Subject: regress
Posted by anniebryant@gmail.com on Tue, 17 Mar 2009 21:45:17 GMT
View Forum Message <> Reply to Message

Greetings,

I have two 2D arrays I would like to regress. Both were created with
the variance command and are intarr(673,274). Does one of these need
to be a vector? Both need to be a vector? If so, is there a way to
convert them from an array to a vector? I watched someone do this
very quickly yesterday, but can not remember how.

I thought this would be quite simple, but I haven't done a regression
in IDL before and am hoping I can figure it out.

Thanks!

Subject: Re: regress
Posted by David Fanning on Tue, 17 Mar 2009 21:50:39 GMT
View Forum Message <> Reply to Message

anniebryant@gmail.com writes:

> I have two 2D arrays I would like to regress. Both were created with
> the variance command and are intarr(673,274). Does one of these need
> to be a vector? Both need to be a vector? If so, is there a way to
> convert them from an array to a vector? I watched someone do this
> very quickly yesterday, but can not remember how.

```
  s = Size(array)
  vector = Reform(array, s[0]*s[1])
```

Cheers,

David

--
David Fanning, Ph.D.
Coyote's Guide to IDL Programming (www.dfanning.com)
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: regress
Posted by David Fanning on Tue, 17 Mar 2009 21:51:56 GMT
View Forum Message <> Reply to Message

---

anniebryant@gmail.com writes:

> I have two 2D arrays I would like to regress. Both were created with
> the variance command and are intarr(673,274). Does one of these need
> to be a vector? Both need to be a vector? If so, is there a way to
> convert them from an array to a vector? I watched someone do this
> very quickly yesterday, but can not remember how.

Whoops, forgot the DIMENSIONS keyword:

  s = Size(array, /Dimensions)
  vector = Reform(array, s[0]*s[1])

Cheers,

David
--
David Fanning, Ph.D.
Coyote's Guide to IDL Programming (www.dfanning.com)
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Subject: Re: regress
Posted by anniebryant@gmail.com on Tue, 17 Mar 2009 22:02:01 GMT

On Mar 17, 3:51 pm, David Fanning <n...@dfanning.com> wrote:
> anniebry...@gmail.com writes:
>> I have two 2D arrays I would like to regress. Both were created with
>> the variance command and are intarr(673,274). Does one of these need
>> to be a vector? Both need to be a vector? If so, is there a way to
>> convert them from an array to a vector? I watched someone do this
>> very quickly yesterday, but can not remember how.
>
> Whoops, forgot the DIMENSIONS keyword:
>
>   s = Size(array, /Dimensions)
>   vector = Reform(array, s[0]*s[1])
>
> Cheers,
>
> David
> --
> David Fanning, Ph.D.
> Coyote's Guide to IDL Programming (www.dfanning.com)
> Sepore ma de ni thui. ("Perhaps thou speakest truth.")

David,

OK, the vector was created no-problem. Thanks.

However, I am still not getting the regression to work.

Here are my two variables:

varimg =

integer mean   std dev   minimum   maximum n_elements
    262.89   205.93   0.0000   936.00 (678,237) =
160686

vector=

integer mean   std dev   minimum   maximum n_elements
    868.24   763.27   0.0000   9299.0 (160686) =
160686

Here is the expression I have input and the error I get.

IDL> reg = regress(varimg, vector)
% REGRESS: X and Y have incompatible dimensions.
% Error occurred at: REGRESS   126 /Applications/ENVI_IDL/itt/
idl70/lib/regress.pro
% $MAIN$
% Execution halted at: $MAIN$

Any thoughts?

Subject: Re: regress
Posted by David Fanning on Tue, 17 Mar 2009 22:48:30 GMT
View Forum Message <> Reply to Message

anniebryant@gmail.com writes:

> OK, the vector was created no-problem. Thanks.
>
> However, I am still not getting the regression to work.

I'm probably not the one you should be talking to,
because I don't believe I have ever used the REGRESS
function, but after reading the documentation, I have
a sneaking suspicion it may not be the routine you
think it is. What exactly is it you are trying to
do with your data?

