SigmaXL® Version 9.0

Workbook

SigmaXL: Pronóstico de Series de Tiempo

Traducción al Español de Ivan Falconi

Contenido

Pronóstico de Series de Tiempo y Cartas de Control para Datos Autocorrelacionados 3
Introducción – Pronóstico de Series de Tiempo 3
Introducción – Cartas de Control para Datos Autocorrelacionados 6
Resumen de Funciones de Pronóstico de Series de Tiempo y Cartas de Control para Datos Autocorrelacionados 7
Carta de Corridas 9
Gráficos de Autocorrelación (FAC/FACP) 26
Gráficos de Correlación Cruzada (FCC) 37
Diagrama de Densidad Espectral 40
Gráficos de Descomposición de Tendencias Estacionales 46
Pronóstico de Suavización Exponencial 60
Pronóstico de Suavización Exponencial – Descomposición Estacional Múltiple (DEM) 102
Carta de Control de Suavización Exponencial 117
Cartas de Control Suavización Exponencial Descomposición Estacional Múltiple (DEM) 129
Pronóstico ARIMA 137
El Pronóstico ARIMA con Predictoras 166
Pronóstico Descomposición Estacional Múltiple ARIMA – (DEM) 192
Cartas de Control ARIMA 207
Cartas de Control ARIMA con Predictoras 223
ARIMA Carta de Control Descomposición Estacional Múltiple (DEM) 228
Utilitarios – Datos en Diferencia 237
Utilitarios – Datos de Rezagos 240
Utilitarios – Interpolar los Valores Ausentes 242
SigmaXL Apéndice 245
SigmaXL Versión 9.0 Feature List Summary 246
Distribuciones 250
Transformaciones 262

Pronósticos de Series de Tiempo y Cartas de Control para Datos Autocorrelacionados 266
Autocorrelación (FAC), Autocorrelación Parcial (FACP) y Correlación Cruzada (FCC) 266
La Prueba de Ljung-Box 268
Transformación Box-Cox 268
Frecuencia Estacional 269
Descomposición de Tendencia Estacional 269
Densidad Espectral y Detección Automática de la Frecuencia Estacional 270
Interpolación Lineal de Valores Ausentes Ajustados Estacionalmente 270
Criterio de Información para la Comparación del Modelo 271
Precisión del Pronóstico 272
Suavización exponencial - ETS 273
Promedio Móvil Autorregresivo Integrado - ARIMA 279
Evaluación de la Precisión de Pronósticos a través de Concursos de Pronósticos 283
Cartas de Control para Datos Autocorrelacionados 284
Referencias del Pronóstico de Series Temporales y Cartas de Control para Datos Autocorrelacionados 286
Acerca de SigmaXL, Inc. 288

Pronóstico de Series de Tiempo y Cartas de Control para Datos Autocorrelacionados

Introducción – Pronóstico de Series de Tiempo

De https://en.wikipedia.org/wiki/Time_series:

> Una serie de tiempo es una serie de datos indexados en una lista o gráfico ordenados en función del tiempo. Con frecuencia, una serie de tiempo es una secuencia espaciada a intervalos regulares de puntos en el tiempo.
>
> El análisis de series de tiempo comprende métodos que analizan los datos de las series de tiempo con el fin de extraer un significado estadístico y otras características de los datos. El pronóstico de las series de tiempo es el uso de modelos para predecir valores futuros observados anteriormente

El libro online de libre acceso de Rob Hyndman y George Athanasopoulos, "Forecasting: Principles and Practice," (https://otexts.com/fpp2/) es un excelente material de introducción para el análisis de las series de tiempo (a partir de ahora se lo define "fpp2"). El libro hace mención a la librería forecast de R (https://cran.r-project.org/web/packages/forecast/index.html), cuyas respuestas estimadas por SigmaXL se asemejan a aquellas que provienen del referido libro. Tener en cuenta que SigmaXL no se relaciona con R. Las pequeñas diferencias son producto de mínimas diferencias en la selección del algoritmo y los métodos de optimización.

Los siguientes artículos, cuya fuente es Wikipedia, son recursos útiles:

https://en.wikipedia.org/wiki/Exponential_smoothing

https://en.wikipedia.org/wiki/Autoregressie_integrated_moving_average

Box y Jenkins, et al., 2016, es el clásico libro sobre modelización con ARIMA (a partir de ahora se lo define como "Box y Jenkins"):

> Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2016). *Time Series Analysis, Forecasting and Control*, 5th edition, Wiley.

Los siguientes datos se utilizarán:

1. Concentración de Un Proceso Químico – Series A (Box y Jenkins)
2. Aerolínea Mensual de Pasajeros – Series G (Box y Jenkins)
3. Pasajeros Mensuales de la Aerolínea - Valores
4. Pasajeros Mensuales de la Aerolínea – Modificado para Cartas de Control

5. Demanda Diaria de Electricidad con Predictoras – ElecDaily (fpp2)

6. Ventas con Indicador – Serie M Modificada (Box y Jenkins)

7. Demanda Media-Hora de Electricidad Estacional Múltiple – Taylor (Pronóstico en R)

SigmaXL resuelve los siguientes métodos de pronóstico de series de tiempo:

- Suavización Exponencial
- Descomposición Estacional Múltiple (DEM) – Suavización Exponencial
- ARIMA – Box-Jenkins Promedio Móvil Autorregresivo Integrado
- ARIMA con Predictoras
- ARIMA – MSD

Por lo general, se puede utilizar la suavización exponencial o ARIMA. Resulta útil probar ambos para ver cuál ofrece un mejor modelo o usar el promedio del pronóstico de ambos métodos.

Si los datos son estacionales (es decir, influenciados por factores estacionales), SigmaXL requiere que se especifique la frecuencia estacional. La frecuencia es el número de observaciones por unidad de "ciclo" de tiempo, por lo que las ventas mensuales se especificarían como frecuencia estacional = 12 (observaciones por año). Los ingresos trimestrales se especificarían como frecuencia estacional = 4. Los datos por hora serían 24 (observaciones por día). Obsérvese que la "frecuencia estacional" también se conoce como "período estacional", pero SigmaXL sigue la convención como se estila en Hyndman y Athanasopoulos (fpp2).

La suavización exponencial se limita a una frecuencia estacional máxima de 24. Para frecuencias más altas, use la Descomposición Estacional Múltiple (DEM) - Suavización Exponencial -. En la DEM, el componente estacional se elimina primero mediante la descomposición, un modelo de suavización exponencial no estacional con un ajuste del residual (tendencia +), y luego se agrega el componente estacional nuevamente. Para el pronóstico, se usa un pronóstico naïve estacional para el componente estacional. Como su nombre lo indica, la descomposición estacional múltiple (DEM) es también flexible para incluir la estacionalidad múltiple.

El ARIMA no tiene un límite teórico respecto a la frecuencia, pero debido a la eficiencia computacional y para minimizar la pérdida potencial de observaciones a través de la diferenciación, se recomienda el uso de ARIMA - DEM para una frecuencia estacional superior a 52 (o con frecuencias múltiples). Si los datos incluyen predictoras continuas o categóricas, use ARIMA con predictoras. No está disponible el ARIMA con predictoras – DEM.

ARIMA parte del supuesto que la serie temporal es estacionaria, es decir, tiene la propiedad de que la estructura de la media, la varianza y la autocorrelación no varían con el tiempo. Si la media de una serie temporal no es estacionaria (por ejemplo, la tendencia), esto se puede corregir con la diferenciación, calculando las diferencias entre las observaciones consecutivas

para datos no estacionales y entre períodos consecutivos para los datos estacionales (por ejemplo, enero de 2019 - enero de 2018, etc.). Si la variación cambia con el tiempo, se puede aplicar una transformación Box-Cox para lograr una variación constante. La suavización exponencial no requiere estacionariedad.

En el Apéndice se proporcionan detalles sobre los métodos y fórmulas de pronóstico de series temporales utilizados en SigmaXL: **Pronóstico de Series de Tiempo y Cartas de Control para Datos Auto correlacionados**. Las referencias se dan en: Referencias para **Pronósticos de Series Temporales y Cartas de Control para Datos Auto correlacionados**.

Introducción – Cartas de Control para Datos Auto correlacionados

El control estadístico de procesos para procesos auto correlacionados ha utilizado normalmente el modelo de pronóstico de un paso hacia adelante EWMA (promedio móvil ponderado exponencialmente). El modelo de series de tiempo pronostica el movimiento en la media y grafica una carta de control de Individuales de los residuos para la detección de causas asignables. SigmaXL amplía este concepto que incluye todos los métodos de pronóstico mencionados anteriormente.

Woodall y Faltin proporcionan algunas pautas útiles para tratar la autocorrelación [20]:

- Si es posible, primero se debe intentar eliminar la fuente de la autocorrelación.
- Si no se puede eliminar directamente la fuente de autocorrelación, entonces puede ser posible modelar la autocorrelación y utilizar un esquema de control de retroalimentación para reducir la variabilidad sobre un valor objetivo especificado.
- Si la fuente de la autocorrelación no puede eliminarse directamente, y el control de retroalimentación no es una opción viable, entonces es importante vigilar el proceso con gráficos de control que no den señales que se repiten debido a la presencia de la autocorrelación.

Si no se tiene en cuenta la autocorrelación, se producirán límites demasiado estrechos que darán lugar a excesivas falsas alarmas, o límites demasiado amplios que provocarán errores. Un pronóstico preciso para su serie de tiempo significa que los residuos muy a menudo tendrán las propiedades correctas para aplicar correctamente una carta de control, lo que conducirá a una mejorada carta de control con reducidas falsas alarmas y fallas.

Además de crear una carta de control de Individuales para los residuos, se dispone de una carta de límites móviles, que utiliza la predicción de un paso como la línea central, por lo que los límites de control se mueven con la línea central.

Las funciones "Add Data", "Show Last 30" y "Scroll" en las Herramientas de las Cartas de SigmaXL están disponibles para estas cartas de control. Para "Add Data", los modelos de series temporales no se reajustan, sino que se utilizan para calcular los valores de los residuos de los nuevos datos.

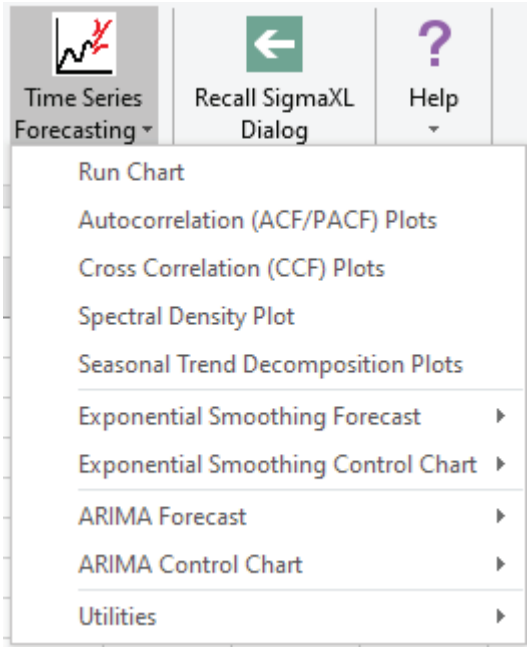
Revise el Apéndice para detalles adicionales y referencias: **Cartas de Control para Datos Autocorrelacionados**.

Resumen de Funciones de Pronóstico de Series de Tiempo y Cartas de Control para Datos Autocorrelacionados

- Cartas de Corrida
- Gráficos de la Función de Autocorrelación (FAC)/Autocorrelación Parcial (FACP)
- Gráficos de Correlación Cruzada (FCC) con la opción de Pre blanqueo de Datos
- Gráficos de Descomposición de Tendencia Estacional
- Gráficos de Densidad Espectral con Detección de Frecuencia Estacional
- Suavización Exponencial:
    - Pronósticos con Intervalos de Predicción
    - Modelos de Suavización Exponencial con la taxonomía de Rob Hyndman:
        - Error Aditivo/Multiplicativo
        - Tendencia Aditiva/Aditiva Reducida
        - Estacional Aditiva/Multiplicativa
        - Esto incluye los clásicos modelos de suavización exponencial tales como: Simple/Single/EWMA, Doble y Holt-Winters
    - Opción de Descomposición Estacional Múltiple (DEM)
        - Útil para datos de alta y/o múltiple frecuencia tales como mensual con frecuencia = 12, Diario con frecuencia = 7 y Horario con frecuencia = 24
- Cartas de Control de Residuos de Suavización Exponencial para datos auto correlacionados:
    - Carta de Individuales y Límites Móviles (con Pronóstico de Un Paso Hacia Adelante)
    - Opciones de Agregado de Datos, Mostrar los Últimos 30 Datos, Desplazar
    - Opción DEM
- Promedio Móvil Autorregresivo Integrado (ARIMA):
    - Pronóstico con Intervalos de Predicción
    - Pronóstico ARIMA con Predictoras (Continuas y/o Categóricas)
    - Opción DEM
- Utilidades:
    - Datos en Diferencia
    - ARIMA Datos de rezago
    - Interpola Valores Ausentes (interpolación lineal de ajuste estacional)
- Características del modelo de Pronóstico de Series Temporales:
    - ARIMA y los modelos de suavización exponencial son completamente automáticos o especificados por el usuario
    - Utiliza modernos modelos de espacio de estado y filtro de Kalman para una estimación precisa de los parámetros
    - ARIMA estima valores perdidos con el filtro de Kalman; la suavización exponencial utiliza interpolación lineal ajustada estacionalmente
    - Transformación Box-Cox Automática

  - Detección Automática de frecuencia estacional
- Diagnósticos de modelos:
  - Gráficos ACF/PACF
  - Valores p de Ljung-Box
  - Verosimilitud Logarítmica, AIC, AICc, BIC, DesvEst de los Residuos
  - Gráfico de Residuos (histograma, probabilidad normal, residuos versus ajustes, residuos versus orden)
- Precisión de pronóstico:
  o Errores de pronóstico de un paso hacia adelante en la muestra (estimación) (RMSE, MAE, MASE, MAPE)
  o Errores de pronóstico de un paso hacia adelante fuera de la muestra (retención)
  o Errores de pronóstico de múltiples pasos hacia adelante fuera de la muestra (retención)
  o Evaluado utilizando el pronóstico del estándar de referencia M4 con datos de competencia, un total de 100,000 conjuntos de datos con datos anuales, trimestrales, mensuales, semanales, diarios y por hora. Utilizando un promedio híbrido de suavización exponencial automático y ARIMA, SigmaXL (extraoficialmente) ocupó el décimo lugar entre 60 en el puntaje de precisión del pronóstico promedio ponderado general, por delante de tres conocidos paquetes de software comerciales de pronóstico.

Menú Pronóstico Series de Tiempo



Carta de Corridas

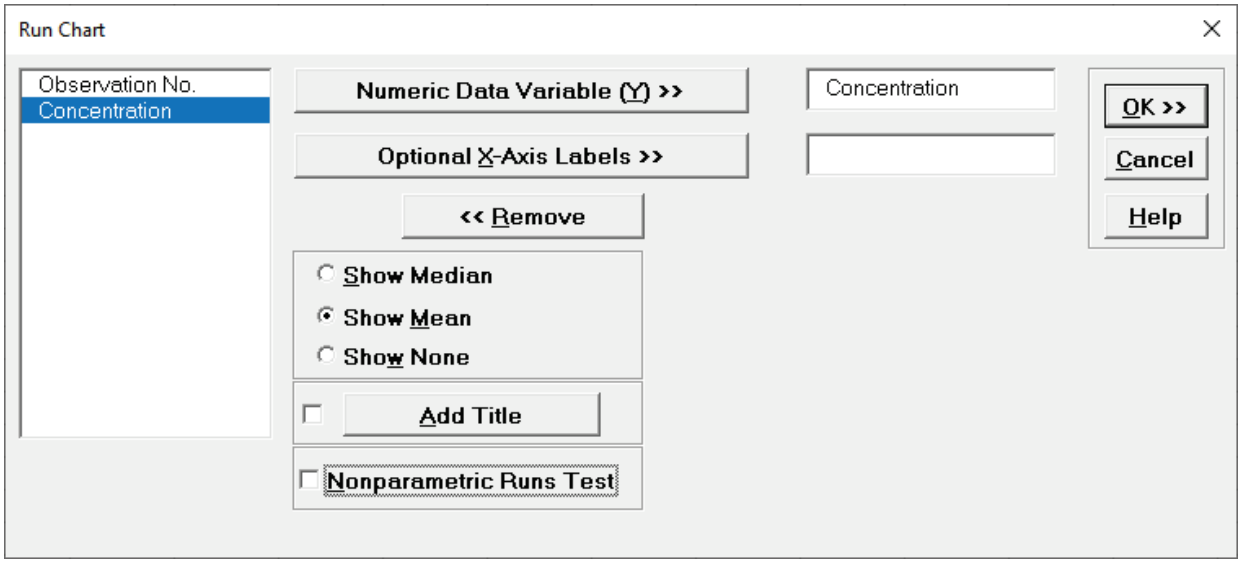
La Carta de Corridas en el Pronóstico de Series de tiempo es la misma herramienta empleada en Graphical Tools > Run Chart (Véase Parte H - Cartas de Corrida). Se demostrará el uso de las Cartas de Corrida para el análisis exploratorio de los datos de los ejemplos de series de tiempo, así también como el uso de los utilitarios de las diferencias e interpolación y el empleo de las herramientas gráficas para los datos con predictoras.

Concentración de un Proceso Químico - Serie A

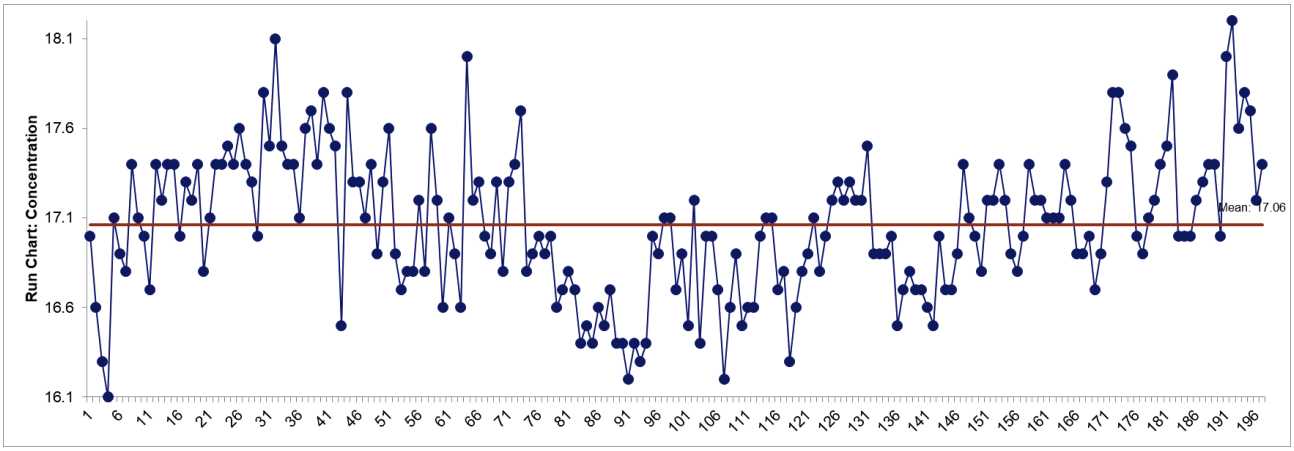
1. Abrir **Chemical Process Concentration – Series A.xlsx** (pestaña **Hoja 1**). Estos datos son una Serie A de Box y Jenkins, un conjunto de 197 valores de la concentración de un proceso químico que se han obtenido a intervalos de dos horas.

2. Hacer click en **SigmaXL > Time Series Forecasting > Run Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marque **Use Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Concentration*, hacer click en **Numeric Data Variable (Y) >>**. Seleccione **Show Mean**. Desmarcar **Nonparametric Run Test**.



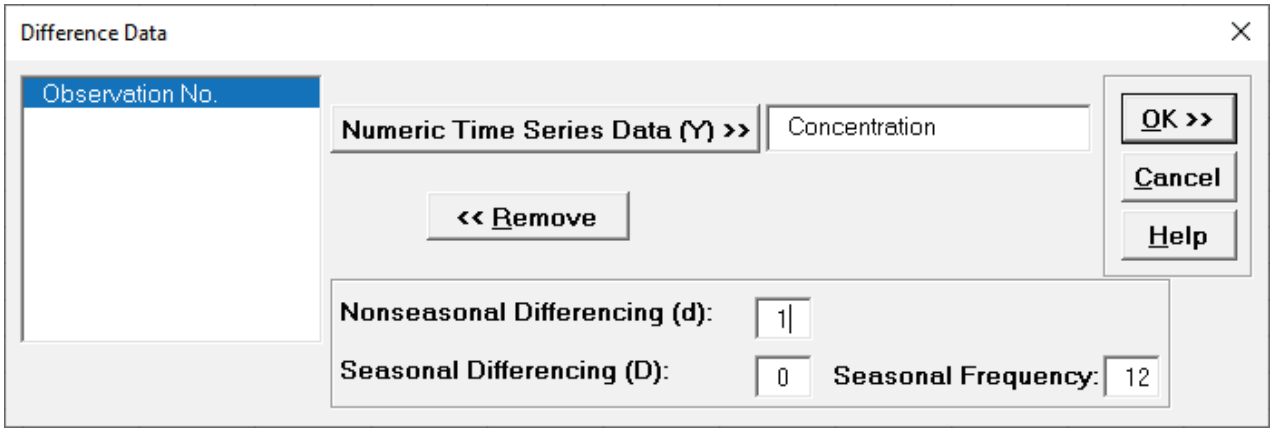
4. Hacer click en **OK**. Se genera una Carta de Corrida de la Concentración con una línea central de la Media.



5. Evidentemente, esto es un proceso no estacionario con una "media fluctuante". La suavización exponencial puede modelar esto, pero ARIMA requiere un proceso estacionario. Se demostrará cómo la diferenciación puede convertir un proceso estacionario. (Tomar nota que el Pronóstico ARIMA lo resolverá automáticamente).

6. Hacer click en la pestaña de la **Hoja 1**. Hacer click en **SigmaXL > Time Series Forecasting > Utilities > Difference Data.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marque **Use Entire Data Table**. Hacer click en **Next**.

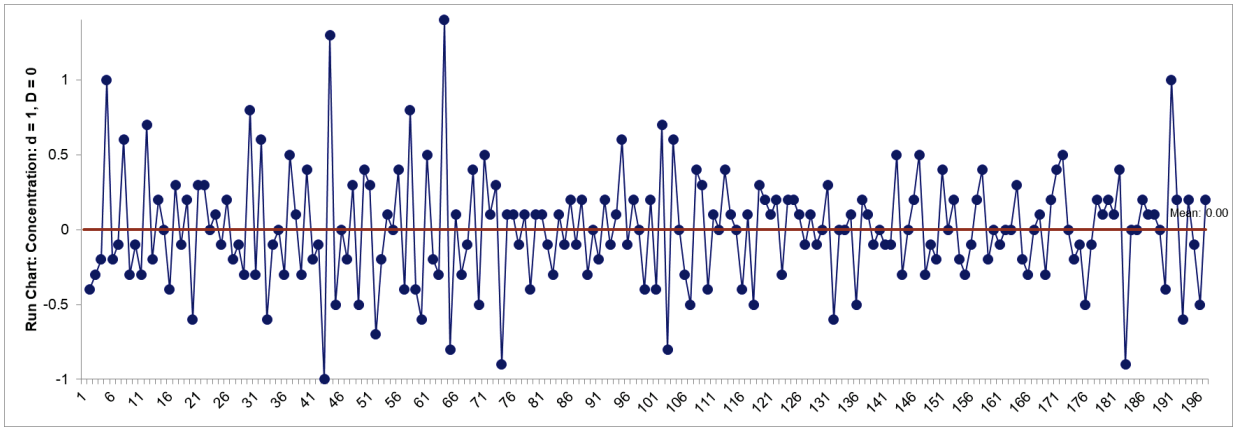
7. Seleccionar *Concentration*, hacer click en **Numeric Time Series Data (Y) >>**. Registrar 1 en **Nonseasonal Differencing (d)**.



8. Hacer click en OK. Se genera una nueva hoja de los datos en diferencias con el orden d = 1 (Y_2 - Y_1, Y_3-Y_2, …).

Concentration	Concentration: d = 1, D = 0	
17		
16.6	-0.4	
16.3	-0.3	
16.1	-0.2	
17.1	1	
16.9	-0.2	
16.8	-0.1	
17.4	0.6	
17.1	-0.3	
17	-0.1	
16.7	-0.3	
17.4	0.7	

9. Repetir la Carta de Corrida para los datos en diferencias:

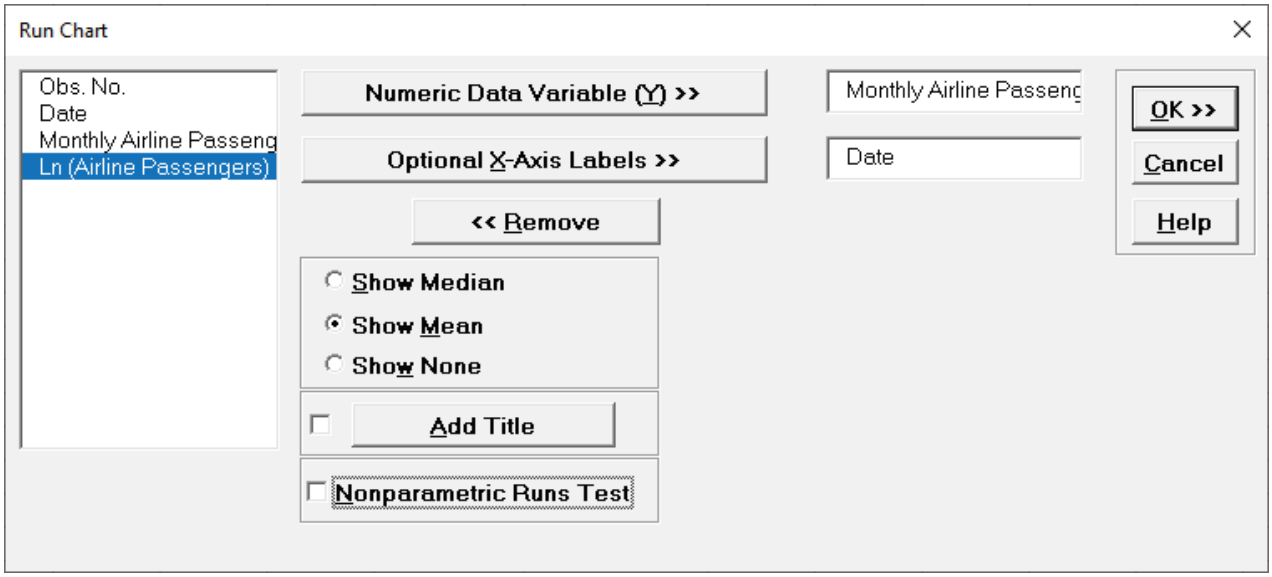


Aparenta ser un proceso estacionario y puede modelarse con ARIMA. Grabar la hoja de ejercicio con los datos en diferencias para los Gráficos de Autocorrelación.

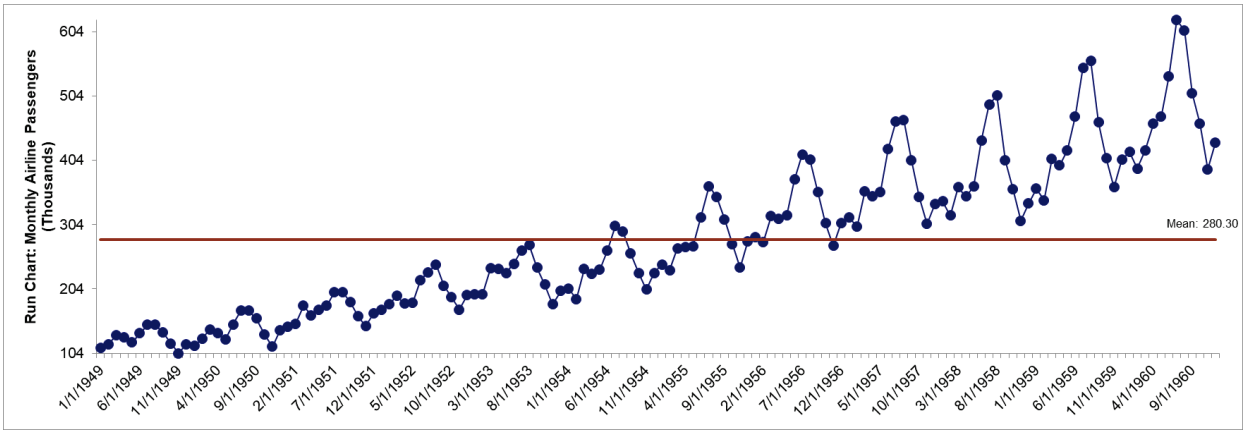
10. Abrir **Monthly Airline Passengers - Series G.xlsx** (pestaña **Hoja 1**). Estos son los datos de una Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional desde enero 1949 hasta diciembre 1960 y es uno de los datos más generalizados que se utiliza en la introducción a los pronósticos de series de tiempo.

11. Hacer click en **SigmaXL > Time Series Forecasting > Run Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marque **Use Entire Data Table**. Hacer click en **Next**.

12. Seleccionar *Monthly Airline Passengers (Thousands)*, hacer click en **Numeric Data Variable (Y) >>**; seleccionar *Date*, hacer click en **Optional X-Axis Labels >>**. Seleccionar **Show Mean**. Desmarcar **Nonparametric Runs Test**.

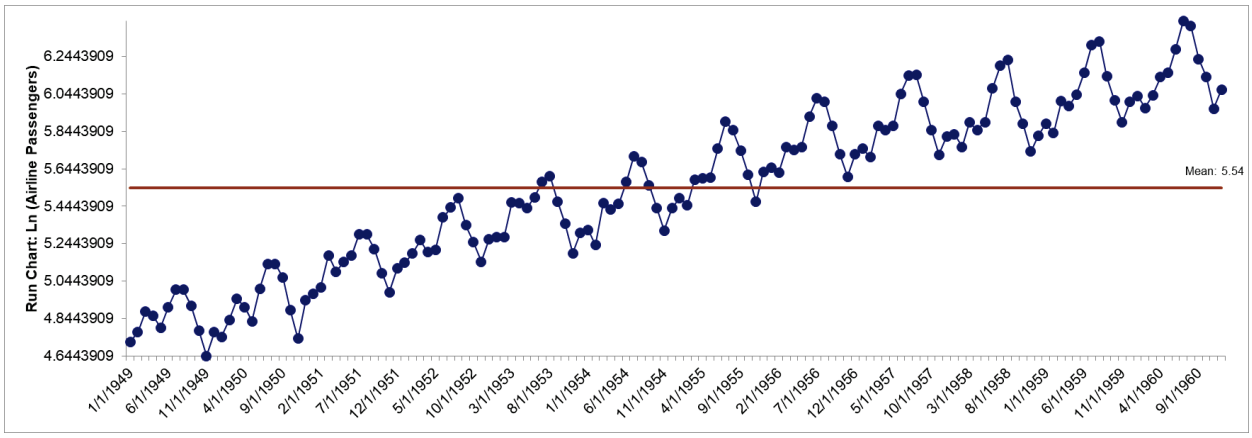


13. Hacer click en **OK**. Se genera una Carta de Corrida de Pasajeros Mensuales de la Aerolínea con una línea central de la Media.



14. Se observa aquí una fuerte tendencia positiva, así como un efecto estacional mensual (ciclo anual de 12 meses), con un aumento en la variación en el tiempo. Esta no estacionariedad en la varianza se puede abordar mediante una transformación Ln de los datos. Más adelante, se demostrará el uso de una transformación Box-Cox para determinar automáticamente la mejor transformación, pero por el momento se emplearán los datos Ln (Pasajeros de la Aerolínea).

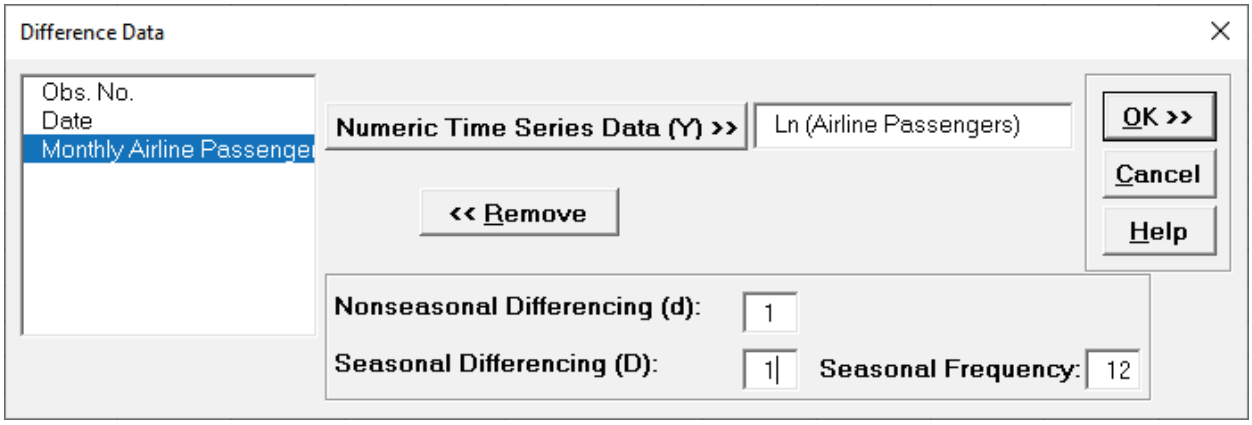
15. Hacer click en la pestaña de la **Hoja 1**. Repetir la Carta de la Corrida para *Ln (Airline Passengers)*:



16. Esto direcciona el cambio en la variación a lo largo del tiempo, pero todavía necesita ser diferenciada para que sea estacionaria para el ARIMA. (Tenga en cuenta que ARIMA Forecast se encargará automáticamente de esto).

17. Hacer click en la pestaña de la **Hoja 1**. Hacer click en **SigmaXL > Time Series Forecasting > Utilities > Difference Data.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marque **Use Entire Data Table**. Hacer click en **Next**.

