Subject: Re: Statistic codes: Significance level
Posted by R.G. Stockwell on Tue, 02 Mar 2004 00:35:21 GMT
View Forum Message <> Reply to Message

"Andry William (Please remove ".spam")" <andry@ya.com.spam> wrote in message news:4043EB95.1020400@ya.com.spam...
> Dear all:
>
> I am processing some long term time series of various data.
> I computed the long term trends using the REGRESS fuction.
> However, I need to also give the significance level of the
> trends. I have been looking around for a code to do it but
> have not found anything.
>
> Has anybody done similar calculation? Is there a function or
> procedure I can use for that purpose?
>
> Thanks for any help and suggestion,
>
> Andry
>

Perhaps the sigma values will give you what you want. I don't remember offhand, but 1-sigma is the 67% significance level (or is that 90%?)

You can say things like "the long term trend is "result" +- sigma."
Or "the long term trend is between result + sigma and result-sigma with a confidence of 90% (or is that 67%?).

Cheers,
bob

From the help:
SIGMA
Set this keyword to a named variable that will contain the 1-sigma uncertainty estimates for the returned parameters.

Subject: Re: Statistic codes: Significance level
Posted by Craig Markwardt on Tue, 02 Mar 2004 09:32:44 GMT
View Forum Message <> Reply to Message

"Andry William (Please remove \".spam\")" <andry@ya.com.spam> writes:

> Dear all:
>
> I am processing some long term time series of various data.
> I computed the long term trends using the REGRESS fuction.

> However, I need to also give the significance level of the
> trends. I have been looking around for a code to do it but
> have not found anything.

What do you mean by "significance level of the trend?" Do you mean you want to estimate a confidence interval of the slope parameter to your fit? Can you supply uncertainty estimates for your time series? If yes to both, then routines like LINFIT or MPFITFUN can do this (Bob points you in the right direction). If you can't supply uncertainty estimates, then any estimated significance levels are irrelevant. Also, if your trend model is non-linear, then estimated confidence regions provided by these routines are not reliable, and you need to make a chi-square confidence grid, as described in Bevington or Numerical Recipes.

But "significance level" can have other meanings too. Sometimes this can mean the goodness of the fit, in which case you would want to compute the probability of chance occurrence of having a chi-square value greater or less than a certain value (usually you would want to do this when the fit is bad). MPCHITEST can help you there.

"Significance level of the trend" could also mean a comparison of the fit with and without a trend. In that case you will want to perform an F-test for that addition of the trend parameters. You can read about this in Bevington, and use MPFTEST to compute the probabilities. You will need two fitted chi-saure values: one with, and one without, the trend.

Good luck,
Craig

P.S. MP* routines are at

http://cow.physics.wisc.edu/~craigm/idl/idl.html (under fitting)

Subject: Re: Statistic codes: Significance level
Posted by wmconnolley on Tue, 02 Mar 2004 09:45:46 GMT
View Forum Message <> Reply to Message

R.G. Stockwell <noemail@please.com> wrote:
>> I am processing some long term time series of various data.
>> I computed the long term trends using the REGRESS fuction.
>> However, I need to also give the significance level of the
>> trends. I have been looking around for a code to do it but
>> have not found anything.

> Perhaps the sigma values will give you what you want. I don't remember

> offhand, but 1-sigma is the 67% significance level (or is that 90%?)

Here is what I use. Sorry for the messy pro... I was young then.

If you don't want to wade through the details, the important bit is:

```
; Compute the (2-tailed) t-statistic appropriate to the probability
t=t_cvf(prob/2.,reg.n/adj-2)

print,'A confidence interval for the slope is: ',reg.b+[-1,1]*reg.seb*t
```

(seb is standard-error-b, as I recall).

NOTE that if your data are auto-correlated then you will get the wrong
result unless you take this into account

