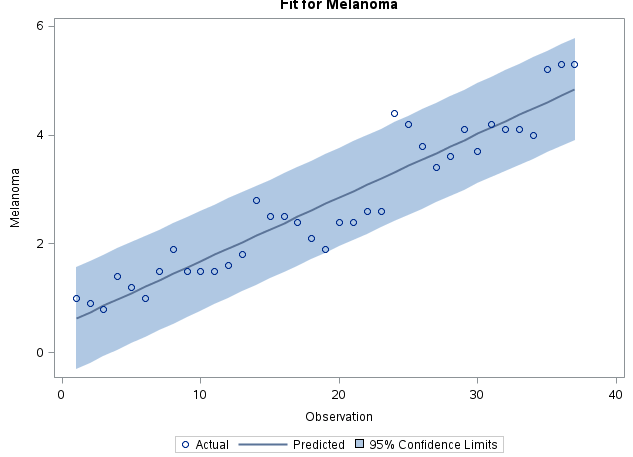
For live session we are going to discuss the general idea of dealing with correlated residuals.  Attached is an excel file that includes Melanoma and Sunspot data over time.  A quick search on google for sunspots may be helpful for reference. The melanoma variable is the rate of melanoma occurrences.

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. The first model runs a regular linear regression of melanoma vs time.  Use the diagnostic graphs 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. Notice here: (I’m technically not running a time series model. I have no options that specify that yet.)

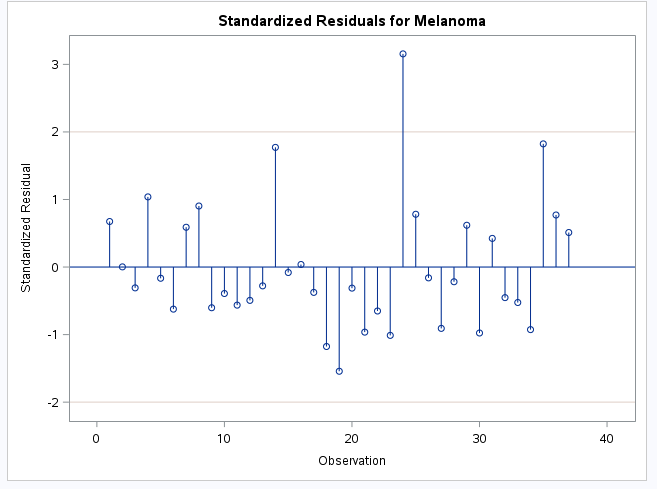
* The first model shows evidence of autocorrelation. For one the residuals show a near sinusoidal pattern around the regression line, and the residuals have many long runs of 4-5 positive or negative residuals in a row( left). Additionally, the autocorrelation residuals for Melanoma show string evidence of a AR-1 relationship (right), additionally a Durbin-Watson value of 1.5 shows strong evidence of autocorrelation

|  |  |
| --- | --- |
|  |  |

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

* The regression line now better represents some of the seasonal trend in the data, and the residuals still show long runs, but they are much smaller. It looks like there is still some autocorrelation left in the model even when the AR1 adjustment has been applied; however, the white noise ratio went from ~.5 in the previous model to ~.05 in the adjusted model

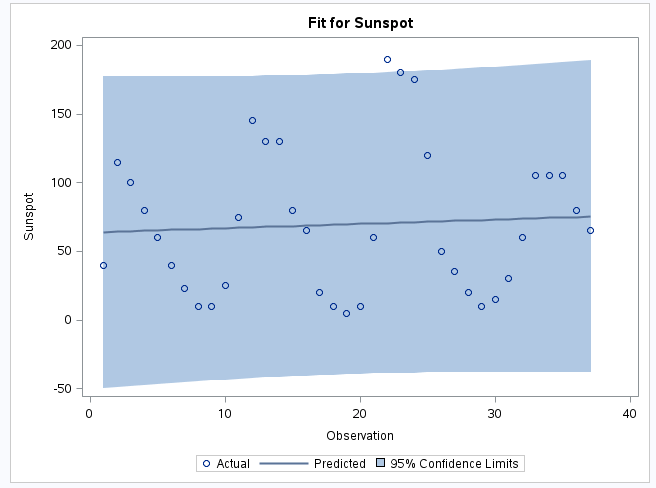
|  |  |
| --- | --- |
|  |  |



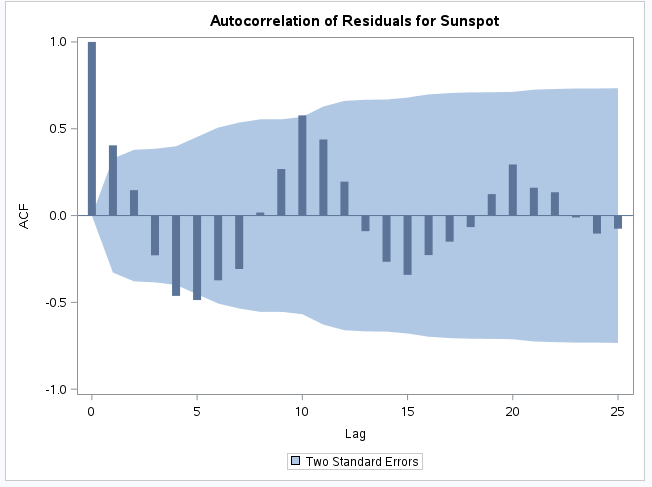
* When regressing on sunspot data (top) vs Melanoma (bottom) sunspots seem to account for much of the seasonal portion of the data, but there is no correlation between sunspots and an increasing trend (as the p-value for sunspots is .7979

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | DF | Estimate | SE | t Value | p |
| **Year** | 1 | 0.4432 | 1.7176 | 0.26 | 0.7979 |
| **Year** | 1 | 0.1173 | 0.009896 | 11.85 | <.0001 |

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
* Autocorrelation residuals for sunspots shows a pattern repeating every 10 years



* 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
     2. Locate the AIC statistic for each of the models and compare them
* Melanoma with AR10 model applied