18. Seleccionar *Ln (Airline Passengers)*, hacer click en **Numeric Time Series Data (Y) >>**. Registrar 1 en **Nonseasonal Differencing (d)**; registrar 1 en **Seasonal Differencing (D)**; especificar 12 en **Seasonal Frequency.**

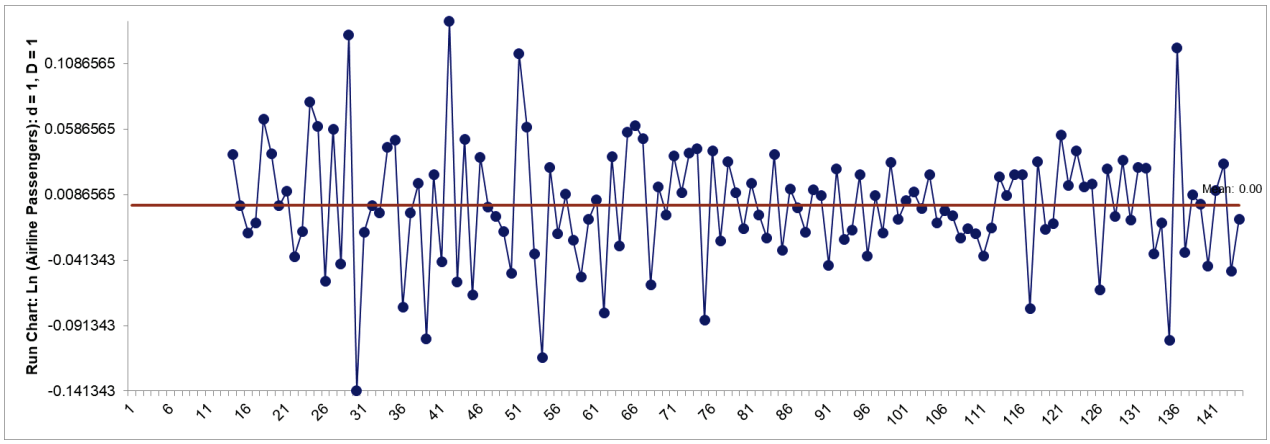


19. Hacer click en **OK**. Se crea una nueva hoja con los datos en diferencias de orden estacional D = 1 (Y_{13} - Y_1, Y_{14}-Y_2, ...) orden no-estacional d=1 ($D_2 – D_1$, $D_3 – D_2$, ...).

Ln (Airline Passengers)	Ln (Airline Passengers): d = 1, D = 1
4.718498871	
4.770684624	
4.882801923	
4.859812404	
4.795790546	
4.905274778	
4.997212274	
4.997212274	
4.912654886	
4.779123493	
4.644390899	
4.770684624	
4.744932128	
4.836281907	0.039164025
4.94875989	0.000360685
4.905274778	-0.020495594
4.828313737	-0.012939182

Nota: Para ver un ejemplo manual de diferencias estacionales y no-estacionales con Excel https://faculty.fuqua.duke.edu/~rnau/Decision411_2007/Class10notes.htm.

20. Repetir la Carta de Corrida para los datos en diferencias:



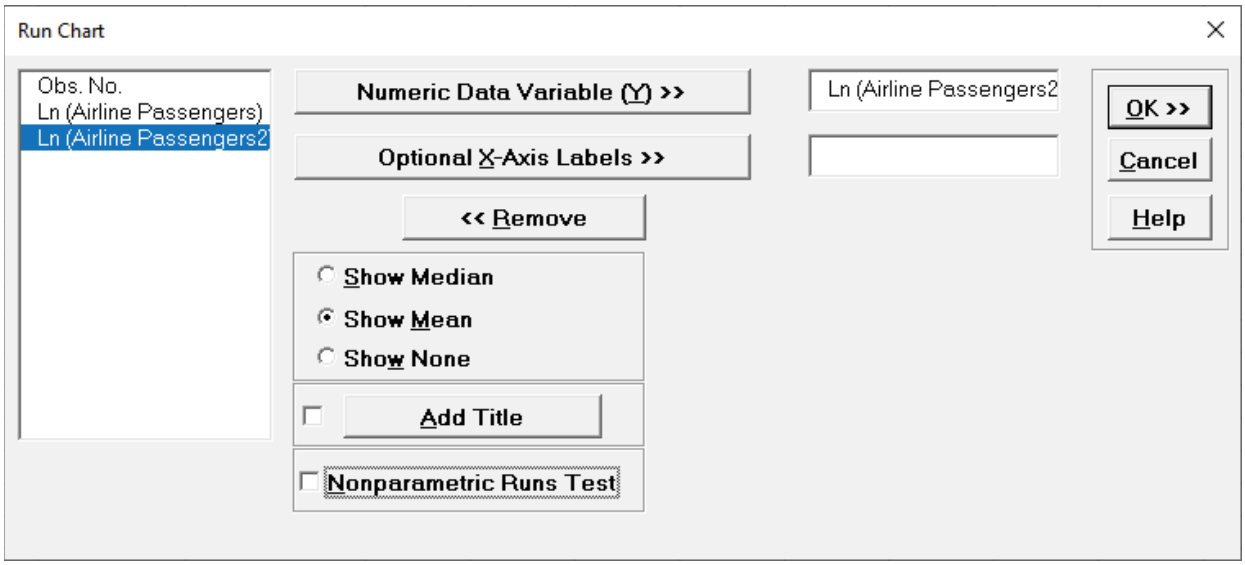
Esto parece ser ahora un proceso estacionario y puede modelarse con ARIMA. Grabar el libro de ejercicios con los datos en diferencias para que se usen con los gráficos de autocorrelación.

Pasajeros Mensuales de la Aerolínea – Valores Ausentes

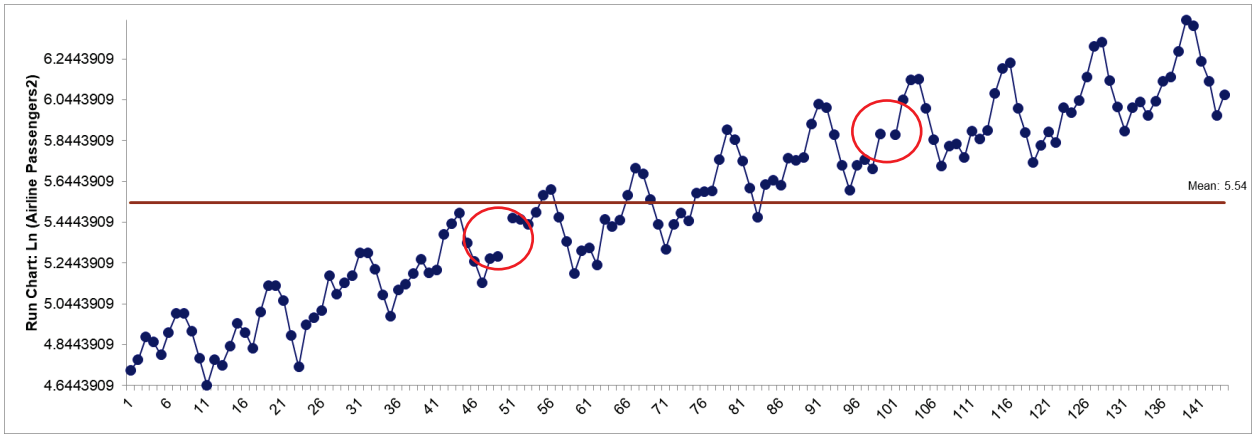
21. Abrir **Monthly Airline Passengers – Missing Values.xlsx** (la pestaña de la **Hoja 1)**. Ln (Airline Passengers2) tiene valores ausentes en las observaciones 50 y 100.

22. Hacer click en **SigmaXL > Time Series Forecasting > Run Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marque **Use Entire Data Table**. Hacer click en **Next**.

23. Seleccionar *Ln (Airline Passengers2)*, hacer click en **Numeric Data Variable (Y) >>**. Seleccionar **Show Mean**. Desmarcar **Nonparametric Runs Test**.



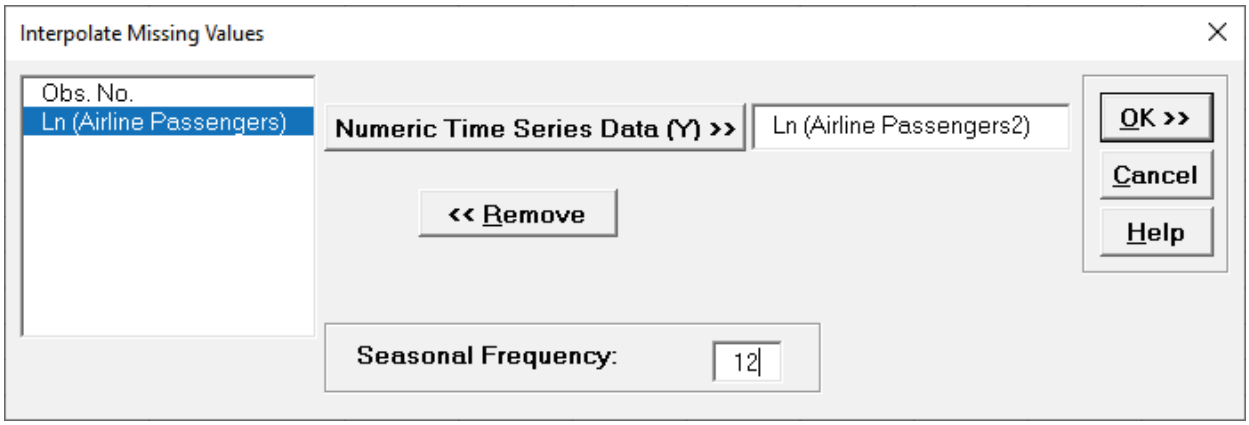
24. Hacer click en **OK**. Se genera una Carta de Corrida de Ln (Airline Passengers2) con la línea central de la Media.



25. Los valores ausentes aparecen en forma de una línea discontinua en la carta de corrida.

26. Hacer click en la pestaña de la **Hoja 1**. Hacer click en **SigmaXL > Time Series Forecasting > Utilities > Interpolate Missing Values.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

27. Seleccionar *Ln (Airline Passengers2)*, hacer click en **Numeric Time Series Data (Y) >>**. **Se especifica 12 como Frecuencia Estacional.**



28. Hacer click en **OK**. Se genera una nueva hoja con los valores interpolados linealmente ajustados estacionalmente y resaltados en amarillo.

Ln (Airline Passengers2)	Y: Interpolated (12)
4.718498871	4.718498871
4.770684624	4.770684624
4.882801923	4.882801923
4.859812404	4.859812404
	5.254047791
	5.846836235

Se estima Y50 en 5.254 (versus el valor original de 5.278). Se estima Y100 en 5.847 (versus el valor original de 5.852).

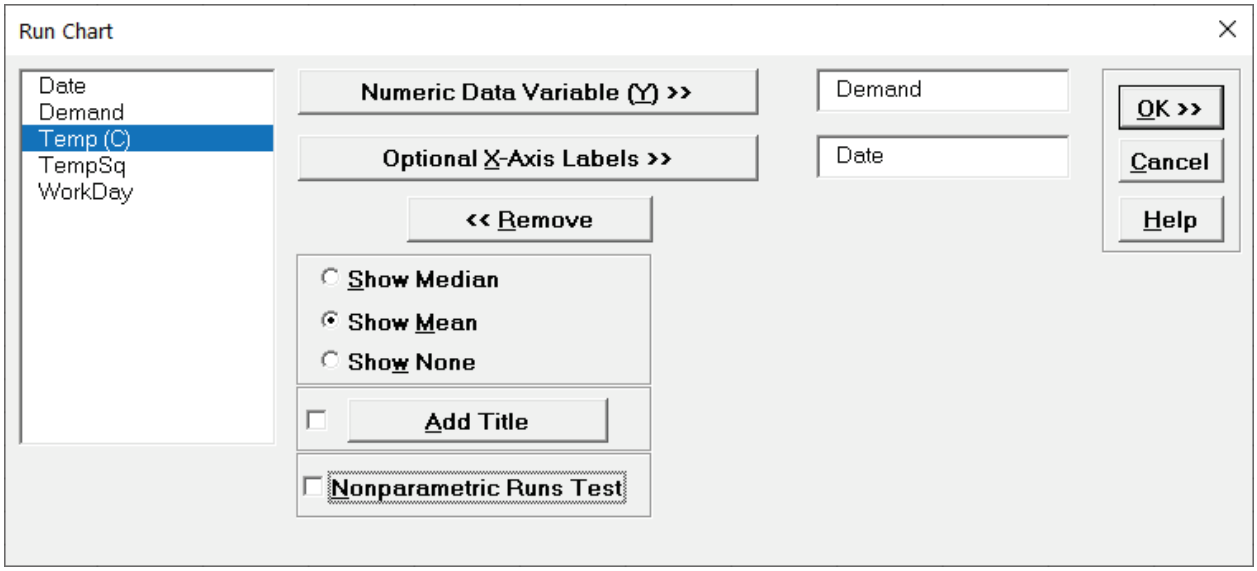
Observar que, si bien existe robustez con respecto a algunos valores ausentes, si el número de valores ausentes es grande, entonces la estimación del modelo y la precisión del pronóstico se degradarán.

Demanda Diaria de Electricidad con Predictoras – ElecDaily

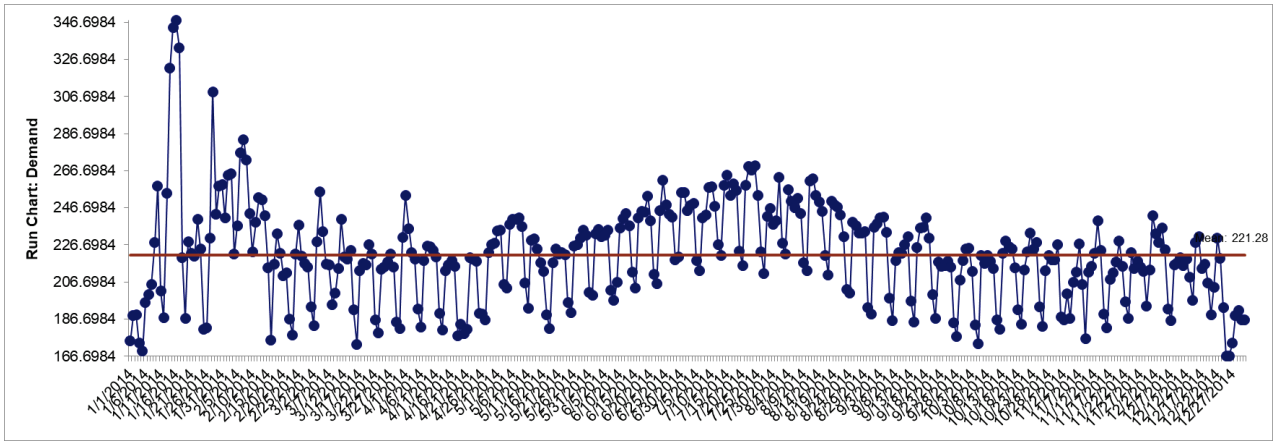
29. Abrir **Daily Electricity Demand with Predictors – ElecDaily.xlsx** (pestaña de la **Hoja 1**). Esta es la demanda diaria de electricidad (GW) para el estado de Victoria, Australia, todos los días durante 2014 (Hyndman, fpp2). La temperatura (C) es la máxima temperatura diaria en grados centígrados para la ciudad de Melbourne. TempSq es la temperatura al cuadrado. WorkDay toma el valor 1 en los días laborables y 0 en los demás. Estos datos tienen una frecuencia estacional = 7 (observaciones por semana).

30. Hacer click en **SigmaXL > Time Series Forecasting > Run Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

31. Seleccionar *Demand*, hacer click en **Numeric Data Variable (Y) >>**; seleccionar *Date*, hacer click en **Optional X-Axis Labels >>**. Seleccionar **Show Mean**. Desmarcar **Nonparametric Runs Test**.

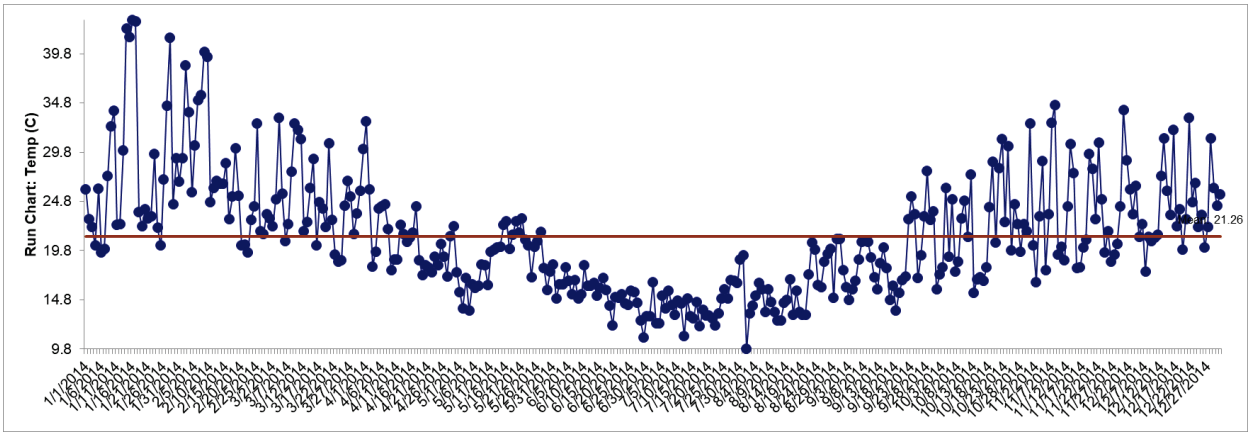


32. Hacer click en **OK**. Se genera una Carta de Corrida de la Demanda de Electricidad con una línea central de la Media.



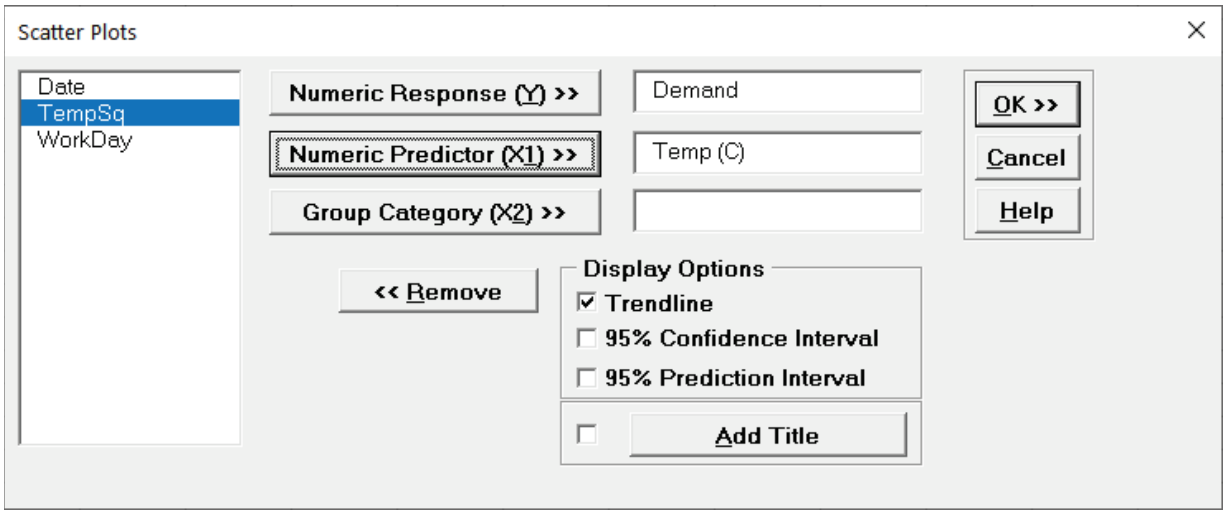
33. Se aprecia la estacionalidad diaria con un ciclo de 7 días, pero también hay un efecto de verano e invierno, con el pico de la demanda a mediados de enero, que es verano en Australia. Luego, se utilizará un diagrama de densidad espectral para confirmar el valor de la frecuencia estacional de 7.

34. Hacer click en la pestaña de la **Hoja 1**. Repetir la Carta de Corrida para Temp (C):

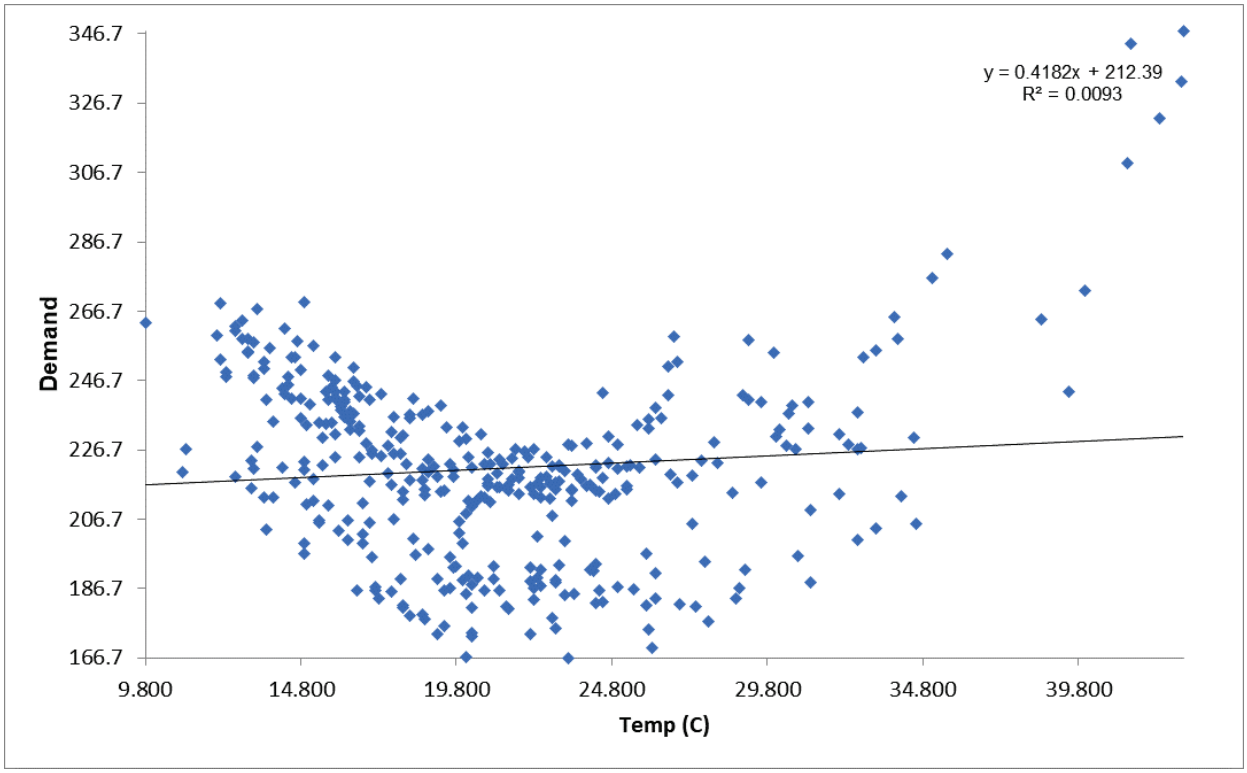


35. Ahora se explorará la relación entre la Demanda y la Temperatura. Hacer click en la pestaña de la **Hoja 1**. Hacer click en **SigmaXL > Graphical Tools > Scatterplots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

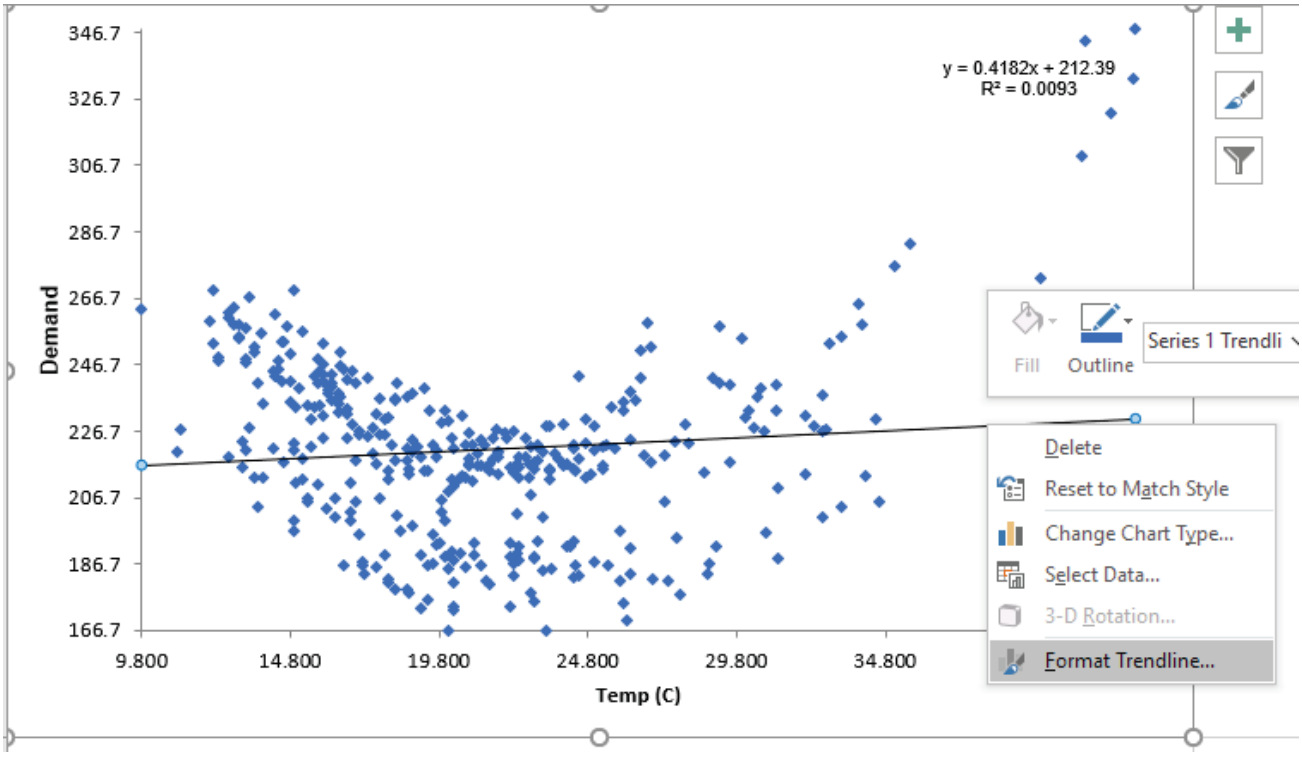
36. Seleccionar *Demand*, hacer click en **Numeric Response (Y) >>**; seleccionar *Temp (C)*, hacer click en **Numeric Predictor (X1) >>**. Marcar **Trendline**.



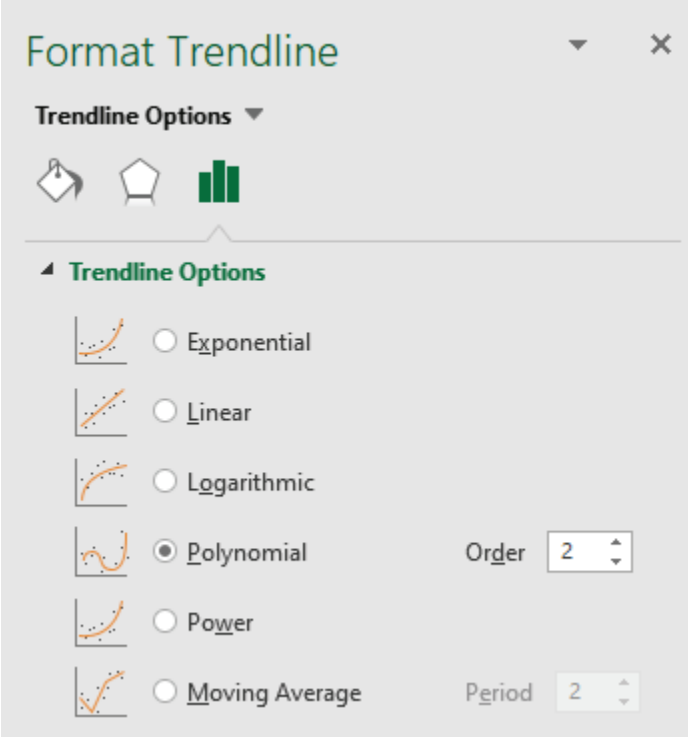
37. Hacer click en **OK**. Se genera un diagrama de dispersión de la Demanda de Electricidad versus la Temperatura.



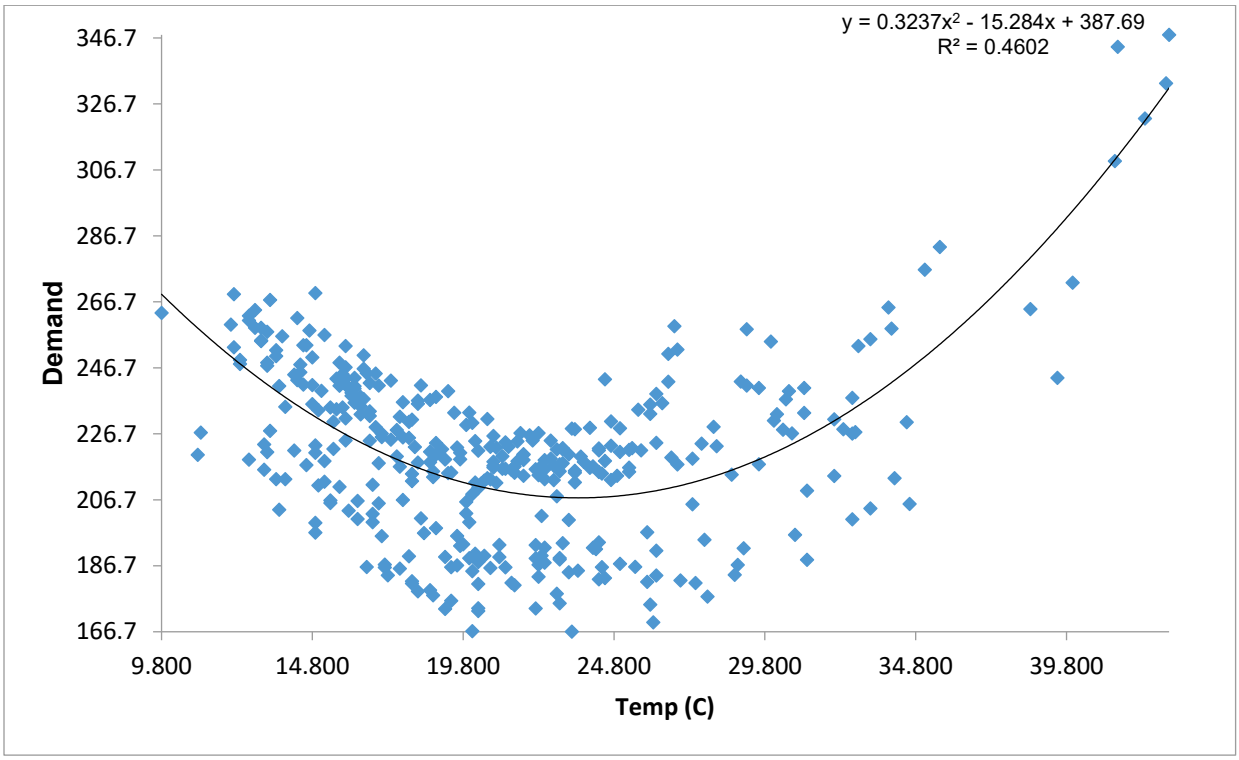
38. Esto muestra una relación cuadrática: las altas temperaturas en el verano causan la demanda de electricidad para el acondicionador de aire, las bajas temperaturas en el invierno causan la demanda de calefacción. Se modifica la línea de tendencia en Excel para un ajuste cuadrático. Hacer click en Trendline, hacer click derecho y seleccionar Format Trendline como se muestra:



Se proporcionan las opciones de **Format Trendline**. Seleccionar **Polynomial** con **Order** 2 como se muestra.



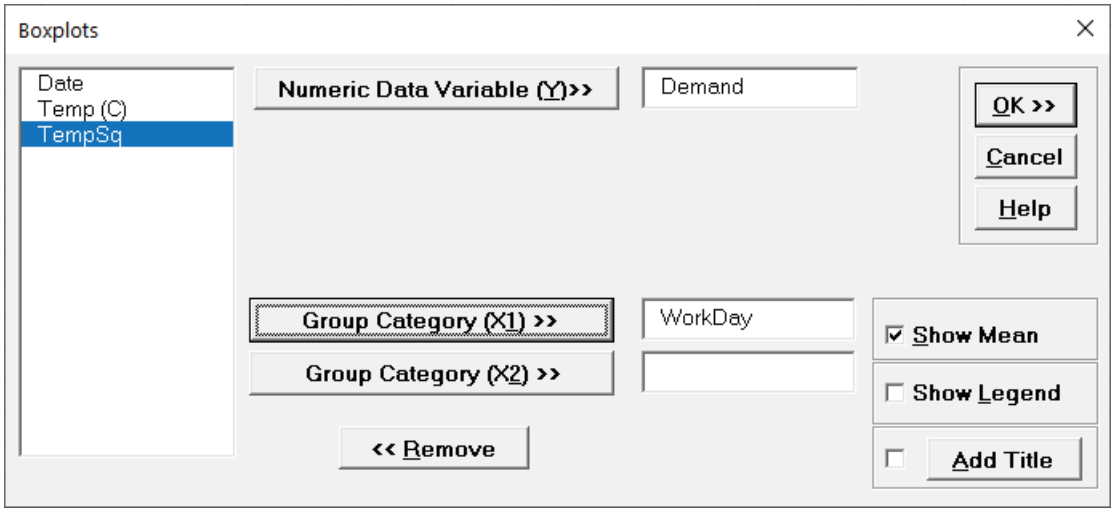
39. La Línea de Tendencia es ahora una función cuadrática como se muestra:



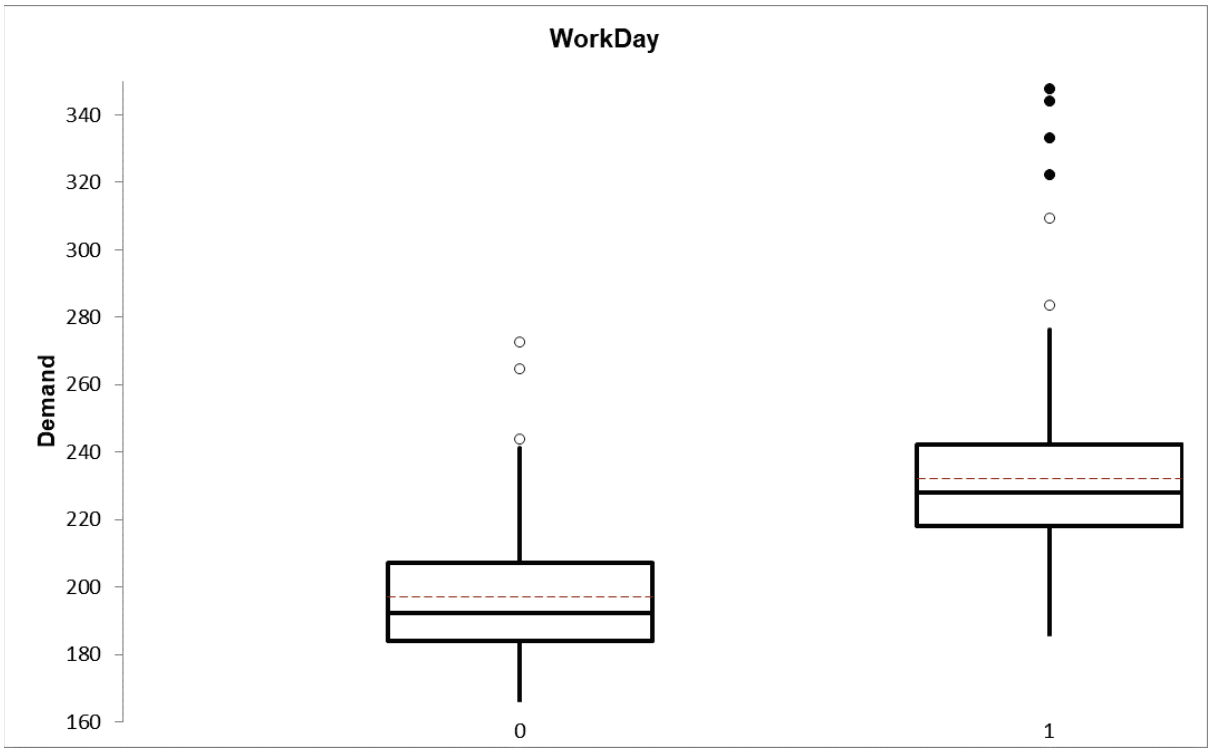
40. Se ha agregado TempSq (Temperature Squared) a los datos de tal manera que se incluya cuando se pronostique ARIMA con los predictoras.

