Definitions for Time Series

Observation: DV score at one time period. The score can be from a single case or an aggregate score from numerous cases

Lag: time periods between two observations. For example, lag 1 is between Y_t and Y_{t-1}. Lag 2 is between Y_t and Y_{t-2}. Time series can also be lagged forward, Y_t and Y_{t+1}.

ARIMA (*p, d, q*): Acronym for an auto-regressive integrated moving average model. The three terms to be estimated in the model are auto-regressive (*p*), integrated (trend—*d*), and moving average (*q*).

Auto-regressive terms (*p*): number of terms in the model that describe the dependency among successive observations. Each term has an associated correlation coefficient that describes the magnitude of the dependency. For example, a model with two auto-regressive terms (p=2) is one in which an observation depends on (is predicted by) two previous observations.

Moving average terms (*q*): number of terms that describe the persistence of a random shock from one observation to the next. A model with two moving average terms (q=2) is one in which an observation depends on two preceding random shocks.

Stationary and Nonstationary series: Stationary series vary around a constant mean level, neither decreasing nor increasing systematically over time, with constant variance. Nonstationary series have systematic trends, such as linear, quadratic, and so on. A nonstationary series that can be made stationary by differencing is called "nonstationary in the homogenous sense."

Differencing: Calculating differences among pairs of observations at some lag to make a nonstationary series stationary.

Trend terms (*d*): terms needed to make a nonstationary time series stationary. A model with two trend terms (d=2) has to be differenced twice to make it stationary. The first difference removes linear trend, the second difference removes quadratic trend, and so on.

Autocorrelation: Correlations among sequential scores at different lags. Because numbers are part of the same series, they will be sequentially related. The lag 1 autocorrelation coefficient is similar to correlation between the pairs of scores at adjacent points in time, $r_{Y_t,Y_{t-1}}$ (e.g., the pair at time 1 and time 2, the pair at time 2 and time 3, and so on). The lag 2 autocorrelation coefficient is similar to correlation between the pairs of scores two time periods apart, $r_{Y_t,Y_{t-2}}$ (e.g., the pair at time 1 and time 3, the pair at time 2 and time 4, and so on).

Autocorrelation function (ACF): pattern of autocorrelations in a time series at numerous lags; the correlation at lag 1, then the correlation at lag 2, and so on.

Partial autocorrelation function (PACF): pattern of partial autocorrelations in a time series at numerous lags after partialing out the effects of autocorrelations at intervening lags. ARIMA models are identified by matching obtained patterns of ACF and PACF plots with idealized patterns (see handout).

Random shock: The random component of a time series. The shocks are independent, normally distributed and reflected by the residuals (or errors) after an adequate model is identified. If a model is good, then all sequential contingencies are removed so that you are left with randomly distributed shocks.