Re: Statistical Ranking for Non−Normal Populations

Source: http://sci.tech−archive.net/Archive/sci.stat.math/2004−10/0287.html

---

From: Peter Hach (*Willy.Yoda_at_gmx.net*)
Date: 10/16/04

```
Date: Sat, 16 Oct 2004 02:07:05 +0000 (UTC)
```

On Thu, 14 Oct 2004 22:36:57 −0400, Richard Ulrich wrote:
>On Wed, 13 Oct 2004 18:30:39 +0000 (UTC), Willy.Yoda@gmx.net (Peter
>Hach) wrote:
>
>> I need to perform (statistical) ranking of a number of large, but
>> finite popolations X[i] = (x[i][1], ... ,x[i][n]) in a scenario
where
>> acquiring each x[i][j] is very expensive. I am looking for the
>> population X[i] with the smallest Sum or Average over the x[i][j]
>> (i.e. I am only interested in the top−ranked one). Furthermore, all
>> x[i][j] are strictly larger than 0.
>>
>> I've started looking into the statistical ranking techniques, and
most
>> work I've seen assumes that samples are generated from normal
>> distributions (and their variances are equal). I suspect this is
>> because in these cases
>>
>> Sample−Mean − True−Mean
>> −−−−−−−−−−−−−−−−−−−−−−− ~ student's t distribution
>> sqrt(Sample−Variance/(n))
>>
>> i.e. one is not reliant on knowing the true Variance...?
>> Unfortunately, my data is non−normal and the X[i]s may differ
>> significantly in variance.
>
>As I read it: You want to be able to state what you assurance
>you may have that a particular 'finite population', among several
>of the same size, will have the largest mean, based on a partial
>sampling.
>
>The original distribution being non−normal is not much
>of a problem, so long as the sum is well−behaved.
>
>Differences in variance may not be as much of a problem
>if the samples are from the same family. Do you know

>they are always finite? Is there much systematic