41. Se explorará ahora la relación entre la Demanda y el día Laborable. Hacer click en la pestaña de la **Hoja 1**. Hacer click en **SigmaXL > Graphical Tools > Boxplots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

42. Seleccionar *Demand*, hacer click en **Numeric Data Variable (Y) >>**; seleccionar *WorkDay*, hacer click en **Group Category (X1) >>**. Marcar **Show Mean**.



43. Hacer click en **OK**. Se genera un diagrama de caja de la Demanda de Electricidad versus el Día Laborable.



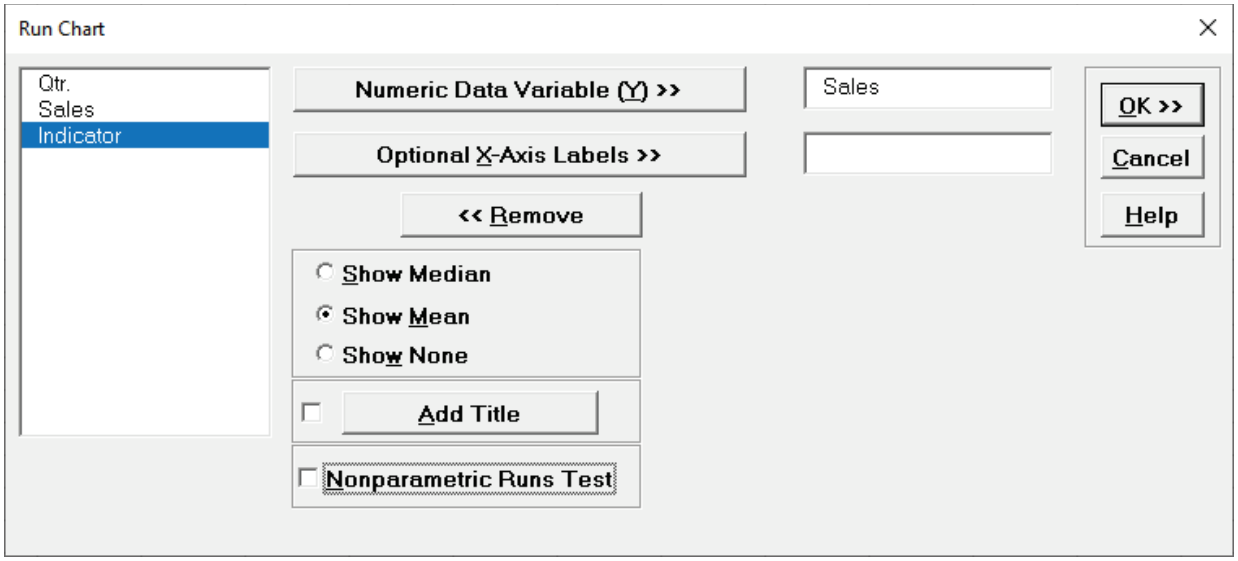
Como era de esperarse, un día laborable (1) tiene una mayor demanda de electricidad que un día no laborable (0). Después, se incluirán Temp (C), TempSq y WorkDay como predictoras en el modelo de pronóstico ARIMA.

Ventas con Indicador – Serie M Modificada

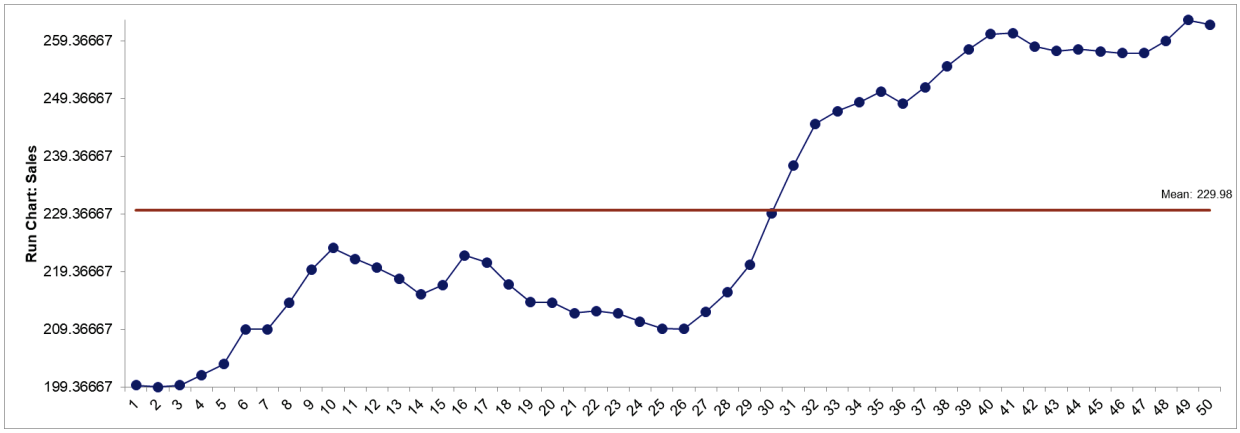
44. Abrir **Sales with Indicator - Modified Series M.xlsx**. (pestaña de la **Hoja 1**). Estos son datos modificados de la Serie M de Box y Jenkins. Originalmente, un conjunto de 150 valores de ventas mensuales corporativas junto con un indicador principal. Los datos se modificaron convirtiéndolos en valores trimestrales al promediárselos cada tres meses, es decir, 50 trimestres. Se realizó esto para simplificar el análisis del indicador principal. Aunque los datos eran mensuales y se resumieron en trimestrales, se los considerará como no estacionales, como lo hicieron Box y Jenkins.

45. Hacer click en **SigmaXL > Time Series Forecasting > Run Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

46. Seleccionar *Sales*, hacer click en **Numeric Data Variable (Y) >>**. Seleccionar **Show Mean**. Desmarcar **Nonparametric Runs Test**.

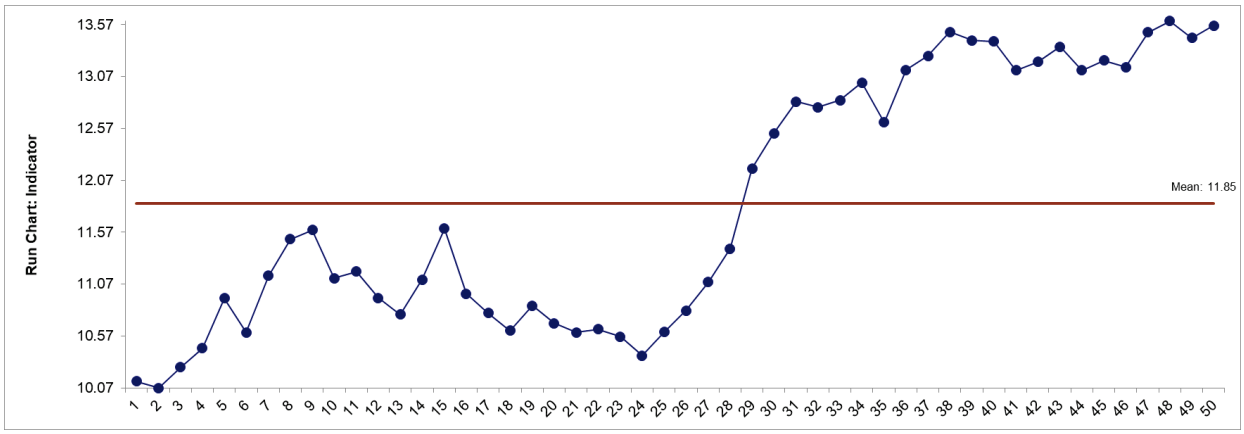


47. Hacer click en **OK**. Se genera una Carta de Corrida de Ventas con una línea central de la Media.



Como se mencionó arriba, no existe una estacionalidad aparente en los datos. Se empleará después un Diagrama de Densidad Espectral para confirmar el valor de la frecuencia estacional de 1.

48. Hacer click en la pestaña de la **Hoja 1**. Repetir la Carta de Corrida para el Indicador:



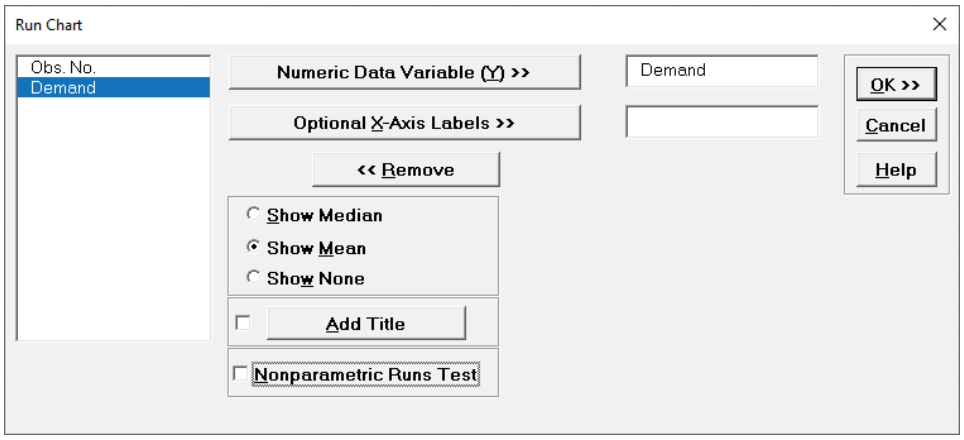
Evidentemente, existe una cercana relación entre las Ventas y el Indicador. Se analizará luego la relación entre la gráfica de la Función de Correlación Cruzada y el modelo de pronóstico ARIMA que incluya al Indicador como predictor.

Demanda de Electricidad Estacional Múltiple de Media Hora – Taylor

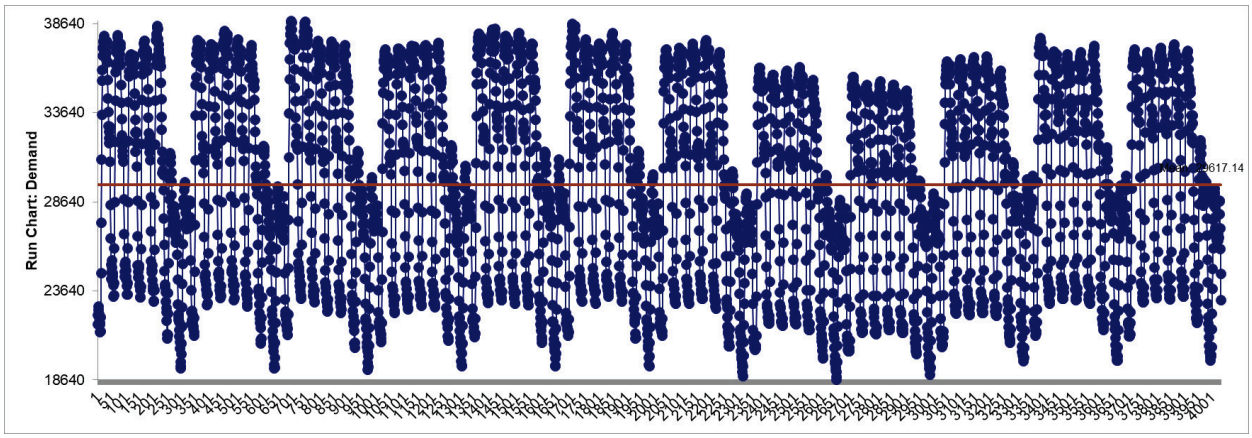
49. Abrir **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (pestaña de la **Hoja 1**). Esta es la demanda de electricidad de media hora (MW) en Inglaterra y Gales desde el lunes 5 de junio de 2000 hasta el domingo 27 de agosto de 2000 (Taylor, R forecast). Estos datos tienen una estacionalidad múltiple con una frecuencia de 48 (observaciones por día) y 336 (observaciones por semana), con un total de 4032 observaciones. No se dispone de predicciones como la temperatura y el día laborable para estos datos. Referencia: Taylor, J.W. (2003) Short-term electricity demand forecasting using double seasonal suavización exponencial. *Journal of the Operational Research Society*, **54**, 799-805.

50. Hacer click en **SigmaXL > Time Series Forecasting > Run Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

51. Seleccionar *Demand*, hacer click en **Numeric Data Variable (Y) >>**. Seleccionar **Show Mean**. Desmarcar **Nonparametric Runs Test**.



52. Hacer click en **OK**. Se genera una Carta de Corrida de la Demanda de Electricidad con una línea centrada en la Media.



Aquí se aprecia el fuerte efecto estacional de media hora con un ciclo diario de 48 observaciones y un ciclo semanal de 48*7 = 336 observaciones. Dado que estos datos tienen múltiples frecuencias (y frecuencias > 24), se requerirá una descomposición estacional múltiple (DEM) para modelizarlo. Más adelante, utilizaremos un Diagrama de Densidad Espectral para confirmar los valores de frecuencia estacional múltiple de 48 y 336.

Gráficos de Autocorrelación (FAC/FACP)

Así como la correlación mide el grado de relación lineal entre dos variables, la autocorrelación mide la relación lineal entre los valores rezagados de los datos. Un gráfico de los datos versus los mismos datos con rezago k puede mostrar una tendencia positiva o negativa. Si la pendiente es positiva, la autocorrelación es positiva; si hay una pendiente negativa, la autocorrelación es negativa. Se empleará el utilitario de rezagos y los gráficos de dispersión para demostrar esto con 3 rezagos usando los datos de la Serie A.

FAC denota la gráfica de la Función de Autocorrelación (llamada a veces Correlograma). FACP denota la Función de Autocorrelación Parcial. Para mayores detalles y fórmulas, ver en el Apéndice: **Autocorrelación (FAC), Autocorrelación Parcial (FACP) y Correlación Cruzada (FCC).**

Se usarán la FAC y la FACP para tener una idea general de qué modelos se deben emplear, pero se permitirá que sean los algoritmos automáticos los que realicen el trabajo pesado para determinar qué modelo es el óptimo. Luego, se utiliza la FAC en los residuos del modelo para ayudar a determinar lo bien que un modelo se ajusta a la estructura de autocorrelación, idealmente con los residuos que no muestren ninguna autocorrelación estadísticamente significativa.

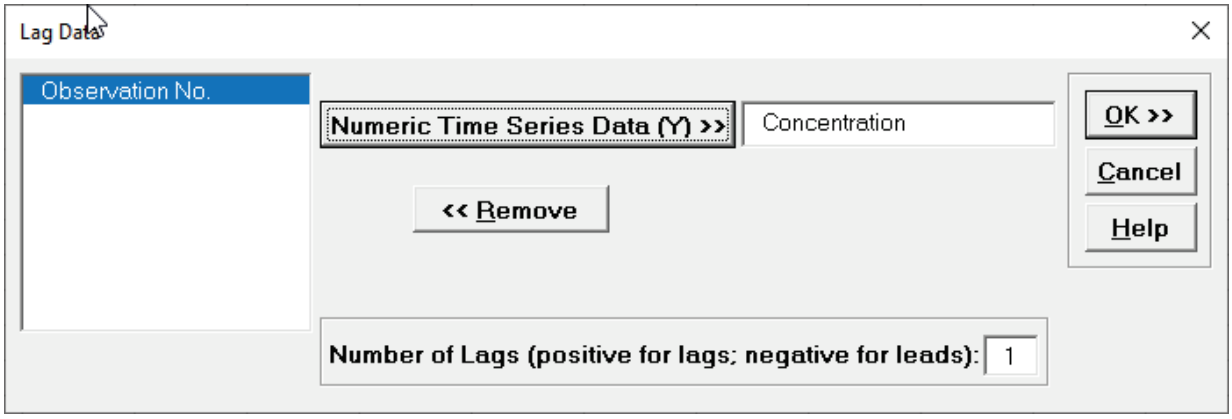
Se examinará también cómo resulta la estacionalidad con la diferenciación en los gráficos de la FAC y la FACP.

Concentración de Proceso Químico - Serie A

1. Abrir **Chemical Process Concentration – Series A.xlsx** (pestaña de la **Hoja 1**). Esta es una Serie A de datos de Box y Jenkins, un conjunto de datos de 197 valores de concentración de un proceso químico que toma intervalos de dos horas.

2. Hacer click en **SigmaXL > Time Series Forecasting > Utilities > Lag Data**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Concentration*, hacer click en **Numeric Time Series Data (Y) >>**. Use **Number of Lags** = 1 por defecto.



4. Hacer click en **OK**. Se genera una nueva hoja con los datos del Rezago 1.

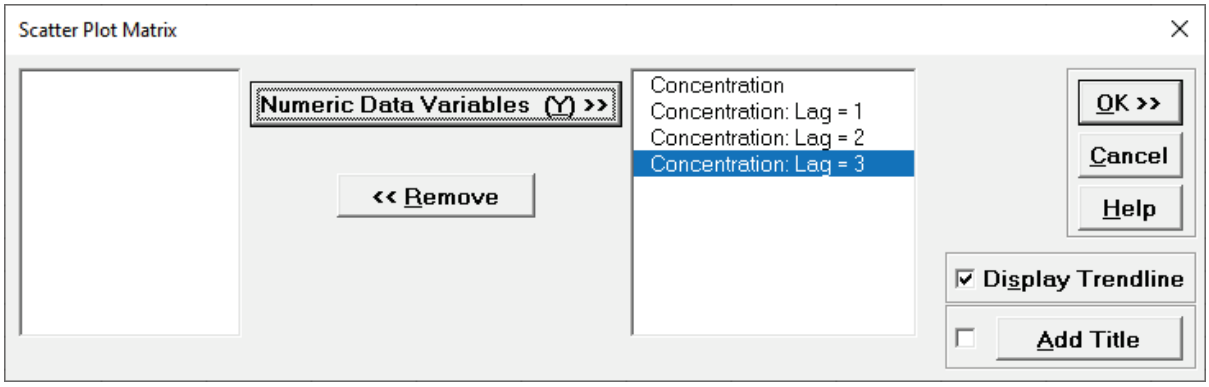
Concentration	Concentration: Lag = 1
17	
16.6	17
16.3	16.6
16.1	16.3
17.1	16.1
16.9	17.1

5. Hacer click en el menú **Recall SigmaXL Dialog** o presione **F3** para repetir el último diálogo. Registrar **Lag** = 2. Hacer click en **OK**. Repetir **Lag** = 3. Las columnas de los rezagos se agregan como sigue:

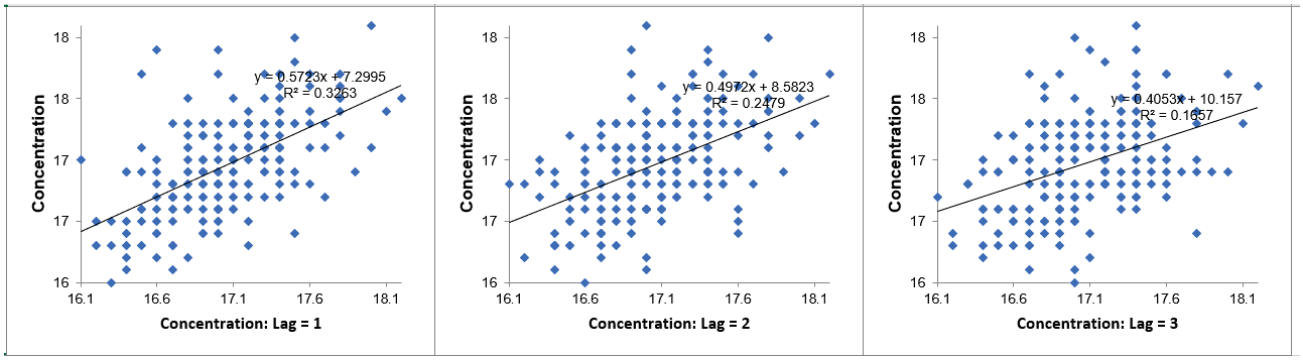
Concentration	Concentration: Lag = 1	Concentration: Lag = 2	Concentration: Lag = 3
17			
16.6	17		
16.3	16.6	17	
16.1	16.3	16.6	17
17.1	16.1	16.3	16.6
16.9	17.1	16.1	16.3
16.8	16.9	17.1	16.1

6. Hacer click en **SigmaXL > Graphical Tools > Scatter Plot Matrix**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

7. Seleccionar todas las cuatro columnas, hacer click en **Numeric Data Variables (Y) >>**. Marcar **Display Trendline**.

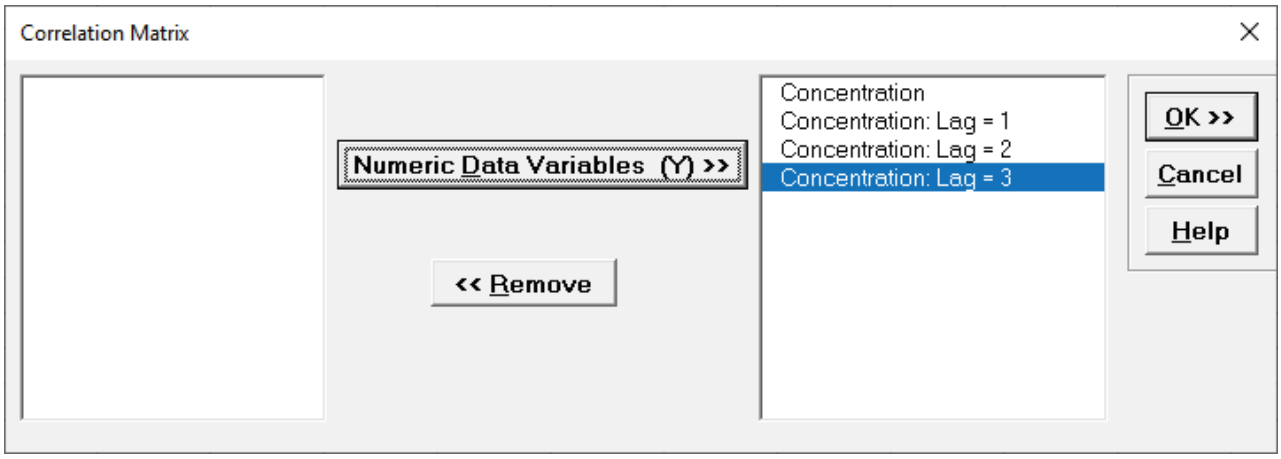


8. Hacer click en **OK**. Se generan los diagramas de dispersión para los Rezagos del 1 al 3, mostrando una tendencia positiva, es decir, una autocorrelación positiva.



9. Hacer click en la pestaña de la hoja **Lag Data**. Hacer click en **SigmaXL > Statistical Tools > Correlation Matrix**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar Use Entire Data Table. Hacer click en Next.

10. Seleccionar todas las cuatro columnas, hacer click en **Numeric Data Variables (Y) >>**.



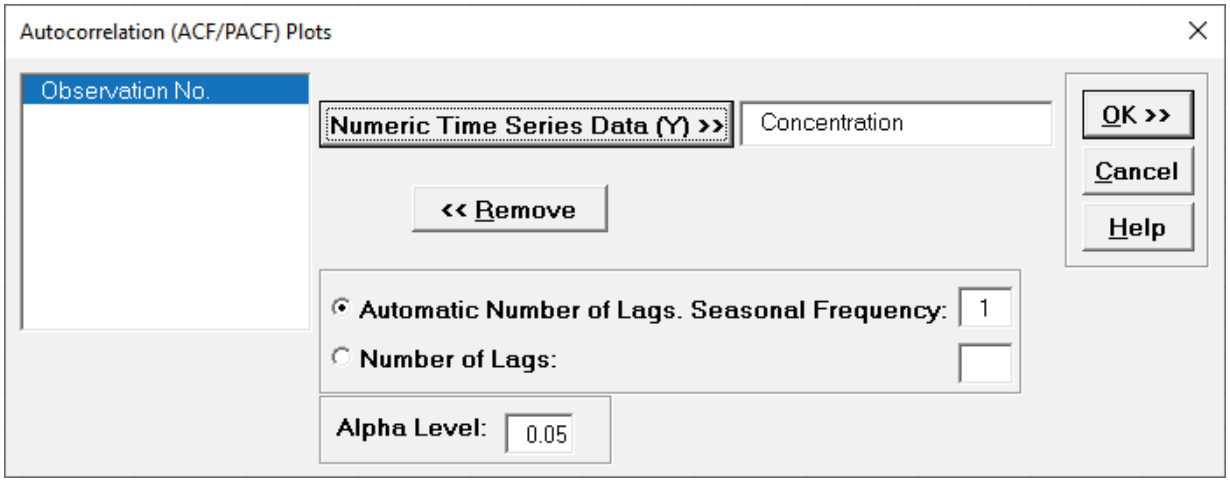
11. Hacer click en **OK**. Se genera la Matriz de Coeficientes de Correlación.

Pearson Correlations	Concentration	Concentration: Lag = 1	Concentration: Lag = 2	Concentration: Lag = 3
Concentration	1	0.5713	0.4979	0.4070
Concentration: Lag = 1		1	0.5710	0.4952
Concentration: Lag = 2			1	0.5721
Concentration: Lag = 3				1

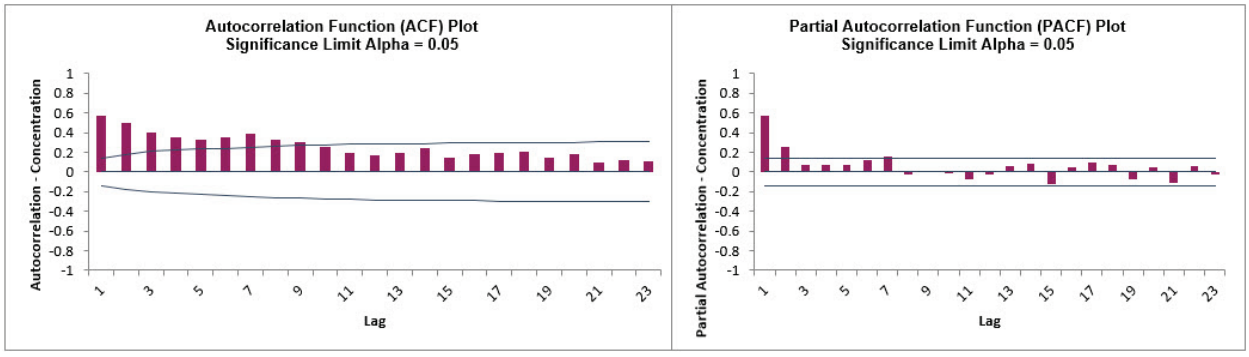
La correlación del Rezago 1 es 0.57, del Rezago 2 es 0.5, y del Rezago 3 es 0.41. Las correlaciones en rojo (y los valores-p en negritas) indican que las correlaciones son significativas.

12. Se generará ahora los gráficos FAC/FACP. Hacer click en la pestaña de la **Hoja 1**. Hacer click en **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

13. Seleccionar *Concentration*, hacer click en **Numeric Time Series Data (Y) >>**. Usar **Automatic Number of Lags. Seasonal Frequency** = 1. **Alpha Level** = 0.05, por defecto.



14. Hacer click en **OK**. Se generan los gráficos FAC y FACP.

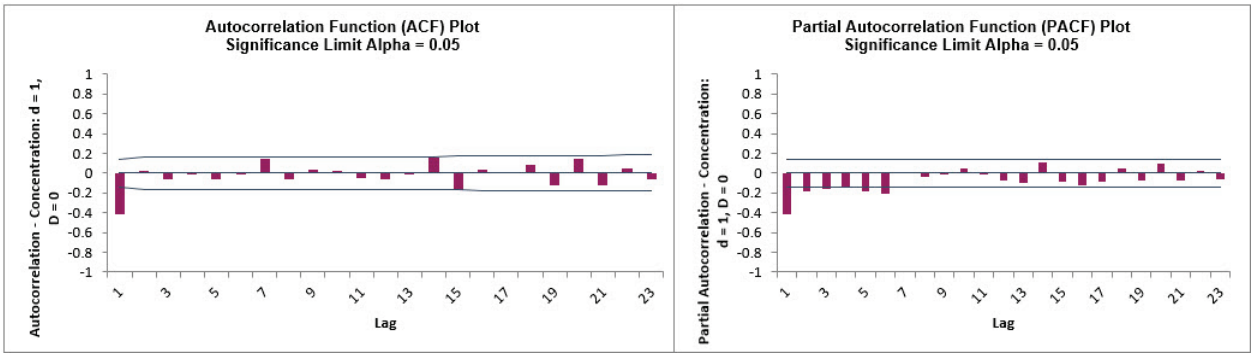


Mover el cursor del ratón alrededor de la barra del Rezago 1 de la FAC para ver el valor de la correlación = 0.57, Rezago 2 = 0.5, y Rezago 3 = 0.4. Los valores son aproximadamente los mismos que los obtenidos manualmente (con algunas pequeñas diferencias debido a la forma en que se calculan las correlaciones).

Los rezagos que superan las líneas de los límites de significancia son estadísticamente significativos con un alfa = 0.05, por lo que hay una estructura de autocorrelación en los datos que se puede modelar utilizando la Suavización Exponencial o ARIMA.

15. La lenta disminución de la FAC a medida que aumentan los rezagos se debe a la no estacionalidad, por lo que ahora se comparará con los datos en diferencias. Seleccionar la hoja **Difference Data** (o reproducir con **SigmaXL > Time Series Forecasting > Utilities > Difference Data**, con **Nonseasonal Differencing (d)** = 1).

16. Repetir los gráficos FAC/FACP para los datos en diferencias:



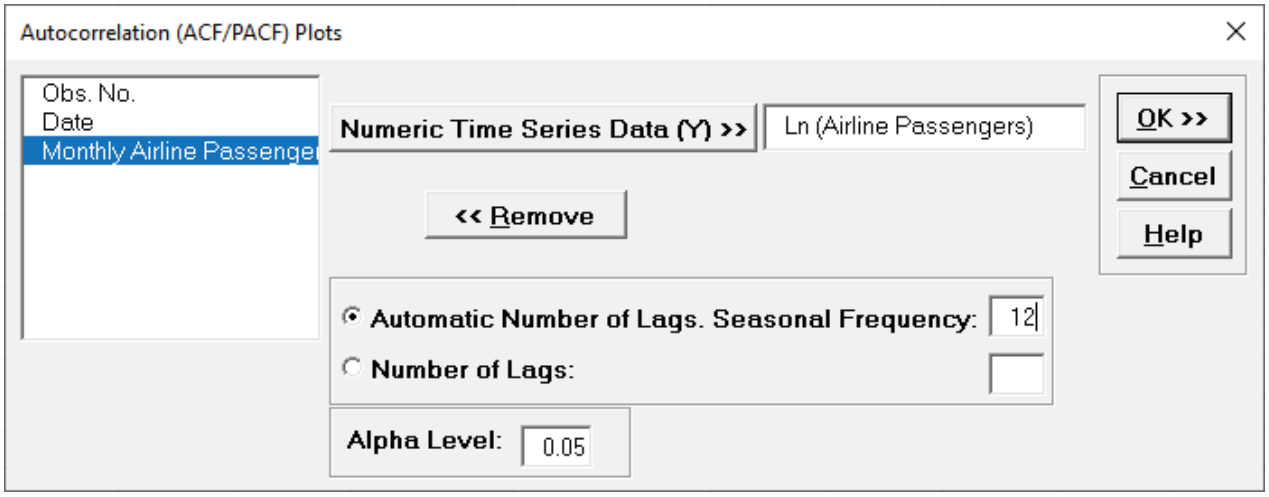
Después de aplicar la diferenciación en el Rezago 1 ahora muestra una autocorrelación negativa, pero el Rezago 2 y los siguientes tienen autocorrelaciones insignificantes y pequeñas autocorrelaciones parciales. Esto concuerda con los resultados de Box y Jenkins (2016, págs. 185-186), quienes también sugieren que, tras la diferenciación, el modelo podría ser un promedio móvil de orden 1. Se revisará esto más adelante en el Pronóstico ARIMA.

Pasajeros Mensuales de la Aerolínea - Serie G

17. Abrir **Monthly Airline Passengers - Series G.xlsx** (pestaña de la **Hoja 1**). Estos son datos de una Serie G de Box y Jenkins, total mensual de los pasajeros de una aerolínea internacional desde enero de 1949 a diciembre de 1960.

