Lab Assignment 1, Math 590C
Due Nov. 3, 2005

Data: $T = 250$ observations on grinding wheel profile data from a mechanical engineering experiment are in the file "grind1.dat" and can be obtained at "http://mendota.umkc.edu/teaching/m561.html".

Objective: The objective of this assignment is first to determine an AR(p) model which "best" fits the data. Afterwards, also find a model within the class of the MA(q) or the ARMA(p,q) models which "best" fits the data. Finally, give an indication of which model fits "best" overall.

Criteria: Your criteria of a good fitting model at this point are basically a model that has the characteristics:

1. The residuals $\hat{\varepsilon}_t$ should behave like a white noise series (zero autocorrelations at all lags).
2. The model should have small variance of the residuals (small value of MSE) compared to other models.
3. The model should have as small a number of estimated parameters as is necessary for an adequate fit, the estimates should be statistically "significant", and hence the model should not involve a number of unnecessary parameters.

Comment: Points (2) and (3) above are to some extent in conflict because as additional parameters are included in the model, the residual variance usually becomes (at least slightly) smaller. As a compromise between (2) and (3), the use of "model selection criteria" has been suggested by many as a possible guideline. That is, among competing models, models are to be preferred that have smaller values of the criteria:

$$AIC = -2\ln(likelihood)/T + 2r/T = \ln(\tilde{\sigma}^2) + 2r/T$$

$$BIC = -2\ln(likelihood)/T + \ln(T)r/T = \ln(\tilde{\sigma}^2) + \ln(T)r/T$$

$$FPE = \hat{\sigma}^2(1 + r/T)$$

where T is the (effective) series length, r is the number of estimated parameters in the model (including the constant term), $\tilde{\sigma}^2 = \sum_{t=1}^{T} \hat{\varepsilon}_t^2/T = (T-r)MSE/T$, and $\hat{\sigma}^2 = MSE = \sum_{t=1}^{T} \hat{\varepsilon}_t^2/(T-r)$, which is an approximately unbiased estimator of $\sigma^2 = Var(\varepsilon_t)$.

To-do List: To carry out the above analysis, you should do (at least) the following using Minitab:

1. make a time series plot of the data (Graph > Time Series Plot),
2. obtain summary statistics (mean, standard deviation, etc.) of the data (Stat > Basic Statistics > Display Descriptive Statistics),
3. obtain the sample ACF of the data (Stat > Time Series > Autocorrelation),
4. obtain the sample PACF of the data (Stat > Time Series > Partial Autocorrelation),
5. specify and fit an appropriate model using ARIMA(p,d,q)(Stat > Time Series > ARIMA). Your choice of model(s) to be estimated should be based in part on the features of the sample ACF and PACF of the series. For the AR(p) models, at least fit AR(2) and AR(3). For the MA(q) models, at least fit MA(2), MA(3) and MA(4). For ARMA(p,q), at least fit ARMA(1,1), ARMA(1,2), ARMA(2,1) and ARMA(2,2).

   Make a table of AIC, BIC and FPE for the above models and found the best model according to each criterion.

($d > 0$ is needed if the series appears nonstationary such as random walk or trend behavior; nonstationary is assessed from the features of the time series plot of the series and by the behavior of the sample ACF fo the series; for this example, you can assume the series is stationary and use $d = 0$. In cases where the series appears nonstationary, the sample ACF and PACF of the differenced series will need to be examined to determine an appropriate model for the stationary differenced series.)

6. obtain the sample ACF of the residuals $\{\hat{\varepsilon}_t\}$.

7. make scatter plots of the residuals vs fitted values and of the residuals vs data order; also make a histogram of residuals, checking for extreme residual values (outliers) or unusual patterns, and make normal scores plot to check for normality.

In checking a fitted model for adequacy, a major consideration is that the residuals should behave like white noise. Hence, in particular, the residual ACF should be examined to check if it is similar to that of a white noise series, i.e., zero autocorrelations at all lags. Consider using $\pm 2/\sqrt{T}$ limits as guides in checking the sample ACF of residuals for white noise characteristics, and also use overall Ljung-Box Q-statistics as checks on model adequacy.

Report Writing: Your data analysis should be written in a *short* report which describes your results and which includes the necessary information to support your results. Provide only the most relevant computer output; all pages of the report must be on 8 $1/2'' \times 11''$ paper, so any computer output must be of this size. The written report should be brief and to the point, and should describe the steps that lead to the model selection, the methods used, and the justification of the results. The report should be in the form of a brief written report with reference to plots, graphs, and so on, followed by the necessary plots, graphs, etc, which should be clearly labeled. The report should begin with a summary page that includes your name, the title of the data series analyzed, and a statement of the final model fitted (including estimated values of all parameters, including the residual variance), and a brief statement to support the selection of the model.