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. I will discuss melanoma in class but it is optional for the prelive assignment

Sunspot Data

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 predictors present.
   * Plot Sunspot versus Years. Visually do you think this time series is stationary. Again do your best based on what you got out of the videos. We will take a deeper dive in class.
   * Using the code below 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. Note that with no nlag option, it is just fitting a regression model with an intercept and nothing more.
   * 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 for each model. What do you make of them, say AR(1) model compared to the AR(4)?
     2. Locate the AIC statistic for each of the models and compare them
   * Try to forecast the next 10-20 years using the model that has the lowest AIC. 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> Hint: Note that you need to merge an extra few “future” observations to the data set with missing observations of which you want to forecast.

SAS code below to get you started.

 data Melanoma;

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

input Year Melanoma Sunspot;

run;

\*Modeling sunspot with simple regression, intercept only;

proc autoreg data=Melanoma all plots(unpack);

model Sunspot= ;

run;quit;

    \*Modeling sunspot with AR(1);                                                                                                                                                                                                                                         proc autoreg data=Melanoma all plots(unpack);

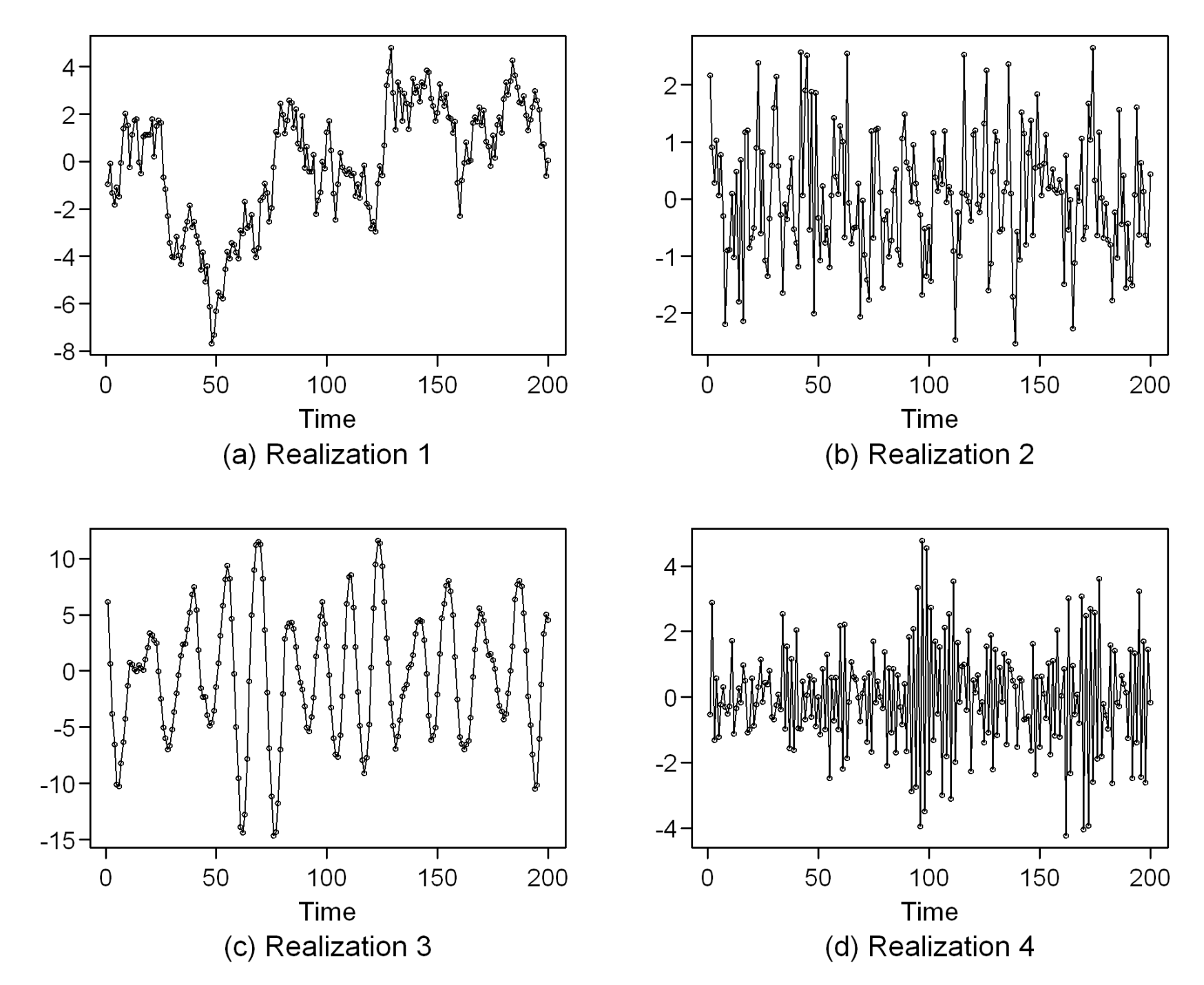
model Sunspot= / nlag=1;