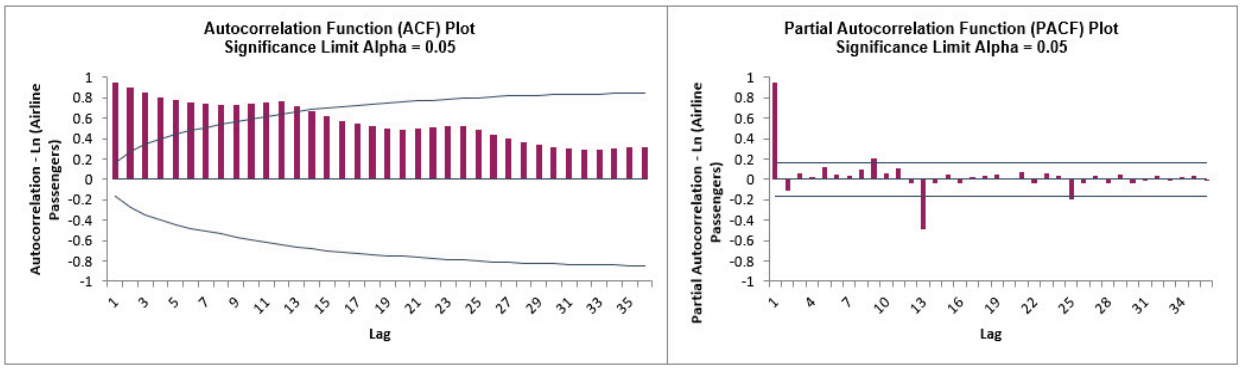
18. Hacer click en **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

19. Seleccionar *Ln (Airline Passengers)*, hacer click en **Numeric Time Series Data (Y) >>**. Seleccionar **Automatic Number of Lags.** Especificar **Seasonal Frequency** = 12 y **Alpha Level** = 0.05.



El número automático de rezagos será (como mínimo) la Frecuencia Estacional * 3 = 36 lo que permite observar los patrones estacionales en la FAC.

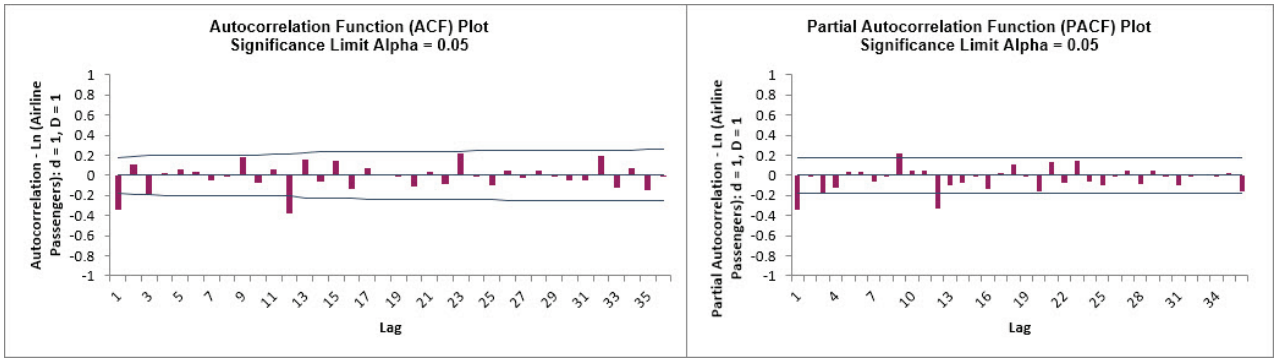
20. Hacer click en **OK**. Se generan los gráficos FAC y FACP.



La lenta disminución de la FAC a medida que aumentan los rezagos se debe a la tendencia, mientras que la forma en “onda" se debe a la estacionalidad.

21. Se comparará ahora con los datos en diferencias para Ln (Airline Passengers). Seleccionar la hoja de **Difference Data** (o reproducir con **SigmaXL > Time Series Forecasting > Utilities > Difference Data**, con **Nonseasonal Differencing (d)** = 1 y **Seasonal Differencing (D)** = 1 con **Seasonal Frequency** = 12).

22. Repetir los gráficos FAC/FACP para los datos en diferencias:



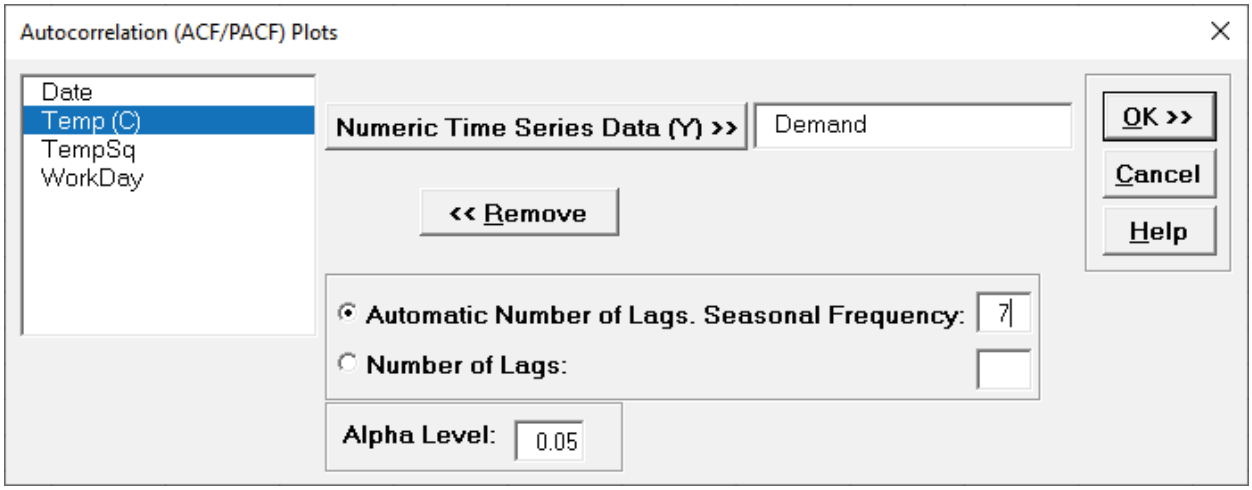
Después de aplicar la diferenciación en el Rezago 1 ahora muestra una autocorrelación negativa, pero el Rezago 2 y los siguientes tienen autocorrelaciones insignificantes y pequeñas autocorrelaciones parciales. El Rezago 12 se debe a la estacionalidad. Esto concuerda con los resultados de Box y Jenkins (2016, Capítulo 9, "Análisis de Series de Tiempo Estacionales", pág. 319), quienes también sugieren que, tras la diferenciación, el modelo podría ser un promedio móvil de orden 1 y promedio móvil estacional de orden 1. Se revisará esto más adelante en el Pronóstico ARIMA.

Demanda Diaria de Electricidad con Predictoras – ElecDaily

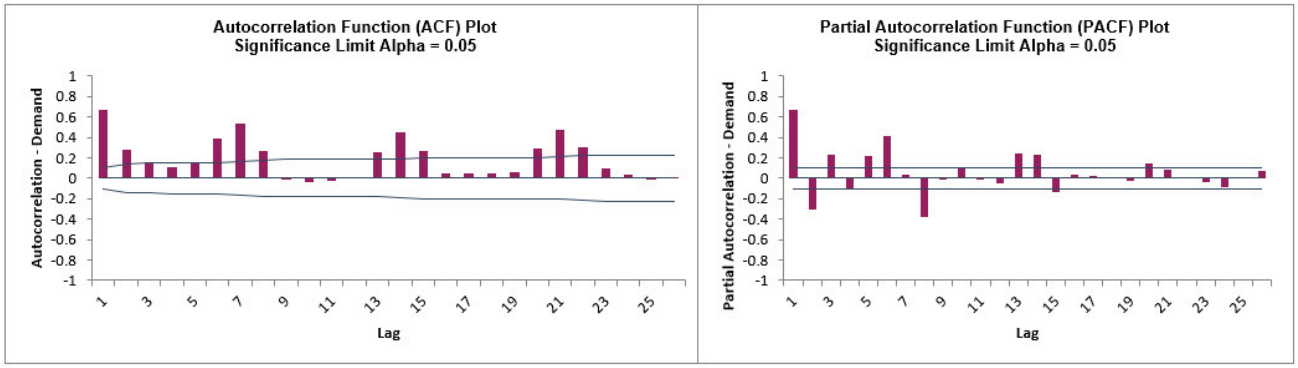
23. Abrir **Daily Electricity Demand with Predictors – ElecDaily.xlsx** (pestaña de la **Hoja 1**). Esto es la demanda diaria de electricidad (GW) para el estado de Victoria, Australia, a diario desde 2014 (Hyndman, fpp2). Estos datos tienen una frecuencia estacional = 7 (observaciones por semana).

24. Hacer click en **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

25. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**. Seleccionar **Automatic Number of Lags.** Especificar **Seasonal Frequency** = 7 y **Alpha Level** = 0.05.



26. Hacer click en **OK**. Se generan los gráficos FAC y FACP.



Se puede observar aquí un patrón estacional de 7 días. No se realizará diferenciación manual para este ejemplo, sin embargo, el algoritmo automático de ARIMA realizará una diferenciación no estacional (d) = 1 y dejará la diferenciación estacional como (D) = 0.

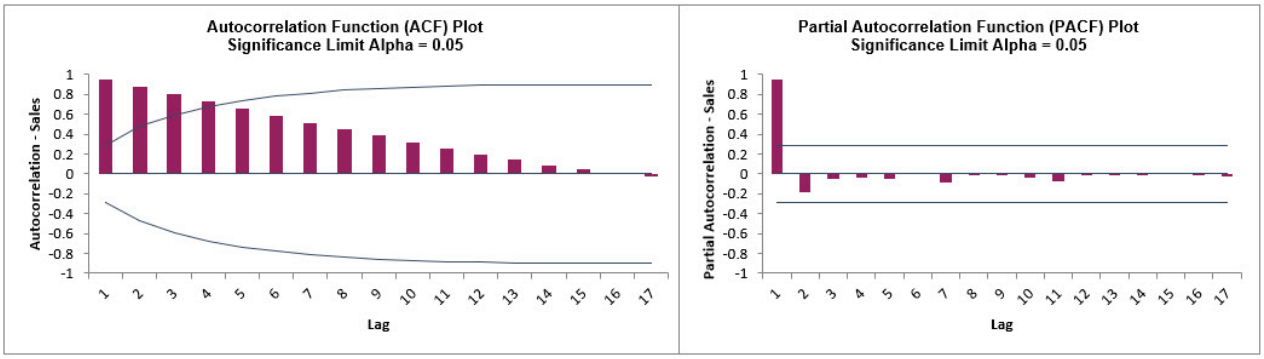
Ventas con Indicador - Serie M Modificada

27. Abrir **Sales with Indicator - Modified Series M.xlsx**. (pestaña de la **Hoja 1**). Esta es una Serie M de datos modificados de Box y Jenkins, con los valores de 50 ventas trimestrales corporativas con un indicador principal. Se consideran los datos como no estacionales, tal como lo hicieron Box y Jenkins.

28. Hacer click en **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

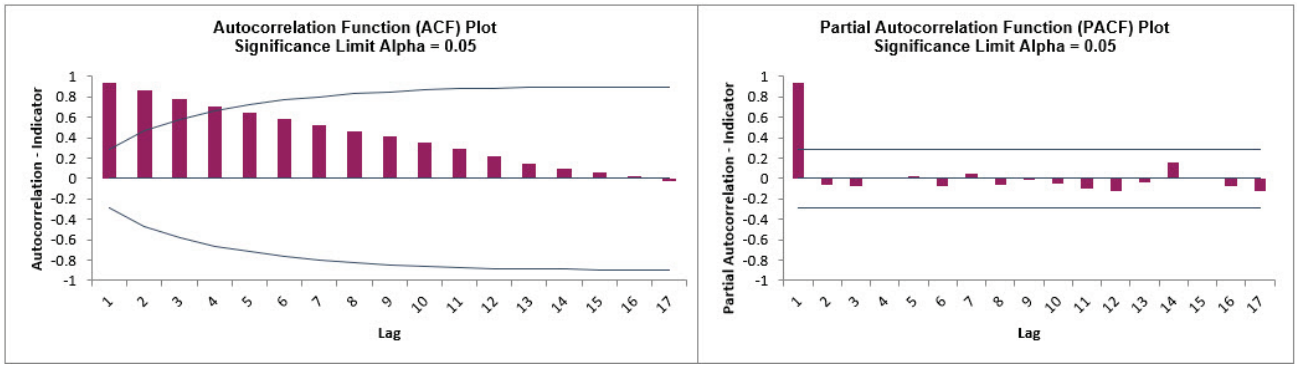
29. Seleccionar *Sales*, hacer click en **Numeric Time Series Data (Y) >>**. Seleccionar **Automatic Number of Lags.** Especificar **Seasonal Frequency** = 1 y **Alpha Level** = 0.05.

30. Hacer click en **OK**. Se generan los gráficos FAC/FACP.



Se puede observar aquí una fuerte autocorrelación con un lento declive debido a la tendencia.

31. Hacer click en la pestaña de la **Hoja 1**. Repetir los gráficos FAC/FACP para el Indicador:



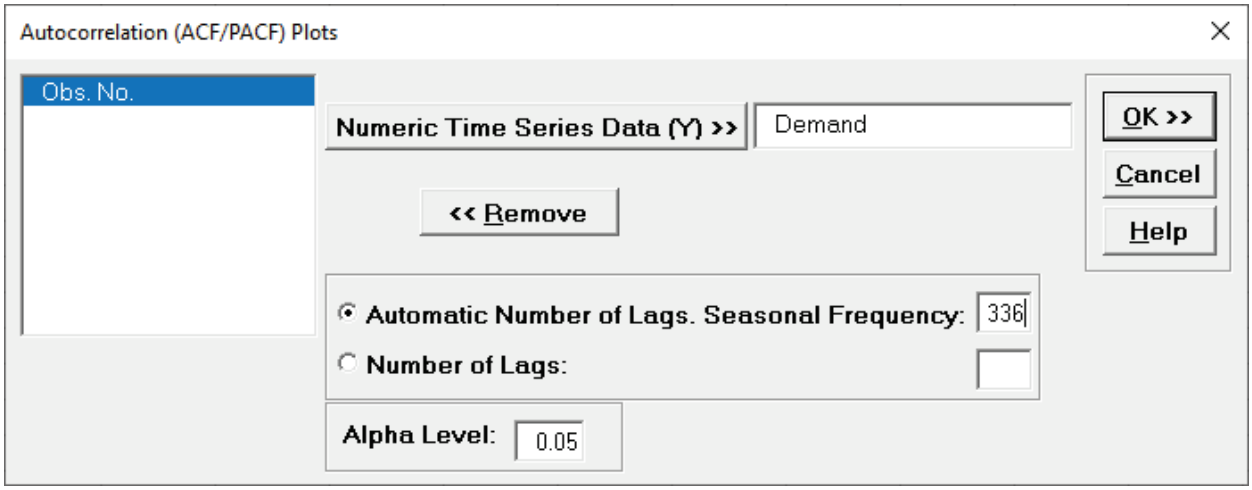
Como se esperaba, la autocorrelación para el Indicador es muy similar al de las Ventas. Se abordará aquello cuando se generen los gráficos de la FCC.

Demanda de Electricidad Estacional Múltiple de Media Hora – Taylor

32. Abrir **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (pestaña de la **Hoja 1**). Esta es la demanda de electricidad de media hora (MW) en Inglaterra y Gales desde el lunes 5 de junio de 2000 hasta el domingo 27 de agosto de 2000 (Taylor, R forecast). Estos datos tienen una estacionalidad múltiple con una frecuencia de 48 (observaciones por día) y 336 (observaciones por semana), con un total de 4032 observaciones.

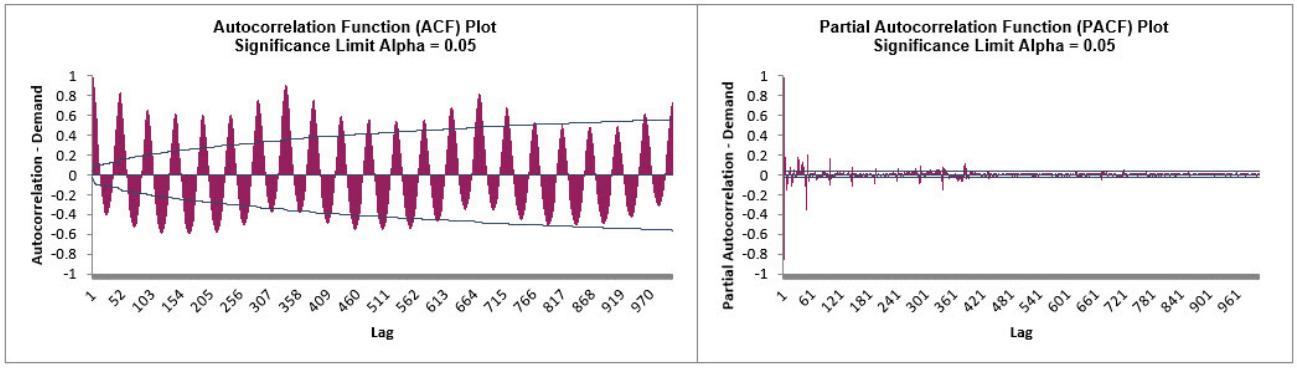
33. Hacer click en **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

34. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**. Seleccionar **Automatic Number of Lags.** Especificar **Seasonal Frequency** = 336 y **Alpha Level** = 0.05.



El número automático de rezagos será (como mínimo) la Frecuencia Estacional * 3 = 1008 lo que permite observar los patrones estacionales en la FAC.

35. Hacer click en **OK**. Se generan los gráficos FAC/FACP.



Se puede apreciar aquí los patrones estacionales de media hora, 48 diarios y 336 por semana.

Gráficos de Correlación Cruzada (FCC)

La correlación cruzada es similar a la autocorrelación, pero las correlaciones se calculan sobre dos variables de series temporales relacionadas, típicamente un proceso de entrada y una salida. Un gráfico de los datos X vs. los datos Y en el rezago k puede mostrar una tendencia positiva o negativa. Si la pendiente es positiva, la correlación cruzada es positiva; si hay una pendiente negativa, la correlación cruzada es negativa. Esto ayuda a identificar los rezagos importantes (o pistas) en el proceso y sería útil para su aplicación con las predictoras en un modelo ARIMA.

Si la entrada X se autocorrelaciona, la FCC se ve afectada por su estructura de la serie temporal y cualquier tendencia "en común" que las series X e Y puedan tener a lo largo del tiempo. El preblanqueo resuelve este problema eliminando la autocorrelación y las tendencias.

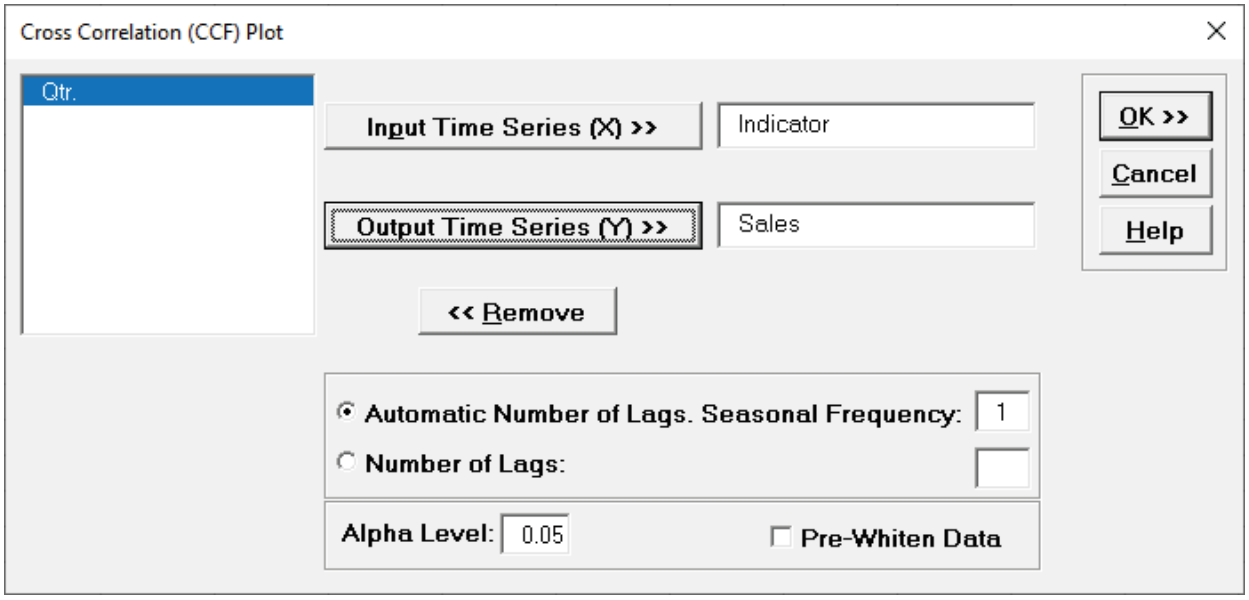
Para mayores detalles y fórmulas, véase el Apéndice: **Autocorrelación (FAC), Autocorrelación Parcial (FACP) y Correlación Cruzada (FCC).**

Ventas con Indicador - Serie M Modificada

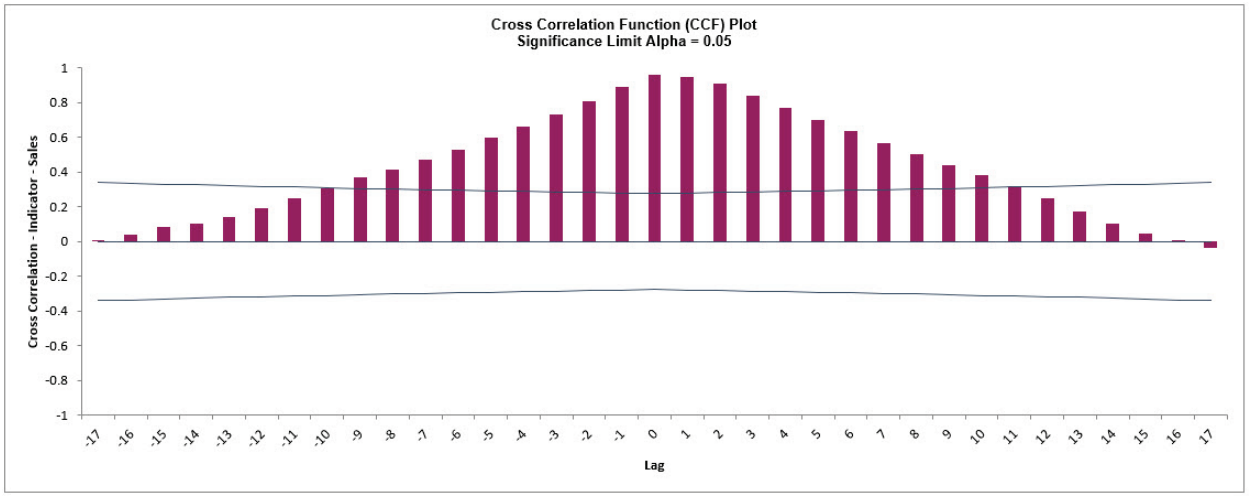
1. Abrir **Sales with Indicator - Modified Series M.xlsx**. (pestaña de la **Hoja 1**). Esta es una Serie M de datos modificados de Box y Jenkins, con los valores de 50 ventas trimestrales corporativas con un indicador principal. Se consideran los datos como no estacionales, tal como lo hicieron Box y Jenkins.

2. Hacer click en **SigmaXL > Time Series Forecasting > Cross Correlation (CCF) Plot**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Indicator*, hacer click en **Input Time Series (X) >>**. Seleccionar *Sales*, hacer click en **Output Time Series (Y) >>**. Usar por defecto **Automatic Number of Lags. Seasonal Frequency =** 1 y **Alpha Level** = 0.05.

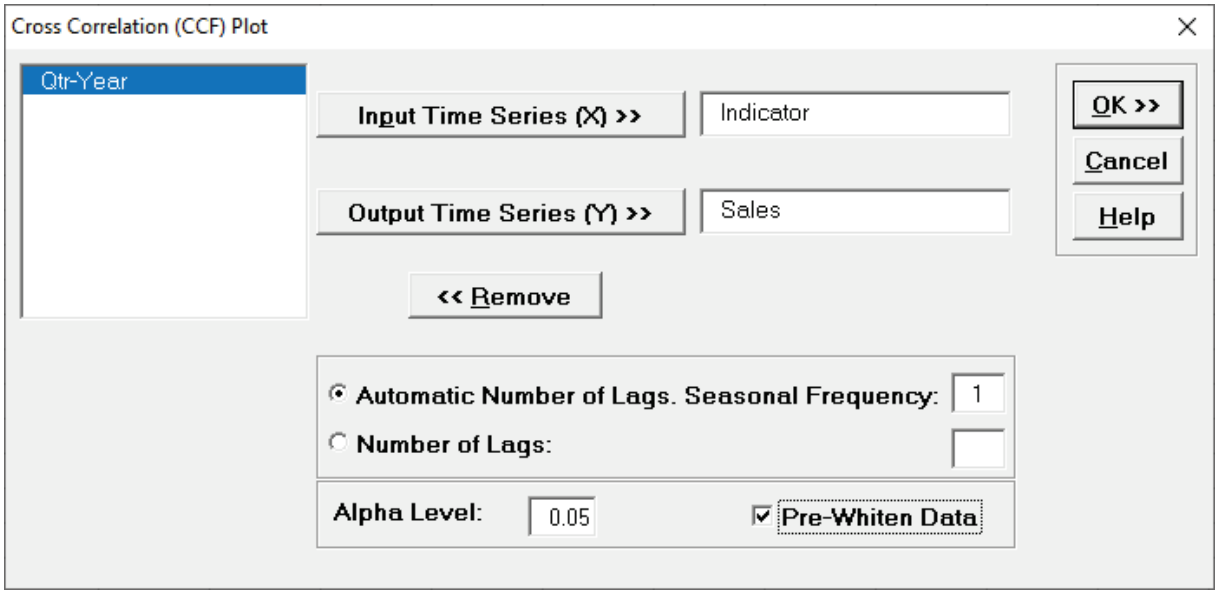


4. Hacer click en **OK**. Se genera el gráfico de la FCC.

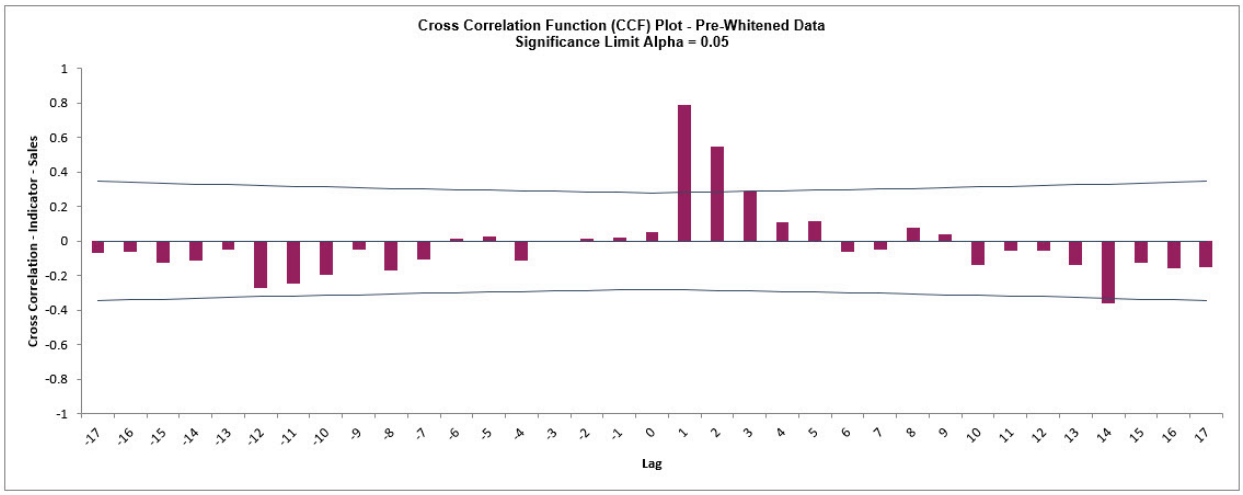


Este gráfico de la FCC muestra una correlación cruzada significativa desde el retardo = -9 a +9, con un pico en 0, sin embargo, la autocorrelación en los datos de X e Y está enmascarando la verdadera naturaleza de la correlación cruzada.

5. Se repetirá el gráfico FCC Plot con la opción de preblanqueo. Hacer click en el menú **Recall SigmaXL Dialog** o presione **F3** para repetir el último diálogo. Marcar **Pre-Whiten Data**.



6. Hacer click en **OK**. Se genera el gráfico de la FCC.



El preblanqueo de los datos ha alterado en forma dramática el gráfico de la FCC, lo que permite apreciar el patrón de correlación cruzada subyacente. Los rezagos 1 y 2 son significativamente positivos, y el rezago 3 cae justo sobre la línea.

Tomar en cuenta, si bien X es conocido como un indicador principal, X le precede a Y en el tiempo, el rezago positivo significa que la variable X está rezagada respecto a la variable Y en términos de la estructura de la correlación. SigmaXL hace uso de esta convención tal como en Box y Jenkins (2016, pp. 437-440).

El gráfico de la FCC será útil cuando se modelen los datos de las Ventas con el Pronóstico ARIMA y los Predictoras.

Diagrama de Densidad Espectral

El diagrama de Densidad Espectral se utiliza para identificar toda la frecuencia estacional dominante en los datos de series temporales mediante un análisis espectral con transformaciones rápidas de Fourier. El algoritmo utilizado aquí es el mismo que se utiliza en la opción del modelo de pronóstico para detectar automáticamente la frecuencia estacional. Si hay estacionalidad múltiple, se identificarán hasta tres frecuencias enteras. Si la frecuencia pico no es un número entero, se redondea.

El eje Y es la Densidad Espectral, el eje X es la Frecuencia Estacional. La gráfica de la Densidad Espectral también se conoce como Periodograma. Obsérvese que el uso que hace SigmaXL del término "frecuencia estacional" es el inverso de lo que se usa típicamente en las transformaciones de Fourier de "período estacional", como se mencionó en la Introducción.

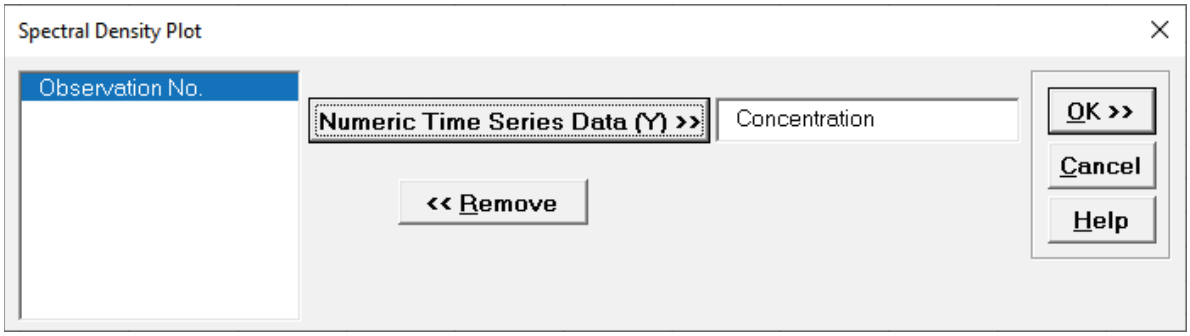
Para mayores detalles y fórmulas, véase el Apéndice: **Densidad Espectral y Detección Automática de Frecuencia Estacional.**

Concentración de Proceso Químico - Serie A

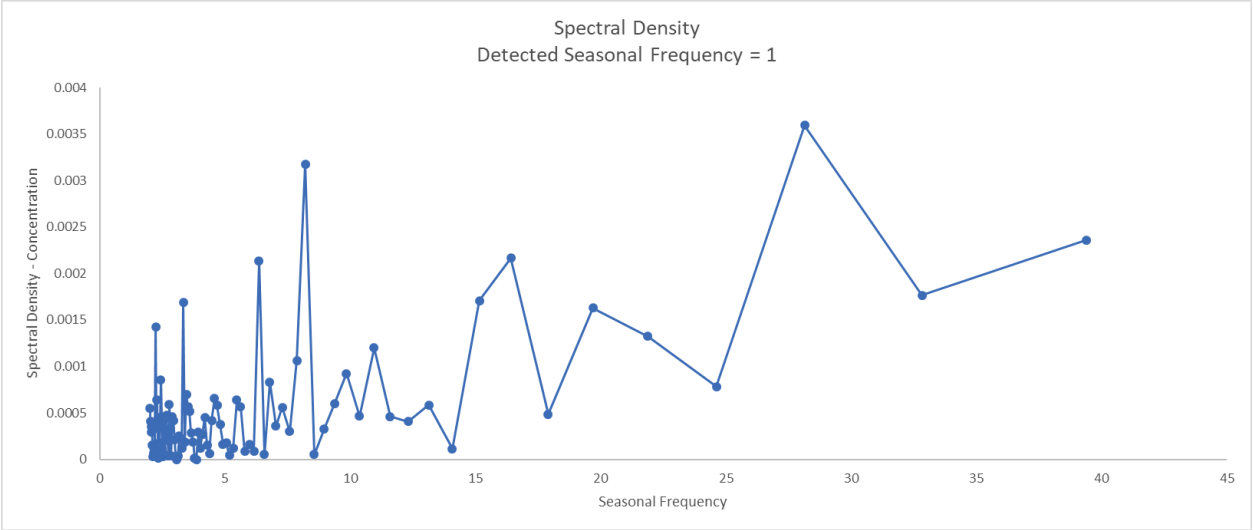
1. Abrir **Chemical Process Concentration – Series A.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie A de Box y Jenkins, un conjunto de 197 valores de la concentración de un proceso químico tomados a intervalos de 2 horas.

2. Hacer click en **SigmaXL > Time Series Forecasting > Spectral Density Plot.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Concentration*, hacer click en **Numeric Time Series Data (Y) >>**.



