Subject: Identifying outliers in data
Posted by siumtesfai on Fri, 26 Jun 2015 02:03:32 GMT
View Forum Message <> Reply to Message

Hi All,

I am using cgboxplot.pro to identify outliers in my data. It is nice program that I see I have outliers in my data

Next step I would like to store my good data to an array and continue processing them.

My data is two dimension wind data

wind = Array(number of days, pressure levels)

e.g wind= Array(31, 17)

Once I am able to exclude the outliers from my daily dataset, I am interested to make monthly mean data set

Can anyone suggest me how I would solve my problem

Thank you in advance

Best regards

Subject: Re: Identifying outliers in data
Posted by David Fanning on Fri, 26 Jun 2015 14:02:09 GMT
View Forum Message <> Reply to Message

siumtesfai@gmail.com writes:

> I am using cgboxplot.pro to identify outliers in my data. It is nice program that I see I have outliers in my data
>
> Next step I would like to store my good data to an array and continue processing them.

It would be a mistake to use a plotting program for an analytical solution to your problem. Open cgBoxPlot up and see if you can learn HOW it identifies outliers. Then, write your own program that does exactly that, saving the non-outliers in whatever way makes sense to you.

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.idlcoyote.com/
Sepore ma de ni thue. ("Perhaps thou speakest truth.")

Subject: Re: Identifying outliers in data
Posted by siumtesfai on Fri, 26 Jun 2015 18:57:36 GMT
View Forum Message <> Reply to Message

On Thursday, June 25, 2015 at 10:03:34 PM UTC-4, siumt...@gmail.com wrote:
> Hi All,
>
> I am using cgboxplot.pro to identify outliers in my data. It is nice program that I see I have outliers in my data
>
> Next step I would like to store my good data to an array and continue processing them.
>
>
> My data is two dimension wind data
>
> wind = Array(number of days, pressure levels)
>
> e.g wind= Array(31, 17)
>
> Once I am able to exclude the outliers from my daily dataset, I am interested to make monthly mean data set
>
>
>
> Can anyone suggest me how I would solve my problem
>
> Thank you in advance
>
> Best regards

I would think i can do this

```
; Draw outliners if there are any.
IF maxcount GT 0 THEN BEGIN
  outliermax=fltarr(maxcount)
  FOR k = 0,maxcount-1 do outliermax(k)=imax(k)
  print,'outliermax'
```

```
  print,outliermax

   FOR j=0,maxcount-1 DO PLOTS, xlocation, data[imax[j]], $
      PSYM=cgSymCat(9), COLOR=cgColor(outliercolor), NOCLIP=0
ENDIF
IF mincount GT 0 THEN BEGIN

  outliermin=fltarr(mincount)
FOR kk = 0,mincount-1 do outliermin(kk)=imin(kk)
   print,'outliermin'
   print,outliermin
   FOR j=0,mincount-1 DO PLOTS, xlocation, data[imin[j]], $
      PSYM=cgSymCat(9), COLOR=cgColor(outliercolor), NOCLIP=0
ENDIF
```

But the problem would be the original data have been sorted . I would have a problem locating the location or index of the outlier in the original data.

I found in the above step is the index or location from the already sorted data.

Best regards

Subject: Re: Identifying outliers in data
Posted by Jeremy Bailin on Sat, 27 Jun 2015 17:06:10 GMT
View Forum Message <> Reply to Message

On Friday, June 26, 2015 at 2:57:37 PM UTC-4, siumt...@gmail.com wrote:
> On Thursday, June 25, 2015 at 10:03:34 PM UTC-4, siumt...@gmail.com wrote:
>> Hi All,
>>
>> I am using cgboxplot.pro to identify outliers in my data. It is nice program that I see I have outliers in my data
>>
>> Next step I would like to store my good data to an array and continue processing them.
>>
>>
>> My data is two dimension wind data
>>
>> wind = Array(number of days, pressure levels)
>>
>> e.g wind= Array(31, 17)
>>
>> Once I am able to exclude the outliers from my daily dataset, I am interested to make monthly mean data set
>>

>>
>>
>> Can anyone suggest me how I would solve my problem
>>
>> Thank you in advance
>>
>> Best regards
>
>
>
>
> I would think i can do this
>
> ; Draw outliners if there are any.
> IF maxcount GT 0 THEN BEGIN
> outliermax=fltarr(maxcount)
> FOR k = 0,maxcount-1 do outliermax(k)=imax(k)
> print,'outliermax'
> print,outliermax
>
> FOR j=0,maxcount-1 DO PLOTS, xlocation, data[imax[j]], $
> PSYM=cgSymCat(9), COLOR=cgColor(outliercolor), NOCLIP=0
> ENDIF
> IF mincount GT 0 THEN BEGIN
>
> outliermin=fltarr(mincount)
> FOR kk = 0,mincount-1 do outliermin(kk)=imin(kk)
> print,'outliermin'
> print,outliermin
> FOR j=0,mincount-1 DO PLOTS, xlocation, data[imin[j]], $
> PSYM=cgSymCat(9), COLOR=cgColor(outliercolor), NOCLIP=0
> ENDIF
>
>
>
> But the problem would be the original data have been sorted . I would have a problem locating the location or index of the outlier in the original data.
>
> I found in the above step is the index or location from the already sorted data.
>
> Best regards

What are imin and imax? How do they get defined? Once you understand that, then you will know how they relate to your original data.

-Jeremy.

Subject: Re: Identifying outliers in data
Posted by David Fanning on Sat, 27 Jun 2015 17:16:53 GMT
View Forum Message <> Reply to Message

Jeremy Bailin writes:

> What are imin and imax? How do they get defined? Once you understand that, then you will know how they relate to your original data.

Right. iMin and iMax *are* the locations of the outliers in the original
data, assuming your original data didn't contain "missing" values.

Cheers,

David
--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.idlcoyote.com/
Sepore ma de ni thue. ("Perhaps thou speakest truth.")