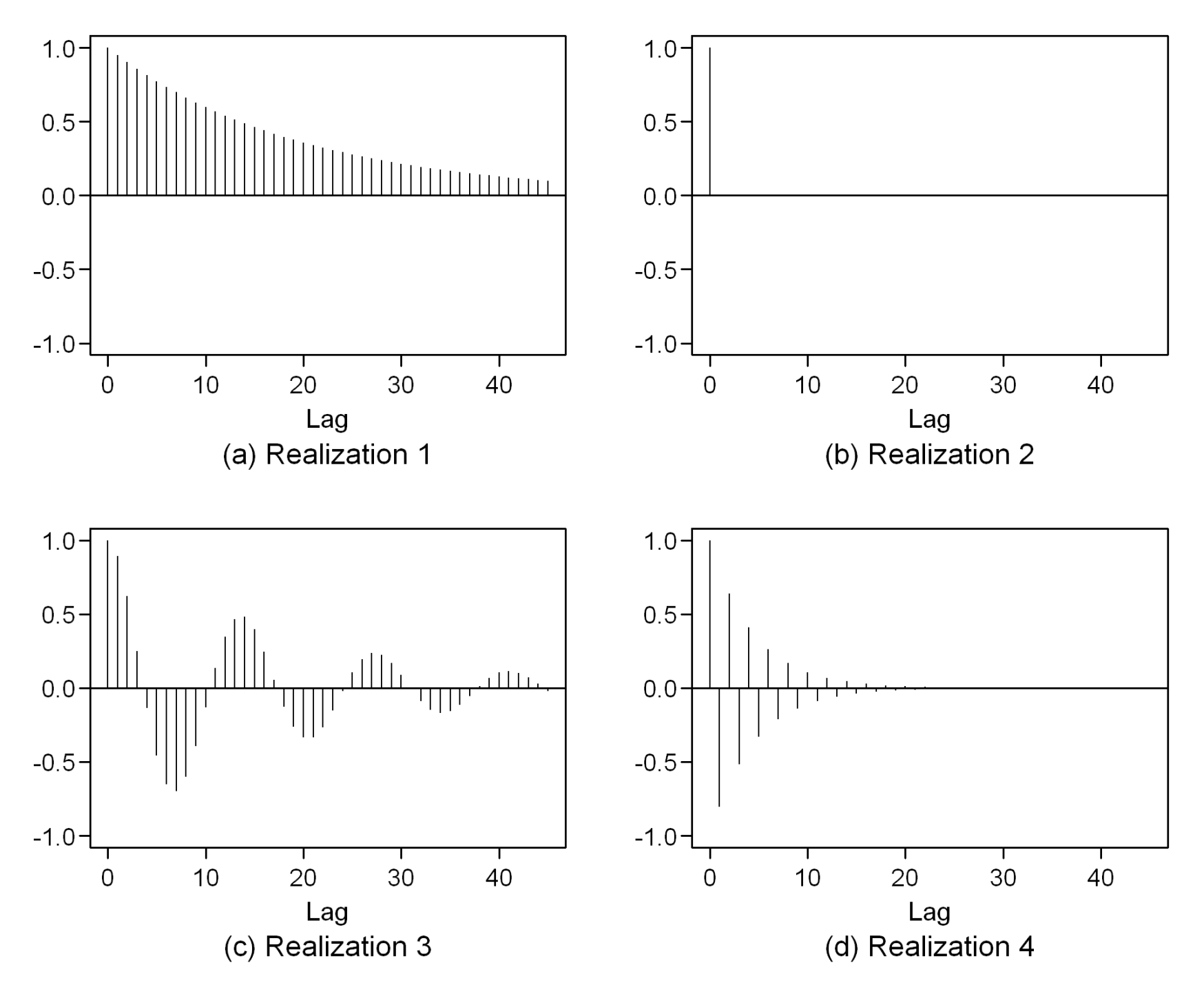