-W.

```
; Explain the components of a regression generated by pp_regress
;
; Input
; reg - anonymous structure generated by pp_regress
; Options
; prob=x - probability value for 2-tailed test. Default is 0.05 (95% conf)
; /quiet - suppress info
; /simple - don't try to account for autocorr if it exists
; Output
;  sig - return value, 1 if slope, b, is sig. diff from 0; 0 otherwise
;  - Also writes info to the screen
; ac - value of first five auto correlations at lags 1...5
;       acs     - ac-sig: index of sig autocorrs, empty if there are none
; sig_ac  - threshold used for significance
;
; A reasonable test of this routine is:
; n=24 & ntest=1000
; a=0 & x=indgen(n)
; for i=1,ntest do  a=reg_explain(pp_regress(x,randomn(seed,n),/arr),prob=.1,/qu )+a
; print,a/float(ntest)
; Which should produce a number something like 0.10
;
; Nerdy example of use:
; print,'reg slope '+(['is not','is'])(reg_explain(/qu,pp_regress(x,y,/arr)))+' diff from zero'

function reg_explain1,reg,prob=prob,quiet=quiet,ac,acs,sig_ac

@comm_error
```

```
; Set up confidence level
if (not keyword_Set(prob)) then begin
  prob=0.05
  if (not keyword_Set(quiet)) then message,/cont,'Setting prob to 0.05 (2-tailed), ie 95%
confidence'
endif

***some auto-corr stuff removed***
adj=1.0

; Compute the (2-tailed) t-statistic appropriate to the probability
t=t_cvf(prob/2.,reg.n/adj-2)

if (not keyword_Set(quiet)) then begin
  print,'The series had '+shtstr(reg.n)+' elements & the x-range is: ',reg.xr
  print,'The regression equation is: y='+shtstr(reg.b)+' * x + '+shtstr(reg.a)
  print,'The r and r^2 value is: ',shtstr(reg.r)+', '+shtstr(reg.r^2)
  print,'A confidence interval for the slope is: ',reg.b+[-1,1]*reg.seb*t
  print,'The t-value used was: ',shtstr(t)
endif

; Decide if B is sig diff from zero, and return 1 if so else 0
if (abs(reg.b) gt reg.seb*t) then begin
  if (not keyword_Set(quiet)) then print,'B is sig. diff. from zero'
  return,1
endif else begin
  if (not keyword_Set(quiet)) then print,'B is not sig. diff. from zero'
  return,0
endelse

end

; ------------------------

function reg_explain,regs,ac,acs,sig_ac,_extra=e

for i=0,n_elements(regs)-1 do $
  if (i eq 0) then  ret=replicate(reg_explain1(regs(i),ac,acs,sig_ac,_extra=e),n _elements(regs)) $
else $
  ret(i)=reg_explain1(regs(i),ac,acs,sig_ac,_extra=e)

if (n_elements(ret) eq 1) then ret=ret(0)
return,ret
end
```

--
William M Connolley | wmc@bas.ac.uk | http://www.antarctica.ac.uk/met/wmc/

Climate Modeller, British Antarctic Survey | Disclaimer: I speak for myself
I'm a .signature virus! copy me into your .signature file & help me spread!

Subject: Re: Statistic codes: Significance level
Posted by wmconnolley on Tue, 02 Mar 2004 20:44:00 GMT
View Forum Message <> Reply to Message

In article <4044FD8C.9020906@ya.com.spam> you wrote:
> First of all, thank you Bob, Craig, and William for the direction.

> I am now looking at each suggestion to see which one(s) fit(s) the
> statistics I am asked to perform (I am still kind of a beginner
> with respect to massive data statistics). I will try to sumarize
> my findings once I get some results.

>> ; for i=1,ntest do a=reg_explain(pp_regress(x,randomn(seed,n) $
>>   ,/arr),prob=.1,/qu)+a

> With respect to the suggestion made by William Connolley, I am
> afraid I don't know what PP_REGRESS is. I beleive it should be
> another function which returns some Percentil-Percentil result
> or something similar.

Hi. Sorry about that... pp_regress is of course one of mine.
Basically it just wrap up REGRESS because I could never remember
how to use it, and works for a particular type of field too
(pp-fields). Find it at:

http://www.antarctica.ac.uk/met/wmc/idl/pp_regress.pro

You'll want to use it with arrays (not structures) and use /array
for your purposes. The line above is just a Monte-Carlo thingy
to test that the statistics coming out are about correct (a useful
test for those of us who can't do Big Maths since our brains
decayed).

BTW, if anyone is interested in break statistics in time series
(Lund/Wang) see:

http://www.antarctica.ac.uk/met/wmc/idl/lund/

-W.

--
William M Connolley | wmc@bas.ac.uk | http://www.antarctica.ac.uk/met/wmc/
Climate Modeller, British Antarctic Survey | Disclaimer: I speak for myself
I'm a .signature virus! copy me into your .signature file & help me spread!

Subject: Re: Statistic codes: Significance level
Posted by Andry William (Please on Tue, 02 Mar 2004 21:33:00 GMT
View Forum Message <> Reply to Message

First of all, thank you Bob, Craig, and William for the direction.

I am now looking at each suggestion to see which one(s) fit(s) the statistics I am asked to perform (I am still kind of a beginner with respect to massive data statistics). I will try to sumarize my findings once I get some results.

> ; for i=1,ntest do a=reg_explain(pp_regress(x,randomn(seed,n) $
> ,/arr),prob=.1,/qu)+a

With respect to the suggestion made by William Connolley, I am afraid I don't know what PP_REGRESS is. I beleive it should be another function which returns some Percentil-Percentil result or something similar.

Thanks for the explaining more about the PP_REGRESS

Andry