

# Stat 510 Take Home Exam

## Due Date

Monday May 1, by midnight U.S. Eastern Daylight Time.

## Scoring

There are two data analysis questions. The two questions count equally.

## Instructions

Please submit a MS Word document or PDF file to the designated Drop Box in the Drop Boxes folder. Collaboration with other students is NOT allowed. Please do NOT discuss exam questions on the message board.

## Data Analysis Questions

1. Use the e2x.dat and e2y.dat datasets from the Datasets folder. In R, the dataset could be read using something like `x=scan("e2x.dat")` and `y=scan("e2y.dat")`. The data are 150 simulated values in sequential order.
  - (a) Do an ordinary simple linear regression relating e2y to e2x. Give the estimated intercept and slope along with their standard errors as the answer to this part.
  - (b) Examine the ACF and PACF of the residuals from the regression in part (a). For this part, show the graphs and write a brief explanation of what ARIMA model may be appropriate for the residuals.
  - (c) Using the arima command with the xreg option, estimate the coefficient(s) of the regression model and the ARIMA model for the residuals identified in part (b) with maximum likelihood. Give the estimated coefficient(s) and their test statistics as the answer to this part.
  - (d) Discuss whether the model in part (c) is a suitable fit. If your choice is not suitable, repeat step (c) with an amended ARIMA model and refit using the arima command with the xreg option to improve the fit. You need only provide output for your final model.
  
2. Use the dataset e2q2.txt in the Exam 2 Module. (You may read in a .txt file just as a .dat file.)
  - (a) Plot the data, interpret the plot, and comment on two possible threshold values: -82 and -86. Based on the TS plot alone, which do you think is a better choice and why? Would you suggest another value?
  - (b) For threshold  $c = -82$ , estimate an AR(1) model for the original data in each of the two regions. Provide the model output and discuss significance of terms.

Note: Because an AR(1) is requested, you will need to make adjustments to the code provided in Lesson 13.2:

```
model = ts.intersect(y, lag1y=lag(y,-1))
P = model[,2]
```

Because `P` now contains just one column, you may replace `P[,1]` with `P` in the rest of your code.

- (c) For threshold  $c = -86$ , estimate an AR(1) model for the original data in each of the two regions. Provide the model output and discuss significance of terms.
- (d) Comment on the suitability of each model and compare the actual and predicted values. Is either threshold value preferable?