4. Hacer click en **OK**. Se genera un diagrama de densidad espectral para la Concentración.



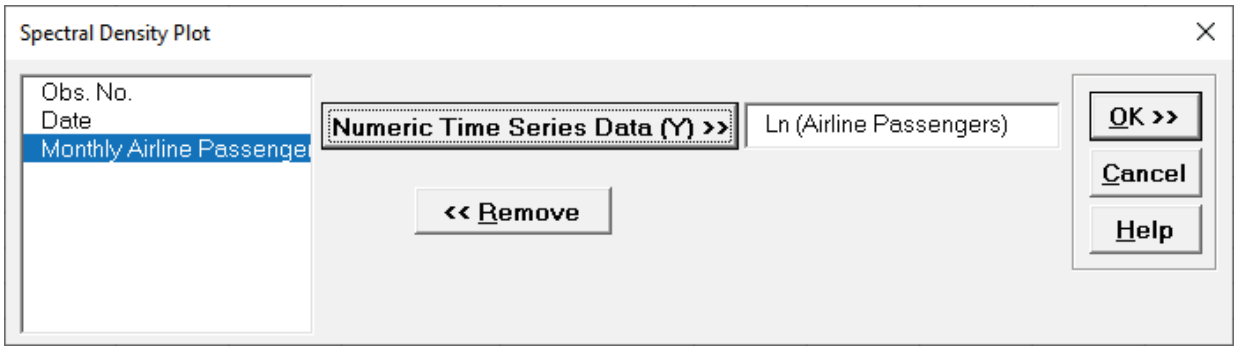
La frecuencia estacional detectada es 1, lo que significa que es un proceso no estacional. El pico a los 28 no tiene suficiente "fortaleza" estacional para que se considere su uso como una frecuencia estacional en un modelo de series de tiempo.

Pasajeros Mensuales de la Aerolínea - Serie G

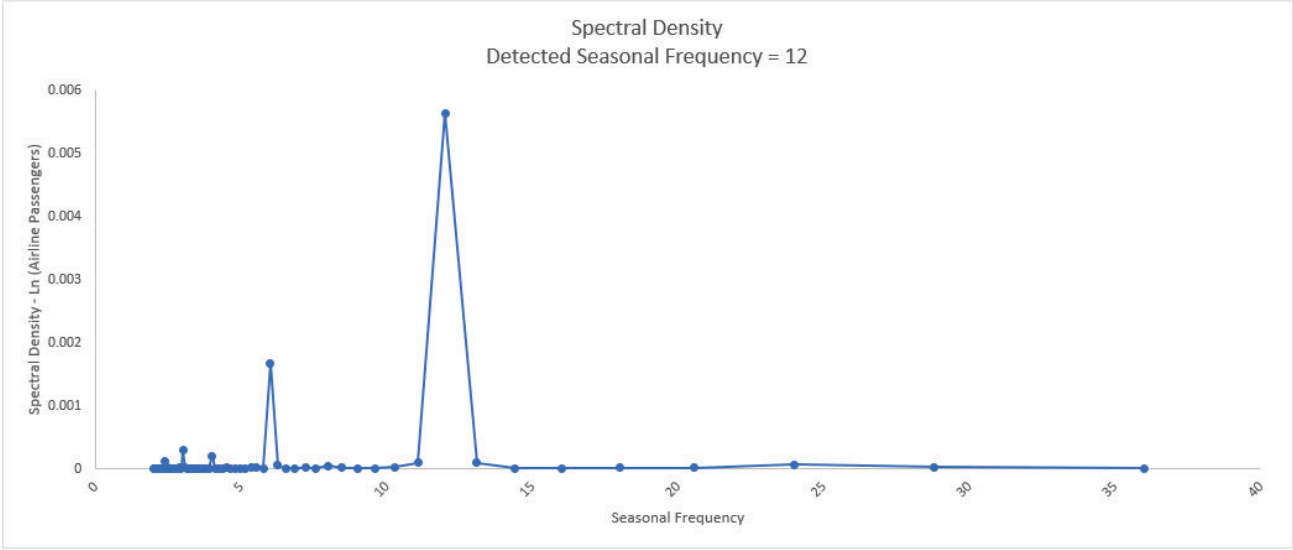
5. Abrir **Monthly Airline Passengers - Series G.xlsx** (pestaña de la **Hoja 1**). Estos datos son de la Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional desde enero de 1949 hasta diciembre de 1960.

6. Hacer click en **SigmaXL > Time Series Forecasting > Spectral Density Plot.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

7. Seleccionar *Ln (Airline Passengers)*, hacer click en **Numeric Time Series Data (Y) >>**.



8. Hacer click en **OK**. Se genera un Diagrama de Densidad Espectral para Ln (Airline Passengers).



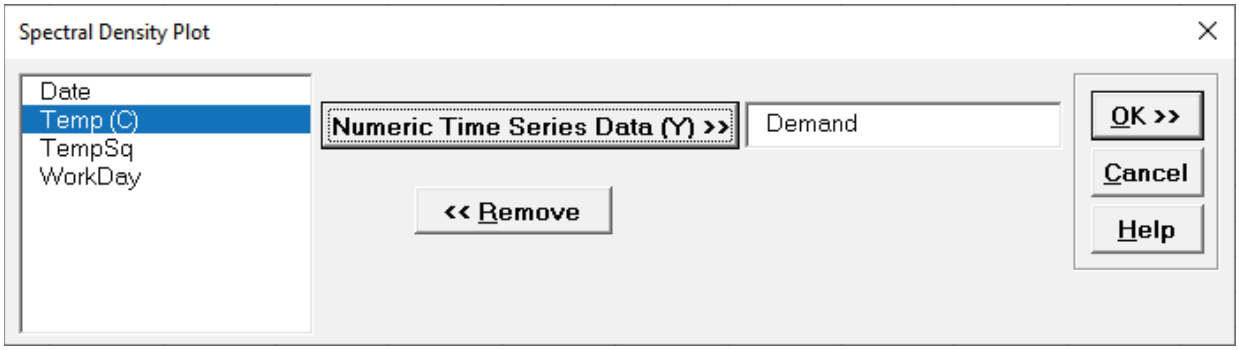
Como era de esperarse, la frecuencia estacional detectada para los datos mensuales es 12. El pico a los 6 no tiene suficiente "fortaleza" para que se lo considere como una segunda frecuencia estacional.

Demanda Diaria de Electricidad con Predictoras – ElecDaily

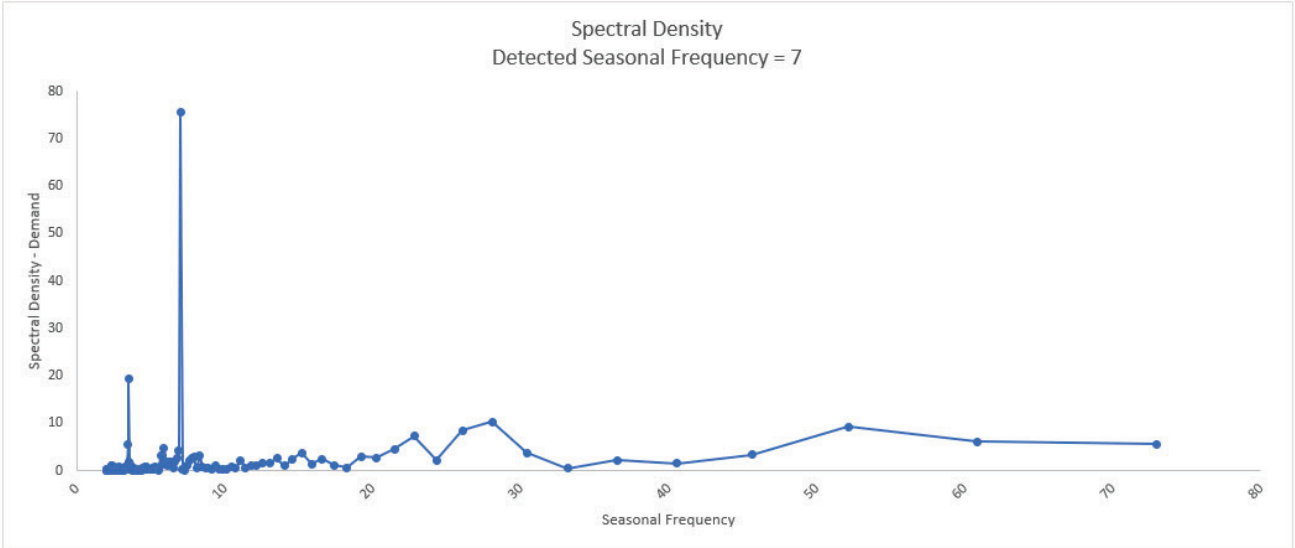
9. Abrir **Demanda Diaria de Electricidad con Predictoras – ElecDaily.xlsx** (pestaña de la Hoja 1). Esta es la demanda diaria de electricidad (GW) para el estado de Victoria, Australia, todos los días durante 2014 (Hyndman, fpp2). Estos datos tienen una frecuencia estacional = 7 (observaciones por semana).

10. Hacer click en **SigmaXL > Time Series Forecasting > Spectral Density Plot.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

11. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**.



12. Hacer click en **OK**. Se genera un Diagrama de Densidad Espectral para la Demanda.



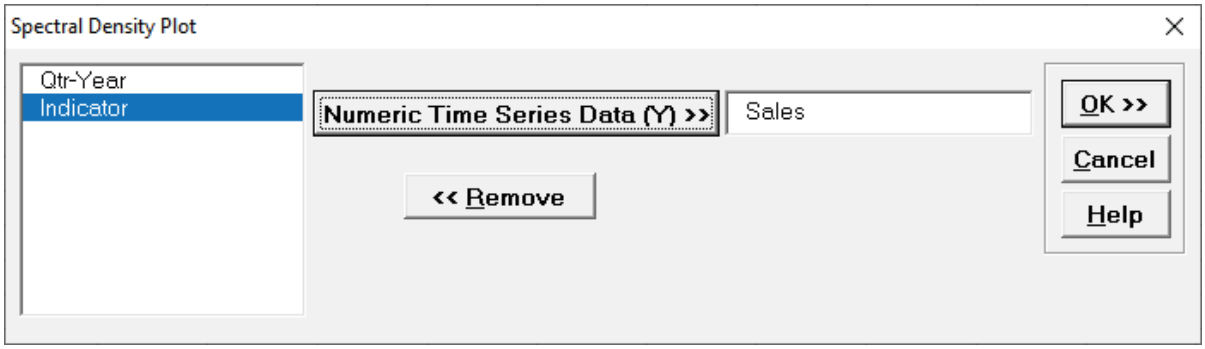
Como era de esperarse, la frecuencia estacional detectada para los datos diarios es 7.

Ventas con Indicador – Serie M Modificada

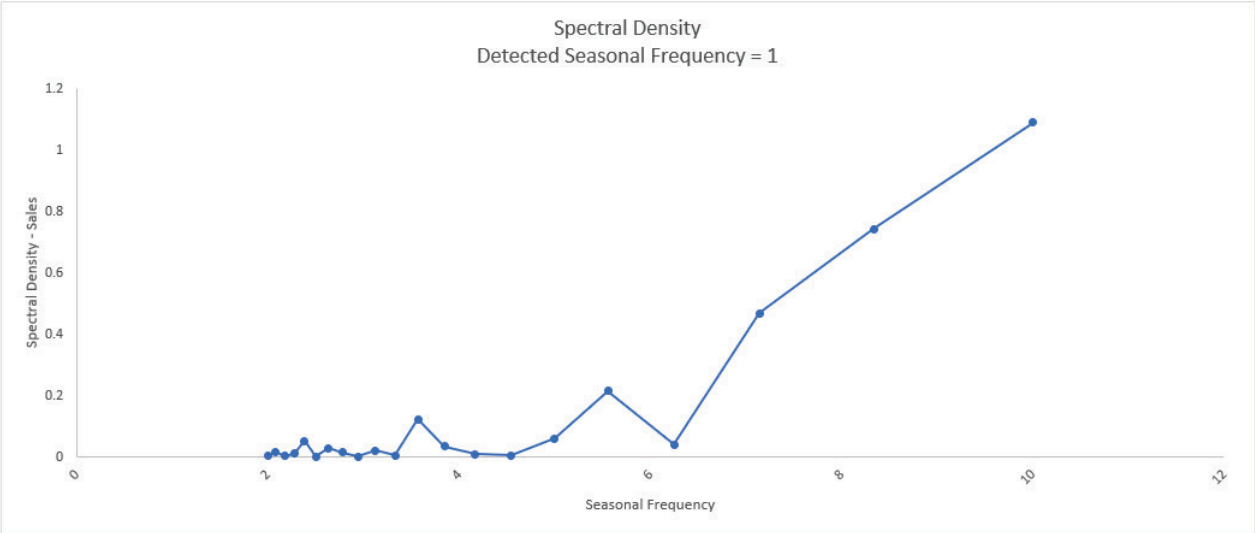
13. Abrir **Sales with Indicator - Modified Series M.xlsx**. (pestaña de la Hoja 1). Esta es una Serie M de datos modificados de Box y Jenkins, con los valores de 50 ventas trimestrales corporativas con un indicador principal. Se consideran los datos como no estacionales, tal como lo hicieron Box y Jenkins.

14. Hacer click en **SigmaXL > Time Series Forecasting > Spectral Density Plot.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

15. Seleccionar *Sales*, hacer click en **Numeric Time Series Data (Y) >>**.



16. Hacer click en **OK**. Se genera un Diagrama de Densidad Espectral para las Ventas.



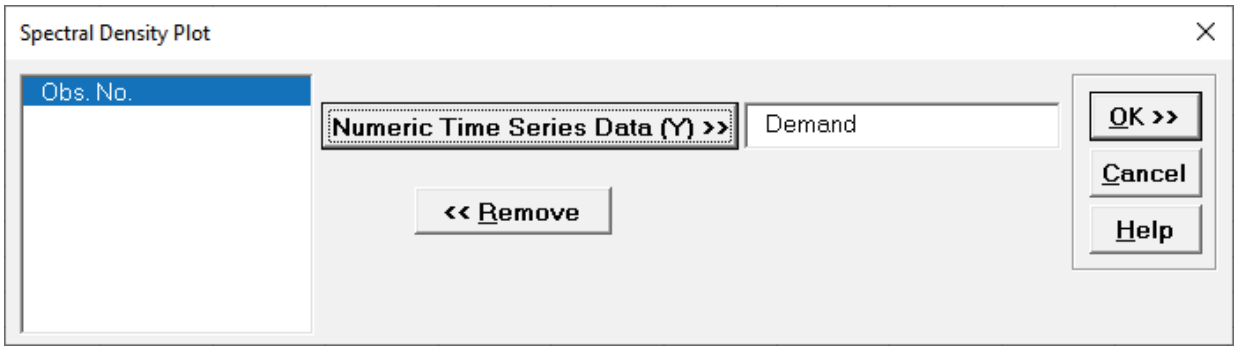
La frecuencia estacional detectada es 1, lo que confirma ser un proceso no estacionario. El pico a los 10 no tiene suficiente “fortaleza” para que se considere su uso como una segunda frecuencia estacional en los modelos de series temporales.

Demanda de Media Hora de Electricidad Estacional Múltiple – Taylor

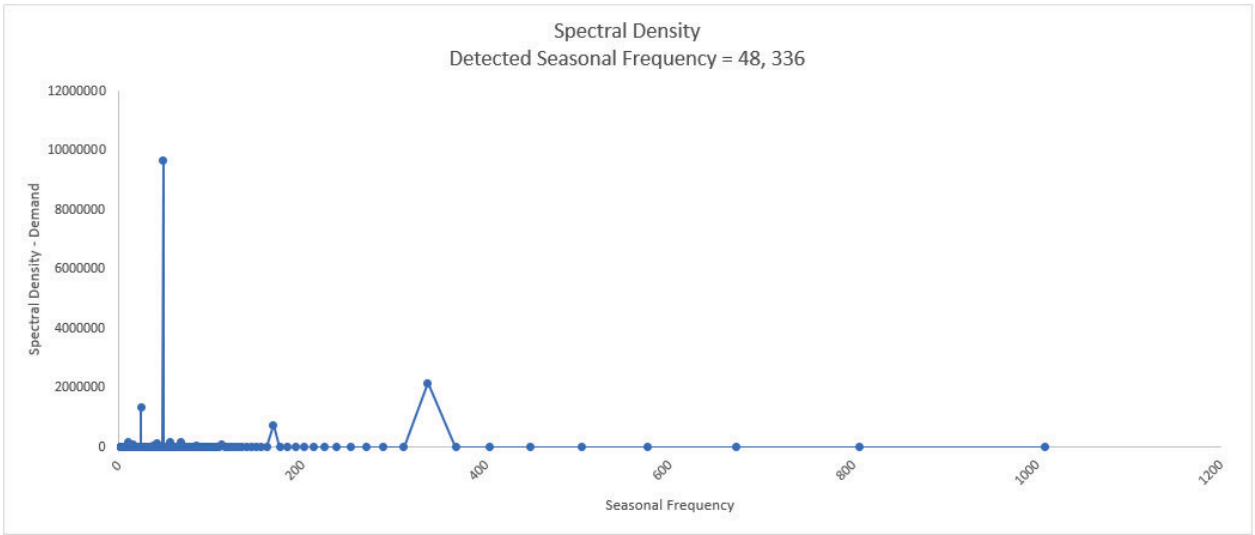
17. Abrir **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (pestaña de la **Hoja 1**). Esta es la demanda de electricidad de media hora (MW) en Inglaterra y Gales desde el lunes 5 de junio de 2000 hasta el domingo 27 de agosto de 2000 (Taylor, R forecast). Estos datos tienen una estacionalidad múltiple con una frecuencia de 48 (observaciones por día) y 336 (observaciones por semana), con un total de 4032 observaciones.

18. Hacer click en **SigmaXL > Time Series Forecasting > Spectral Density Plot.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

19. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**.



20. Hacer click en **OK**. Se genera un Diagrama de Densidad Espectral para la Demanda.



Se confirma que se detecta en 48,336 la frecuencia estacional múltiple.

Gráficos de Descomposición de Tendencias Estacionales

Los gráficos de Descomposición de Tendencias Estacionales son útiles para distinguir visualmente las tendencias y los componentes estacionales en los datos de series de tiempo. Si la Frecuencia Estacional está desmarcada, la primera carta que se genera es un Gráfico de Descomposición de Tendencias, mostrando los datos sin procesar y la tendencia. El componente de la tendencia utiliza la suavización de los datos, en lugar de una tendencia lineal, de manera que puede mostrar una tendencia lineal o patrones cíclicos. Si se especifica una única frecuencia estacional, se genera un gráfico de Descomposición de Tendencia Estacional, que muestra los datos, la tendencia suavizada y el componente estacional. Si se especifica una frecuencia estacional múltiple, se genera un gráfico de Descomposición de Tendencia Estacional Múltiple, que muestra los datos, la tendencia suavizada y el componente estacional múltiple.

El segundo gráfico muestra sólo la tendencia suavizada; el tercero (si procede) muestra sólo el componente estacional o el componente multiestacional. El último gráfico es el componente residual.

Se trata de un modelo de descomposición aditivo, de modo que la suma del valor de la tendencia + valor(es) estacional(es) + valor de los residuos dan como resultado el valor original de los datos. Se puede obtener un equivalente multiplicativo especificando la transformación Box-Cox con Lambda = 0, que es una transformación Ln, pero los gráficos mostrarán los datos transformados para mantener un modelo aditivo. También se puede utilizar Lambda redondeado u óptimo, pero sólo se considerará el rango de valores 0 a 1 (este enfoque conservador se emplea en el pronóstico de series temporales, a diferencia del Box-Cox regular en SigmaXL que utiliza un rango de -5 a 5). Véase el Apéndice: **Transformación Box-Cox.**

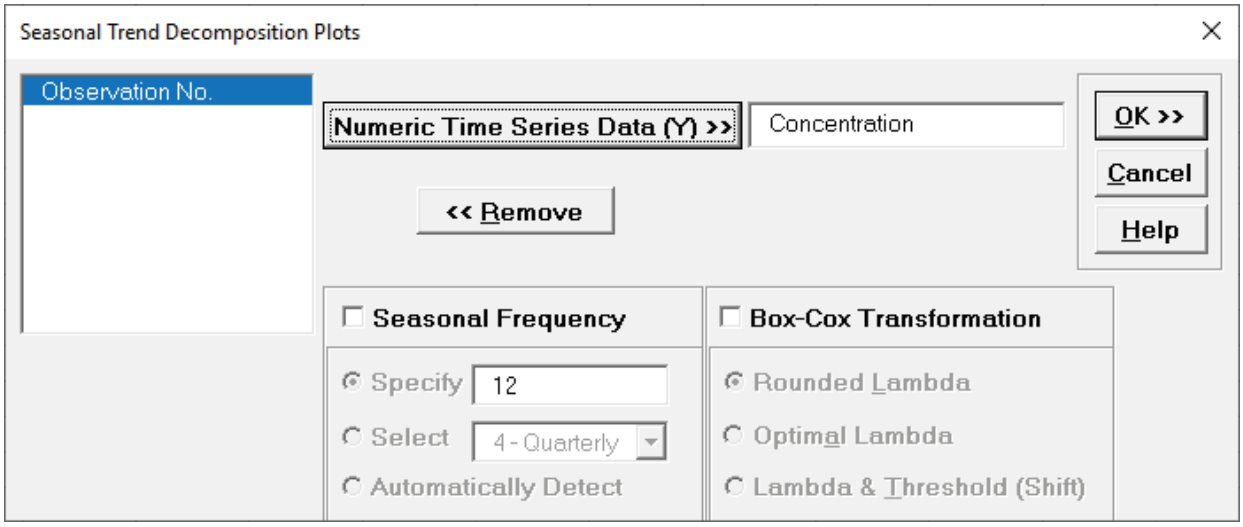
Los algoritmos de descomposición que se utilizan aquí son los mismos que se utilizan en la Suavización Exponencial - Descomposición Estacional Múltiple (DEM), y ARIMA - DEM. El componente estacional se elimina primero a través de la descomposición, un modelo de suavización exponencial no estacional ajustado al residual + tendencia, y luego se vuelve a añadir el componente estacional. Esto se utiliza principalmente para series temporales de alta frecuencia estacional y/o de frecuencia estacional múltiple. Para más detalles y fórmulas, véase el Apéndice: **Descomposición de Tendencia Estacional.**

Concentración de Proceso Químico – Serie A

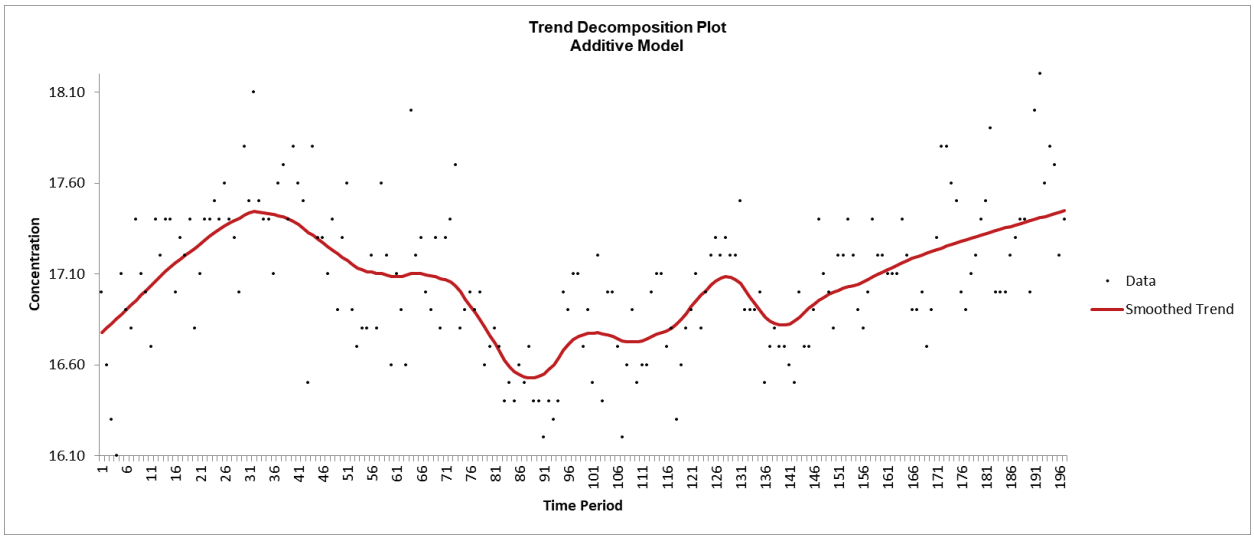
1. Abrir **Chemical Process Concentration – Series A.xlsx** (pestaña de la **Hoja 1**). Estos son los datos de una Serie A de Box y Jenkins, un conjunto de datos de 197 valores de concentración de un proceso químico que toma intervalos de dos horas.

2. Hacer click en **SigmaXL > Time Series Forecasting > Seasonal Trend Decomposition Plots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Concentration*, hacer click en **Numeric Time Series Data (Y) >>**. Desmarcar **Seasonal Frequency** y **Box-Cox Transformation** como se muestra.



4. Hacer click en **OK**. Se genera un gráfico de la Descomposición de la Tendencia para la Concentración.



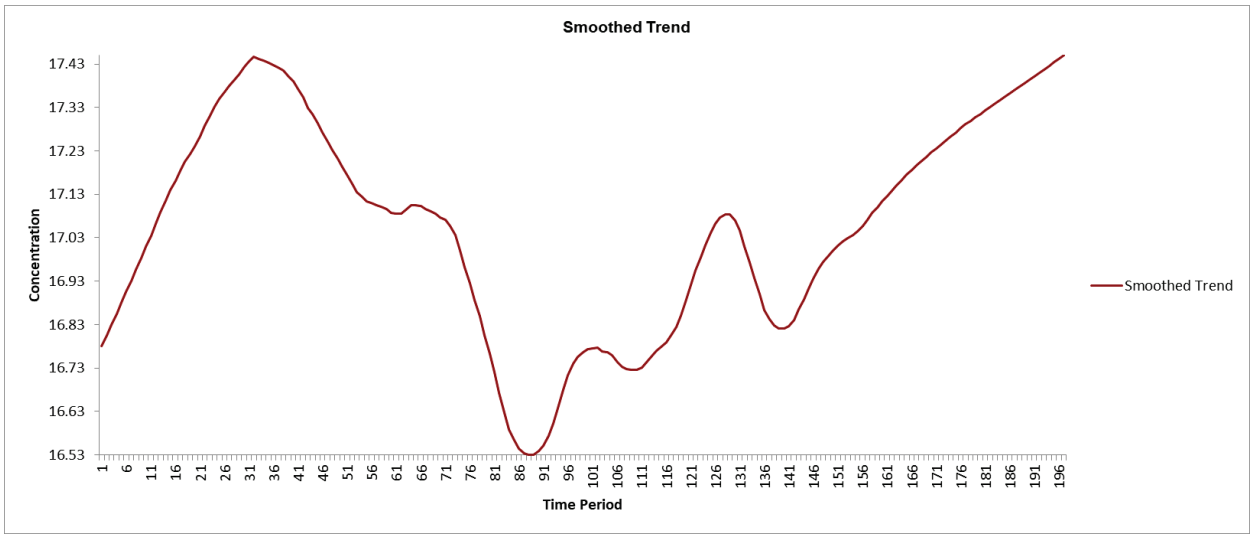
Evidentemente, se puede apreciar "la media fluctuante" en este proceso. Como se mencionó anteriormente, esto se puede modelar con suavización exponencial o con ARIMA después de aplicar diferencias.

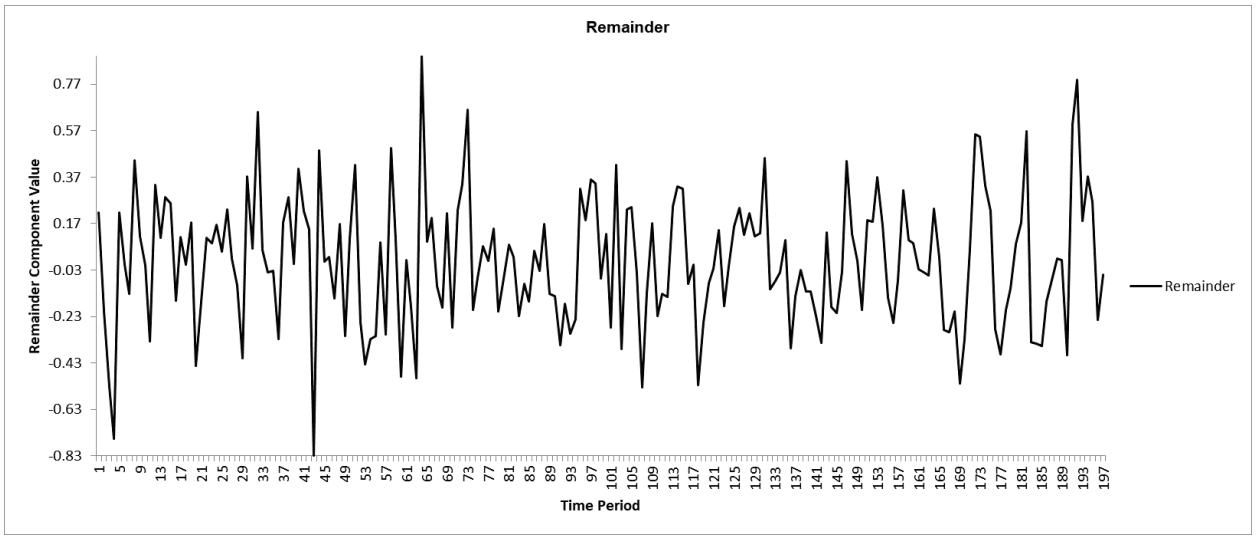
5. Se incluye un informe del Resumen de la Descomposición en el lado derecho de la gráfica, que indica la frecuencia estacional y los parámetros Box-Cox (si procede)

Decomposition Summary	
Seasonal Frequency	1
Box-Cox Transformation	N/A
Lambda	
Threshold	

La Frecuencia Estacional = 1 denota un proceso no estacional.

6. Aparecen también los gráficos de la Tendencia Suavizada y de los Residuos:



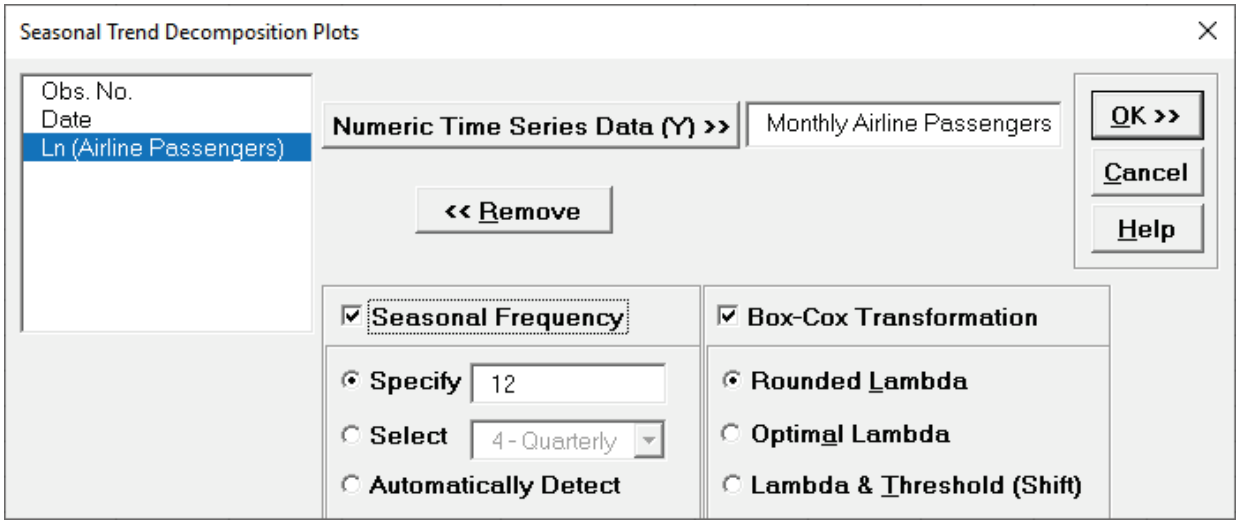


Pasajeros Mensuales de la Aerolínea – Serie G

7. Abrir **Monthly Airline Passengers - Series G.xlsx** (pestaña de la **Hoja 1**). Estos datos son de la Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional desde enero de 1949 hasta diciembre de 1960.

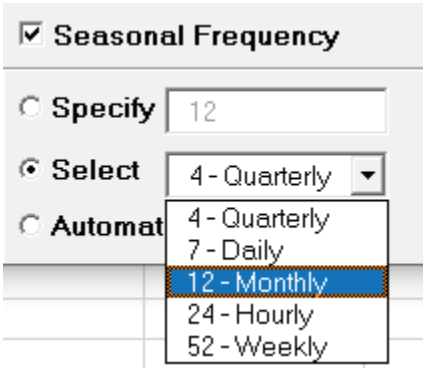
8. Hacer click en **SigmaXL > Time Series Forecasting > Seasonal Trend Decomposition Plots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

9. Seleccionar *Monthly Airline Passengers*, hacer click en **Numeric Time Series Data (Y) >>**. Marcar **Seasonal Frequency** con **Specify** = 12**.** Marcar **Box-Cox** con **Rounded Lambda** (se ha seleccionado porque **Carta de Corrida** ha mostrado un incremento en la varianza estacional a lo largo del tiempo).



- **Seasonal Frequency Specify** permite el ingreso de frecuencias múltiples

- **Seasonal Frequency Select** ofrece una lista desplegable de las frecuencias estacionales más comunes:



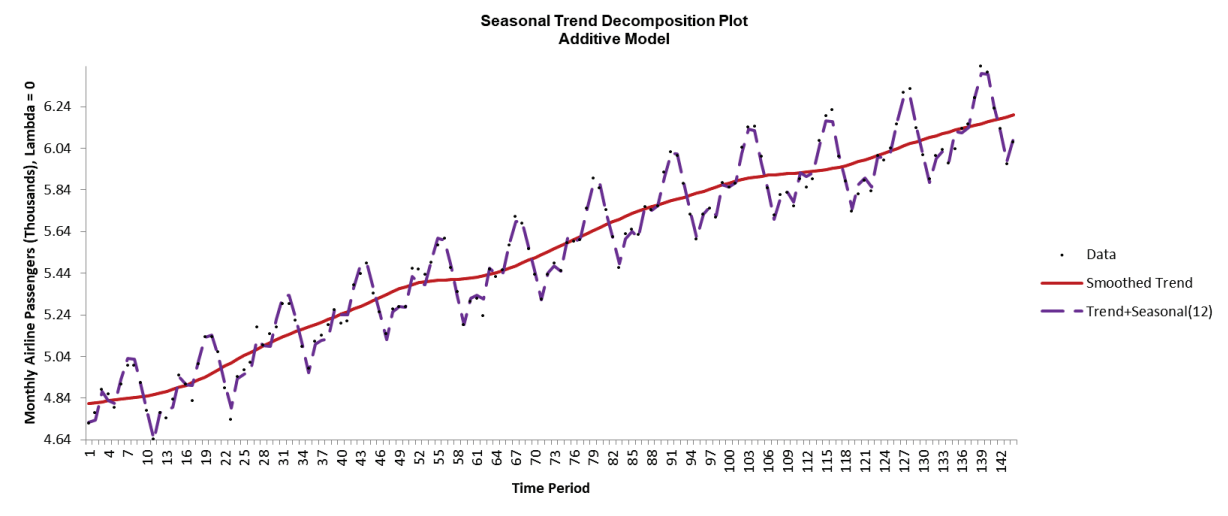
- **Seasonal Frequency Automatically Detect** debería usarse si no está seguro cuál es el valor de la frecuencia estacional (o realice un Diagrama de Densidad Espectral antes del Gráfico de Descomposición de la Tendencia Estacional).

