Subject: t-testing in IDL
Posted by mariamalene on Wed, 23 Jan 2008 15:52:51 GMT
View Forum Message <> Reply to Message

I have two sample data sets with 20 years of 12 months simulations in a lon/lat grid of 128 x 64. , i.e. 2x(20,12,128,64). I would like to find the difference between the two simulations and find out which differences that are significant at a 95% level. I have calculated the annual average of my 20-year simulation, and also the standard-deviation, but how do I perfom the t-test? I have looked at the tm_test but haven't figured out how to apply it on my 4-D approach..
Thanks

Subject: Re: t-testing in IDL
Posted by mmiller3 on Thu, 24 Jan 2008 14:52:40 GMT
View Forum Message <> Reply to Message

>>>> > "mariamalene" == mariamalene <mariamalene@hotmail.com> writes:

> I have two sample data sets with 20 years of 12 months
> simulations in a lon/lat grid of 128 x 64. ,
> i.e. 2x(20,12,128,64). I would like to find the difference
> between the two simulations and find out which differences
> that are significant at a 95% level. I have calculated the
> annual average of my 20-year simulation, and also the
> standard- deviation, but how do I perfom the t-test? I have
> looked at the tm_test but haven't figured out how to apply
> it on my 4-D approach..

I'm not entirely clear on what you want to do, but I'll take a shot.

If you have two data sets, A and B, and you want to test the hypothesis that the means are the same, you can use tm_test as in tm_test(A,B). It the data sets are paired, meaning A[i] is correlated with B[i], you can use tm_test(A,B,/paired) to test the hypothesis that the mean difference is zero. I suspect that this might be what you want. If you want to test the hypothesis that the mean difference is zero and that the difference does not have a spatial dependence, you could use a mixed effect model.

If you are going to do stats, I'd suggest you use a statistics package, such as R (www.r-project.org) instead of IDL. For example, if a t-test is to be useful, your data must be normally distributed. That is very easy to test with R, and relatively easy to test with IDL, but once you get beyond very basic statistics, you'll get beyond IDL's build-in capabilities. There

are relatively easy to use mixed effects modeling tools in R.

Mike

P.S. For the record: this posting should not be taken as an opinion against IDL, which I love/hate and use constantly. It's just that R, which I love/hate and use constantly, has a different set of strenghts and weaknesses. Give me grep, awk, python, IDL and R and an emacs to run it all in and I can shoot myself in the other foot every time!

P.P.S. Getting even more off topic: this list needs to be updated: http://burks.bton.ac.uk/burks/language/shoot.htm

---

Subject: Re: t-testing in IDL
Posted by mariamalene on Thu, 24 Jan 2008 15:31:37 GMT
View Forum Message <> Reply to Message

On Jan 24, 3:52 pm, mmill...@iupui.edu (Michael A. Miller) wrote:
>>>> >> "mariamalene" == mariamalene <mariamal...@hotmail.com> writes:
>
>> I have two sample data sets with 20 years of 12 months
>> simulations in a lon/lat grid of 128 x 64. ,
>> i.e. 2x(20,12,128,64). I would like to find the difference
>> between the two simulations and find out which differences
>> that are significant at a 95% level. I have calculated the
>> annual average of my 20-year simulation, and also the
>> standard- deviation, but how do I perfom the t-test? I have
>> looked at the tm_test but haven't figured out how to apply
>> it on my 4-D approach..
>
> I'm not entirely clear on what you want to do, but I'll take a
> shot.
>
> If you have two data sets, A and B, and you want to test the
> hypothesis that the means are the same, you can use tm_test as in
> tm_test(A,B). It the data sets are paired, meaning A[i] is
> correlated with B[i], you can use tm_test(A,B,/paired) to test
> the hypothesis that the mean difference is zero. I suspect that
> this might be what you want. If you want to test the hypothesis
> that the mean difference is zero and that the difference does not
> have a spatial dependence, you could use a mixed effect model.
>
> If you are going to do stats, I'd suggest you use a statistics
> package, such as R (www.r-project.org) instead of IDL. For
> example, if a t-test is to be useful, your data must be normally
> distributed. That is very easy to test with R, and relatively

> easy to test with IDL, but once you get beyond very basic
> statistics, you'll get beyond IDL's build-in capabilities. There
> are relatively easy to use mixed effects modeling tools in R.
>
> Mike
>
> P.S. For the record: this posting should not be taken as an
> opinion against IDL, which I love/hate and use constantly. It's
> just that R, which I love/hate and use constantly, has a
> different set of strenghts and weaknesses. Give me grep, awk,
> python, IDL and R and an emacs to run it all in and I can shoot
> myself in the other foot every time!
>
> P.P.S. Getting even more off topic: this list needs to be
> updated:http://burks.bton.ac.uk/burks/language/shoot.htm

You are guessing right about my reasons for doing this. I'll check out R, although I hope I don't have to go further into statistics than this! Thank you.

Subject: Re: t-testing in IDL
Posted by Brian Larsen on Thu, 24 Jan 2008 16:03:55 GMT
View Forum Message <> Reply to Message

> If you are going to do stats, I'd suggest you use a statistics
> package, such as R (www.r-project.org) instead of IDL. For
> example, if a t-test is to be useful, your data must be normally
> distributed. That is very easy to test with R, and relatively
> easy to test with IDL, but once you get beyond very basic
> statistics, you'll get beyond IDL's build-in capabilities. There
> are relatively easy to use mixed effects modeling tools in R.

I agree with all of this but my issue always is that as soon as I start mixing tools (IDL and R) things start to get confusing as well as I seem to spend all my time doing file io instead of analysis. To this end I started making some stats routines in IDL, some are general purpose and some are really specific.

For this example you want to know if the data is normally distributed, for this I like normal probability plots and the Blom Position test, both of which I implemented in IDL (and would love fixes, feedback, etc). There are other IDL stats resources on the web too, it would be good to pull those together into an IDLstats package...

See:
http://people.bu.edu/balarsen/IDLdoc/
blom_position_test.pro

normal_prob_plot.pro

Cheers,

Brian

---------- ----------
Brian Larsen
Boston University
Center for Space Physics

Subject: Re: t-testing in IDL
Posted by Jean H. on Thu, 24 Jan 2008 17:25:04 GMT
View Forum Message <> Reply to Message

> P.P.S. Getting even more off topic: this list needs to be
> updated: http://burks.bton.ac.uk/burks/language/shoot.htm

IDL:

You aim at your foot, but due to a rounding error you end up shooting
yourself in the head as your body wraps around.

:-)
Jean
