Unit 4 Pre-Live

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 pre-live assignment.

# Sunspot Data

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.

## Question 1

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.

A screenshot of a cell phone

Description automatically generated

\*Plot Sunspot versus Years;

**proc** **sgscatter** data = Melanoma;

plot Sunspot \* Year;

**run**;

**No, I don’t believe that the time series is stationary.**

## Question 2

Using the code below as an example, fit a simple regression model to Sunspot with just an intercept. 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.

A picture containing screenshot

Description automatically generatedA screenshot of a social media post

Description automatically generated

\*Modeling sunspot with simple regression, intercept only;

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

model Sunspot= ;

**run**;

**quit**;

**The ACF plot shows a geometric decay, while the PACF plot shows also appears to be geometric, at least through lag=10. Would that make this a ARMA model?**

## Question 3

Fit an AR(1), AR(2), AR(3), and AR(4) model by specifying the nlag option to 1, 2, 3, or 4

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)? Locate the AIC statistic for each of the models and compare them.

|  |  |  |  |
| --- | --- | --- | --- |
| AR(1) | AR(2) | AR(3) | AR(4) |
| 373.081849 | **360.626686** | **355.053179** | **355.550762** |
| A screenshot of a cell phone  Description automatically generated | A screenshot of a social media post  Description automatically generated | A screenshot of a social media post  Description automatically generated | A screenshot of a social media post  Description automatically generated |
| A screenshot of a social media post  Description automatically generated | A screenshot of a social media post  Description automatically generated | A screenshot of a social media post  Description automatically generated | A screenshot of a social media post  Description automatically generated |

\*Modeling sunspot with AR(1);

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

model Sunspot= / nlag=**1**;

**run**;

**quit**;

\*Modeling sunspot with AR(2);

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

model Sunspot= / nlag=**2**;

**run**;

**quit**;

\*Modeling sunspot with AR(3);

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

model Sunspot= / nlag=**3**;

**run**;

**quit**;

\*Modeling sunspot with AR(4);

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

model Sunspot= / nlag=**4**;

**run**;

**quit**;

**AR(1) shows a geometric decay in its ACF plot, while AR(4) only shows one significant lag. I found the way the AIC values progressed interesting, although I’m not sure I understand why. It goes from ~373 to ~360 to finally ~355 for the final two plots.**

## Question 4

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 a few extra “future” observations to the data set with missing observations of which you want to forecast.

\*Forecasting data;

**data** Melanoma2;

input Year Melanoma Sunspot;

datalines;

1973 . 50

1974 . 35

1975 . 20

1976 . 30;

**run**;

\*Append Melanoma2 to Melanoma;

**proc** **append** base=Melanoma data=Melanoma2;

**run**; **quit**;

\*Check data;

**proc** **print** data = Melanoma;

**run**;

**quit**;

\*Proc autoreg;

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

model Sunspot= / nlag=**3**;

