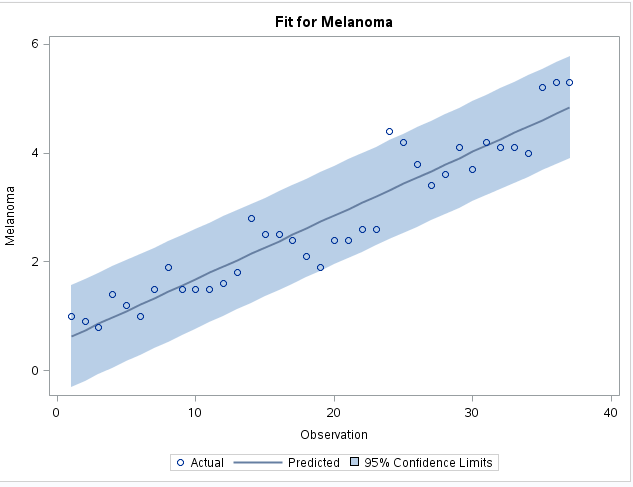
For live session we are going to discuss the general idea of dealing with correlated residuals.  Attached is an excel file that includes Melanoma and Sunspot data over time.  A quick search on google for sunspots may be helpful for a reference. The melanoma variable is the rate of melanoma occurrences.

We are going to use PROC Autoreg here, but I will discuss some additional models that can be done using PROC ARIMA so be aware of this that other procs exist to deal with more modeling scenarios.

Melanoma

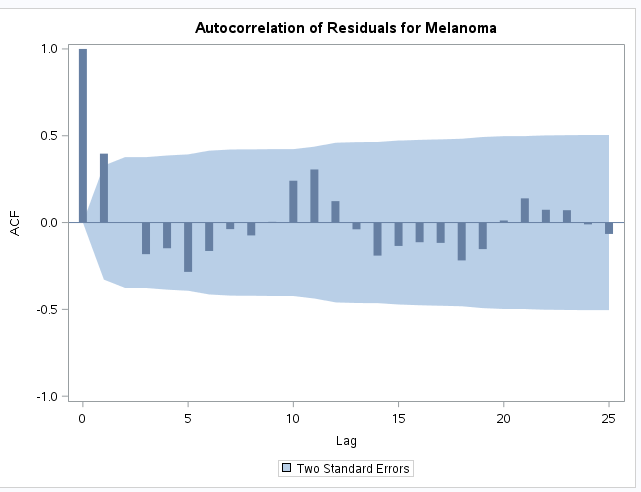
1. Plot Melanoma versus Years



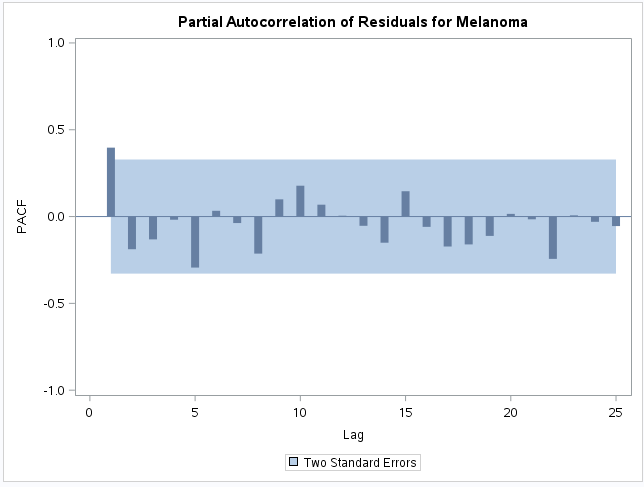
1. Take a look back at what it means for a time series to be stationary. Does the melanoma time series look stationary to you?

In order for it to be stationary, the mean, variance, and autocorrelation should be constant over time which the graph above does not appear to be stationary.