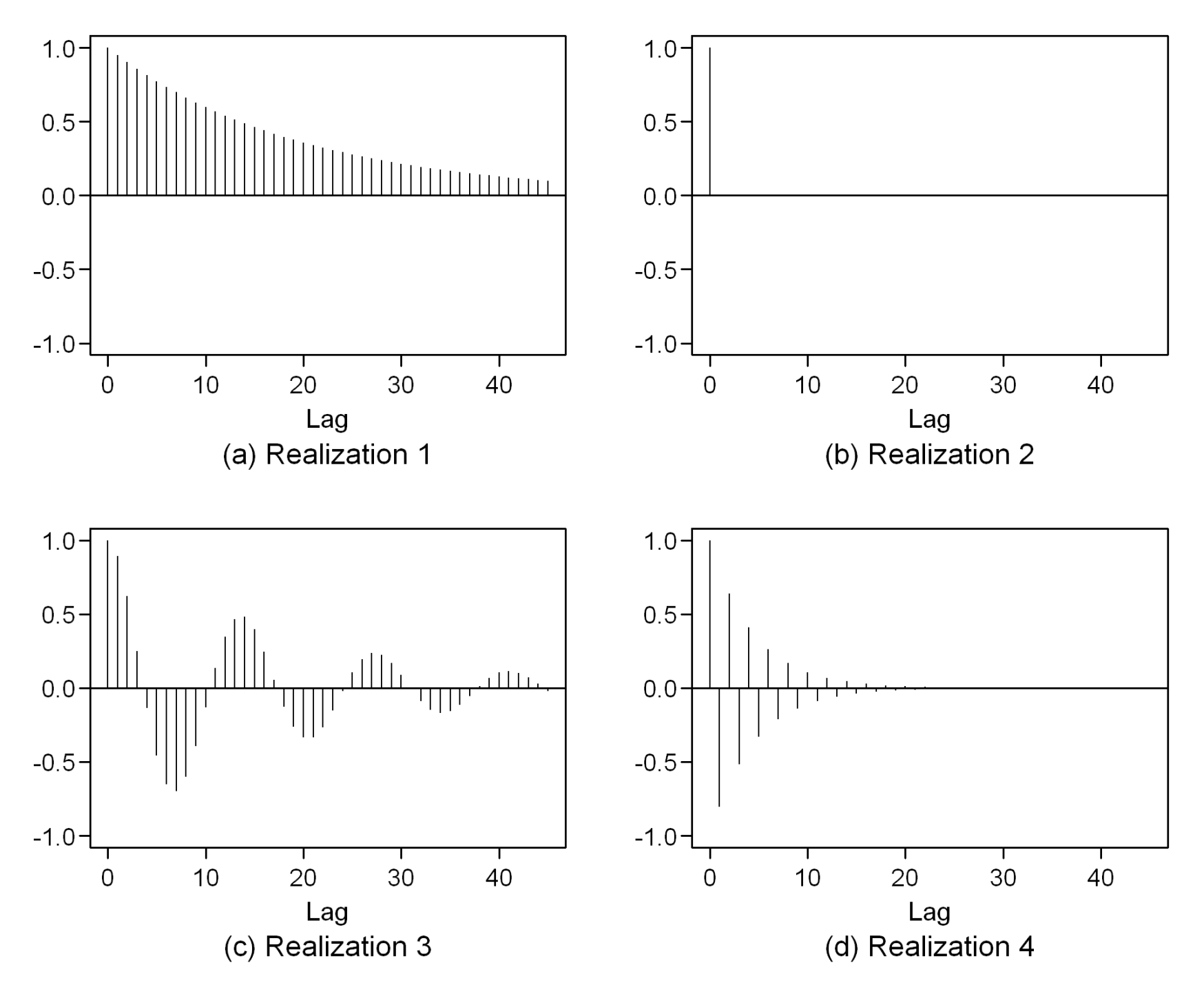
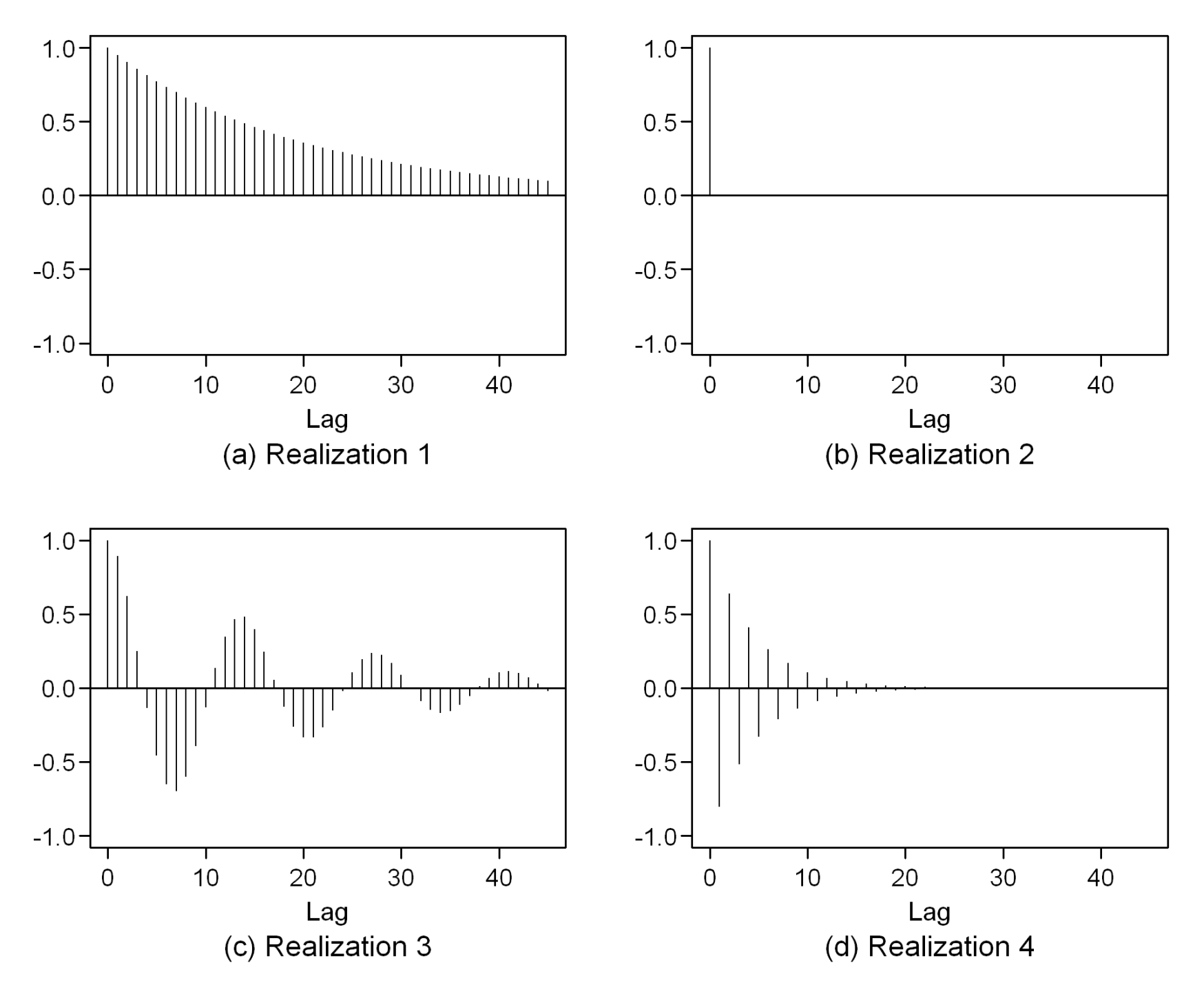