**run**;

**quit**;

A screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

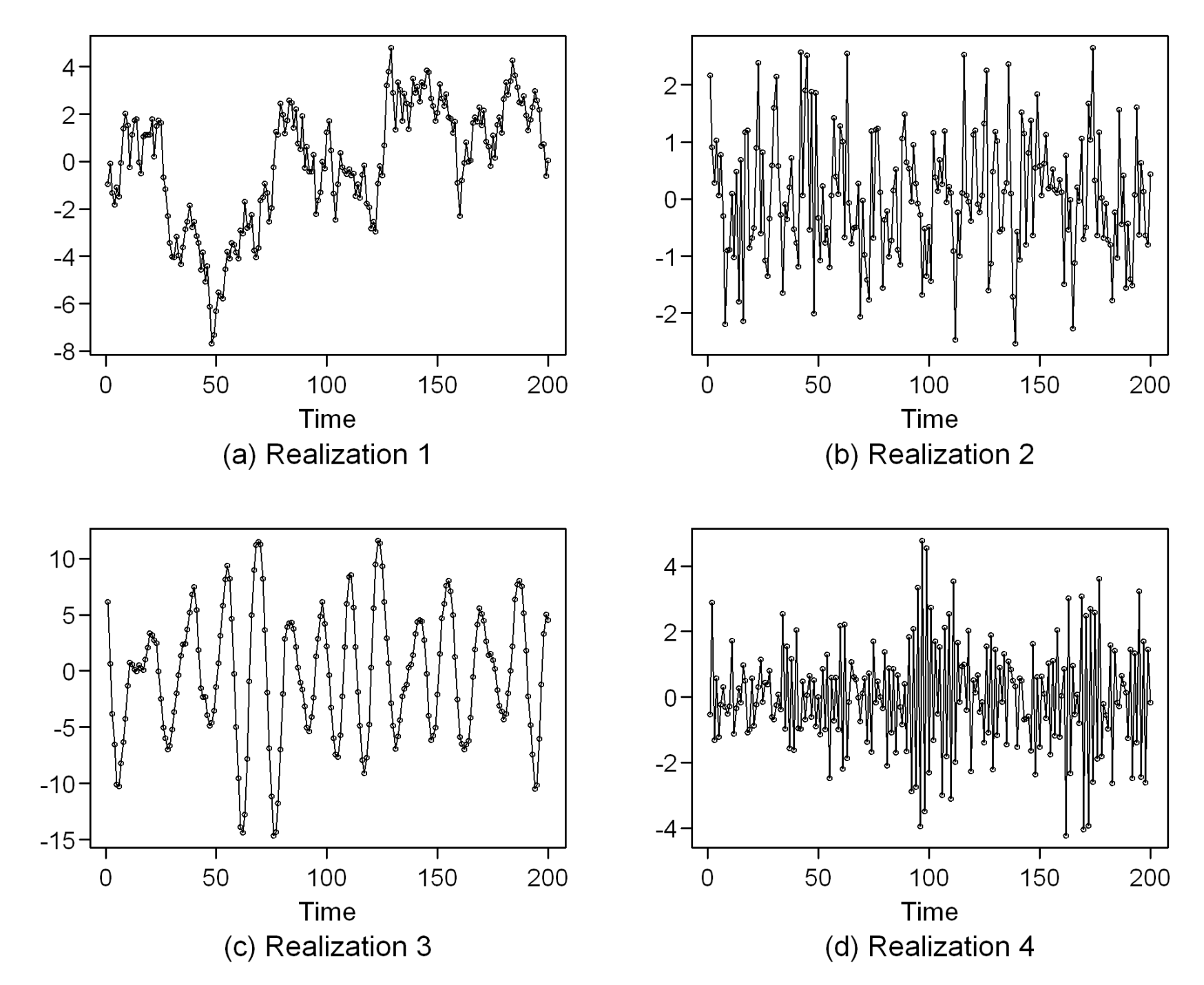
Description automatically generated