1. The first model below runs a regular linear regression of melanoma vs time without any timeseries modeling. Its just a regular regression run.  Use the diagnostic graphs (ACF and PACF plots) to assess if there is any evidence of autocorrelation (aka correlated errors).  Be prepared to discuss your basic understanding of what the graph is telling you and what you wished it looked like to have independent data.

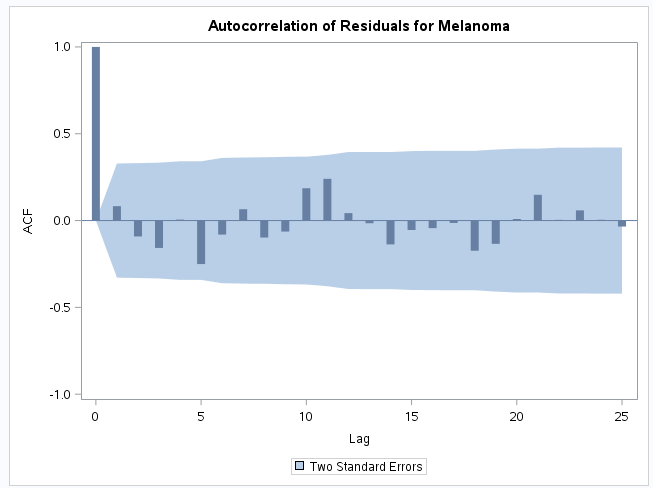


There appears to be a decay as lag decreases at first, but the algebraic signs for the autocorrelations alternate between positive and negative suggesting that there is a negative autocorrelation coefficient (a1).



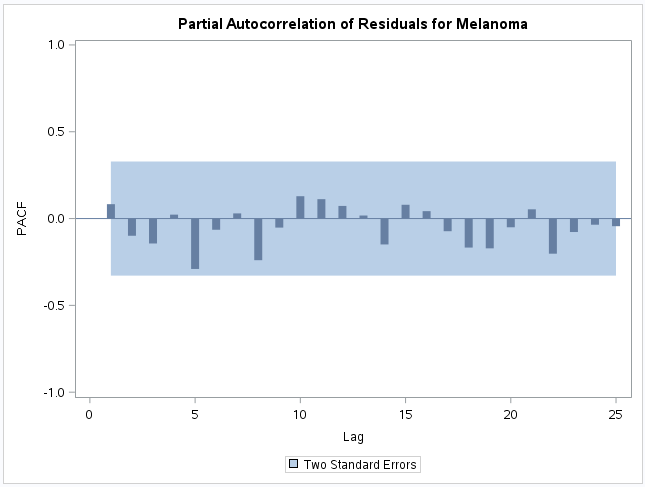
The PACF starts as positive which suggests that this should be an AR1 model and the blue area suggests that these autocorrelations are zero.

1. The second model runs a regression on time but now with an autoregressive process assumed.
   1. Check the residual diagnostic ACF and PACF plot. What do you make of it compared to #1?



I can’t tell a big difference between plot 1 and 2. Both seem to have a negative autocorrelation, but the spikes in the positives are not as high as the first plot.

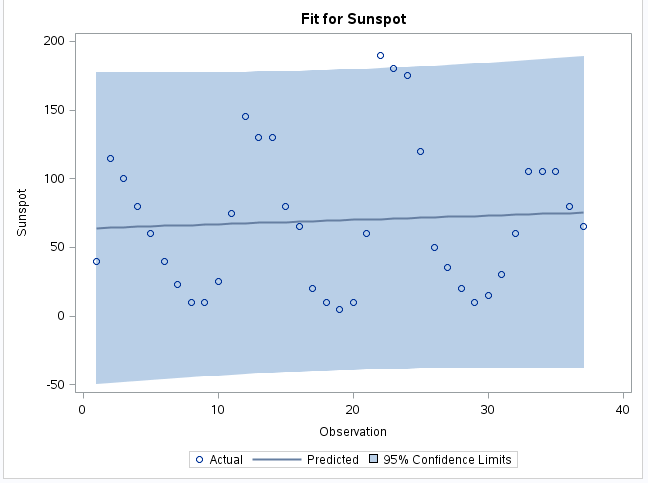
* 1. Check the regression coefficient and standard error on the “time” predictor and compare it to what is reported in #1.