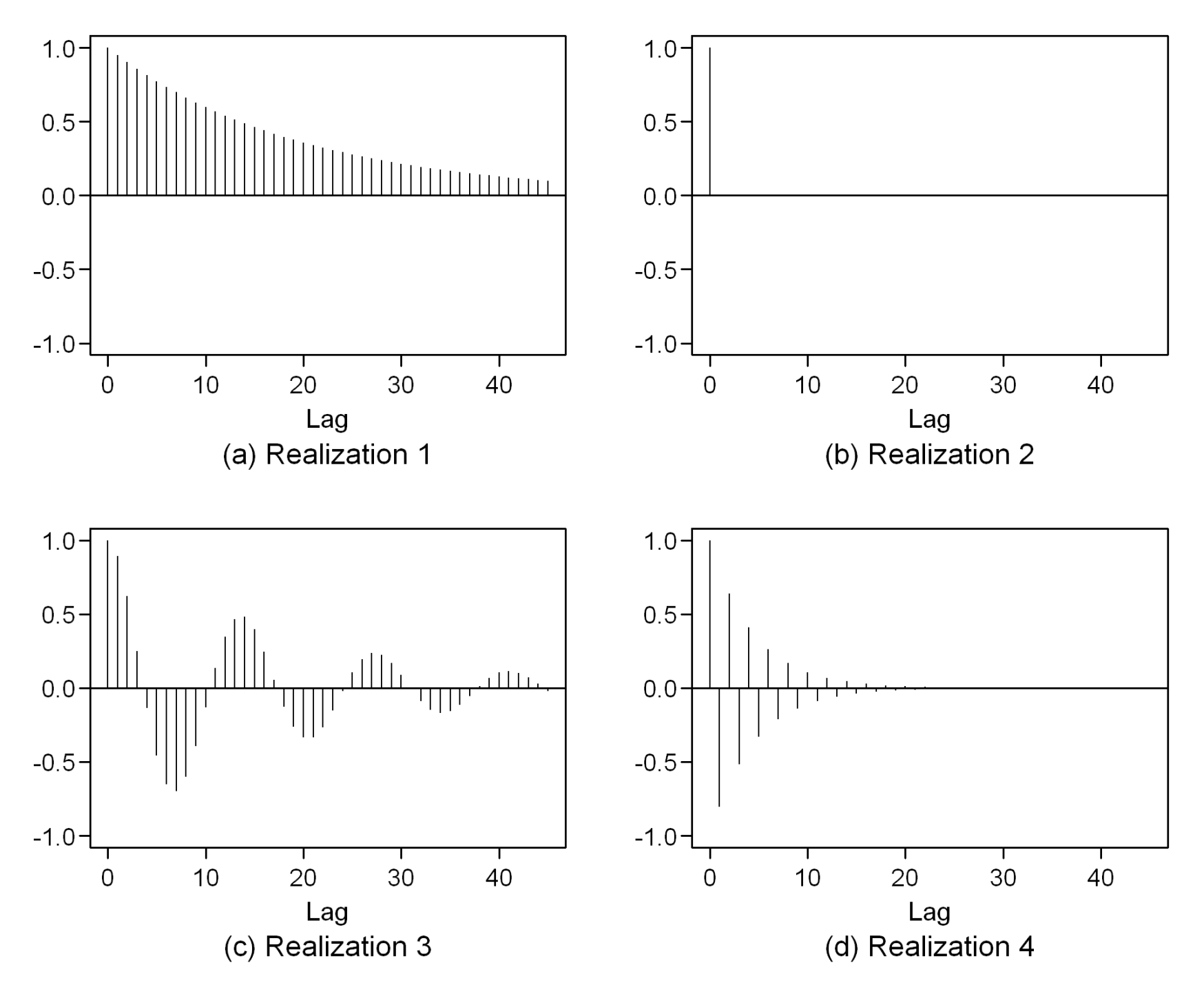
run;quit;

