STA 317 Test 2 Fall 2019 Name:

This test is to be completed independently. You should ask questions only to Dr. Miller. It is due by Monday night (by 11:59 pm) Nov 25. You are allowed your notes, textbook, etc. Essentially, you may use whatever resources you wish, other than another person. Good luck.

Please respond to all exercises and submit your completed test as an email attachment to [millercar@nku.edu](mailto:millercar@nku.edu) by the due date.

1. The Göta River in Sweden was observed for average annual discharge rates (volume of water in m3/s) between 1807 and 1957. The data file Gota.txt contains the values. Read these into R using a start date of 1807. You do not need to fit any trend or to take any differences for this data.
2. Create ACF and PACF plots for the data. Discuss what patterns you see and what models are suggested from these two plots. (10 points)
3. Consider an AR(1) model and an MA(1) model. Here, the goal is to fit a model where you are limited to a single parameter (other than the process mean – called the “intercept” in arima output). Fit these two models and discuss which is better based on the results. (10 points)
4. For the better model from part B, use approximate 95% confidence interval to estimate the unknown parameters. Notice one will be the “process mean” or the average annual discharge rate. The other is the parameter associated with the AR or the MA component.

Note: R negatives MA parameters, so if an output says theta-hat is 0.6789, then theta is estimated to be -0.6789, or Yt =  + et – (-0.6789)et-1 =  + et + 0.6789et-1 . An AR parameter is used as demonstrated in examples from class (no negating). (10 points)

1. The MOM (Method of Moments) estimate for the parameter is -0.6555. Does this value lie in the interval from part C for the comparable parameter? Which estimate do you believe is better to use and WHY? (10 points)
2. Forecast the next 2 years of discharge volumes for the Göta River. Standard errors for the first and second forecasts are and . Using your output, find the standard errors. Fill in the table with your values. (You can check work in R, but results may be slightly different because R reports ARIMA output as rounded. Just be sure values seem close to what you calculate.) (10 points)

Forecast 1 (1957) =

Forecast 2 (1958) =

For any other forecasts, the last line will be the forecast and the standard error, hence no need to venture further into the future.

1. Bovine blood sugar levels were recorded and placed in the file Cows.txt. This data consists of daily blood sugar readings for a single cow, measured in mg/100 ml of blood. (The cow was given 10 mg of dexamethasone intermuscularly in an attempt to increase milk production.) Fit all models using the method=’ML’ (ie: maximum likelihood) option in the arima() function. Do not use any differencing for this data.
2. Use BoxCox.ar() to determine if a transformation is needed. Include your graph showing that the log likelihood suggests (which option): (5 points)

if the interval is nearly centered at and includes 0, use log(Yt) as a transformation

if the interval is nearly centered at and includes 0.5, use sqrt(Yt) as a transformation

if the interval is nearly centered at and includes 1, do not use a transformation

1. Looking at the EACF, what model would you fit? Fit this ARIMA(p,d,q) model. Save the results so you can fill in a table in part C. (5 points)
2. Overfit the model using both an ARIMA(p+1,d,q) and ARIMA(p,d,q+1) option. For each model, identify the estimated parameter values and standard errors to complete the table below. Circle or highlight any which are NOT significant. Based on the results, which model seems most appropriate? (15 points)

Model All parameters and standard errors

ARIMA(p,d,q)

ARIMA(p+1,d,q)

ARIMA(p,d,q+1)

1. For your chosen model from part C, obtain the residuals. If you used cowmodel<-arima() in fitting the model, then cowmodel%residuals can be used to access the residuals. Complete a test to determine if the residuals appear to follow a normal distribution. Also determine if the independence assumption appears to be met for the residuals. (10 points)
2. Provide the forecast for the next week of readings, that is provide 7 forecast blood sugar levels for the cow. Give the actual value and provide a plot showing the blood sugar level from day 150 using n1=150 in the plot command. (10 points)
3. If we followed the forecast “for a really long time”, what would the eventual forecast become? Give the result based on you model output from part B. (5 points)