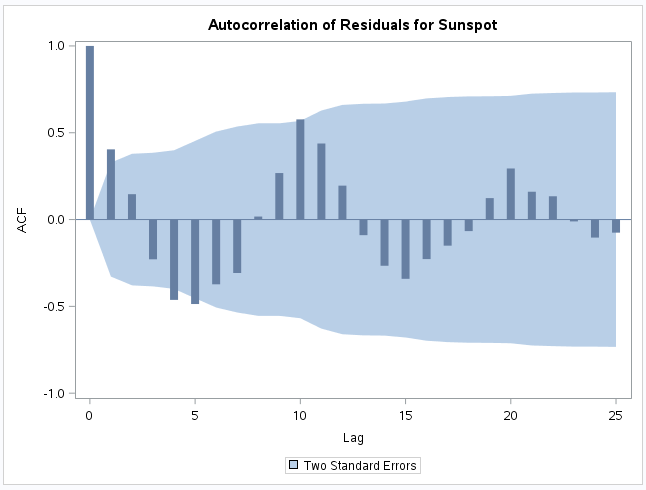
The PACF starts as positive which suggests that this should be an AR1 model and the blue area suggests that these autocorrelations are zero/statistically insignificant.

Sunspot

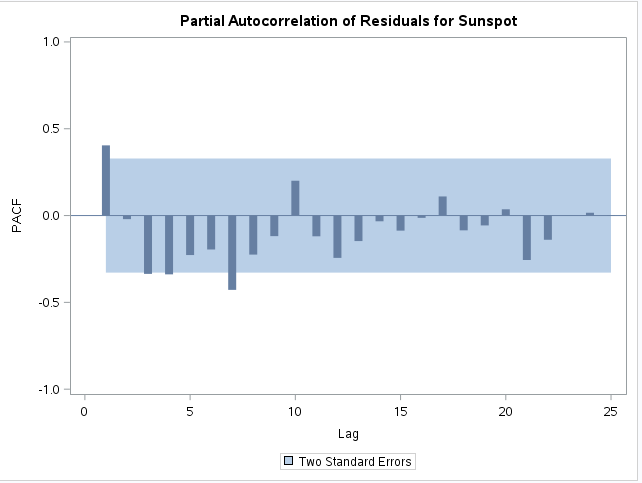
1. The sunspot data has a cyclical behavior. What we are going to do here is explore how an Autoregressive model can actually capture the cyclical behavior without any covariates present.
   1. Plot Sunspot versus Years



* 1. Using the code for melanoma as an example, fit a simple regression model to Sunspot with just an intercept (model sunspot= / nlag= in SAS) Comment on the ACF and PCF plots



