Subject: Finding the index of the median
Posted by Dean Schulze on Fri, 25 Oct 1996 07:00:00 GMT
View Forum Message <> Reply to Message

The median function will return the median value of an
array. Is there anyway to obtain the index into the array
of the median value? For example, if I find the median in an
array of measured values how can I get the corresponding value
from the companion noise array?

Dean Schulze

== ==========
 "You plot the growth of the NEA [National Education Association] and
 the dropping of SAT scores, and they're inversely proportional.
 The problems are unions in the schools. The problem is bureaucracy.
 I'm one of these people who believes the best thing we could ever do
 is go to the full voucher system."

 Steve Jobs
 Wired Magazine, Feb. 96
 http://www.hotwired.com/wired/4.02/features/jobs.html
== =========

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

Dean Schulze wrote:

>
>   That is exactly why the WHERE() function won't work.
> I need to know which one of those locations is returned
> by the MEDIAN() function.
>
> Dean Schulze

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.

 ~~~~~~ ~~~~~~
  David S. Foster   Univ. of California, San Diego

Programmer/Analyst   Brain Image Analysis Laboratory
foster@bial1.ucsd.edu Department of Psychiatry
(619) 622-5892   8950 Via La Jolla Drive, Suite 2200
                La Jolla, CA 92037
                [UCSD Mail Code 0949]
~~~~~~ ~~~~~~

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

David Fanning wrote:

> Be aware that there can be multiple locations in your
> array that are equal to the median value. The index that
> is returned by the WHERE function will be an array of
> all of those values.

   That is exactly why the WHERE() function won't work.
I need to know which one of those locations is returned
by the MEDIAN() function.

Dean Schulze

== ==========
 "You plot the growth of the NEA [National Education Association] and
 the dropping of SAT scores, and they're inversely proportional.
 The problems are unions in the schools. The problem is bureaucracy.
 I'm one of these people who believes the best thing we could ever do
 is go to the full voucher system."

 Steve Jobs
 Wired Magazine, Feb. 96
 http://www.hotwired.com/wired/4.02/features/jobs.html
== ========

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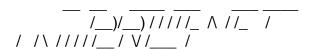View Forum Message <> Reply to Message

In article <32769981.73B9@cassini.lpl.arizona.edu>,
Dean Schulze <schulze@cassini.lpl.arizona.edu> wrote:
> Joseph M. Zawodny wrote:
>>
>> Dean Schulze wrote:

>>>
>>> David Fanning wrote:
>>>
>>>> Be aware that there can be multiple locations in your
>>>> array that are equal to the median value. The index that
>>>> is returned by the WHERE function will be an array of
>>>> all of those values.
>>>
>>>   That is exactly why the WHERE() function won't work.
>>> I need to know which one of those locations is returned
>>> by the MEDIAN() function.
>
>> MEDIAN does not return an element or LOCATION, it returns a
>> VALUE which may be held by one or more elements.
>
>   Sorry, careless writing on my part. I should have said
> "I need to know the location of the value that is returned
> by MEDIAN()".
>
 >

 (snipped to save space)

>
>   An unusual situation, but one that MEDIAN() and WHERE() don't
> seem to work with.
>
>
> Dean Schulze
>

Dean
Implement your own median

let data = your data
   noise = associated noise array
   n = n_elements(data)
   s = sort(data)
   median = data(s(n/2))
   associated_noise_value = noise(s(n/2))
  This is pretty much what median does.

I think that does it.
Let me know if I'm wrong.
WHD
--
  -- ------------------

William Daffer
daffer@primenet.com

```
 __  __   ____  ___      ___  ____
/__)/__) / / / /_  /\  / /_   /
/  /\ / / / /__/ \/ /___  /
```

Subject: Re: Finding the index of the median
Posted by davidf on Tue, 29 Oct 1996 08:00:00 GMT

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

Subject: Re: Finding the index of the median
Posted by Dean Schulze on Wed, 30 Oct 1996 08:00:00 GMT
View Forum Message <> Reply to Message

Thanks to all of you who have responded to this thread.

    As some of you have pointed out it only makes sense to
distinguish between multiple median values if there is some
other criterion, which in my case is the location on a CCD.
(Temperature gradients and edge effects depend on location.)

