nature neuroscience

Running the numbers

Brains are extremely sensitive pattern detectors—so sensitive, in fact, that they often detect patterns that do not actually exist[1]. Daniel Kahneman, professor of psychology at Princeton University, won the 2002 Nobel Prize in economics for his studies of the cognitive shortcuts that lead to such errors. Most scientists are aware that people do poorly at detecting biases in conclusions based on incomplete or unrepresentative data, and thus evaluate their hypotheses by formal statistical testing. Unfortunately, the same cognitive mechanisms that lead to pattern detection errors in everyday life, such as a tendency to interpret new information as supporting one's current beliefs, can also lead to faulty intuitions about the correct application of statistical tests.

To help assure readers that the conclusions of our papers do not reflect such biases, the Nature journals have developed a set of guidelines for the analysis and reporting of statistics in our pages, which can be found on our website at http://www.nature.com/neuro/authors/index.html, along with a checklist for authors. Some of the guidelines simply aim to ensure that the statistical evidence for each finding is clearly described: what tests were used, how many samples were evaluated in each condition, which comparisons were done, and what significance level was found (reported as the actual P value, not merely "$P < 0.05$"). Graphs should include error bars, clearly labeled as standard error or standard deviation. The Methods section of all papers that include statistical testing should contain a subsection describing the analysis. We will make sure that all the required information is presented in the final version of the paper.

We are also instituting a standard set of requirements for the statistical analysis itself that editors and referees will evaluate before a paper is accepted for publication. In particular, all data sets should be summarized with descriptive statistics, including a measure of center, such as the mean or median, and a measure of variability, before further analyses are done. Authors will be asked to justify their choice of analysis and the exclusion of any data points, and to confirm that their data conform to the assumptions underlying the tests that were used. We invite referees to point out any potential areas of statistical concern and let the editors know if they feel that a particular paper needs to be evaluated by a statistics expert.

Following the new guidelines should help authors avoid several common statistical errors. One of the most widespread is the use of multiple comparisons, which increases the risk of false-positive results. For example, carrying out a series of pairwise comparisons by t-tests gives a higher chance of a falsely 'significant' result (because each test has a 1-in-20 risk of a false positive at $P < 0.05$, one would expect 1 false positive out of every 20 tests performed) than evaluating the same data with a single analysis of variance (ANOVA) at the same significance level. Along the same lines, analyses of functional imaging data should be corrected for multiple comparisons to account for testing across multiple voxels. Another common error is the failure to recognize that most parametric tests require the data to be normally distributed. If this assumption is not valid for a particular data set, a nonparametric test should be used instead. Another point that is less widely recognized is that ANOVAs require approximately equal variance across the different groups or conditions examined. If this assumption is violated (as it frequently is, because variance often scales with the mean), a nonparametric test or appropriate transformation of the data is necessary. Finally, researchers should take care to choose the correct statistical tests for small data sets (roughly $n < 10$), for which ranges are a more appropriate measure of variability than standard deviations or standard errors.

No set of general guidelines can protect against all possible sources of statistical error, of course. Needless to say, all scientists should understand the reasoning behind their analysis, rather than leaving the choice of test to the discretion of their favorite statistical software package. Segregating data into subgroups is another common source of error and bias. Researchers should select data to analyze as a subgroup with care, preferably based on an independent variable, rather than, for instance, sorting their data into 'high-responding' and 'low-responding' subject groups for further analysis. Negative findings should be stated with caution, and if critical to the conclusions, supported by a power analysis that indicates that the number of subjects would be adequate to detect an effect of the expected size. In many universities, statistical experts are available for consultation with researchers in other departments who need help in designing experiments and analyses, preferably before the data are collected.

Even careful neuroscientists have a tough task, however, because they often work with data that are mathematically complex and thus difficult to analyze correctly. Spike rates tend not to be normally distributed and not to have equal variances across groups. They also violate an underlying assumption of cross-correlation analysis: that data are 'stationary'—meaning that their stochastic properties do not change with time[2]. Another assumption often violated in neuroscience is that data points are independent of one another; because neurons are highly interconnected, physiological responses are often correlated.

There is nothing shameful about cognitive shortcuts; during our evolution, it has often been more adaptive to be able to evaluate a situation and respond quickly than to get the precisely correct answer. Using rigorous analysis methodology allows scientists to bypass the potential bad consequences of this tendency and get the right answer most of the time. We hope that our new statistical guidelines will contribute to this effort, and welcome feedback on the new policy, which can be sent to the editors at neurosci@natureny.com. ∎

1. Gilovich, T. *How We Know What Isn't So: The Fallibility of Human Reason in Everyday Life* (Free Press, New York, 1993).
2. Brown, E.N., Kass, R.E. & Mitra, P.P. *Nat. Neurosci.* **7**, 456–461 (2004).