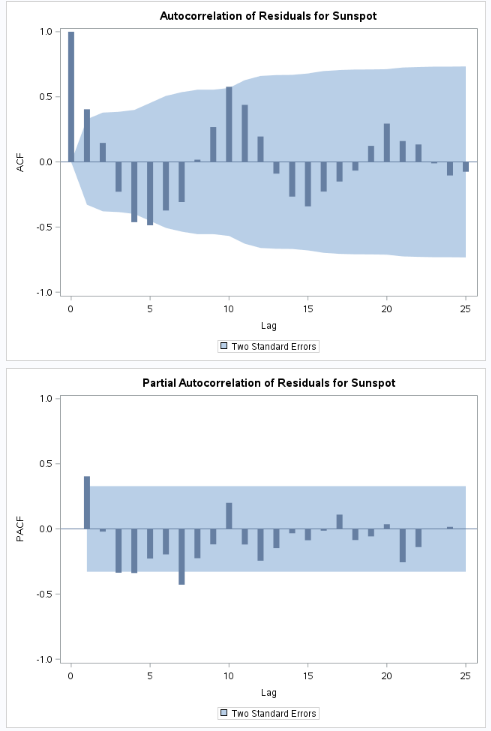
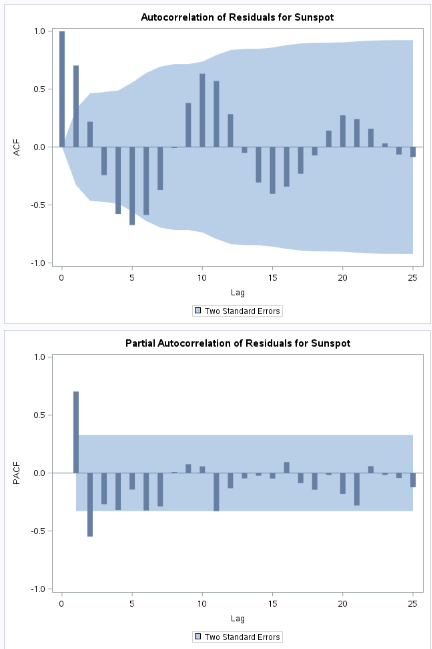
There appears to be a decay as lag decreases at first, but the algebraic signs for the autocorrelations alternate between positive and negative suggesting that there is a negative autocorrelation coefficient.



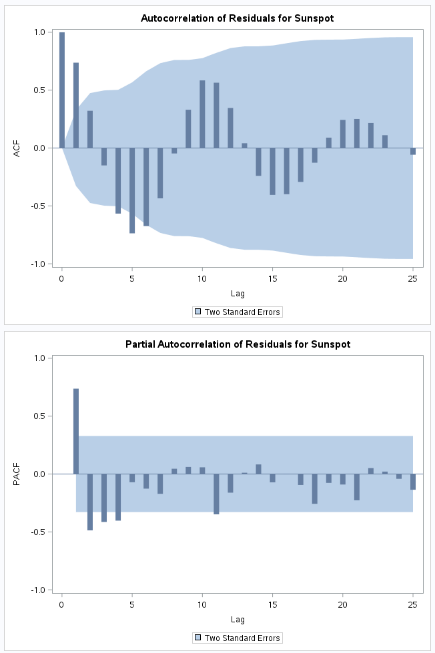
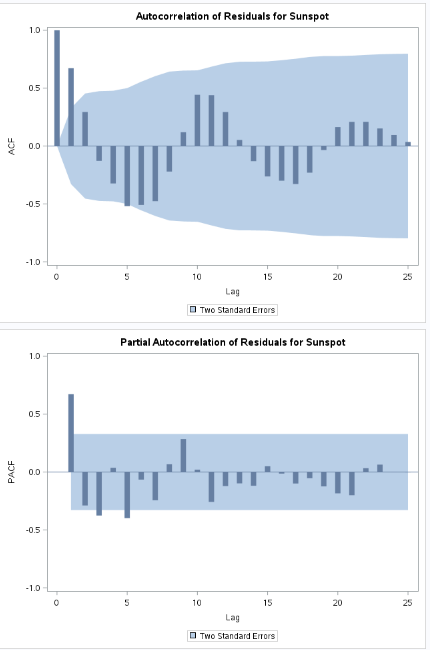
This looks like an AR1 model, but a little concerned with the fact that at lag 3 and 4, there seems to be a negative correlation that is significantly greater than zero. Not sure what the means when this occurs when it is negative. Possibly AR4?