    I was being too pedantic in asking for the location
along with the value from the MEDIAN() function since I
can assign any of the corresponding noise values to it
that WHERE() returns and still be correct - for that one
location.

Dean Schulze

 === ==========
  "You plot the growth of the NEA [National Education Association] and
  the dropping of SAT scores, and they're inversely proportional.
  The problems are unions in the schools. The problem is bureaucracy.
  I'm one of these people who believes the best thing we could ever do
  is go to the full voucher system."

   Steve Jobs
   Wired Magazine, Feb. 96
   http://www.hotwired.com/wired/4.02/features/jobs.html
 === =========

Subject: Re: Finding the index of the median
Posted by haferman on Wed, 30 Oct 1996 08:00:00 GMT
View Forum Message <> Reply to Message

"Numerical Recipes" (Second Edition, section 8.5) discusses "selection",
which is what Dean originally inquired about. It's a good, short read.

Subject: Re: Finding the index of the median
Posted by Joseph M Zawodny on Wed, 30 Oct 1996 08:00:00 GMT
View Forum Message <> Reply to Message

David Fanning wrote:
>
> 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.

Dave,
 This was exactly my point in my original response. He
needed to add some additional selection criteria to allow the
selection of only one of the possible array elements which were
equal to the median value. Until he fully understands his own
requirements in this respect neither he nor anyone else can come
up with the required code. Maybe he has already sorted the data
in some meaningful way and he simply wants the "middle median"
(still an illposed question if there are an even number of elements
which equal the median value), but he must acknowledge the existance
of the need for additional selectivity. Simply stating that "I want
the element that IDL selects as representative of the median value"
assumes that the median function actually does distinguish between
otherwise equivalent elements, which I doubt it does. Even your
own example becomes ambiguous if there were two accounts with
exactly $1M and the same number of transactions in the last month
(in which case I'd decide to steal $2M ;-).

--
Work: Dr. Joseph M. Zawodny   Play: Joe Zawodny

NASA Langley Research Center   KO4LW@amsat.org
E-mail: J.M.Zawodny@LaRC.NASA.gov   zawodny@exis.net
(757) 864-2681 (757) 864-2671 FAX

Subject: Re: Finding the index of the median
Posted by meron on Wed, 30 Oct 1996 08:00:00 GMT
View Forum Message <> Reply to Message

In article <32769981.73B9@cassini.lpl.arizona.edu>, Dean Schulze
<schulze@cassini.lpl.arizona.edu> writes:
> Joseph M. Zawodny wrote:
>>
>> Let's try this again: There may be more than one value which
>> equals the median value. Therefore the question you ask here
>> makes no sense at all without some additional information by
>> which to prioritize or otherwise sort those values which are
>> equal to the median. Having said that, you should use the
>> where function to find pointers to the median values and then
>> perform some other task or assessment on those values.
>>
>> Maybe we would all understand your question better if you tell
>> us which element in the following arrays is your "median element"
>>
>> [1,2,2,3]
>>
>> [1,2,2,2,3]
>>
>> One of the above arrays must have at least two "median elements"
>> by your definition. How will you choose?
>>
>>   Or am I just dense?
>
>   No, but a part of my original question has been left out. I
> said that there are two arrays, one containing data and another
> containing noise (noise is not just dependant on N in this case).
> I can let MEDIAN() return any of the equivalent median values
> in the data array it chooses, but I need to know the location of
> the one it chooses so I can get the corresponding value out of
> the noise array.
>
>   Consider a CCD with a temperature gradient that is known or
> can be modeled. There may be several pixels that have the median
> value from different points on the CCD, and each of those points
> would have a different dark current value due to the T gradient.
> In order to determine the S/N ratio of the value returned by MEDIAN()
> I need to know the location of that pixel to get the right value
> from a noise array (or get the right value of T from a T array).

I'm afraid you still don't quite understand the way median works. For
the case mentioned above (by Zavodny) of the array [1,2,2,2,3], MEDIAN
will return 2, but this value doesn't correspond to a specific
location. You cannot ask "which of the three 2s was returned?", this
is meaningless. The most you can ask is "at which location the value
equals the median value?" This you'll get with

 WHERE(array eq MEDIAN(array))

Mati Meron | "When you argue with a fool,
meron@cars.uchicago.edu | chances are he is doing just the same"