Cheers,

David

--
David Fanning, Ph.D.
Coyote's Guide to IDL Programming (www.dfanning.com)
Sepore ma de ni thui. ("Perhaps thou speakest truth.")

On Mar 17, 4:48 pm, David Fanning <n...@dfanning.com> wrote:
> anniebry...@gmail.com writes:
>> OK, the vector was created no-problem. Thanks.
>
>> However, I am still not getting the regression to work.
>
> I'm probably not the one you should be talking to,
> because I don't believe I have ever used the REGRESS
> function, but after reading the documentation, I have
> a sneaking suspicion it may not be the routine you
> think it is. What exactly is it you are trying to
> do with your data?
>
> Cheers,
>
> David
>
> --
> David Fanning, Ph.D.
> Coyote's Guide to IDL Programming (www.dfanning.com)
> Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Hey David,

I am trying to do a basic linear regression. My variables are
variance of ground-cover and variance of topography. All I want to
determine is if topographic variance is a good predictor of ground-
cover variance, i.e. does a high topo variance value do a good job of
predicting a high ground-cover variance value? You don't think
REGRESS is the way to go?

Thanks

Annie

Subject: Re: regress
Posted by Jeremy Bailin on Tue, 17 Mar 2009 23:47:36 GMT

On Mar 17, 5:51 pm, David Fanning <n...@dfanning.com> wrote:
> anniebry...@gmail.com writes:
>> I have two 2D arrays I would like to regress. Both were created with
>> the variance command and are intarr(673,274). Does one of these need
>> to be a vector? Both need to be a vector? If so, is there a way to
>> convert them from an array to a vector? I watched someone do this
>> very quickly yesterday, but can not remember how.
>
> Whoops, forgot the DIMENSIONS keyword:
>
>   s = Size(array, /Dimensions)
>   vector = Reform(array, s[0]*s[1])
>
> Cheers,
>
> David
>
> --
> David Fanning, Ph.D.
> Coyote's Guide to IDL Programming (www.dfanning.com)
> Sepore ma de ni thui. ("Perhaps thou speakest truth.")

Incidentally, I've often wondered if that was any more or less
efficient than

vector = (s)[*]

Anyone have any thoughts?

-Jeremy.

Subject: Re: regress
Posted by David Fanning on Wed, 18 Mar 2009 01:03:17 GMT

anniebryant@gmail.com writes:

> I have two 2D arrays I would like to regress. Both were created with
> the variance command and are intarr(673,274). Does one of these need
> to be a vector? Both need to be a vector? If so, is there a way to
> convert them from an array to a vector? I watched someone do this
> very quickly yesterday, but can not remember how.
>
> I thought this would be quite simple, but I haven't done a regression

> in IDL before and am hoping I can figure it out.

OK, if you just want to do a linear fit through the data,
I guess you can use REGRESS. I always have used LINFIT, but
in a quick test case, they appear to return the same results.

You will have to reformat both your images into vectors.
So, using LINFIT:

```
r = Linfit(Reform(d1, 673*274), Reform(d2, 673*274), YFIT=yfit
Print, 'Slope: ', r[1], '  Intercept: ', r[0]
Plot, Reform(d1, 673*274), Reform(d2, 673*274), PSYM=3
OPlot, Reform(d1, 673*274), yfit
```

Using REGRESS:

```
slope = Regress(Reform(d1, 673*274), Reform(d2, 673*274), $
   YFIT=yfit, CONST=intercept)
Plot, Reform(d1, 673*274), Reform(d2, 673*274), PSYM=3
OPlot, Reform(d1, 673*274), yfit
```

Cheers,

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

Subject: Re: regress
Posted by jameskuyper on Wed, 18 Mar 2009 11:26:56 GMT
View Forum Message <> Reply to Message

anniebryant@gmail.com wrote:

...
> I am trying to do a basic linear regression. My variables are
> variance of ground-cover and variance of topography. All I want to
> determine is if topographic variance is a good predictor of ground-
> cover variance, i.e. does a high topo variance value do a good job of
> predicting a high ground-cover variance value? You don't think
> REGRESS is the way to go?