* 1. Fit an AR(1), AR(2), AR(3), and AR(4) model by specifying the nlag option to 1,2,3, or 4.
     1. Examine and compare the ACF and PACF plots

AR1: AR2:

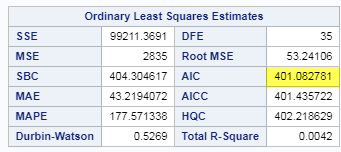
AR3: AR4:

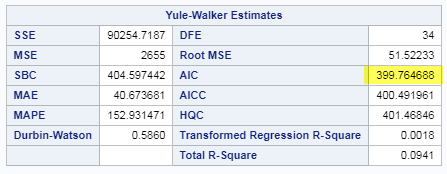
The ACF plots are very similar. However, we see a difference when fitting a model to various AR models. For the PACF plots, we can see that the AR order seems to change a bit when we fit a different model.

* + 1. Locate the AIC statistic for each of the models and compare them

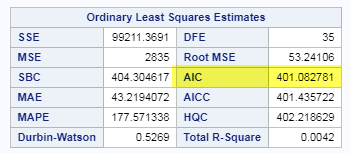
AR1:



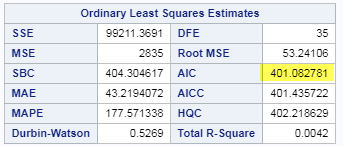
AR2:



AR3:



AR4:



The AICs are exactly the same between the AR1, AR2, and AR4 models at 401.08, but it actually improves on the AR2 model to 399.76.

* 1. Try to forecast the next 10-20 years using the model that has the lowest AIC from part c. Once you have the predictions, try to add them to your plot from part A so we can see what is going on. If you are stumped on how to predict future values of the time series, check out the Output and Predicted statement and options within Proc Autoreg or some of the examples:

<http://support.sas.com/documentation/cdl/en/etsug/63939/HTML/default/viewer.htm#etsug_autoreg_sect042.htm>

Code:

data Melanoma;

infile '/home/marinfamily1010/sasuser.v94/AppliedStats/Data/Melanomatimeseries.csv' dlm=',' firstobs=2;

input Year Melanoma Sunspot;

run;

proc autoreg data=Melanoma all plots(unpack);

model Melanoma=year;

run;quit;

proc autoreg data=Melanoma all plots(unpack); proc autoreg data=Melanoma all plots(unpack);

model Melanoma=Year / nlag=1;

run;quit;

proc autoreg data=Melanoma all plots(unpack);

model Sunspot=year;

run;quit;

proc autoreg data=Melanoma all plots(unpack); proc autoreg data=Melanoma all plots(unpack);

model Sunspot=Year / nlag=(2);

run;quit;

data Melanoma2;

Sunspot = .;

Melanoma = .;

do Year = 1973 to 1993; output; end;

run;

data Melanoma3;

merge Melanoma Melanoma2;

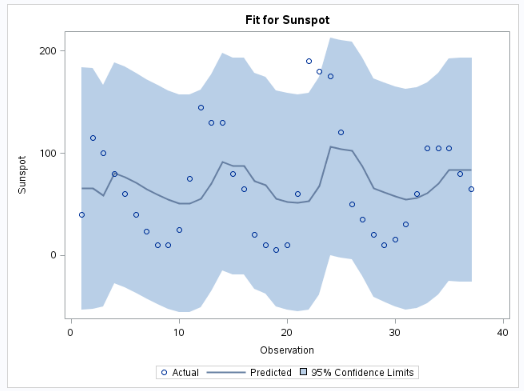
by Year;

run;

proc autoreg data=Melanoma3 all plots(unpack); proc autoreg data=Melanoma all plots(unpack);

model Sunspot=Year / nlag=(2);

run;quit;



Looks like the mean is changing rather than being held constant.

 SAS code below to get you started.

 data Melanoma;

infile 'C:\Users\e80100\Desktop\Melanomatimeseries.csv' dlm=',' firstobs=2;

input Year Melanoma Sunspot;

run;

proc autoreg data=Melanoma all plots(unpack);

model Melanoma=year;

run;quit;

                                                                                                                                                                                                                                             proc autoreg data=Melanoma all plots(unpack);

model Melanoma=Year  / nlag=1;

run;quit;