- **Box-Cox Transformation** con **Rounded Lambda** seleccionará Lambda = 0 (Ln), 0.5 (SQRT) o 1 (Sin transformar). Threshold (Shift) se procesa en forma automática si la serie de tiempo incluye 0 o valores negativos, de lo contrario es 0.

- **Box-Cox Transformation** con **Optimal Lambda** usa el rango de 0 a 1 para Lambda. El umbral (threshold) se procesa en forma automática si la serie de tiempo incluye 0 o valores negativos.

- **Box-Cox Transformation** con **Lambda & Threshold (Shift)** si se lo deja vacío, procesará el óptimo de lambda y el umbral. El usuario puede especificar también Lambda y el Threshold. Se puede especificar Lambda fuera del rango de 0 a 1, pero para el análisis de las series de tiempo, se debería limitar entre -1 y 2. El umbral, por lo general, es 0, no obstante, si los datos de la serie temporal incluyen 0 o valores negativos, se debería registrar un valor negativo para el umbral que sea más pequeño que el valor mínimo. Se restará este valor de los datos que resulten de los valores positivos de la serie temporal.

10. Hacer click en **OK**. Se genera un gráfico de la Descomposición de la Tendencia Estacional para los Pasajeros Mensuales de la Aerolínea.

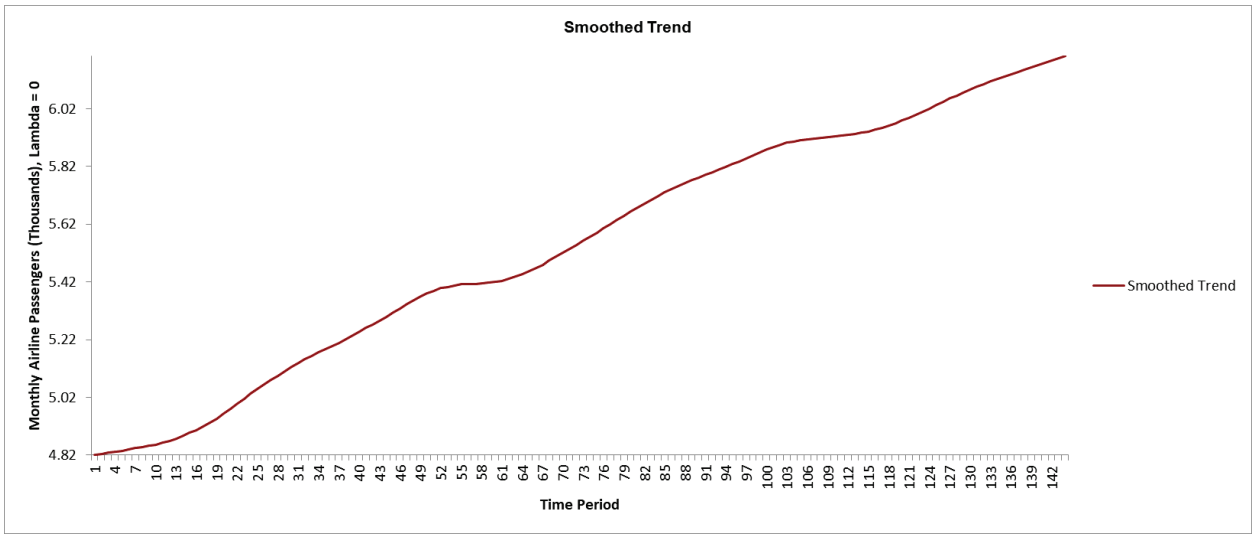


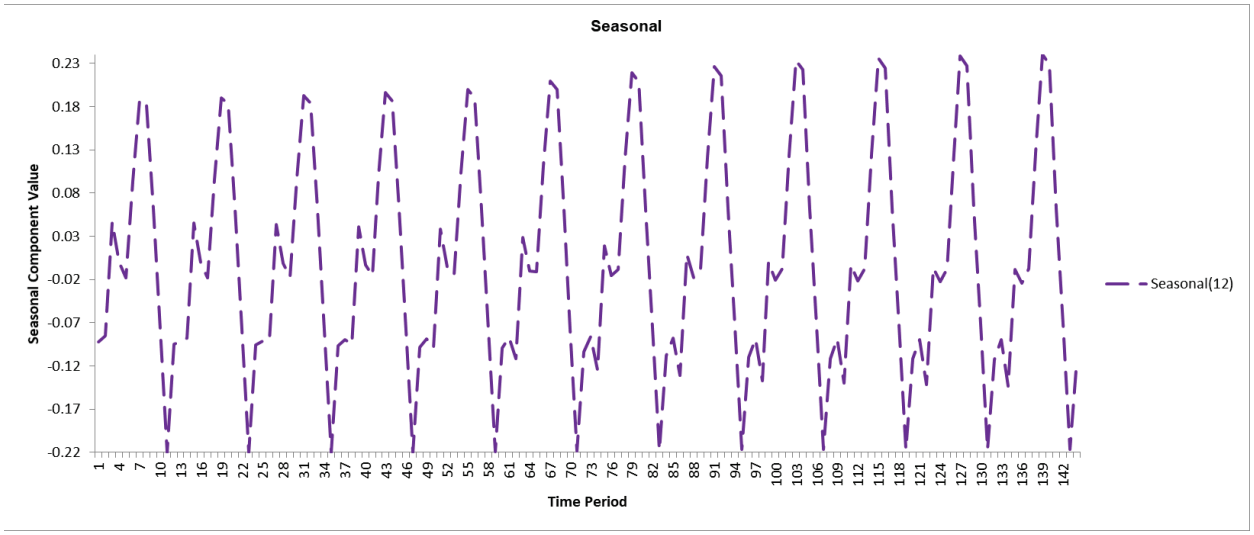
Se aprecia aquí una fuerte tendencia positiva, así como también un efecto mensual estacional. Tomar en cuenta que este es el dato con Lambda = 0 (Ln transformed). La información sobre la transformación Box-Cox aparece en el informe del Resumen de la Descomposición, al lado derecho de la gráfica:

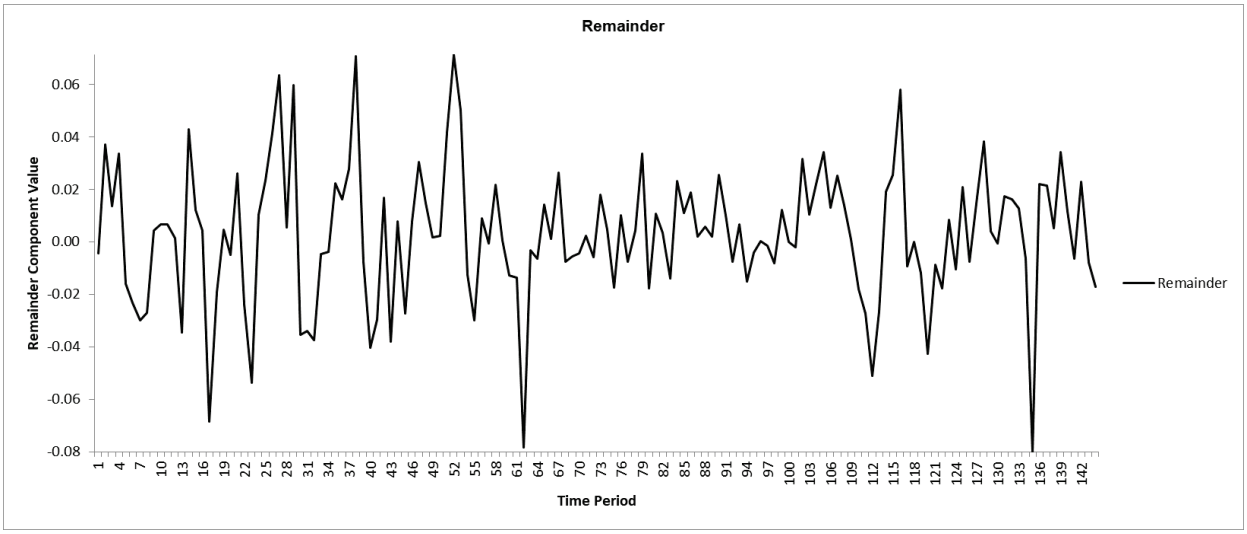
Decomposition Summary	
Seasonal Frequency	12
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

La Frecuencia Estacional = 12 y Lambda =0. Los datos transformados con Ln se visualizan para mantener un modelo aditivo, el cual es más sencillo que interpretar un modelo multiplicativo.

11. Aparecen también los gráficos de la Tendencia Suavizada y de los Residuos:





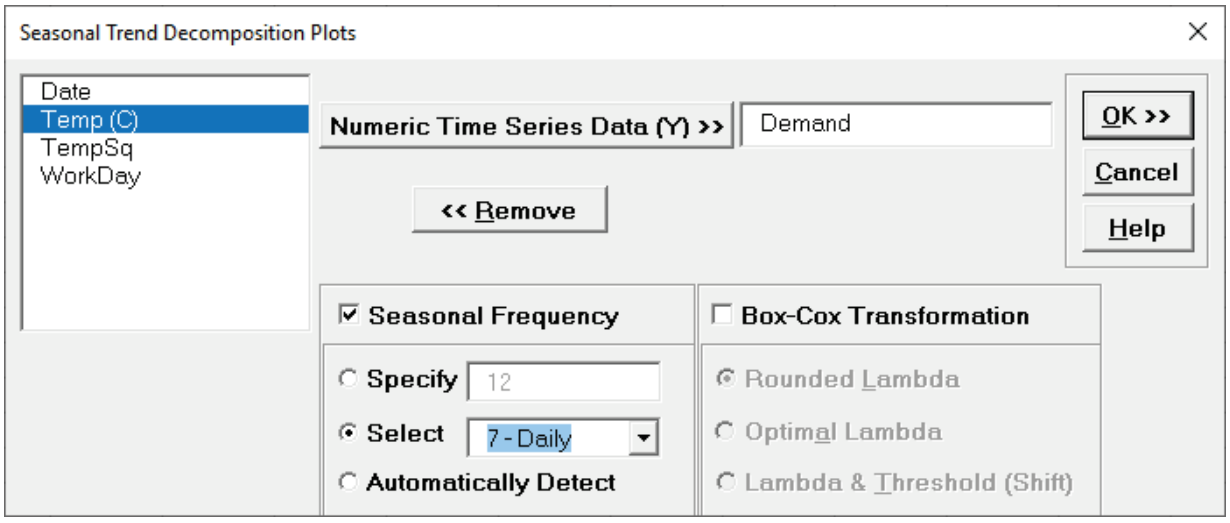


Demanda Diaria de Electricidad con Predictoras – ElecDaily

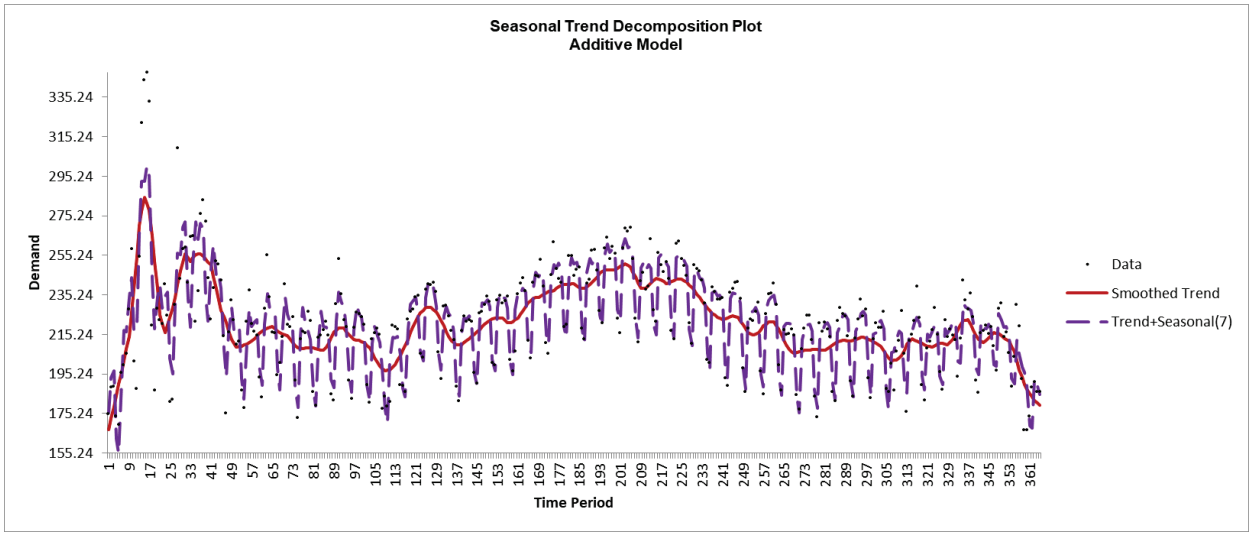
12. Abrir **Demanda Diaria de Electricidad con Predictoras – ElecDaily.xlsx** (pestaña de la **Hoja 1**). Esto es la demanda diaria de electricidad (GW) para el estado de Victoria, Australia, a diario durante el 2014 (Hyndman, fpp2). Estos datos tienen una frecuencia estacional = 7 (observaciones por semana).

13. Hacer click en **SigmaXL > Time Series Forecasting > Seasonal Trend Decomposition Plots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

14. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**. Marcar **Seasonal Frequency** con **Select** = *7-Daily* desde el listado despegable**.** Desmarcar **Box-Cox Transformation**.

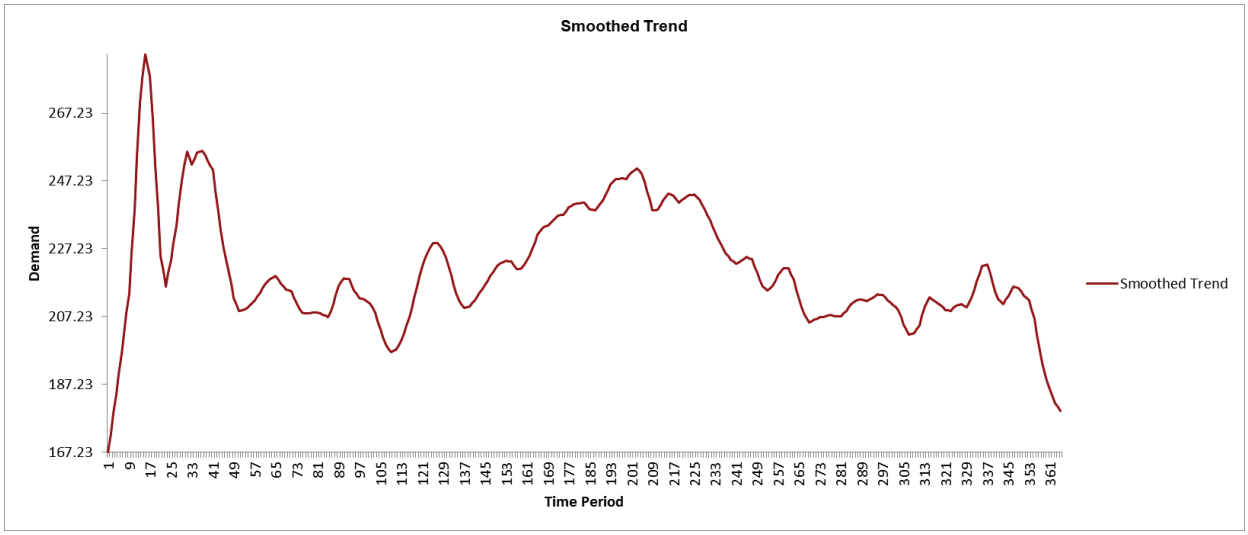


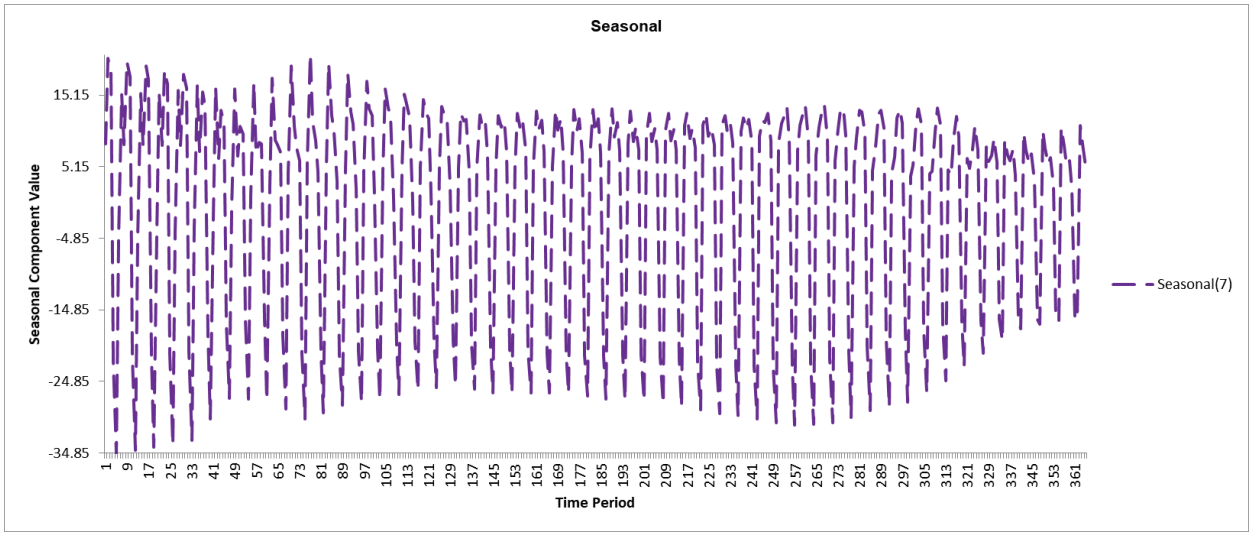
15. Hacer click en **OK**. Se genera un gráfico de Descomposición de la Tendencia Estacional para la Demanda.

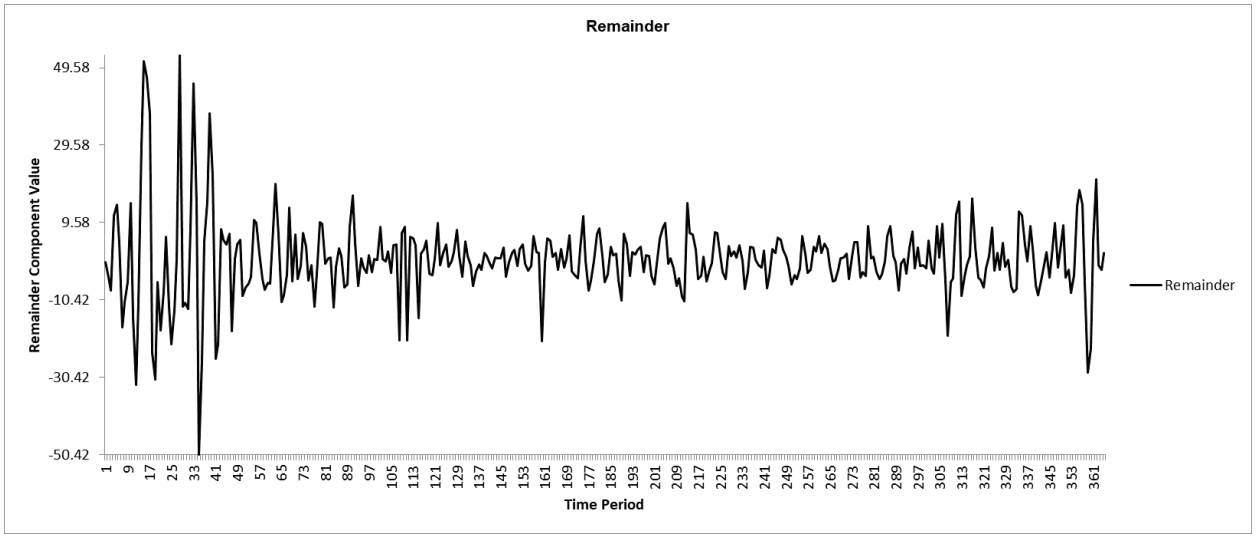


Se aprecia aquí el efecto estacional diario. Como se mencionó anteriormente, algunos de los patrones de la tendencia se pueden explicar por las predictoras Temp (C), TempSq y WorkDay.

16. Se generan los gráficos de la Tendencia Suavizada, Estacional y los Residuos como se muestra:




Seasonal
Seasonal Component Value
15.15
5.15
-4.85
-14.85
-24.85
-34.85
Time Period
Seasonal(7)

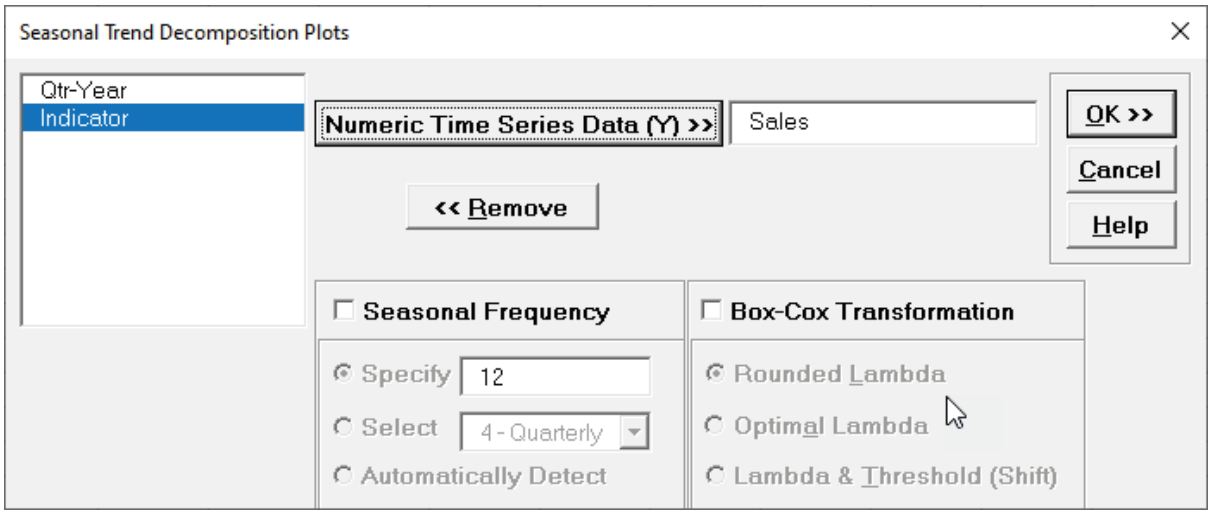

Remainder
Remainder Component Value
49.58
29.58
9.58
-10.42
-30.42
-50.42
Time Period
Remainder

Ventas con Indicador - Serie M Modificada

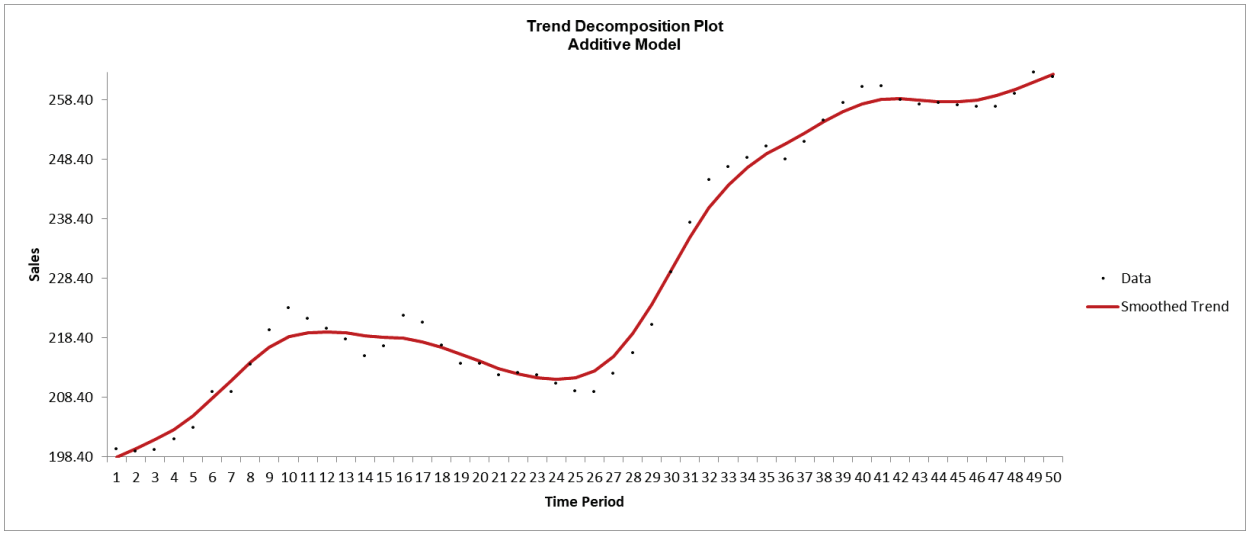
17. Abrir **Sales with Indicator - Modified Series M.xlsx**. (pestaña de la **Hoja 1**). Estos son los datos de una Serie M modificada de Box y Jenkins, con 50 valores de las ventas trimestrales corporativas con un indicador principal. Se consideran los datos no estacionales, tal como lo hicieron Box y Jenkins.

18. Hacer click en **SigmaXL > Time Series Forecasting > Seasonal Trend Decomposition Plots**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

19. Seleccionar *Sales*, hacer click en **Numeric Time Series Data (Y) >>**. Desmarcar **Seasonal Frequency** y **Box-Cox Transformation** como se muestra.

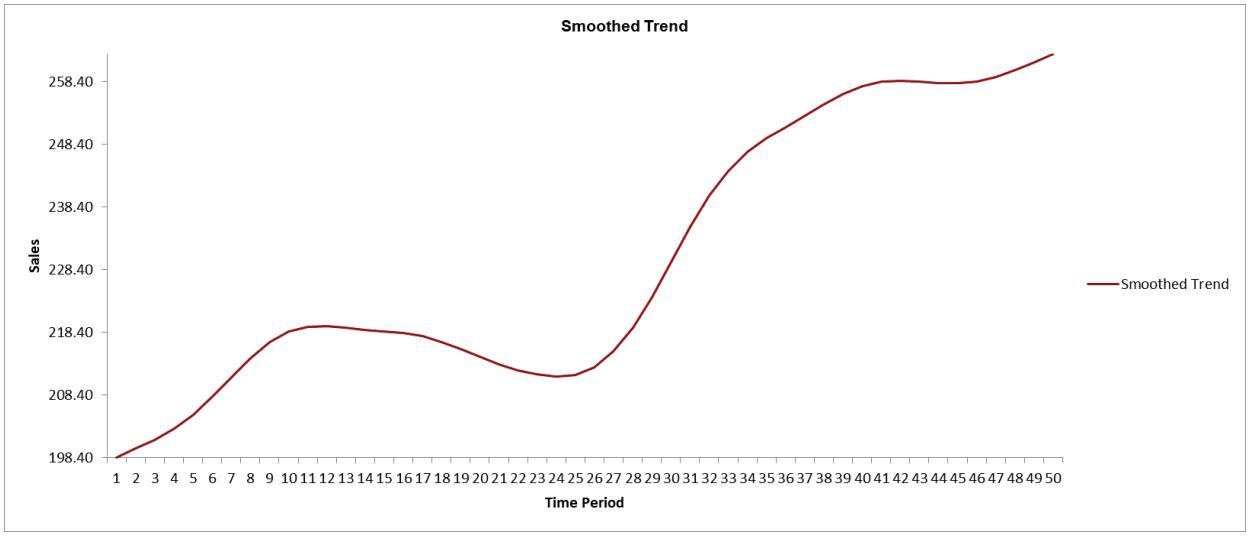


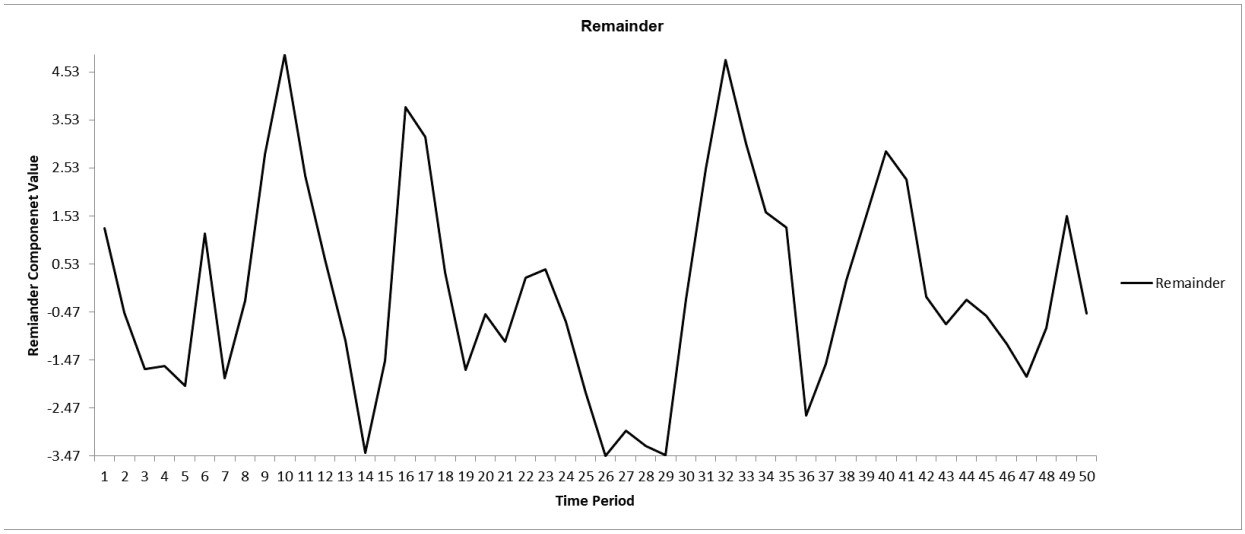
20. Hacer click en **OK**. Se genera un gráfico de la Descomposición para la Concentración.



Se aprecia aquí una tendencia general positiva en los 50 trimestres.

21. Se generan los gráficos de la Tendencia Suavizada, Estacional y los Residuos como se muestra:



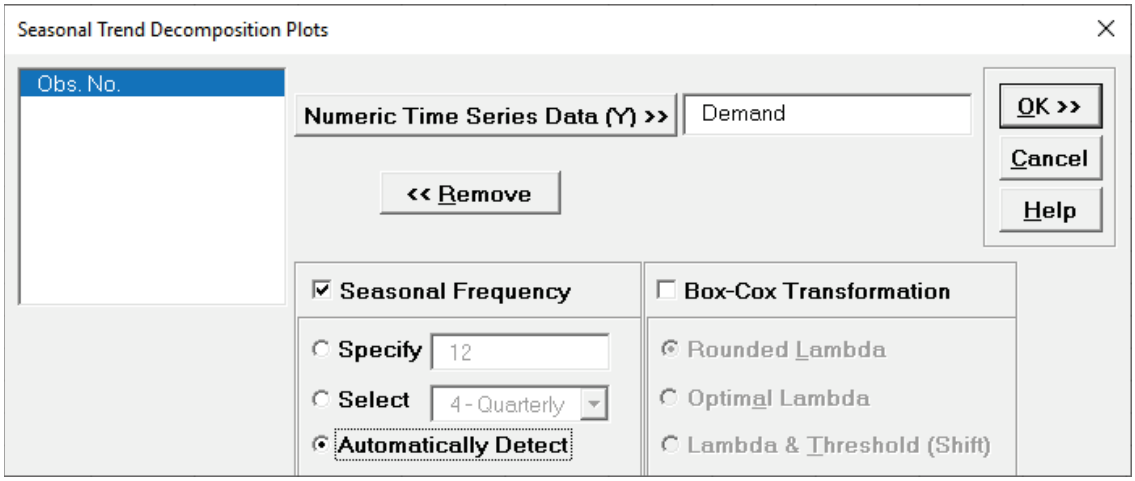


Demanda de Electricidad Estacional Múltiple de Media Hora – Taylor

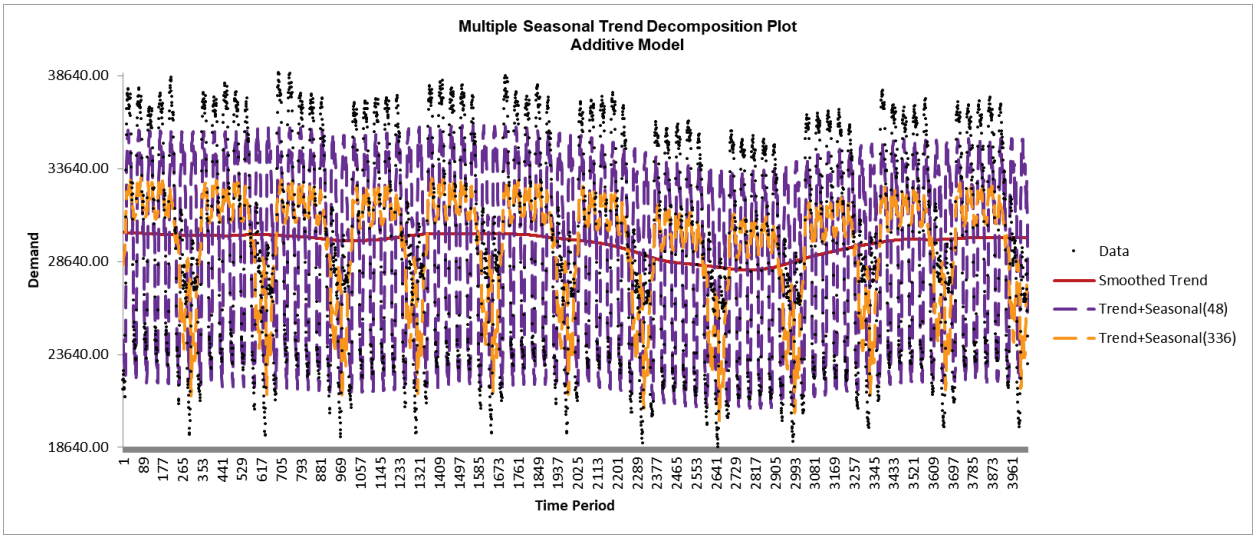
22. Abrir **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (pestaña de la Hoja 1). Esta es la demanda de electricidad de media hora (MW) en Inglaterra y Gales desde el lunes 5 de junio de 2000 hasta el domingo 27 de agosto de 2000 (Taylor, R forecast). Estos datos tienen una estacionalidad múltiple con una frecuencia de 48 (observaciones por día) y 336 (observaciones por semana), con un total de 4032 observaciones.

