

Re: Determining Confidence Interval

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2004-09/0570.html>

From: Paige Miller (paige.miller_at_kodak.com)

Date: 09/29/04

Date: Wed, 29 Sep 2004 08:40:32 -0400

Roger Sherman wrote:

- > *Thanks for indulging my ignorance. Put as succinctly as I can and*
- > *allowing for my limited knowledge of statistics, given a "specific*
- > *result" (number of events or instances out of a total number of*
- > *occurrences), how large a sample of total occurrences will be required*
- > *to obtain a high level of confidence in the "specific result."*
- >
- > *More to the point, in the play of cards, "specific events" take place.*
- > *For the sake of this exercise, I have determined arbitrarily that out*
- > *of the total number of occurrences (not defined at this point), the*
- > *"specific events" take place 20 % of the time. I similarly could have*
- > *expressed the situation in this way: Out of 100 total occurrences,*
- > *the "specific events" occurred 20 times. Can I have a high level of*
- > *confidence that this result reflects a consistent trend or would I*
- > *need more than 100 total occurrences (200, 3001000) in order to*
- > *have confidence that the "specific events" take place 20% percent of*
- > *the time ??*
- >
- > *Finally, I would appreciate any input you could provide regarding the*
- > *methodology that could be used to sort out the issues raised above.*

Let's back up. Confidence intervals are appropriate when you have some data, and you estimate something (often a mean or a percent) from the data.

So, when you say ... a certain event in a game of cards occurred 20 times out of 100 ... that certainly could be the thing you want a confidence interval for. However, the questions you are asking still are way off the mark for what statistics can tell you.

A confidence interval would tell you that although you observed 20 out of 100, repeated sampling would provide results in the range x to y 95% of the time (or some other percent). An hypothesis test would tell you if the data you actually observed was consistent (or not) with a true event rate of z times out of 100.

sci.stat.math: Re: Determining Confidence Interval

Your question: "Can I have a high level of confidence that this result reflects a consistent trend or would I need more than 100 total occurrences (200, 300 ... 1000) in order to have confidence that the 'specific events' take place 20% percent of the time ??" This is meaningless from a statistical point of view. We don't say we have confidence in an event, or in a sequence of events; sometimes we say that we have a certain level of confidence in an hypothesis test, or that we have a confidence interval about our observation. If your hypothesis is that the true probability of this event is 0.2 (20 out of 100), then you can have as high a level of confidence as you desire, but this is backwards ... we don't observe a value and then make that value our hypothesis, we state a hypothesis first and then see how well the data conforms to that hypothesis.

Basically, any introductory statistical text will cover hypothesis testing, confidence intervals and sample size in much more detail than I can cover it here.

--

Paige Miller
Eastman Kodak Company
paige dot miller at kodak dot com
<http://www.kodak.com>

"It's nothing until I call it!" -- Bill Klem, NL Umpire
"When you get the choice to sit it out or dance, I hope you dance"
-- Lee Ann Womack