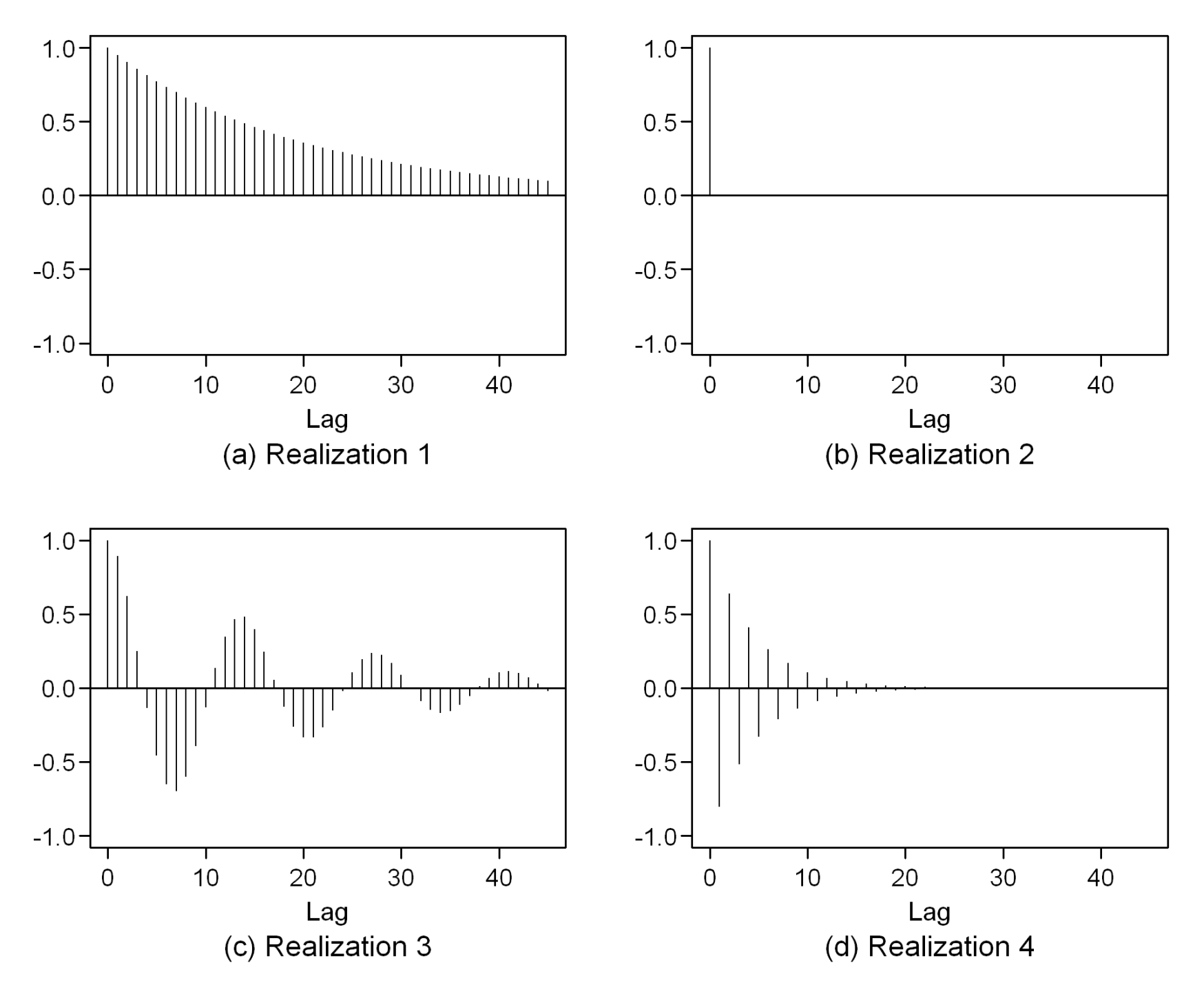
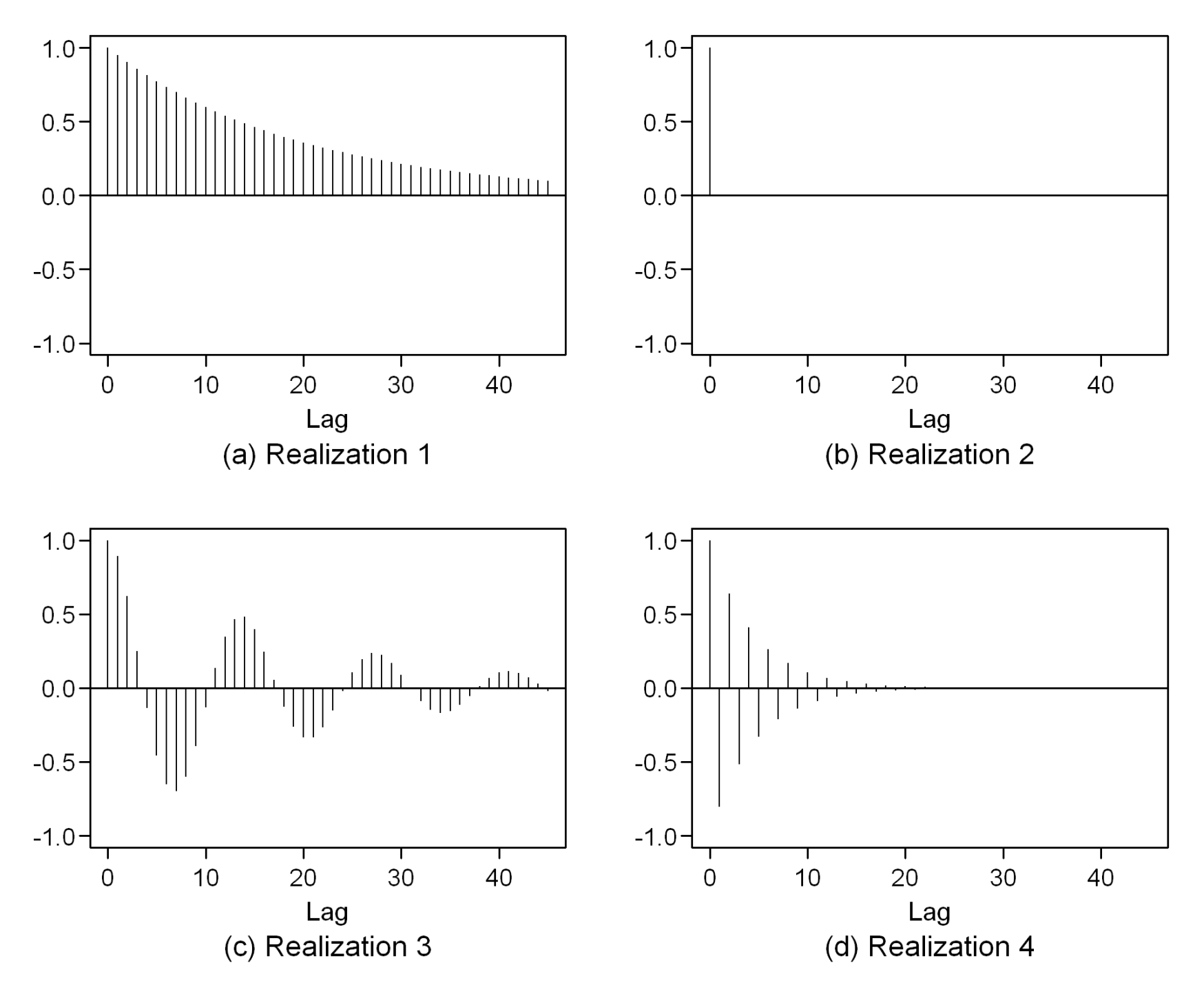
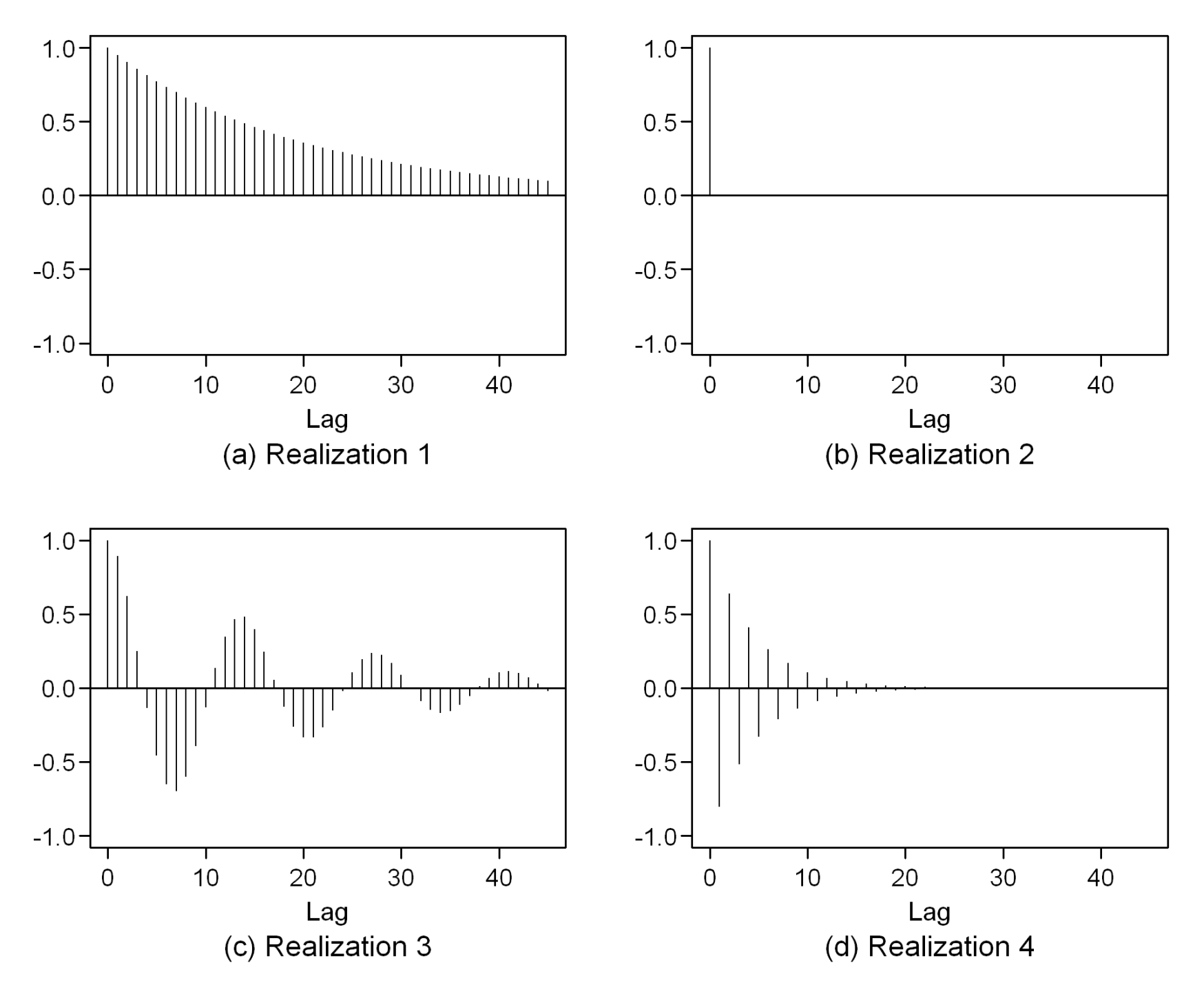
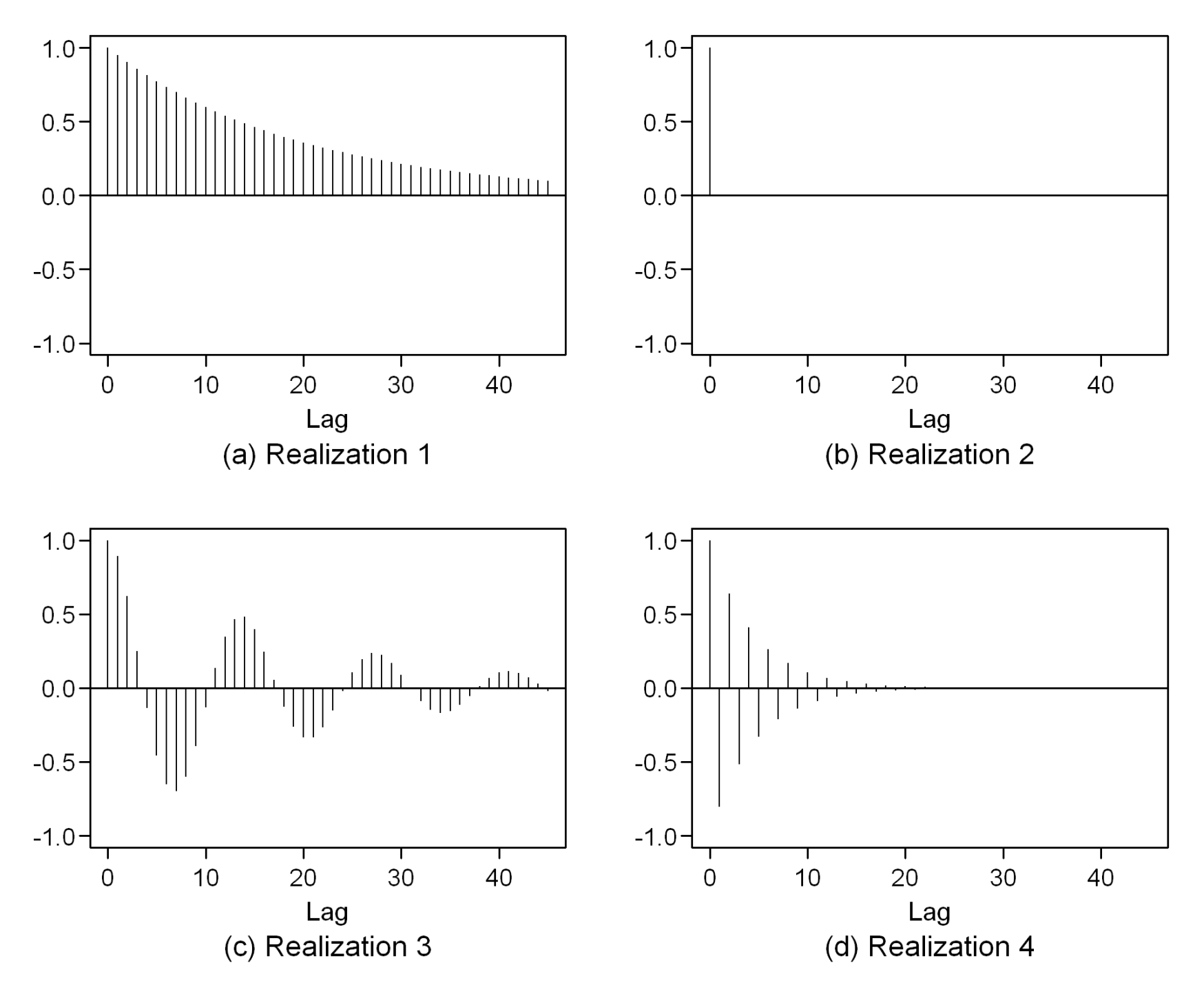
# Autocorrelation Plot

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.





1

4

3

2

1. **ACF 1 = Time Series D**
2. **ACF 2 = Time Series C**
3. **ACF 3 = Time Series B**
4. **ACF 4 = Time Series A**

# Extra Practice

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. It’s 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.