Complete the exam, include **all your code and all outputs** from R. Submit your solutions on blackboard or through email, by 11:59pm Saturday, May 1st. You are to work completely independently on this exam; however you may use notes, your textbook, etc.

1. (30pt) Consider the yearly global temperature data in period 1900-1997. The data set is given on blackboard. You may use the following to read in the data and make it a time series:
dt=read.table("your directory/globaltemps.txt", header=T)
temps=ts(dt$Temps, start=1900)

(a) Fit a simple linear regression model to the data, where y_t is the yearly global temperature x_t is time. Report the ANOVA table and summary for the model coefficients. Plot of the data with the least squares regression line overlaid.
(b) Examine the residuals from your fitted model for normality and independence. Display the sample ACF. Do the residuals look to resemble a normal, zero mean white noise process?
(c) Conduct a Durbin-Watson test on the residuals. Comment on your conclusion.
(d) Use one iteration of the Cochrane-Orcutt procedure to estimated the regression coefficients. Also calculate the standard errors of the coefficients. Are the standard errors (from the Cochrane-Orcutt procedure) larger than the ones from simple linear regression?
(e) Instead of using a deterministic trend model, consider a model from the ARIMA(p, d, q) family. Choose a potential model and explain/defend your selection. Fit the model of your choice to the data and write out the the full model with estimated parameters.
(f) Use your model to forecast the global temperature for 1998-2003. Plot your forecast along with the prediction intervals.

2. (20pt) Tuberculosis, commonly known as TB, is a bacterial infection that can spread through the lymph nodes and bloodstream to any organ in your body. The dataset has the number of TB cases (per month) in the United States from January 2000 to December 2009. The data can be found on blackboard. You may use the following to read in the data and make it a time series:
dt=read.table("your directory/TB.txt", header=T)
tb=ts(dt$TB, start=2000, frequency=12)

(a) Construct a time plot of the data and describe any patterns in terms of overall trend and seasonality. Also construct ACF ad PACF plots. Describe any patterns you notice on ACF and PACF.
(b) Fit an additive model using the Holt-Winters method. Let the function choose the optimal smoothing parameters automatically. Report the smoothing parameters and coefficients. Superimpose the fitted values on the time plot.
(c) Consider an alternative approach using seasonal ARIMA (SARIMA) modeling, i.e. in ARIMA$(p, d, q)\times(P, D, Q)_s$ class. Choose a potential model (i.e. the values of p, d, q, P, D, Q, and s) and explain/defend your selection. Fit the model of your choice and write out the full model with estimated parameters.
(d) Check residuals from your SARIMA model for normality (histogram, qq-plot), for independence (ACF and the Ljung-Box test). Comment on your findings.
(e) Choose a final model (with the smallest AIC) between Holt-Winters and SARIMA. Report your selection and calculate the forecasts with prediction intervals for 5 future values. Display the forecasts and prediction bands visually.