23. Hacer click en **SigmaXL > Time Series Forecasting > Seasonal Trend Decomposition Plots.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

24. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**. Marcar **Seasonal Frequency** with **Automatically Detect.** Desmarcar **Box-Cox Transformation**.

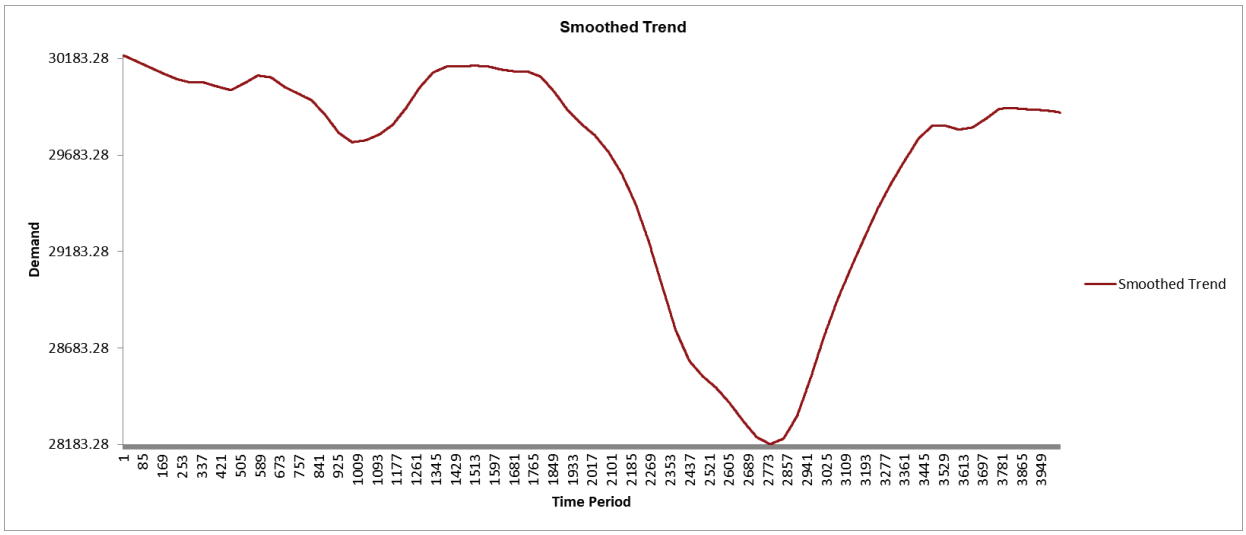


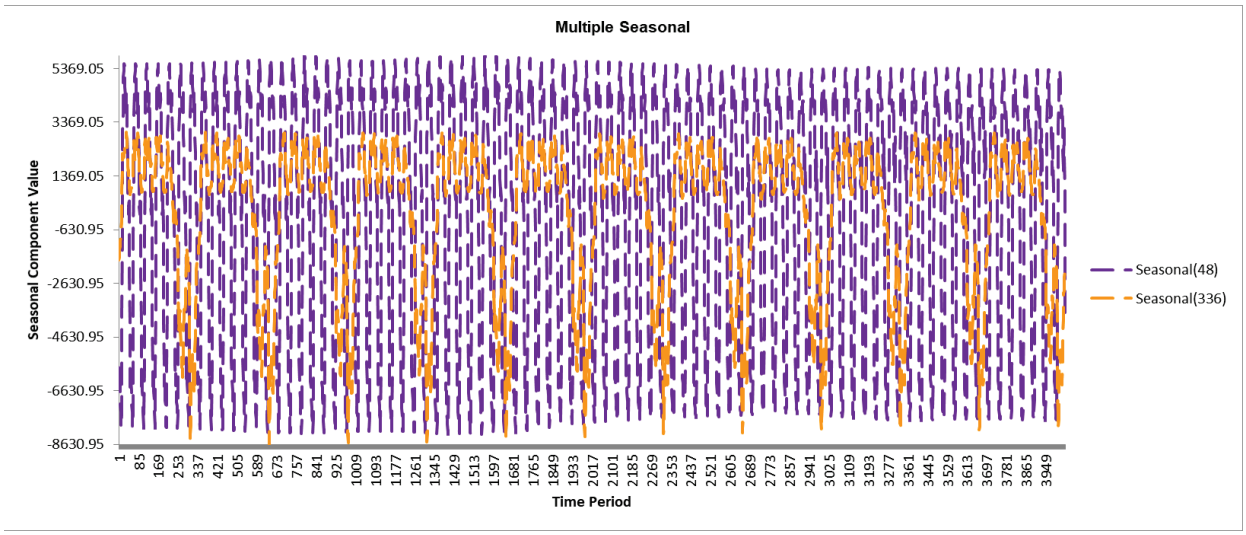
25. Hacer click en **OK**. Se genera el gráfico de la Descomposición de la Tendencia Estacional para la Demanda.

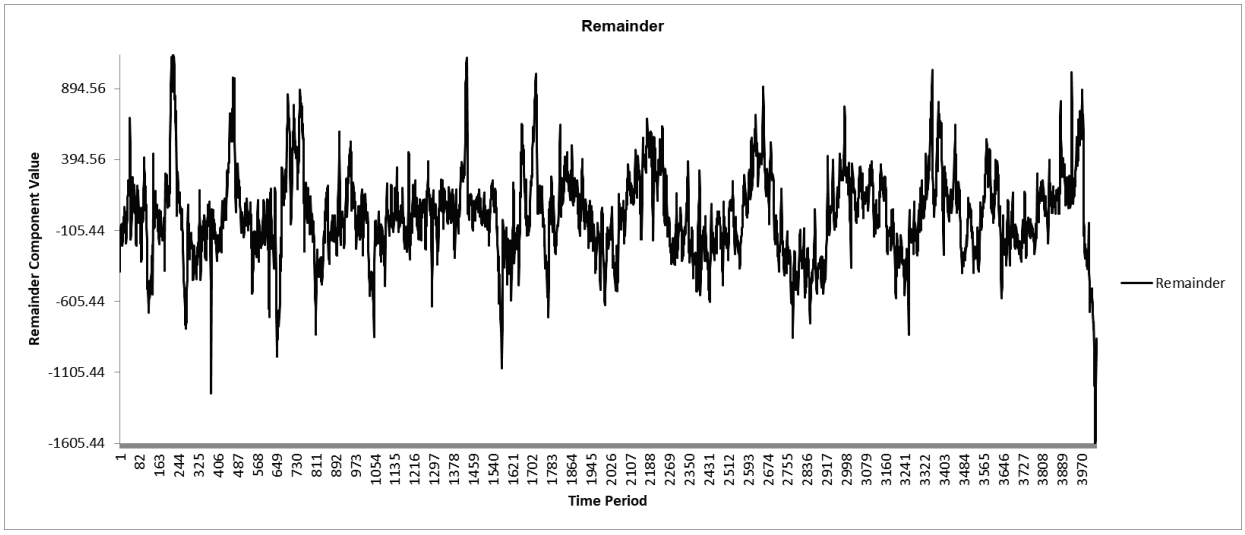


La detección automática de la frecuencia estacional da 48, 336, tal como se obtuvo anteriormente con el Diagrama de Densidad Espectral. Se aprecia aquí el efecto estacional múltiple cada media hora con 48 observaciones diarias y 336 observaciones por semana.

26. Se generan los gráficos de la Tendencia Suavizada, Estacional y los Residuos como se muestra:






Remainder
Remainder Component Value
894.56
394.56
-105.44
-605.44
-1105.44
-1605.44
1
82
163
244
325
406
487
568
649
730
811
892
973
1054
1135
1216
1297
1378
1459
1540
1621
1702
1783
1864
1945
2026
2107
2188
2269
2350
2431
2512
2593
2674
2755
2836
2917
2998
3079
3160
3241
3322
3403
3484
3565
3646
3727
3808
3889
3970
Time Period
Remainder

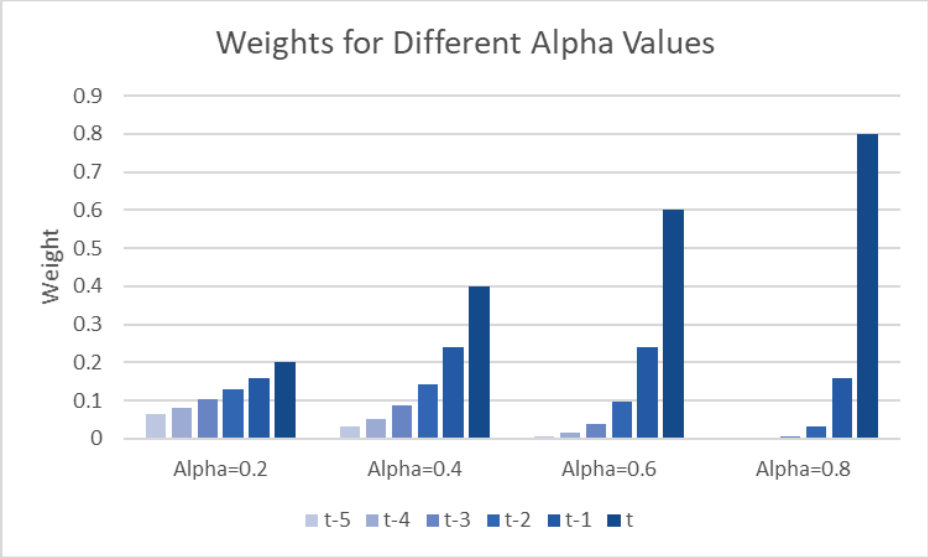
Pronóstico de Suavización Exponencial

Suavización Exponencial Simple

Los pronósticos de Suavización Exponencial Simple se calculan utilizando promedios ponderados, donde los pesos disminuyen exponencialmente en la medida que las observaciones vienen de más lejos en el pasado con los pesos más pequeños asociados a las observaciones más antiguas:

$$\hat{y}_{t+1} = \alpha\ y_t + \alpha(1-\alpha)\ y_{t-1} + \alpha(1-\alpha)^2\ y_{t-2} + \cdots$$

de donde $0 \leq \alpha \leq 1$ es el parámetro de suavización. También se lo conoce como Promedio Móvil Ponderado Exponencial (por sus siglas en inglés, EWMA). Se muestran los gráficos de los diferentes valores de los pesos:



Una fórmula equivalente para la suavización exponencial simple es:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

Con el valor inicial de pronóstico (nivel inicial) $\hat{y}_1$ a ser estimado y que se denota como el nivel l_0.

El parámetro de suavización α y el nivel inicial l_0 se determinan al minimizar la suma de los errores del pronóstico al cuadrado (residuos):

$$\text{SSE} = \sum_{t=1}^{T}\ (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T}\ e_t^2.$$

Modelos de Error, Tendencia, Estacional (ETS)

Los modelos de error, tendencia, estacional (ETS) se expanden en una simple suavización exponencial para acomodar los componentes de tendencia y estacionales, así como los errores aditivos o multiplicativos. La Suavización Exponencial Simple es un modelo de error. El modelo de error, de tendencia, es el lineal de Holt, también conocido como suavización exponencial doble. El Modelo de Error, de Tendencia Estacional, es el Holt-Winters, también conocido como suavización exponencial triple.

Rob Hyndman ha desarrollado una taxonomía completa que describe todas las combinaciones de los modelos de suavización exponencial de forma coherente (véase fpp2):

- Error:
  - Aditivo o Multiplicativo. El error multiplicativo es un error relativo: $e_t = \frac{y_t - \hat{y}_t}{\hat{y}_t}$
  - Los pronósticos puntuales desarrollados por los modelos son idénticos si utilizan los mismos valores de los parámetros de la suavización. Sin embargo, el multiplicativo generará diferentes intervalos de pronósticos para acomodar el cambio en la varianza.
  - Una alternativa al multiplicativo es el uso de la transformación Ln (transformación Box-Cox con Lambda = 0).
  - Los modelos de error incluyen el parámetro de suavización α y el valor del nivel inicial.

- Tendencia:
  - Ninguna, Aditiva, Aditiva Amortiguada
  - No se recomienda la tendencia multiplicativa ya que tiende a producir malos pronósticos, por lo que no se incluye en SigmaXL
  - Los modelos de tendencia agregan un parámetro de suavización β y un valor de tendencia inicial.
  - Los modelos de tendencia amortiguados agregan un parámetro de suavización φ que "amortigua" la tendencia a una línea plana en algún momento del futuro.

- Estacional:
  - Ninguno, Aditivo, Multiplicativo
    - Se prefiere el método aditivo cuando las variaciones estacionales son aproximadamente constantes a lo largo de la serie, en cambio es preferible el método multiplicativo cuando las variaciones estacionales cambian de manera proporcional con respecto al nivel de la serie.

- Los modelos estacionales agregan un parámetro de suavización *γ* y valores estacionales iniciales.
  - La máxima frecuencia estacional es 24. Para frecuencias más altas opte por la Suavización Exponencial - Descomposición Estacional Múltiple (DEM)
  - # de valores iniciales estimados = frecuencia estacional – 1
  - limitado a la suma a 0 para la frecuencia aditiva o estacional *m* para la multiplicativa

Para mayores detalles, véase en el Apéndice: **Suavización Exponencial - ETS.**

Resumen de los Modelos ETS en SigmaXL:

Abreviado (Error, Tendencia, Estacional)	Método
(A, N, N)	Suavización Exponencial Simple con Errores Aditivos – Promedio Móvil Ponderado Exponencial (EWMA)
(M, N, N)	Suavización Exponencial Simple con Errores Multiplicativos
(A, A, N)	Método de Tendencia Aditiva con Errores Aditivos (Holt Lineal)
(M, A, N)	Método de Tendencia Aditiva con Errores Multiplicativos (Holt Lineal)
(A, A, A)	Método de Tendencia Aditiva, Estacional Aditivo con Errores Aditivos (Holt-Winters)
(M, A, A)	Método de Tendencia Aditiva, Estacional Aditivo con Errores Multiplicativos (Holt-Winters)
(A, N, A)	Método Estacional Aditivo con Errores Aditivos
(M, N, A)	Método Estacional Aditivo con Errores Multiplicativos
(A, Ad, N)	Método de Tendencia Atenuada con Errores Aditivos
(M, Ad, N)	Método de Tendencia Atenuada con Errores Multiplicativos
(A, Ad, A)	Método de Tendencia Atenuada Aditiva Estacional Aditiva con Errores Aditivos
(M, Ad, A)	Método de Tendencia Atenuada Aditiva Estacional Aditiva con Errores Multiplicativos
(M, A, M)	Método de Tendencia Aditiva Estacional Multiplicativo con Errores Multiplicativos (Holt-Winters)
(M, N, M)	Método Estacional Multiplicativo con Errores Multiplicativos

(M, Ad, M)	Additive Damped Trend, Multiplicative Seasonal Method with Multiplicative Errors (Holt-Winters)

Estimación de Parámetros en la Suavización Exponencial, Estadísticas del Modelo y Criterio de Información para la Comparación del Modelo

Los parámetros del modelo se resuelven mediante la maximización no lineal de la función Log-Verosimilitud. La log-verosimilitud está relacionada con - Ln (Error de suma de cuadrados). Los criterios de información AICc, AIC y BIC se calculan con el empleo de -2*Log-Verosimilitud e incorporan una penalización por el número de términos en el modelo, por lo que mientras más pequeño, mejor. Se utilizan en la selección automática del modelo. AICc es el Criterio de Información por defecto, se basa en el performance del error de pronóstico de los datos con los que se confronta. Para mayor información, véase el Apéndice: **Criterio de Información para la Comparación del Modelo**.

Valores Ausentes

La Suavización Exponencial maneja los valores ausentes con una interpolación lineal ajustada estacionalmente. Si bien existe una robustez frente a algunos valores ausentes, si el número de valores ausentes es grande, entonces la estimación del modelo y la precisión del pronóstico se degradarán. Al seleccionar una serie de tiempo con valores ausentes, aparecerá una ventana emergente "Advertencia": Valores ausentes detectados. Se utilizará la interpolación lineal ajustada estacionalmente". Véase el Apéndice: **Interpolación Lineal Ajustada Estacionalmente para Valores Ausentes**.

Métricas de Precisión del Pronóstico

SigmaXL hace uso de las siguientes métricas de precisión del pronóstico:

Raíz cuadrada del error cuadrático medio: las siglas en inglés, RMSE $= \sqrt{\text{media}(e_t^2)}$

Error medio absoluto: MAE $= \text{media}(|e_t|)$

Error porcentual medio absoluto: MAPE $= \text{media}\left(\left|\frac{100e_t}{y_t}\right|\right)$

Error en escala medio absoluto: MASE $= \text{media}(|e_t|)/escala$

donde la *escala* es el MAE del pronóstico naïve de la muestra o del pronóstico estacional naïve (establezca todos los pronósticos en el valor de la última observación/período). Observe que la escala para el no estacional es idéntica a la media del rango móvil de la carta de Rango Móvil de Individuales. Un error de escala es menor que uno si surge de un pronóstico mejor que el promedio de los pronósticos naïve o estacionales naïve. Por el contrario, es mayor que uno si el pronóstico es peor que el pronóstico naïve promedio. Se recomienda el MASE en lugar del popular MAPE, porque el MAPE se convierte en infinito para cualquier y_t=0. Para mayores detalles, véase el Apéndice: **Precisión del Pronóstico**

Los datos dentro de la muestra se refieren a los datos del "Entrenamiento". La misma medida se puede aplicar a los datos Fuera de la Muestra (Retención), también conocidos como los datos de "Prueba" y pueden ser de un paso hacia adelante o de múltiples pasos hacia adelante. Esto se demostrará más adelante. Los datos de fuera de la muestra (retención) no se utilizan en la estimación de los parámetros del modelo, por lo que son un mejor indicador de la verdadera precisión del pronóstico. Las medidas de precisión dentro de la muestra pueden estar sesgadas debido al exceso de ajuste. La escala siempre se calcula utilizando los datos dentro de la muestra.

Demo de la Suavización Exponencial Simple– Concentración

1. Abrir el **Demo of Simple Suavización exponencial – Concentration.xlsx** (pestaña de la **Hoja 1**). Estos son datos de la Serie A de Box y Jenkins, un conjunto de 197 valores de la concentración de un proceso químico tomado a intervalos de dos horas en la Columna B.

	A	B	C	D
1	Obs. No.	Concentration (y)	One-step-ahead forecast (y-hat)	Residuals (e)
2	1	17	17	0
3	2	16.6	17	-0.4
4	3	16.3	16.96	-0.66
5	4	16.1	16.894	-0.794
6	5	17.1	16.8146	0.2854
7	6	16.9	16.84314	0.05686
8	7	16.8	16.848826	-0.048826
9	8	17.4	16.8439434	0.5560566
10	9	17.1	16.89954906	0.20045094
11	10	17	16.91959415	0.080405846

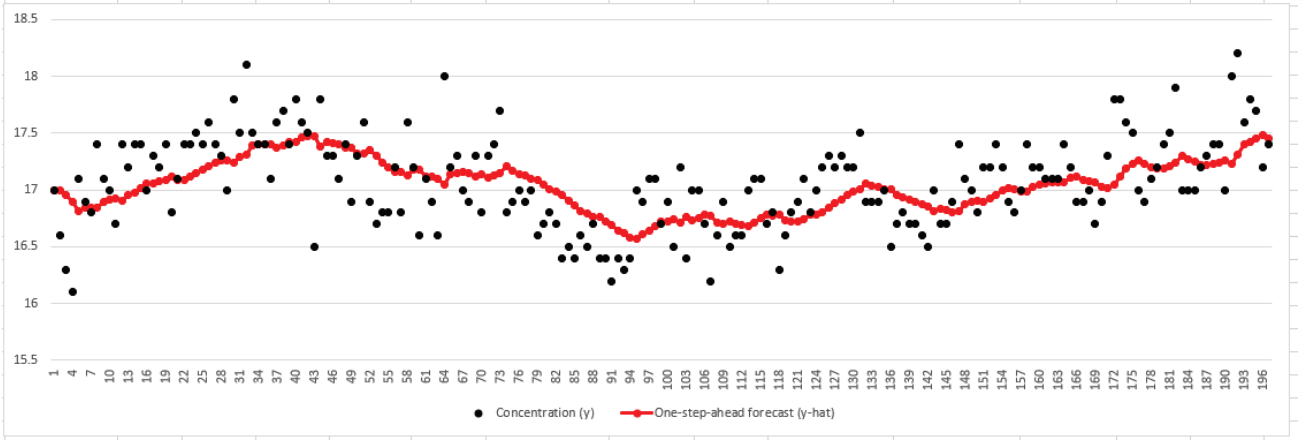
Exponential Smoothing Model Parameters	
alpha	0.1
initial level	17

Use Solver to obtain parameter values that minimize SSE:

SSE	21.8341

Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-303.72228
AICc	613.56892
AIC	613.44457
BIC	623.29418

In-Sample Forecast Accuracy Metrics	
RMSE	0.332916231
MAE	0.271930445
MAPE	1.592107249
Scale (MR-bar)	0.275510204
MASE	0.9870068



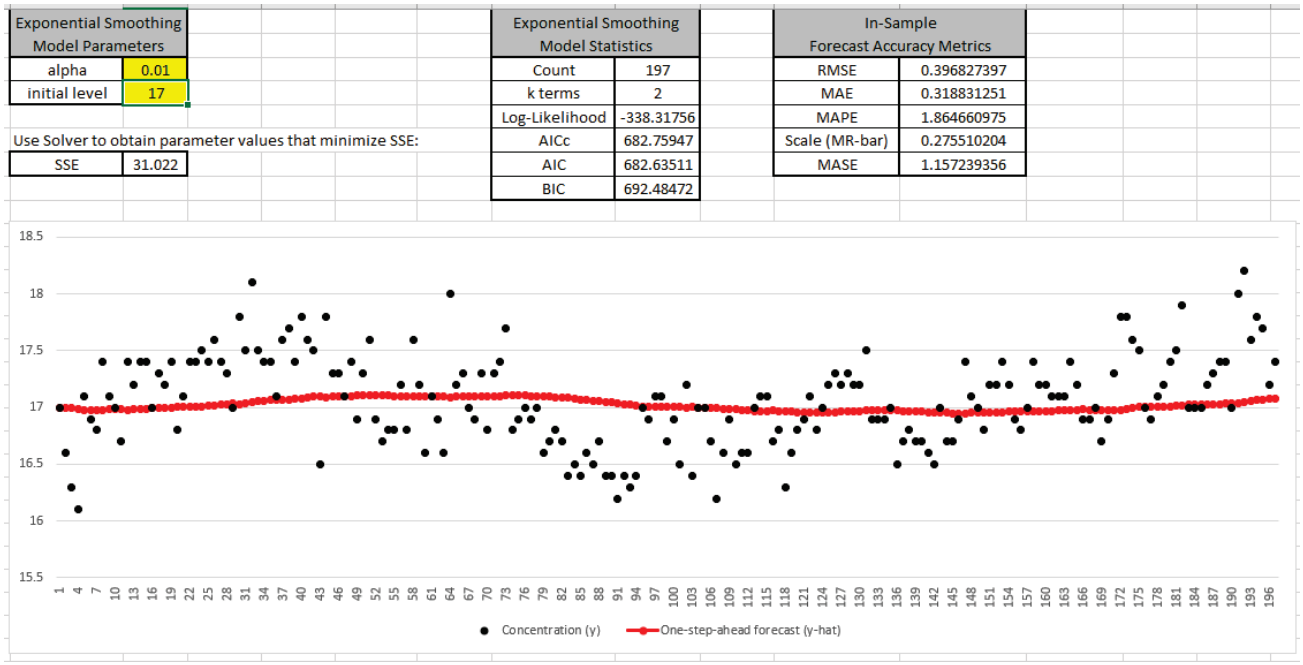
En la celda **G3** está el valor del parámetro alfa de suavización, en la celda **G4** está el nivel inicial. Se los puede Registrar manualmente o se lo resuelve con el Solver. El pronóstico de un-paso-hacia-adelante se calcula a continuación: B2 es el valor del nivel inicial; B3 =alfa*A2+(1-alfa) *B2; B4 =alfa*A3+(1-alfa) *B3, etc.

La celda **G7** es la Suma Cuadrática de los errores Residuos (SSE) y es el valor a minimizar, con

el propósito de generar el pronóstico de un-paso-hacia-adelante más preciso. Se proporcionan las estadísticas del modelo de suavización exponencial para su posterior comparación con el informe de SigmaXL.

Los datos de la concentración sin procesar son los puntos negros del gráfico; los valores del pronóstico de un paso hacia adelante son los puntos rojos. Las estadísticas del modelo y las medidas de precisión de los pronósticos se discutirán más adelante.

2. Antes de usar el Solver para determinar los valores óptimos alfa y de nivel inicial, se registra manualmente algunos valores para ver cómo cambia el gráfico. Registrar alfa = 0.01 como se muestra:



Exponential Smoothing Model Parameters	
alpha	0.01
initial level	17

Use Solver to obtain parameter values that minimize SSE:

SSE	31.022

Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-338.31756
AICc	682.75947
AIC	682.63511
BIC	692.48472

In-Sample Forecast Accuracy Metrics	
RMSE	0.396827397
MAE	0.318831251
MAPE	1.864660975
Scale (MR-bar)	0.275510204
MASE	1.157239356

Este pequeño valor alfa produce un suave ajuste con un SSE = 31.02 que es mayor que aquel con el que se inició (SSE = 21.83).

3. Ingrese ahora alfa = 0.99 como se muestra:

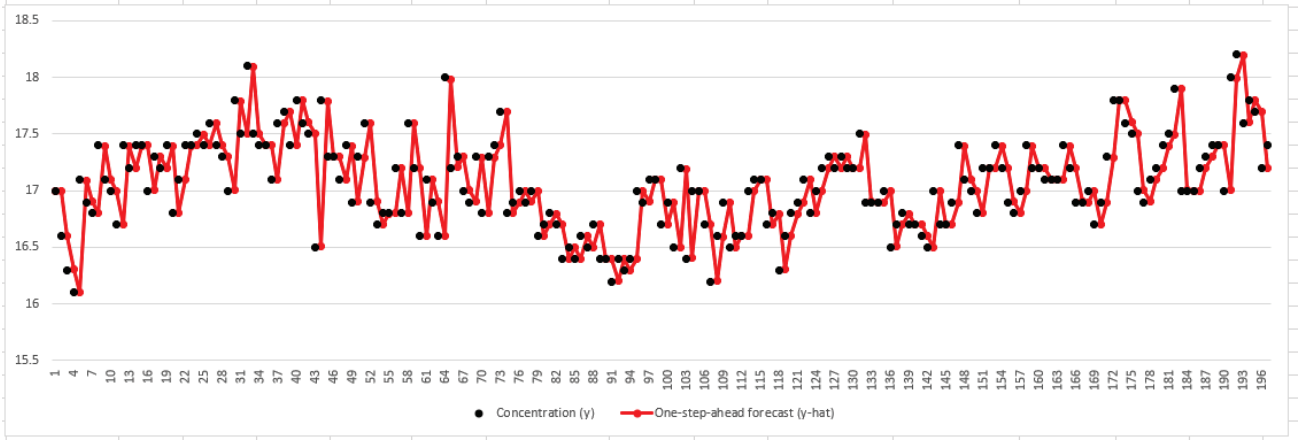
Exponential Smoothing Model Parameters	
alpha	0.99
initial level	17

Use Solver to obtain parameter values that minimize SSE:

SSE	26.5219

Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-322.88029
AICc	651.88493
AIC	651.76058
BIC	661.61019

In-Sample Forecast Accuracy Metrics	
RMSE	0.366918467
MAE	0.273380342
MAPE	1.598165452
Scale (MR-bar)	0.275510204
MASE	0.992269389



Este valor grande de alfa resulta aproximadamente en un pronóstico naïve, con el valor del pronóstico un paso hacia adelante igual al valor actual previo.

4. Registrar alfa = 0.1 y el nivel inicial = 20 como se muestra:

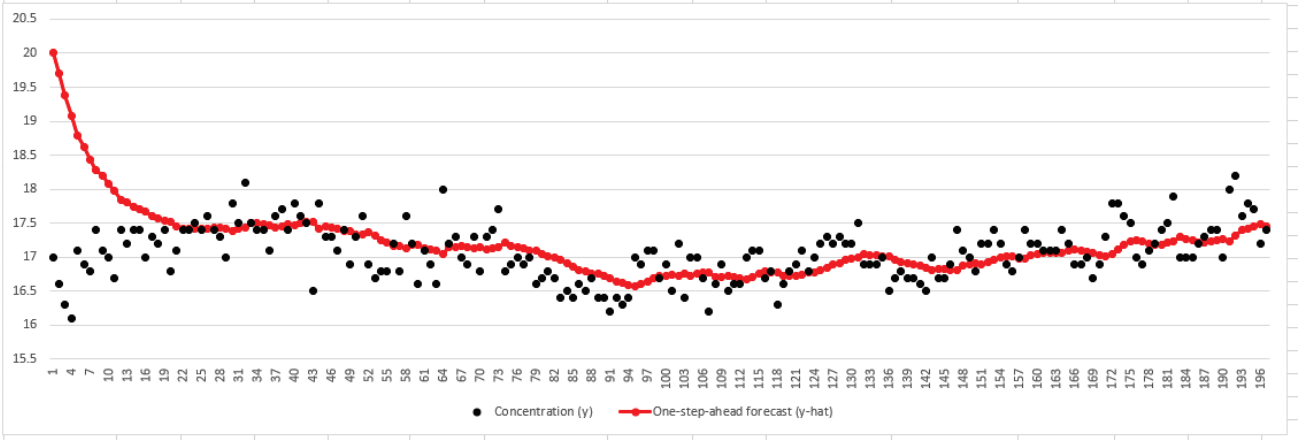
Exponential Smoothing Model Parameters	
alpha	0.1
initial level	20

Use Solver to obtain parameter values that minimize SSE:

SSE	71.0199

Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-419.90158
AICc	845.92751
AIC	845.80316
BIC	855.65277

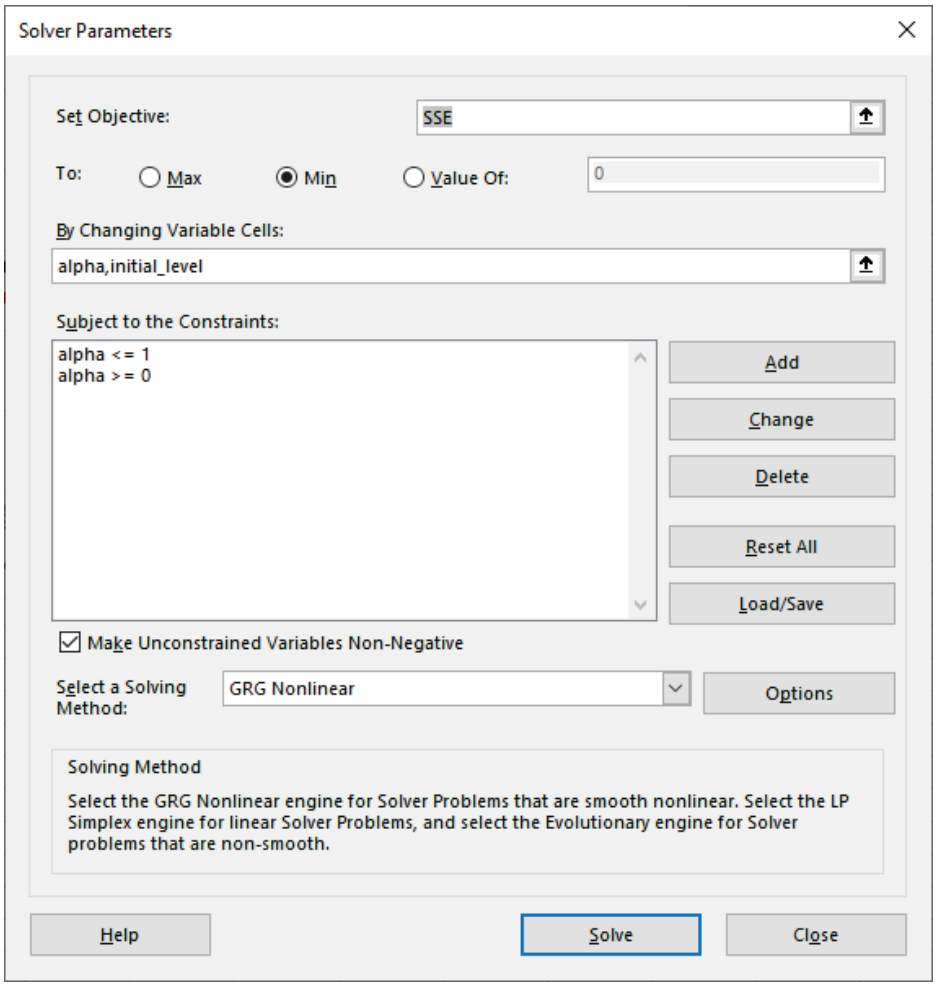
In-Sample Forecast Accuracy Metrics	
RMSE	0.60042246
MAE	0.365989697
MAPE	2.156801395
Scale (MR-bar)	0.275510204
MASE	1.32840705



Este nivel inicial es una mala estimación del valor inicial y resulta en un SSE = 71.02. Después de unas 30 observaciones, la influencia de este pobre valor inicial es insignificante. La duración para la cual el nivel inicial tiene influencia en el pronóstico de un paso hacia adelante depende del alfa. Un valor alfa más pequeño sería de una duración más larga, un alfa más grande sería

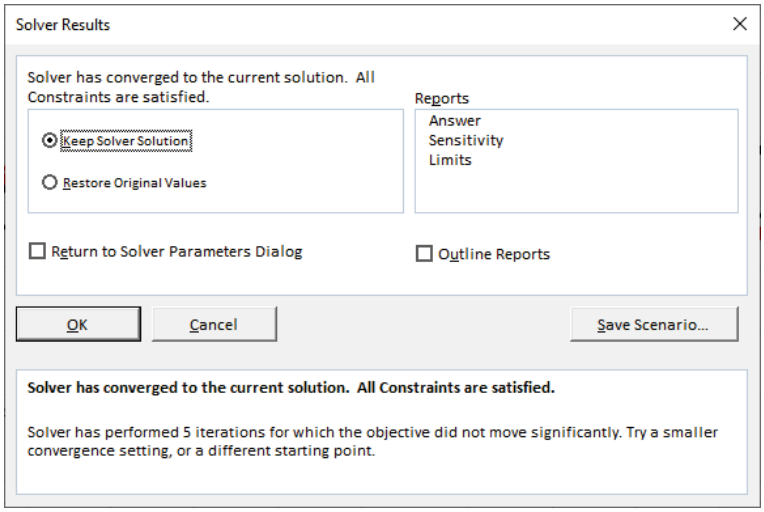
de una duración más corta. Este valor inicial es una pobre estimación del valor inicial y resulta en un SSE = 71.02.