Step one should be to plot your two variables against each other. If
you've got two many data points, you might need to look at a 2-D
histogram, instead. Only use REGRESS if it seems plausible from such a
plot that there is a linear relationship between them. If it looks like

there's some other relationship between them, then you should fit to a
curve that more closely resembles that relationship.

However, if the two variables are only loosely connected, which is what
I'd expect in this case, no curve is going to be a particularly good
fit, and using a fitting routine is not the right way to approach this.
What you should be calculating is the correlation between the two variables.

Subject: Re: regress
Posted by Brian Larsen on Wed, 18 Mar 2009 13:07:08 GMT
View Forum Message <> Reply to Message

> Step one should be to plot your two variables against each other. If
> you've got two many data points, you might need to look at a 2-D
> histogram, instead. Only use REGRESS if it seems plausible from such a
> plot that there is a linear relationship between them. If it looks like
> there's some other relationship between them, then you should fit to a
> curve that more closely resembles that relationship.

I point this out to everyone as much as to the original poster.
Plotting your data really is the key to initial understanding. To
make "looking" at plots more quantitative NIST has a nice statistics
guide that people should know about. See:
http://www.itl.nist.gov/div898/handbook/index.htm

I have written routines for many of the plots shown in the guide. Two
are 4-plot and 6-plot. See:
http://people.bu.edu/balarsen/IDLdoc/stats/fourplot.html
http://people.bu.edu/balarsen/IDLdoc/stats/sixplot.html
and the NIST guide for a description of what the plots show.

For example the 4-plot is testing these underlying assumptions about
regression:
  1. Fixed Location:
     If the fixed location assumption holds, then the run sequence
plot will be flat and non-drifting.

  2. Fixed Variation:
     If the fixed variation assumption holds, then the vertical
spread in the run sequence plot will be the approximately the same
over the entire horizontal axis.

  3. Randomness:
     If the randomness assumption holds, then the lag plot will be
structureless and random.

  4. Fixed Distribution:

If the fixed distribution assumption holds, in particular if the fixed normal distribution holds, then
  1. the histogram will be bell-shaped, and
  2. the normal probability plot will be linear.

The "scary" thing is that if any of these assumptions are violated in a meaningful way then regression is invalid to use on a data set, but it s done anyway....

Subject: Re: regress
Posted by Jeremy Bailin on Thu, 19 Mar 2009 14:37:57 GMT
View Forum Message <> Reply to Message

On Mar 17, 6:47 pm, Jeremy Bailin <astroco...@gmail.com> wrote:
> On Mar 17, 5:51 pm, David Fanning <n...@dfanning.com> wrote:
>
>
>
>> anniebry...@gmail.com writes:
>>> I have two 2D arrays I would like to regress. Both were created with
>>> the variance command and are intarr(673,274). Does one of these need
>>> to be a vector? Both need to be a vector? If so, is there a way to
>>> convert them from an array to a vector? I watched someone do this
>>> very quickly yesterday, but can not remember how.
>
>> Whoops, forgot the DIMENSIONS keyword:
>
>>   s = Size(array, /Dimensions)
>>   vector = Reform(array, s[0]*s[1])
>
>> Cheers,
>
>> David
>> --
>> David Fanning, Ph.D.
>> Coyote's Guide to IDL Programming (www.dfanning.com)
>> Sepore ma de ni thui. ("Perhaps thou speakest truth.")
>
> Incidentally, I've often wondered if that was any more or less
> efficient than
>
> vector = (s)[*]
>
> Anyone have any thoughts?
>
> -Jeremy.

A simple test seems to suggest that there's no speed difference:

```
IDL> n1=512l & n2=256l
IDL> bigarray = fltarr(n1,n2)
IDL> s1=systime(/sec) & for i=1,10000 do vector=reform(bigarray,n1*n2)
& s2=systime(/sec)
IDL> print, s2-s1
      2.8659091
IDL> s3=systime(/sec) & for i=1,10000 do vector=(bigarray)[*] &
s4=systime(/sec)
IDL> print, s4-s3
      2.8190870
```

Not sure about internal memory usage...

-Jeremy.