Give this next question a good college try. I recognize this one is a little tough without hearing me speak directly on it.

The autocorrelation plot tool is an extremely useful tool for diagnosing fits for time series models. My experience in the MSDS program is that this graph and the idea of stationary gives the most fits.

* + Go to my power point slide deck and take a look at slides 9-28 in presentation mode so you can see the automations. I will discuss and record these slides during office hours on Saturday so check it out if you want to hear commentary. The point is to illustrate more step-by-step how the acf graph is created.
  + Once you feel a little bit better about the acf plot. Match the following four time series with their corresponding acf. Zoom in





1

4

3

2

 SAS code below to get you started.

 data Melanoma;

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

input Year Melanoma Sunspot;

run;

\*Modeling sunspot with simple regression, intercept only;

proc autoreg data=Melanoma all plots(unpack);

model Sunspot= ;

run;quit;

    \*Modeling sunspot with AR(1);                                                                                                                                                                                                                                         proc autoreg data=Melanoma all plots(unpack);

model Sunspot= / nlag=1;

run;quit;

If you just want to play around with another data set I will discuss in class, take a look at the following questions for the Melanoma time series. You DO NOT need to do this for the pre live assignment unless you want to for fun and you have time.

Melanoma

* Plot Melanoma versus Years
  + Take a look back at what it means for a time series to be stationary. Does the melanoma time series look stationary to you?
  + 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.
  + 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.