Subject: Re: Finding the index of the median
Posted by davidf on Tue, 29 Oct 1996 08:00:00 GMT
View Forum Message <> Reply to Message

David Foster <foster@bial1.ucsd.edu> writes in this ever lengthening thread:

> The MEDIAN() function isn't returning "one" of those locations.
> The very fact that it's the median of the array means that it's
> likely that there are many indices with that value. It doesn't
> make sense to expect to find THE index. If there are 5 indices
> with the median value, there's no way to distinguish between
> them.

Perhaps we are all missing Dean's point. Take a hypothetical situation. Suppose I have a scheme to steal some money from the bank. But my scheme requires that there be *exactly* one million dollars in the account I steal from. And suppose I have a function like the WHERE function that can tell me, out of all the accounts in the bank, which ones have exactly one million dollars.

Most of the posters to this thread seem to be arguing that it doesn't make one bit of difference which account I steal the money from, one would appear to be just as good as the other.

But of course, it probably does matter. In fact, given the choice, I would rather steal from an account that had little activity since this would give me more time to reach the South Islands without my little scheme being discovered.

I seem to remember that in Dean's original question he had a parallel array that contained some kind of noise estimate. Maybe in my hypothetical example I have a list of all accounts and how many transactions they have had in the past month. What I really want to know is what accounts have exactly one million dollars in them, and of *those* which provides me with the best chance of escaping with the money (i.e, have the fewest transactions in the past month). (Does this sound like a physics problem yet?)

Suppose it looked like this (where the number 10 represents exactly one million dollars):

```
money =    [ 4, 10, 3, 6, 10, 8, 10]
activity = [ 3,   4,  2, 8,  2, 2,   7]
```

I could use IDL to tell me which million dollar account to steal

from like this:

First, find the million dollars accounts:

```
index = WHERE(money EQ 10)
```

Next, sort the activity accounts of these million dollar accounts
based on number of transactions:

```
sorted = SORT(activity(index))
```

Finally, print the number of the account I should steal from:

```
PRINT, (index)(sorted(0))
```

Walla! In this example, the number 4 is printed, which is *exactly*
the index of the account I want (i.e, it has a million dollars in it
and has the fewest number of transactions of all the million dollar
accounts).

Now, you can argue that there might be multiples of this too. But
I can apply the same technique with some other criteria, say the
date of the last transaction, etc.

In principle (now it *really sounds like a physics problem!) I could
narrow it down to a specific account. Or I could finally say it doesn't
make a damn bit of difference and spend the money however I please!

Perhaps what Dean wants to know is, of all the median values,
which one has the smallest amount of noise associated with it.

Chao!

David

--
David Fanning, Ph.D.
Phone: 970-221-0438
Fax: 970-221-4728
E-Mail: davidf@fortnet.org