

Re: How to determine if a number is statistically meaningful

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2005-03/0304.html>

From: Richard Ulrich (*Rich.Ulrich_at_comcast.net*)

Date: 03/22/05

Date: Mon, 21 Mar 2005 19:11:34 -0500

On 21 Mar 2005 14:00:16 -0800, "davegb" <davegb@safebrowse.com> wrote:

- > *I've posted another thread about this situation, but I have a separate*
- > *question concerning the same dataset.*
- > *I have statistical info about state clients for the 64 counties in the*
- > *state. I have the total number of clients over a year, and the number*
- > *of clients who'd had a specific occurrence during that time. The*
- > *statewide average for this occurrence is about 4%. Of course, I have*
- > *some small, rural counties with only a few clients (and a few with*
- > *none, which doesn't matter). If 1 of 4 clients experiences the*
- > *occurrence, it's statistically misleading to report on the spreadsheet*
- > *that the county had a 25% occurrence rate. I'd like to suppress those*
- > *numbers when they would be misleading.*
- > *How do I determine the threshold at which the numbers become*
- > *statistically relevant? Or the inverse, irrelevant?*

There is not any magic in "statistics" to determine a threshold.
The best answer is probably what you get when you Ask an expert.

You might find a little bit of answer, or else some comfort, by reading the notes that the International Movie Data Base provides concerning their own "weighted averages" for rankings of movies. See <http://imdb.com/> and look for "voting"; try to vote for some movie, then look for information.

IMDB provides, among other things, a formula for a "Bayesian average." – Effectively, this regresses each score "toward the mean", but it is hardly noticeable except where the N is small.

- (a) No score is reported if the Movie/ person has fewer votes than some cutoff MIN (which varies, for categories).
- (b) The reported score is a weighted average of the actual votes, as moderated by adding in an additional MIN votes as occurring at the average score.

IMDB will not report any average for a movie unless there are at least 5 votes; it will not report a movie as being one

sci.stat.math: Re: How to determine if a number is statistically meaningful

of the "bottom 100" unless there are 625 votes; it will not report a movie as being one of the "top 250" by using the total number of votes, but only those from "regular voters"; the smallest N that I see is 1765. -- The ratings are most often lower by 0.2 points in the list of 250, compared to their averages for "all voters."

There, in IMDB, you can see a system that works pretty well – it has a number of *arbitrary* cutoffs, but there is not any better way.

> *Any help here would be greatly appreciated. I'm not a statistician, but*
> *have an engineering degree and still remember a little statistics, but*
> *not much! But at least I can figure out the math, and use the functions*
> *in Excel to make all this happen.*

--
Rich Ulrich, wpilib@pitt.edu
<http://www.pitt.edu/~wpilib/index.html>