similarly, the number 3 has to wait till 26 numbers have passed, and so on. Again, the number, 1 is the first to be repeated as a first digit, a hundred times, from 100 to 199.

This feature of numbers that arise in random processes, like in accounting, are not maintained, for instance, when a dishonest accountant, falsifies numbers to siphon money out of the till. A simple, routine check, using computers, of how the first digits in a firm's ledgers are distributed, can reveal malpractice that may otherwise need intense verification to detect.

PLUS POINTS Thinking simply



Researchers at the Indian Institute of Technology, Hyderabad, have developed a simple and innovative method to synthesise a catalyst that can convert chemicals derived from biomass into bio-fuel precursors. The team has developed a novel process that can produce carbon catalysts at room temperatures using simple materials — sugar, sulphuric acid and salt.

This development is important for India, especially for states such as Uttar Pradesh and Andhra Pradesh-Telengana, the first and second largest producers of corn in the country respectively. They produce a large amount of corncob waste, much of which is currently burnt as its potential to be converted to valuable fuels is not harnessed owing to lack of awareness, expertise and technology. This research on cheap and efficient catalysts can enable this conversion, which would trigger the cascade of additional earning opportunity for the corn-farmer, provision of a sustainable energy source and reduction of carbon footprint in the automobile/aviation sector.

The research was led by Atul Suresh Deshpande, assistant professor, department of materials science and metallurgical engineering, IIT Hyderabad, and Sunil Kumar Maity, department of chemical engineering, IIT Hyderabad, along with their research students, D Damodar, K Alekhya and V Mohan. Their paper detailing the synthesis procedure and the catalytic efficiency of the material produced, has been recently published in the peer-reviewed journal, ACS Sustainable Chemistry and Engineering. In order to control the microstructure of carbon during the exothermic dehydration of sugar, the researchers added common salt. Salt helps in controlled dehydration of sugar, which leads to the formation of carbon nanoplates — plate-like structures that are a hundred thousand times smaller than a strand of human hair. Not only is this nanostructure important because of the large surface area, but the surface is also covered by sulfonate groups that makes it an active catalyst. The researchers believe that this approach can be further modified to obtain not just carbon nanoplates but also other types of nanostructured carbon materials that are useful in other commercial applications.



The objective of statistical analysis is to identify a "central tendency" of a collection of measurements, when exact measurement or calculation is not feasible. When things are measured in this way, it is of value to know how reliable this estimate of most likely measurement is, to stand in place of the correct value. Although no one measurement may be exactly the correct value, it is found that a collection of measurements crowd around the exact value, with measurements that are further from the correct value being less frequent.

A representation of the number of instances at the different values is found to have the bell shape, shown in the diagram above. The central value, which is taken as correct, is known as the "mean" of the distribution. When a large number of measurements are made, the "bell shape" of the distribution, which reflects the relative frequency of deviation from the mean, becomes smoother, and the "mean" is more reliable.

The set of three bell shapedcurves, on the right side of the diagram show three levels of "spread" or the uncertainty of a measurement

being close to the correct value. A measure of the spread of the data is a value called the Standard Deviation. This is the range, on the two sides of the mean, within which just over 68 per cent of the data falls. We can see that a low value of the SD indicates data that is close to the mean, while a large SD indicates data that is more spread out.

The "bell shaped" distribution and the SD, depend on the nature of the data and given a number of measurements that have been made, one set of measurements generally shows the same level of spread as another set of measurements. In fact, just as things like the SD help assess how good the mean of a set of random measurements has turned out to be, features of the measurements made also help to assess how truly random, or unbiased, the process of measurement has been.

This reverse process is what Dr Carlisle has been doing, with the data that has been published by researchers and pharmaceutical firms. Statistics are an important tool in the hands of researchers, both of basic physical principles as well as of the

has been affected. And there are ways of establishing whether any change in the behaviour is one caused only by chance or as a result of the changed factors. In the case of a test of a drug, the value measured could be the incidence or severity of disease and the factor that is varied could be the dosage of the drug administered.

But it may happen that a researcher wishes to show some result that the statistics do not really prove to shore up the effectiveness of a drug for a drug manufacturer. In these cases, the researchers are known to "doctor" the results of measurements, so that they show a shift in the mean or a change, which seems to be better than what could have come about by chance. This kind of dishonest reporting could set ongoing research on an incorrect path or, in the case of drugs, lead to wrong use or dosage, with serious consequences.

