A change point analysis to determine if the measles vaccine was working.

The attached data set is a recording of known measles cases in the US over time starting from 1950 to 2008. The main question of interest is focused on when the vaccine for measles was licensed in 1963, the question, do we see a significant decrease in measles cases after this point in time?

***Analysis Question 1***

Examine the the scatter plot, autocorrelation, and partial autocorrelation plots of the raw time series. Does this support evidence of a stationary time series?

***Analysis Question 2***

Using proc auto reg to build an ordinary regression model that could be used to test if a change in the mean of measles cases using 1963 as a cut off. Examine the code below and understand what each regression coefficient represents. (You will need to convert Vaccine to a dummy variable 1=Yes 0=No). Examine the ACF and PACF with this regression model. Does the ACF and PACF support the assumption of a stationary time series now? What is the time series we are actually making a comment on now, is it the original time series or something else?

proc autoreg data=meas all plots(unpack);

class vaccine;

model cases = vacnum / dw=1 dwprob noint method=yw;

output out=p p=yhat pm=ytrend

lcl=lcl ucl=ucl;

run;

***Analysis Question 3***

Use your rule of thumbs and the previous ACF and PACF plot to propose a candidate

model to account for any correlation that exists. Also add the options dw=1 and dwprob

to the model statement of your previous code. What does the Durbin Watson test tell us about

the time series of the current model? Is the conclusion consistent with what is going on

in the ACF and PACF plots?

***Analysis Question 4***

Based on your decision of how to handle any serial correlation present in your regression model run that model and provide diagnostic evidence that you have accounted for the serial correlation present. Not all the residual diagnostics are given in proc auto reg, but given the ones that are there, is there any other concerns on the assumptions before moving on to hypothesis testing? Regardless of any concerns, what would be the conclusion of the test if all assumptions were actually met?

***Analysis question 5***

Plot the forecast of your final model from Question 5. There are numerous ways to potentially deal with some of the heavy tails that the residuals exhibit that questions the assumption of normality. There are numerous potential fixes to solve this problem. Two of them are…

1. Rather than create a single change point at year 1963, it may make sense to recode the yes/no Vaccine variable into 3 groups, and allow for a transitioning period say from 1963-1973. This may not completely deal with the issue but it could provide a better estimate of the final effect of the vaccine. We could also include a continuous predictor for year to model trending behavior in a smooth way rather than just step wise.
2. Recognizing that really large residuals are coming from the large values of measles counts, a log transformation could help dampen the behavior we see in the residuals. We would need to start over from scratch and go through the modeling process again. Also by taking the log transformation the overall trend of the time series may be more linear since it looks like it decreases exponentially.

Take the analysis approach of number 2. Go through a similar approach to build the model and answer the question. This is a really good example of just how iterative the modeling process can be.