36-402/608 Homework #6 due 10:30AM 2/25

1. Global warming (50 points)

   This problem uses the global warming data of Sleuth Chapter 15, case 2 from case1502.csv. See page 438 for a description.

   (a) Read in the data, and fit a simple regression model of temperature on year. Turn in the regression summary, and explain why the p-value for year might be incorrect.

   (b) Turn in a three-panel plot with panels for temperature vs. year, the ACF of the residuals, and the PACF of the residuals. *For this and subsequent parts, be sure to override the* `main=`*,* `xlab=`*, and* `ylab=` *arguments to produce a plot suitable for showing your boss. For example, nothing of the form "dtf$var" should show on the plots.*

   (c) Fit three `arima()` models which include a linear trend in time and use the time series models: ARMA(1,0), ARMA(2,0), and ARMA(0,1). Show your code and the three AIC values, and state which model is best according to the AIC criterion.

   (d) Using the best model from part *c*, turn in a plot with panels for the ACF of the residuals, the PACF of the residuals, and the cumulative periodogram of the residuals (don't worry if the relative size of this plot looks funny). Use the following code to place an appropriate "outer" title over the three sub-plots.

   ```
   par(mfrow=c(3,1), oma=c(0,0,2,0)) # save outer margin space at the top
   ...
   mtext("My outer title", outer=TRUE, cex=1.4)
   ```

   (e) As an aid to learning about complex R objects, turn in the result of applying `names()` to your best `arima()` model object of part *c*. Also turn in the result of `myArimaObject$coef` and `myArimaObject$var.coef` where "myArimaObject" is whatever you called the model object. You may want to explore some other components of the object, as well as examining the "Value" portion of `?arima` to see what is available by using the "$" operator on an "arima" object. Note that not all of these are useful.

   (f) As an aid to learning about complex R objects, turn in the results of these two commands, substituting your arima object for "myArimaObject".

   ```
   class(myArimaObject)
   methods(class="Arima")
   ```

Now you know that you can use, e.g., `coef(myArimaObject)`, and `vcov(myArimaObject)` as an alternate way to get the coefficients and variance covariance matrix. You can also use `tsdiag(myArimaObject)` and `predict(myArimaObject)` to do things not available with the "$" operator. Also note that the message "Non-visible functions are asterisked" indicates which method functions cannot be directly examined, even though they can be run and you can get help on some of them with "?".

(g) Turn in the plot that results from running `tsdiag()` on your best "arima" object of part c. Include an appropriate outer title. Note that finding all of the p-values > 0.05 in the Ljung-Box plot is a good indicator that you have removed all serial correlation.

(h) Turn in the code to calculate the t-value and p-value and your statistical conclusion for a test of no change in temperature over time vs. a change over time *using code rather than directly entering numbers for the t ratio.*

(i) Turn in the t-value from directly entering the numbers that appear in the printout of the arima object. Explain why it is different from the result of part h.

(j) Explain what you might do to check if the temperature pattern over the 108 years is curved rather than linear.

2. Appliance Sales (50 points)

(a) Load the appliance data from "appliances.dat" into a variable called "app", throwing away or ignoring columns 6 and 7. Turn in an EDA plot of the US sales (which are in thousands of appliances) over the available years for dishwashers, garbage disposals, refrigerators, and washing machines. Be sure to include 0 on the y-axis and to use different line types (`lty=`) so that the appliances can be distinguished in a black and white printout.

(b) Which appliance most clearly requires advanced techniques due to a seasonal (yearly repeating) pattern?

(c) We will examine dishwashers and disposals only. Will a linear trend be adequate?

(d) Create columns in the data.frame for a centered version of year (i.e., year minus the mean of year), and the square of that column.

(e) Run an ordinary linear regression for each of the two appliances over time including the square of time. Under what circumstances would the standard error and p-value reported by `lm()` be correct vs. incorrect?

(f) Turn in the ACF, PACF and cumulative periodogram for the residuals of the dishwasher model of part e.

(g) State your best guesses of the appropriate ARMA models for dishwashers and disposals.

(h) Fit the models of part g using `arima()` with `xreg=cbind()` binding together the centered year variable and its square. Turn in the `$coef` components of these arima models for both outcomes.

(i) Examine the ACF, PACF, and the cumulative periodogram of the arima residuals from part h for both outcomes, and turn in your conclusions.

(j) Turn in the p-value for the test of a curved time course (vs. the null hypothesis of just a linear change over time) for dishwashers, along with your statistical conclusion.