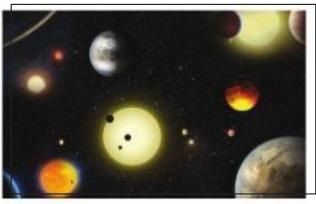
As an anesthetist, Dr Carlisle, and others, became concerned about the findings of one researcher, of the effectiveness of drugs to prevent nausea in patients recovering from surgery. When Dr Carlisle checked the data, he found that there were features in the "spread", which were not what should

be seen in trials that were conducted, in the correct, scientific manner, at random. The measurements were apparently selected in a way that they would support the conclusion that the researcher wished to prove. The University where the researcher worked was alerted, as also the journals that had carried his publications, and after investigation, as many as 183 papers by this researcher were found to be concocted, and were withdrawn.

The article on Dr Carlisle says the work he does, of analysing papers published the world over, has been steadily uncovering fraud after fraud in research. Although there have been instances where suspicions have been raised where they should not have been, Dr Carlisle's work, "together with the work of other researchers who doggedly check academic papers, suggests that the gatekeepers of science — journals and institutions — could be doing much more to spot mistakes," the article says.

The writer can be contacted at response@simplescience.in

Alien laboratory



Nasa has found three new planets including a kind of world unseen in our own solar system. The mysterious planets, part of the TOI-270 system, are "missing link" worlds and could be a huge gift to researchers looking for alien worlds, they said.

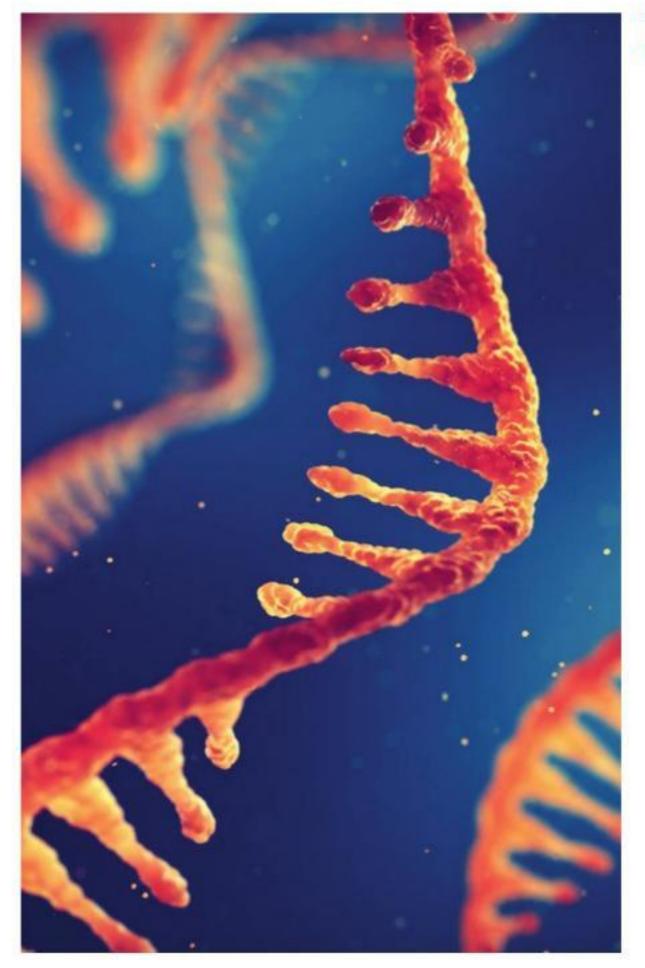
The three planets orbit a star that is only 73 light years away. It makes them among the closest exoplanets ever found, as well as being among the smallest.

They were discovered by researchers using Nasa's Exoplanet Survey Satellite, which was shot into space in 2018 and has been scanning the universe for stars and planets that could support alien life.

TOI-270 has a rocky super-Earth, which is slightly bigger than our planet, and two gaseous planets that are lightly larger. That makes them a "missing link" — sitting between the smaller rocky worlds like our own Earth or Mars, and the much larger gaseous planets such as Saturn and Jupiter. Alongside the Earth-like exoplanet are two gaseous worlds, just over twice the size of our own. One of those, the furthest from the star, is thought to sit in the temperature range that could allow it to support alien life — but its atmosphere is expected to be so thick and dense that it stores lots of heat, and the surface is probably too warm. Scientists hope to learn more about the relatively nearby worlds. As well as being close by and playing host to such unusual worlds, the star is unusually bright, without the solar flares and storms that can sometimes get in the way of observations. "TOI-270 is a true Disneyland for exoplanet science, and one of the prime systems Tess was set out to discover," said lead researcher Maximilian Gunther, from Massachusetts Institute of Technology, "It is an exceptional laboratory for not one, but many reasons — it really ticks all the boxes."

