Misapplication of a Statistical Test

Comment on "Lies, Damned Lies, and Statistics (in Geology)"

PAGE 65

In his Forum, P. Vermeesch (*Eos, 90*(47), 443, doi:10.1029/2009EO470004, 2009) argues that "the strong dependence of p values on sample size makes them uninterpretable" with an example where p values in a hypothesis test using Pearson's chi-square statistic differed by a factor of 10^{16} when the sample size decreased tenfold. The data were a sequence of magnitude 4 or larger earthquake events (N = 118,415) spanning 3654 days [*U.S. Geological Survey,* 2010].

There are two problems with the analysis. First, Vermeesch applied the chi-square test to data with statistical properties that are inconsistent with those assumed in the derivation of the chi-square test. Second, he made an assumption that, using a straightforward calculation, can be shown to be inconsistent with the data. I address here only problems related to the application of statistics without reference to any additional physical processes that may also need to be addressed before statistical analysis is performed (e.g., the physics of how aftershocks are related to main earthquakes).

First, Pearson's chi-square test applies to an experiment in which n independent measurements were made and each measurement falls into one of k categories. An analogous experiment would be n rolls of a seven-sided die, with each side labeled with a day of the week. The probability of two sequential rolls having identical labels is 14%. In the earthquake event list, two events in a row with identical day-of-week labels occur with probability 97%, which is not consistent with the independence assumption. As shown in the online supplement to this *Eos* issue (http://www.agu.org/eos_elec/), such a lack of independence may or may not result in a chi-square test rejecting the hypothesis of a uniform distribution on a histogram that was actually created by draws from a uniform histogram. The result depends on whether or not the events in question give a histogram with amplitudes that are Poisson-distributed, and nonindependent data do not necessarily create a histogram with Poisson-distributed amplitudes.

Second, Vermeesch assumed that if 10 times fewer measurements were used, chi-square would change from 93 to 9.3 when arguing that p values depend on sample size. This assumption is inconsistent with the data. In the case of the earthquake data, each day of the week occurred 522 times in the span of time selected. A new histogram with 10 times fewer data can be created by sampling with replacement only 52 of the available 522 days for each day of the week. For 1000 histograms created this way, chi-square is 70 ± 45 and not 9.3 as assumed.

Reference

U.S. Geological Survey (2010), U.S. Geological Survey Earthquake Data Base, PDE catalog, http://earthquake.usgs.gov/earthquakes/eqarchives/epic/epic_global.php, accessed on 31 March 2010, Reston, Va.

—Robert S. Weigel, Department of Computational and Data Sciences, George Mason University, Fairfax, Va.; E-mail: rweigel@gmu.edu

A Closer Look at Data Independence

Comment on "Lies, Damned Lies, and Statistics (in Geology)"

PAGE 65

In his Forum (*Eos, 90*(47), 443, doi:10.1029/2009EO470004, 2009), P. Vermeesch suggests that statistical tests are not fit to interpret long data records. He asserts that for large enough data sets any true null hypothesis will always be rejected. This is certainly not the case! Here we revisit this author's example of weekly distribution of earthquakes and show that statistical results support the commonsense expectation that seismic activity does not depend on weekday (see the online supplement to this *Eos* issue for details (http://www.agu.org/eos_elec/)).

To test if earthquakes are uniformly distributed over days of the week, we formed the series of daily earthquake occurrences and randomly shuffled its members to compute synthetic histograms of cumulative earthquake occurrences tallied by day of the week. We found that the resulting 95% confidence interval of earthquake tallies (15,897–18,076) contains the observed range of earthquakes, which had, when accumulated over the 10 years of data, a minimum of 16,349 occurrences on Friday and a maximum of 17,752 occurrences on Sunday. Hence, our test fails to reject the null hypothesis of uniform earthquake distribution throughout the week.

Why does the above test produce results different from those of the chi-square testing by Vermeesch? We argue that the effective number of independent observations (n^*) is less than the total number of earthquakes (n = 118,414). Vermeesch implicitly assumed that n^* equals n. However, closer inspection of the earthquakes used in Vermeesch's analysis shows that they exhibit periods of clustering that correspond to nonindependent aftershock sequences of strong earthquakes superimposed on the background of normal seismic activity. Thus, finding n^* meant statistically accounting for the fact that some of the earthquakes were linked and effectively eliminating them from the pool of events being analyzed. Such analysis revealed that only 10% of the events used by Vermeesch were actually independent ($n^* = n/10$; see the online supplement for more details on how we found n^*). Vermeesch's chi-square test that uses this new n^* value supports the notion that earthquakes do not depend on day of the week.

Similar results were obtained when we counted cumulative earthquake occurrences during each hour of the day instead of each day of the week and asked if earthquakes preferentially favored a certain time of day. They do not, as can be shown, for example, via chi-square testing using our estimated n^*, because n^* should not depend on the way the data are binned.

In summary, while the large databases permit identifying small-magnitude phenomena, care should be taken to ensure that the assumptions underlying statistical tests, such as data independence, are satisfied. Failing to do so may result in false rejection of null hypotheses.

—Sergey Kravtsov and Rolando Olivas Saunders, Department of Mathematical Sciences, University of Wisconsin-Milwaukee; E-mail: kravtsov@uwm.edu