



Published on STAT 510 (<https://onlinecourses.science.psu.edu/stat510>)

[Home](#) > Lesson 14: Review

Lesson 14: Review

Assignments:

- Read through the Lesson 14 online notes that follow.
- Complete Lesson 14 Assignment.

14.1: Course Summary

This document outlines the key ideas of the course. We won't give the details again. The purpose is to jar your memory a bit on what we've done and to make some connections between different elements of the course. The summary will be in a loose outline form, with some overlap between sections.

1. General Descriptive Tools for Univariate Time Series

- A. Time series plot of the data – look for trends and seasonality (*Lesson 1*)
- B. Sample ACF and PACF to identify possible ARIMA models (*Lessons 2 to 4* and we'll also outline important features in this document)
- C. Plots of x_t versus lagged values of x_t for various lags
- D. Smoothing to more clearly see trends and seasonality (*Lesson 5*)
- E. Decomposition models (additive and multiplicative) for sorting out trend and seasonality and to estimate seasonally adjusted values (*Lesson 5*)
- F. Periodogram and Spectral Density to identify dominant frequencies and periods/cycles in data (*Lessons 6 and 12*)

2. ARIMA Model Building – (*Lessons 1 to 4*)

- A. Basic structure of ARIMA Models

The basic structure is that a value at time t is a function of past values (AR terms) and/or past errors (MA terms).

A stationary series is one for which the mean, variance, and autocovariance structure remain constant over time. In the presence of non-stationarity, use differencing to create a stationary series.

- B. Identification of ARIMA models

Use the ACF and PACF together to identify possible models. The following table gives some rough guidelines. Unfortunately, it's not a well defined process and some guesswork/experimentation is usually needed.

Combined ACF and PACF Pattern

Possible Model

Sample ACF

Sample PACF

Tapering or sinusoidal pattern that converges to 0, possibly alternating negative and positive signs

Significant values at the first p lags, then non-significant value

AR of order p

Significant values at the first q lags, then non-significant values

Tapering or sinusoidal pattern that converges to 0, possibly alternating negative and positive signs

MA of order q

Tapering or sinusoidal pattern that converges to 0, possibly alternating negative and positive signs

Tapering or sinusoidal pattern that converges to 0, possibly alternating negative and positive signs

ARMA with both AR and MA terms – identifying the order involves some guess work

C. Seasonal Models (*Lesson 4*)

- Seasonal models are used for data in which there are repeating patterns related to times of the year (months, quarters, etc.). Seasonal patterns are modeled with terms connected to seasonal periods – e.g., AR or MA terms at lags of 12, 24, etc, for monthly data with seasonal features.
- It is nearly always the case that seasonal data will require seasonal differencing – e.g., a 12th difference for monthly data with a seasonal pattern.
- To identify seasonal patterns using the sample ACF and PACF, look at the patterns through the seasonal jumps in lags – e.g., 12, 24, etc. for monthly data

D. Model Confirmation

- For a good model, the sample ACF and PACF of the residuals should have non-significant values at all lags. Sometimes we may see a barely significant residual autocorrelation at an unusual lag. This usually can be ignored as a sampling error quirk. Clearly significant values at important lags should not be ignored – they usually mean the model is not right.
- Ideally, the Box-Ljung test for accumulated residual autocorrelation should be non-significant for all lags. This sometimes is hard to achieve. The test seems to be quite sensitive and there is a multiple inference issue. We're doing a lot of tests when we look at all lags so about 1 in 20 may be significant even when all null hypotheses are true.
- Use MSE, AIC and BIC to compare models. You may see two or more models with about the same characteristics. There can be redundancy between different ARIMA models.

E. Prediction and Forecasting with ARIMA models (*Lesson 3*)

Details given in Lesson 3 and software will do the work. The basic steps for substituting values into the equation are straightforward –

For AR type values, use known values when you can and forecasted values when necessary

For MA type values, use known values when you can and 0 when necessary

Exponential smoothing and related methods are sometimes used as more simple forecasting methods.
(Lesson 5)

3. Variations of univariate ARIMA Models

We examined three variations of univariate ARIMA models:

- A. ARCH and GARCH models used for volatile variance changes (Lesson 11)
- B. Fractional differencing as an alternative to ordinary differencing (Lesson 13)
- C. Threshold AR models which allow different AR coefficients for values above and below a defined threshold (Lesson 13)

4. Relationships Between Time Series Variables

A. Ordinary regression with AR errors (Lesson 8) – used when we have a dependent variable (y) and one or more predictors (x -variables), with all variables measured as time series.

We start with ordinary regression methods, then examine the AR structure of the residuals and use that structure to adjust the initial least squares regression estimates.

B. Lagged Regression for relationship between a y -variable and an x -variable (Lessons 8 and 9)

In a lagged regression, we use lags of the x -variable and possibly lags of the y -variable to predict y . The cross correlation function (CCF) is used to identify possible models.

C. Examining the CCF to Identify a Lagged Regression Model

The CCF of the original series may provide what you need for identifying the model, but in some instances you may need to “pre-whiten” the y and x series in some way before looking at the CCF. (Lesson 9)

Pre-whitening steps might be one of the following –

- Difference each series and the look at the CCF for the two differenced series
- Determine an ARIMA model for x . Apply that model to both x and y , to get “residuals” for each series. Look at the CCF for the two residual series

When looking at the CCF –

- Clear spikes at a lag indicate that lag of x may be useful for predicting y
- A tapering or sinusoidal pattern emanating from a clear spike may indicate that a first lag (and/or second lag) of the y -variable may be helpful
- Another tool that may be useful is to examine plots of y versus lagged values of x

D. VAR models (Lesson 11)

Vector autoregressive models are multivariate time series models. A VAR model defines a regression system of models in which each variable is a function of lags of itself and all other variables under consideration. These are useful for relationships between variables which are similar – e.g., rainfall statistics from several different locations or commodities prices at several

different locations.

5. Comparisons of Groups

A. Intervention Analysis (*Lesson 9*)

This is a before/after comparison. We examine how a time series is affected by a new law or procedure that is imposed at some point in time.

B. Repeated Measures/Longitudinal Analysis (*Lesson 10*)

Here, we measure a (usually) short time series on different experimental units, divided into two or more treatment groups. The objective is to compare the treatment groups with respect to how differing treatments affect the response variable over time.

6. Frequency Domain (*Lessons 6 and 12*)

The periodogram and spectral density consider time series in the frequency domain. The underlying structure represents a time series as a sum of cosine (and/or sine) waves of varying frequencies. We look for the dominant frequencies.

Source URL: <https://onlinecourses.science.psu.edu/stat510/node/59>