**////Title: A Statistical Approach to the Reproducibility Crisis**

**////Standfirst:**

Many fields of experimental research are now facing a daunting challenge: despite the fact that separate research teams may work on the same system, the end results of their experiments can be widely varied. In his research, Dr Stanley Luck, at the Science, Technology and Research Institute of Delaware, explores the reasons behind this inherent inability to reproduce experimental results. Based on his analysis, he now suggests two key requirements that must be met to overcome this crisis, which could have a profound influence over the techniques employed by experimental researchers in the future.

**////Main text:**

Currently, many fields of research are undergoing what experimentalists have dubbed a 'reproducibility crisis' in science. As they analyse the characteristics of deeply complex systems, researchers are often finding that their results greatly differ from those gathered in previous studies, even though the same system was being measured.

The consequences of this crisis are felt particularly strongly in medicine and the social sciences – which aim to study the deeply intricate processes taking place within the human mind and body. Yet its effects also extend to many other fields, across biology, chemistry, and physics.

In recent years, statisticians have begun to propose that the mathematical techniques we use to describe experiments on complex systems could lie at the root of the problem. In order to move past the reproducibility crisis, it may be crucial for experimentalists to rethink these methods entirely. In a recent study, Dr Stanley Luck explores this problem in detail. By clearly laying out the statistical challenges involved, his work could open up new routes to reaching this important goal.

…

Within complex systems of variables, close relationships can form between pairs of interacting elements, so that any change in one variable can directly affect another, to which it is inherently linked.

In statistics, the strength of a relationship between two particular variables can be described using a term named their 'effect size'. Another important factor to consider when experimenting on a system is its 'P-value', which relates to the probability of obtaining a result that differs significantly from other measurements of the system. If the value is small, then extreme outcomes to the experiment become highly unlikely.

Although both of these values are key concepts in the statistical analysis of experiments on complex systems, Dr Luck suggests that the methods we use to acquire them could be providing misleading interpretations. Based on this idea, he has now set out two important requirements for ensuring the reproducibility of experimental results: firstly, relating to the influences of unbalanced sample sizes; and secondly, to the number of values involved in the calculation of a system that are allowed to vary.

…

Before they carry out their experiments, researchers must first decide how many observations they should make of the system being analysed. Since the experimental conditions of different studies can vary widely, even when they are measuring the same system, this number of observations made in different experiments – named 'sample sizes', can themselves be highly variable. In turn, this imbalance can have an inconvenient influence on calculations of the effect size – making it far harder for researchers to assess the strengths of relationships between different variables.

Today, this dilemma has been widely discussed within the field of statistics. To express the problem, researchers use a mathematical construct named a 'contingency table': a format that displays the frequency of each measured outcome of an experiment. By summing all entries to the table together, researchers can then determine the system's overall effect size. If the sizes of samples produced in different experiments are more unbalanced, Dr Luck shows measurements of the effect size can vary widely – with significant negative consequences for experimental reproducibility.

…

The second aspect that Dr Luck explores in his study relates to an aspect of complex systems named their 'degrees of freedom'. This term describes the number of values necessary to describe outcomes of an experiment that are free to vary, without violating any stringent constraints on the system. When the measured frequencies of experimental outcomes are again displayed in a contingency table, these variables can be directly related to other aspects of the system that can never change, even if its variables are each multiplied by a common factor.

In real experiments, there is an inherent trade-off between the benefits of varying certain aspects of the variables enabled by a system's degrees of freedom, and the costs required to implement those changes. As they assess whether or not an observed variation is large enough to be meaningful to their experimental results, it is important for researchers to specify this trade-off. As a result, mathematical considerations alone aren't enough to make accurate measurements of the system's effect size, as determined by summing the values contained in its contingency table. Again, this damages the reproducibility of experimental results.

…

From this analysis, Dr Luck has identified two key challenges which must be overcome before the reproducibility crisis can be solved. Firstly, he suggests that researchers must adapt the parameters they use to acquire data, in order to account for imbalanced sample sizes. Secondly, his findings indicate the need for researchers to always account for all possible degrees of freedom when measuring systems, regardless of their experimental requirements.

Ultimately, Dr Luck argues that resolving both of these problems will have broad implications for statistics, and experimental research as a whole. If both requirements are met, researchers could ensure that the results they gather in future experiments can be fully recreated in subsequent studies. In turn, a robust new set of guidelines could be created – providing a crucial first step towards a resolution in the current crisis in reproducibility.

…

This SciPod is a summary of the paper 'Factoring a 2 x 2 contingency table', from *PLOS one*.
doi.org/10.1371/journal.pone.0224460

For further information, you can connect with Dr Stanley Luck at stan.luck@vectoranalytics.ai