Determining order

An almost six decade-old experiment had found that messenger ribonucleic acid guides the synthesis of polypeptide chains



TAPAN KUMAR MAITRA

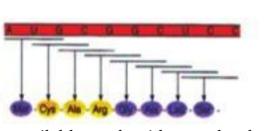
fter the publication of Francis Crick and Sydney Brenner's historic finding in 1961, it Look only five years for the meaning of each of the 64 triplets in the genetic code to be elucidated. Let us first, however, describe the role of ribonucleic acid in the coding system.

As we usually describe it, the genetic code refers not to the order of nucleotides in double-stranded DNA but to their order in the single-stranded mRNA molecules that actually direct protein synthesis. mRNA molecules are transcribed from DNA using a base-pairing mechanism similar to DNA replication, with two significant differences. First, in contrast to DNA replication, where both DNA strands are copied, only one of the two DNA strands — the template strand serve as a template for mRNA formation during transcription. The nontemplate DNA strand, although not directly involved in transcription, is by convention called the coding strand because it is similar in sequence to the single-stranded mRNA molecules that carry the coded message. Second, the mechanism used to copy sequence information from a DNA template strand to a complementary molecule of RNA utilises the same base-pairing rules as DNA replication, with the single exception that the base uracil (U) is employed in RNA where the base thymine (T) would have been incorporated into DNA. This substitution is permitted because U and T can both form hydrogen bonds with the base A. During DNA replication the base A pairs with T, whereas in transcription the base A pairs with U. Hence

Mutation in the DNA addition of one ucieotide pair (b) Overlapping mANA Wid-type

the sequence of an mRNA molecule is not exactly the same as the DNA coding strand, in that mRNA contains the base U anywhere the coding DNA strand has the base T.

How do we know that mRNA molecules, produced by this transcription process, are responsible for directing the order in which amino acids are linked together during protein synthesis? This relationship was first demon-strated experimentally in 1961 by Marshall Nirenberg and J Heinrich Matthei, who pioneered the use of cell-free systems for studying protein synthesis. In such systems, protein synthesis can be studied outside living cells by mixing isolated ribosomes, amino acids, an energy source, and an extract containing soluble components of the cytoplasm. Nirenberg and Matthei found that adding RNA to cell-free systems increased the rate of protein synthesis, raising the question of whether the added RNA molecules were functioning as messages that determined the amino acid sequences of the proteins being manufactured. To address this question, they decided to add synthetic RNA molecules of known base composition to the cell-free system to see if such RNA molecules would influence the type of protein



bles available nucleotides randomly into a linear chain. If only one or two of the four ribonucleotides (ATP, GTP, CTP, and UTP) are provided, the enzyme will synthesise RNA molecules with a restricted base composition. The simplest RNA molecule results when a single kind of nucleotide is used, because the only possible product is an RNA homopolymer consisting of a single



being made. Their initial experiments took advantage of an enzyme called polynucleotide phosphorylase, which can be used to make synthetic RNA molecules of predictable base composition. Unlike the enzymes involved in cellular transcription, polynucleotide phosphorylase does not require a template but simply assem-

repeating nucleotide.

For example, when polynucleotide phosphorylase is incubated with UTP as the sole substrate, the product is a homopolymer of uracil, called poly (U). When Nirenberg and Matthei added poly (U) to a cell-free protein-synthesising system, they observed a marked increase in the incorporation of one particular amino acid, phenylalanine, into polypeptide chains. Synthetic RNA molecules containing bases other than uracil did not stimulate phenylalanine incorporation, whereas poly (U) enhanced the incorporation of only phenylalanine.

From these observations, Nirenberg and Matthei concluded that poly (U) directs the synthesis of polypeptide chains that consist solely of phenylalanine. This observation represented a crucial milestone in the development of the messenger RNA concept, for it was the first demonstration that the base sequence of an RNA molecule determines the order in which amino acids are linked together during protein synthesis.

The writer is associate professor and head, department of botany, Ananda Mohan College, Kolkata

³The independent