5. Se empleará ahora el Solver de Excel para encontrar los valores óptimos alfa y el nivel inicial. Hacer click en **Data > Solver**. (Si el menú del Solver no está disponible, hacer click en **File > Options > Add-Ins**. Hacer click en **Go**… Gestione Excel Add-ins, marcar **Solver Add-in**. Hacer click en **OK**. Hacer click en **Data > Solver**.)

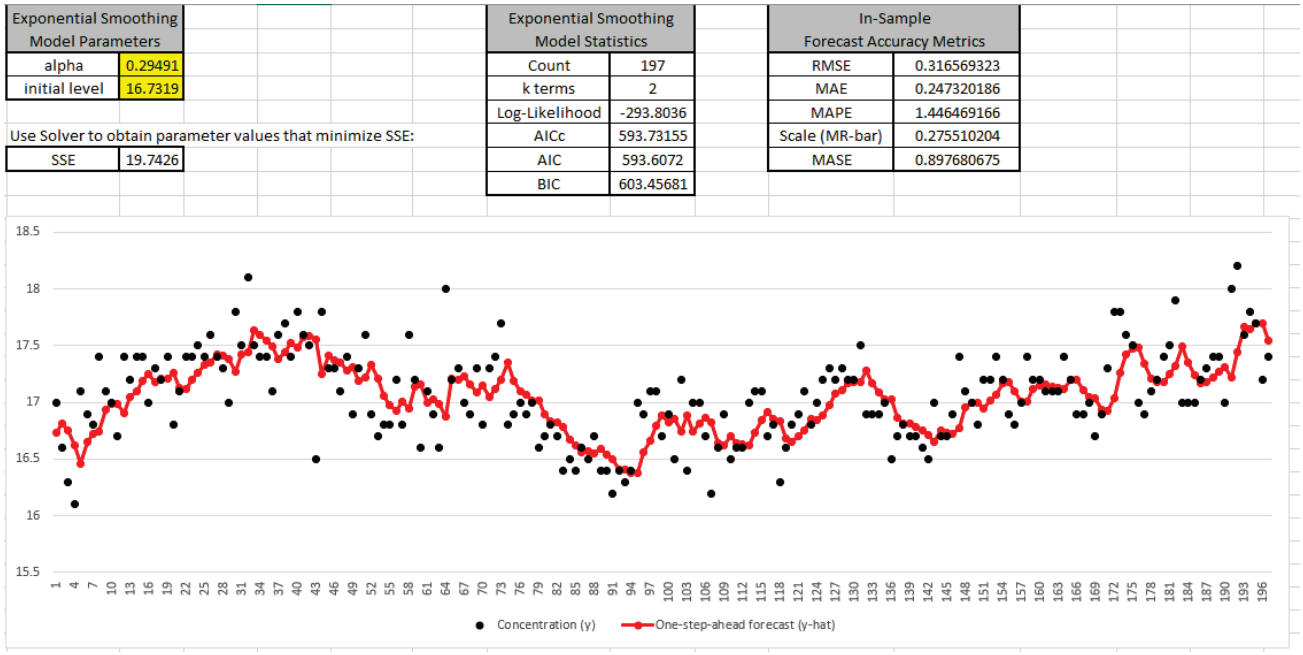


Las localizaciones de las celdas, las restricciones y configuraciones del Solver se han almacenado en el libro de ejercicios, de manera que no se requieren cambios. El Solver minimizará la celda **G7** (SSE), al variar las celdas **G3** (alfa) y **G4** (nivel inicial). El método de solución es GRG Nonlinear; se puede usar el método Evolutionary, pero es más lento.

6. Hacer click en **Solve**. El diálogo de los resultados del Solver indica que se ha encontrado una solución.



7. Hacer click en **OK**. El SSE es ahora 19.74; el valor alfa = 0.295 y el nivel inicial = 16.732.



Exponential Smoothing Model Parameters	
alpha	0.29491
initial level	16.7319

Use Solver to obtain parameter values that minimize SSE:

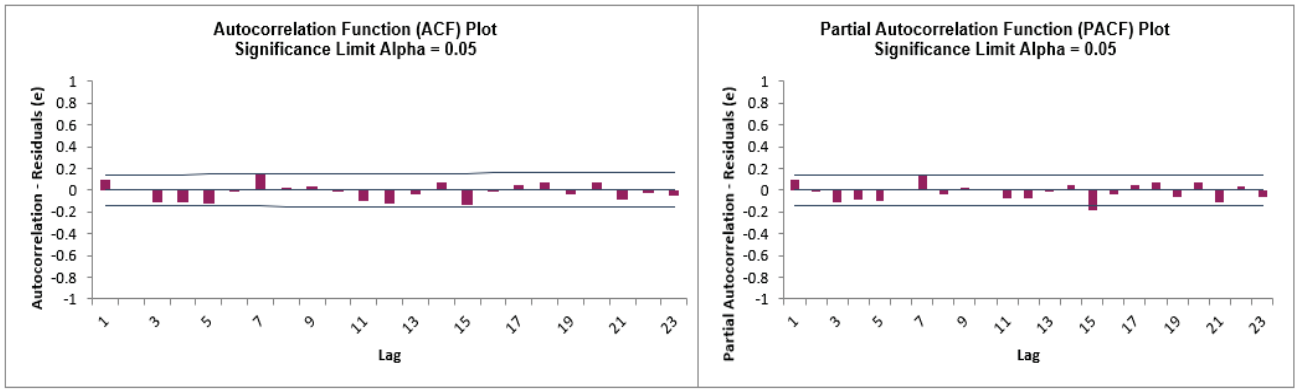
SSE	19.7426

Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-293.8036
AICc	593.73155
AIC	593.6072
BIC	603.45681

In-Sample Forecast Accuracy Metrics	
RMSE	0.316569323
MAE	0.247320186
MAPE	1.446469166
Scale (MR-bar)	0.275510204
MASE	0.897680675

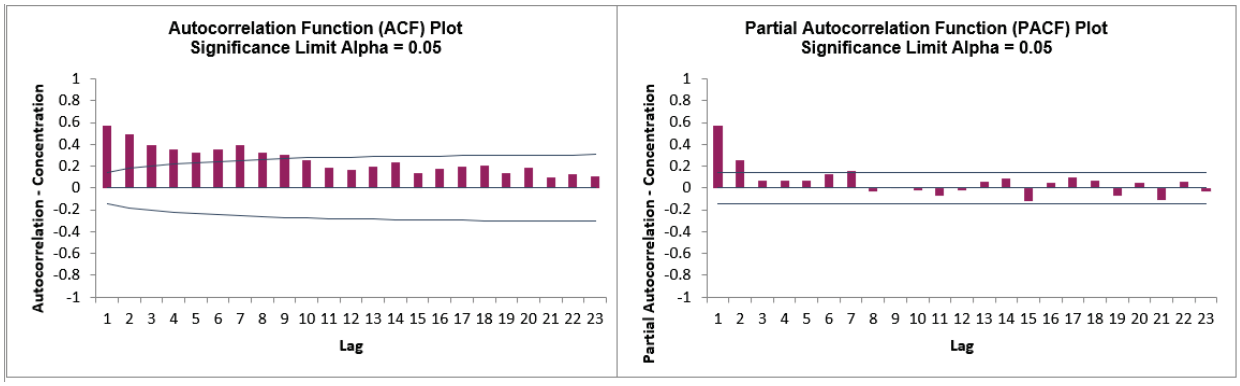
Al minimizar el error cuadrático medio (SSE), se obtiene la mejor estimación de los parámetros para producir los valores del pronóstico un paso hacia adelante.

8. Se generarán ahora los gráficos FAC/FACP para los Residuos. Seleccionar las celdas **D1:D198** y hacer click en **SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots.** Hacer click en **Next**, seleccionar *Residuals (e)*, hacer click en **OK**.

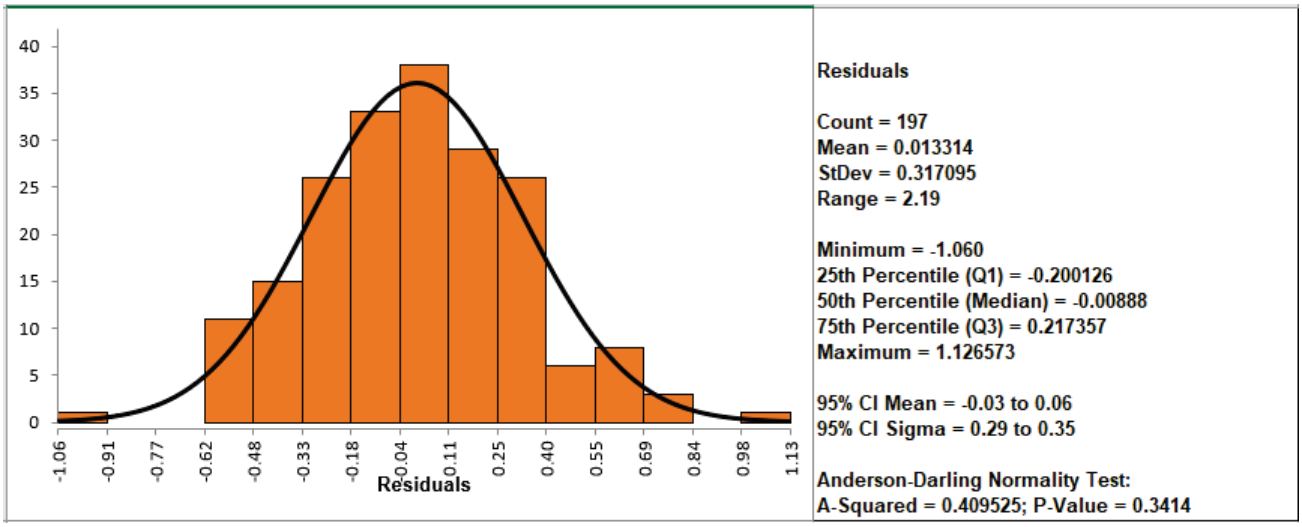


Se puede apreciar que se ha removido toda la autocorrelación por el uso de un modelo de suavización exponencial simple (con la excepción del rezago 15 en la FACP), así que este es un buen ajuste de los datos de series temporales.

9. A modo de comparación, estos son los gráficos de las FAC/FACP para los datos de concentración originales que se han generado.



10. Se procederá ahora a revisar los Residuos utilizando un Histograma. Seleccionar los datos de los Residuos (D1:D198). Hacer click en **SigmaXL > Graphical Tools > Histograms and Descriptive Statistics**. **Next**, seleccionar *Residuals (e)*, hacer click en **OK**.



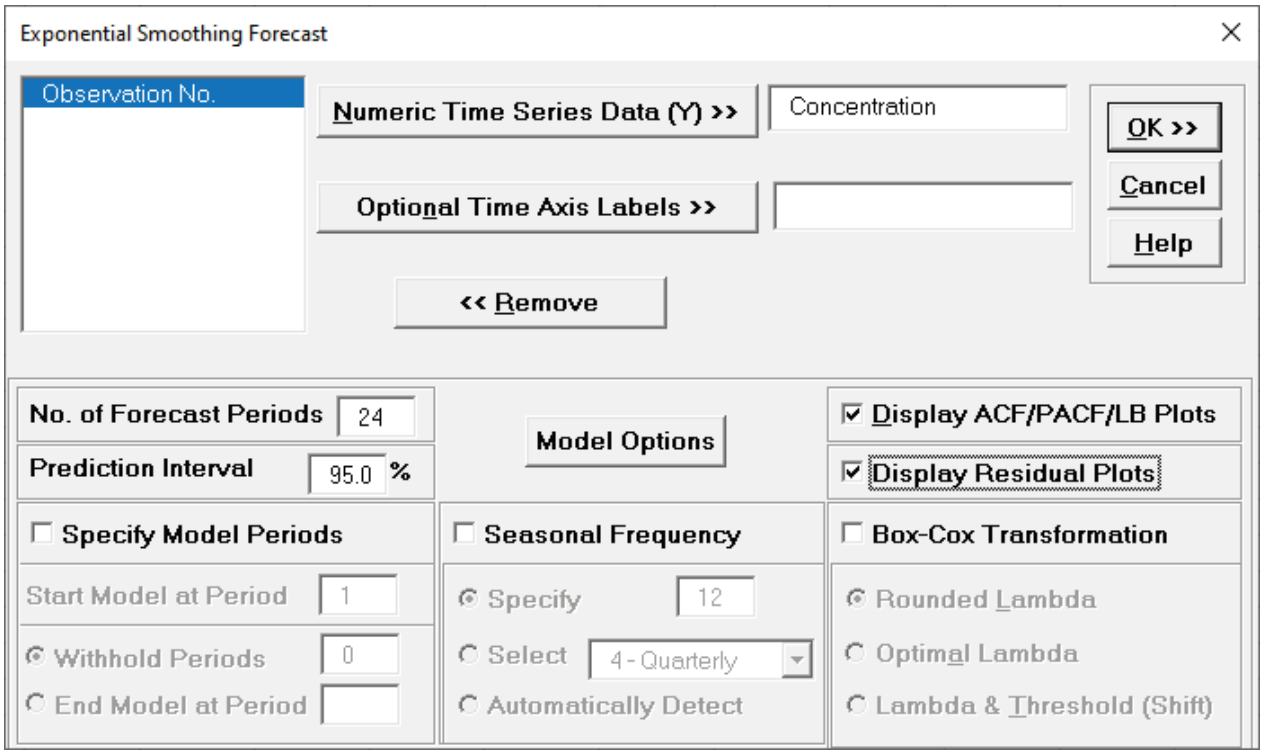
Los residuos están normalmente distribuidos sin valores atípicos extremos.

Concentración de Proceso Químico - Serie A

11. Abrir **Chemical Process Concentration – Series A.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie A de Box y Jenkins, un conjunto de 197 valores de la concentración de un proceso químico tomados a intervalos de 2 horas. Véase para estos datos Carta de Corrida, Gráficos FAC/FACP, Gráficos de Densidad Espectral, Gráficos de Descomposición de Tendencia Estacional.

12. Hacer click en **SigmaXL > Time Series Forecasting > Suavización exponencial Forecast > Forecast.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

13. Seleccionar *Concentration*, hacer click en **Numeric Time Series Data (Y) >>**. Marcar **Display FAC/FACP/LB Plots** y **Display Residual Plots**. No marque **Specify Model Periods, Seasonal Frequency** ni **Box-Cox Transformation.** Se usará por defecto **No. of Forecast Periods** = 24 y **Prediction Interval** = 95.0 %.



- **Optional Time Axis Labels** se mostrará en el eje de tiempo de la carta de pronósticos. Si se utiliza, también se deben incluir las fechas de los períodos del pronóstico, de lo contrario el eje de tiempo quedará en blanco para los períodos del pronóstico.

- **No. of Forecast Periods** son el número de valores de las series temporales que se deben pronosticar (horizonte de pronóstico). El pronóstico más preciso será el primer valor pronosticado (un paso hacia adelante).

- **Prediction Interval %** es el nivel de confianza para las predicciones individuales. Por ejemplo, un intervalo de predicción del 95% contiene un rango de valores que debería incluir el valor futuro real con una probabilidad del 95%. El intervalo se hará más grande cuanto más distante sea la predicción.

- **Model Options** abre otro diálogo que permite fijar las opciones automáticas o especificar un modelo.

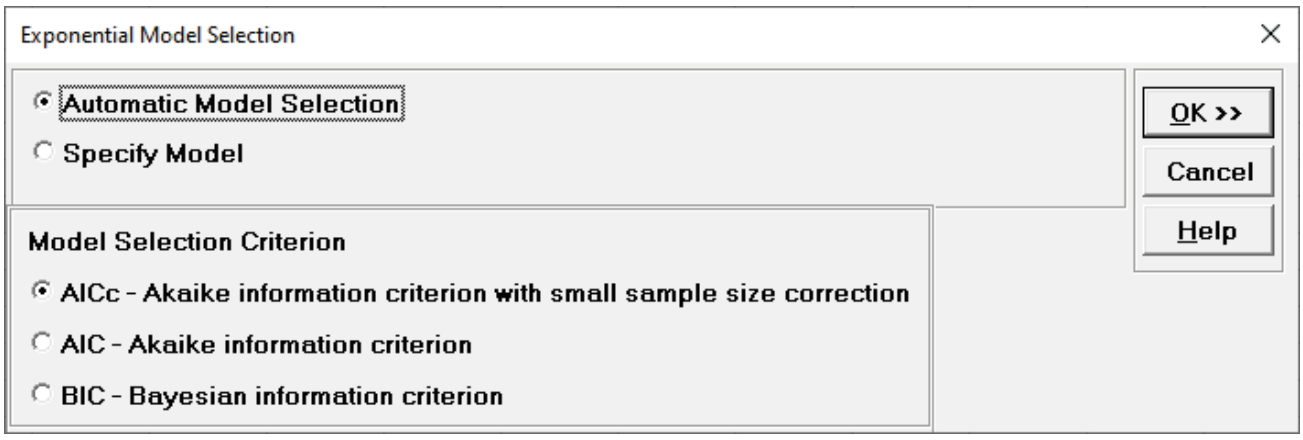
- **Display ACF/PACF/LB option** generará los gráficos de la FAC y FACP para los datos sin procesar, así como para los residuos del modelo. El gráfico LB es una gráfica de los valores p de prueba de Ljung-Box para varios rezagos y se utiliza para determinar si un grupo de autocorrelaciones son significativas, (esto es, las autocorrelaciones no provienen de una serie de ruido blanco). Para mayores detalles, véase **Test Ljung-Box**.

- **Display Residual Plots** generará una tabla de residuos del modelo y las habituales gráficas de residuos del modelo: histograma, gráfica de probabilidad normal, residuos frente al orden de los datos, y valores de los residuos vs aquellos de los del pronóstico. Nótese que si se aplica una transformación de Box-Cox, los residuos que se transforman no serán iguales a los actuales del pronóstico.

- **Specify Model Periods** se utilizan para especificar un período de inicio, un período final o para retener la muestra. Los datos que se retienen no se utilizan en la estimación del modelo, por lo que son muy útiles para la validación y comparación del modelo. Esto se utilizará en un ejemplo posterior.

- **Seasonal Frequency** y **Box-Cox Transformation** se usarán en un posterior ejemplo.

14. Hacer click en **Model Options**.

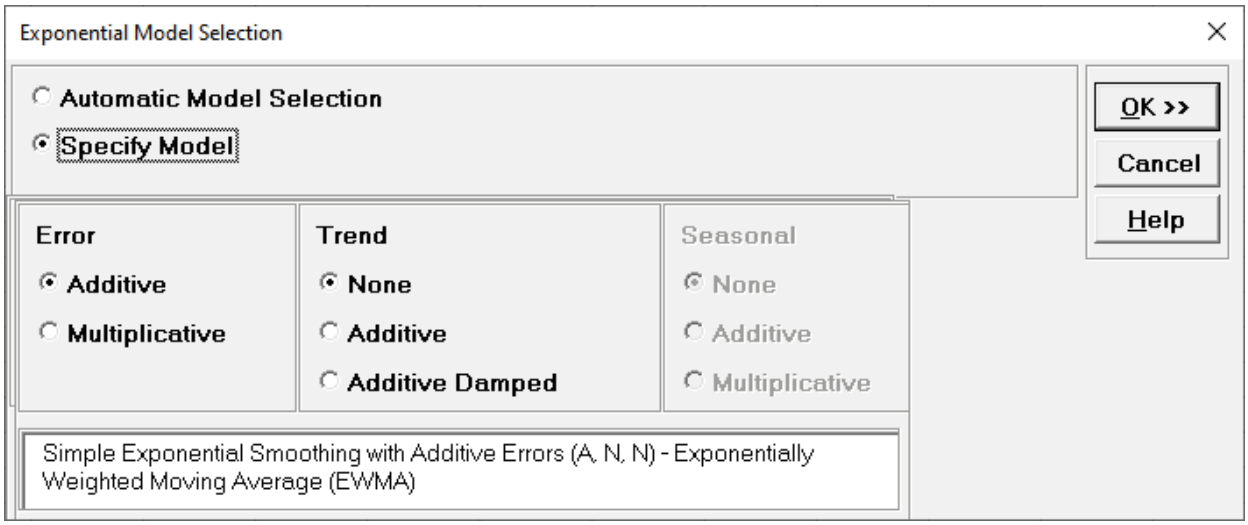


- **Automatic Model Selection** se utilizará posteriormente. Es la Selección por defecto.

- **Model Selection Criterion** es la medida del criterio de información que se utilizará en la Automatic Model Selection. AICc es la Selección por defecto.

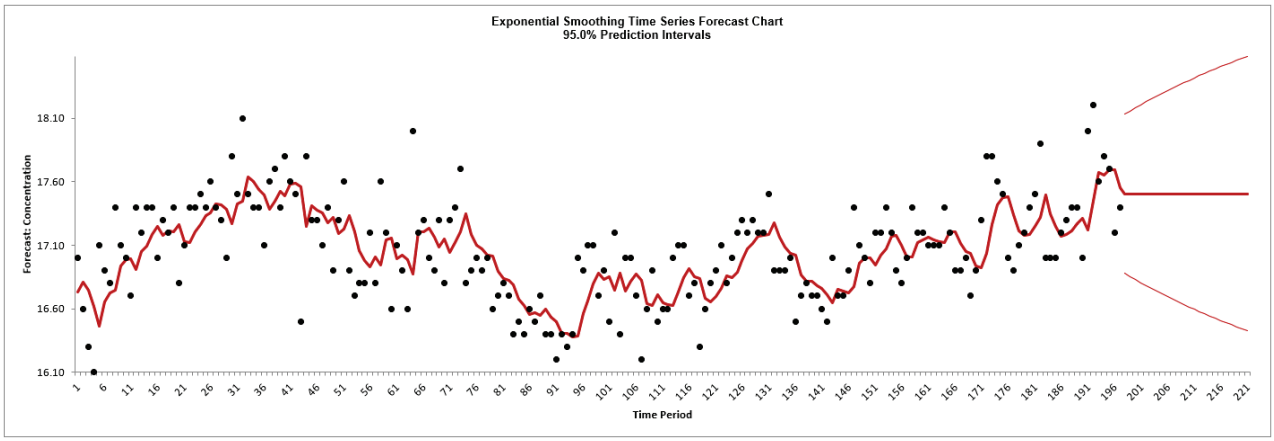
- Al hacer click en **OK** se acepta la configuración y retorna al diálogo previo. Al hacer click en **Cancel** cancelará cualquiera de los cambios y retorna al diálogo previo.

15. Seleccionar **Specify Model**.



- **Specify Model** permite especificar en forma manual el modelo Error-Tendencia-Estacional. Las opciones Estacional no están habilitadas porque la opción **Seasonal Frequency** en el diálogo principal está desmarcada. Se proporciona una descripción del modelo. Se incluirá en el informe del pronóstico.

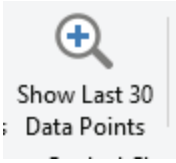
16. Se empleará por defecto **Error: Additive** y **Trend: None**, el cual es un modelo de suavización exponencial simple, mismo que se mostró anteriormente. Hacer click en **OK** para regresar al diálogo del Pronóstico de Suavización Exponencial. Hacer click en **OK**. Se genera el gráfico del pronóstico de suavización exponencial:

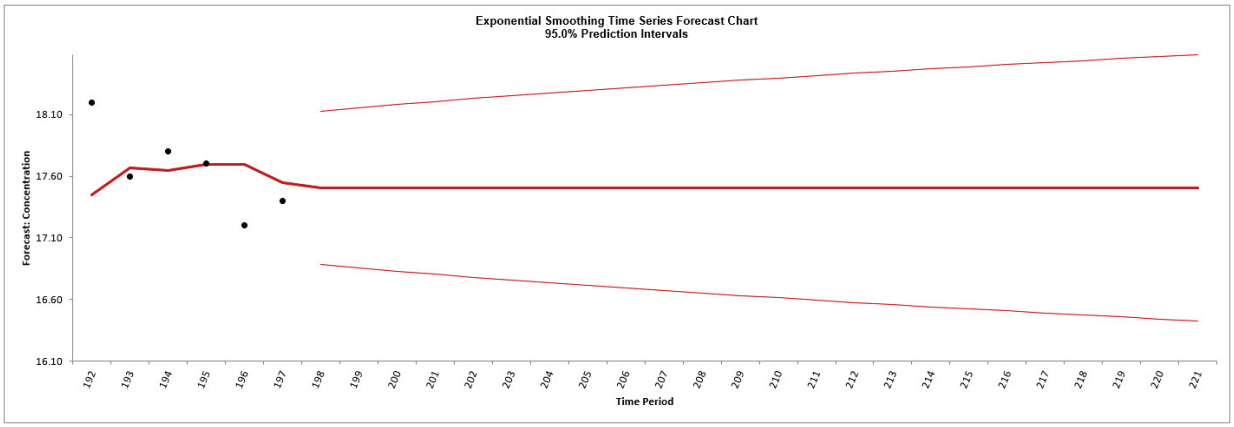


Esta es muy similar a la gráfica exponencial con alisamiento mostrada anteriormente, que muestra los datos de concentración sin procesar (negro) y los valores de pronóstico de un paso

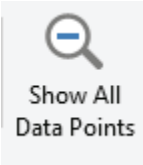
hacia adelante (rojo), pero con la adición de un pronóstico de 24 períodos y el intervalo de predicción del 95%.

17. Puede ampliar la imagen para observar los últimos 30 puntos en la gráfica del Pronóstico. Hacer click en **SigmaXL Chart Tools > Show Last 30 Data Points**.

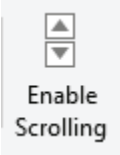




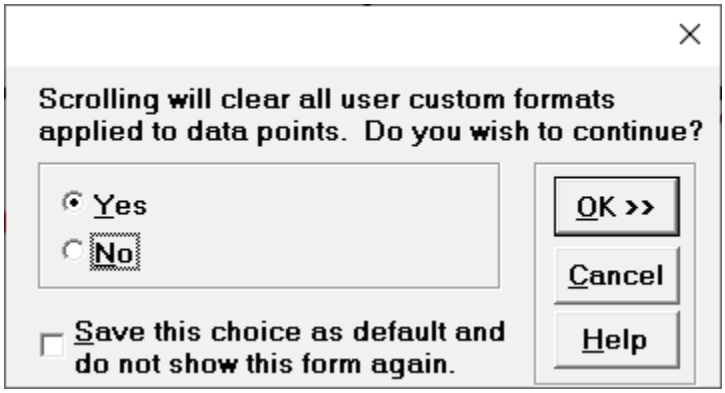
18. Para reestablecer la gráfica, hacer click en **SigmaXL Chart Tools > Show All Data Points**.



19. También puede desplazar los datos. Hacer click en **SigmaXL Chart Tools > Enable Scrolling**

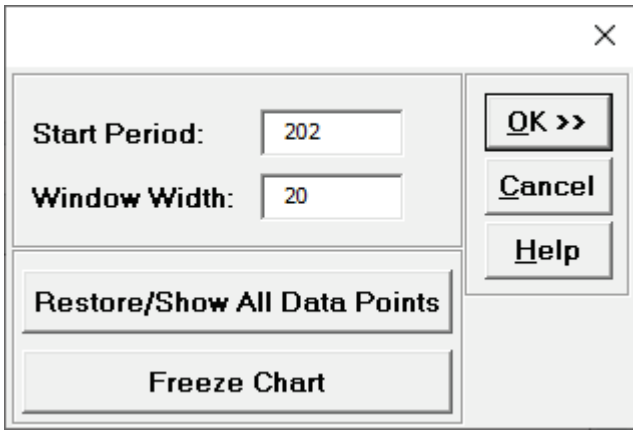


Se abre un mensaje de advertencia sobre la opción para individualizar el formato que va a ser eliminado:



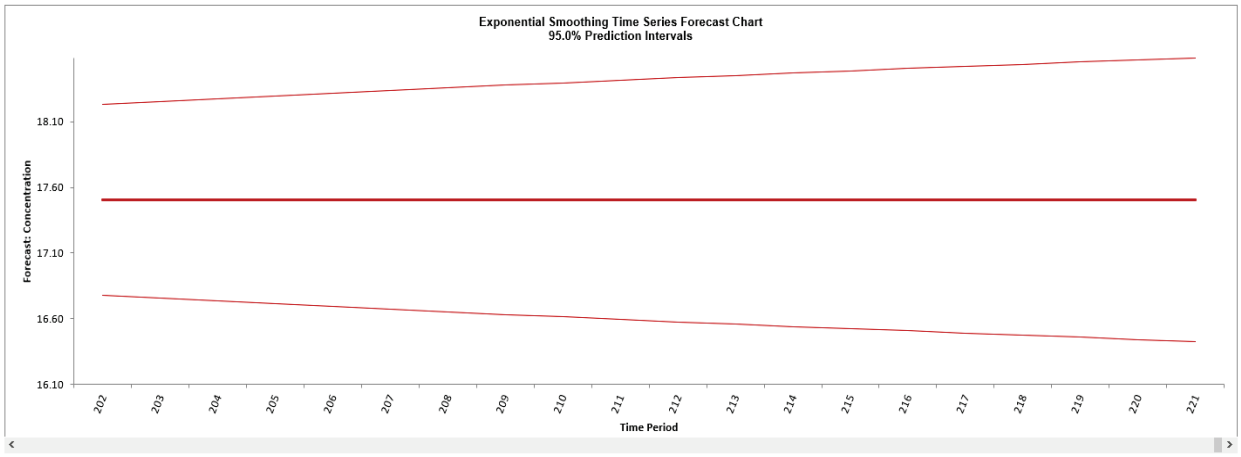
Puede descartar ver esta advertencia al marcar **Save this choice as default and do not show this form again**.

20. Hacer click en **OK.** Aparece el diálogo del desplazamiento que permite especificar el **Start Period** y el **Window Width**:



En cualquier punto, puede hacer click en **Restore/Show All Data Points** o **Freeze Chart**. Al congelar la gráfica va a remover el desplazamiento y descargar el diálogo. El diálogo del desplazamiento se descargará si cambia la hoja de trabajo. Para reestablecer el diálogo, hacer click en **SigmaXL Chart Tools > Enable Scrolling**.

21. Hacer click en **OK.** Una barra de desplazamiento aparece debajo de la gráfica del pronóstico. Puede cambiar también el **Start Subgroup** y el **Window Width** y el **Update**.



Se puede desplazar al hacer click a la derecha o la izquierda, con el ancho especificado de la ventana de 20. Hacer click con el izquierdo para apreciar la gráfica como se muestra.

22. Hacer click en **Cancel para** abandonar el diálogo de desplazamiento.

23. Desplazarse hacia abajo para visualizar el título del Modelo de Suavización Exponencial:

Exponential Smoothing Forecast: Simple Exponential Smoothing with Additive Errors (A, N, N) - Exponentially Weighted Moving Average (EWMA) - User Specified Model
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

El modelo de **Suavización Exponencial Simple con Errores Aditivos (A, N, N) – Promedio Móvil Ponderado Exponencial (EWMA)** es el modelo especificado por el usuario. Esta es la misma información del modelo que se visualizó en el diálogo de Model Selection.

Si se hubiera desmarcado en Specify Model Periods en el diálogo principal, la Selección del inicio, fin o la retención se hubiera resumido aquí, también.

24. Se genera la Información del Modelo de Suavización Exponencial como:

Exponential Smoothing Model Information	
Seasonal Frequency	1
Model Selection Criterion	Specified
Box-Cox Transformation	N/A
Lambda	
Threshold	

Esto es un resumen de la información del modelo con una Frecuencia Estacional = 1 (no estacional); Model Selection Criterion = “Specified” porque el modelo fue especificado por el usuario; y Box-Cox Transformation = “N/A” porque se desmarcó Box-Cox Transformation.

25. Las Estimaciones de los Parámetros son:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.294785988
l (initial level)	16.73121246

Las estimaciones de los parámetros se asemejan a los valores que se obtuvieron anteriormente con la demostración del Solver:

alpha:	0.29491
initial level:	16.7319
SSE:	4.44326

Las leves diferencias en los resultados de los parámetros se originan debido a las diferencias en el método de optimización.

26. Las Estadísticas del Modelo de Suavización Exponencial son:

Exponential Smoothing Model Statistics	
No. of Observations	197
DF	195
StDev	0.31819
Variance	0.10124
Log-Likelihood	-293.804
AICc	593.732
AIC	593.607
BIC	603.457

Grados de libertad (DF) = n – 2 (términos en el modelo). Véase: Estimación de Parámetros de Suavización Exponential, Estadísticas del Modelo y Criterio de Información para Comparación del Modelo

Los resultados coinciden con aquellos que se muestran en la demostración con el Solver:

Exponential Smoothing Model Statistics	
Count	197
k terms	2
Log-Likelihood	-293.8036
AICc	593.73155
AIC	593.6072
BIC	603.45681

Se pueden ver las ecuaciones al hacer click en las celdas **M5**, **M6**, **M7** o **M8**.

27. Las métricas de la Precisión del Pronóstico en la Muestra son:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	197	
RMSE	0.316569334	
MAE	0.247329038	
MAPE	1.446520183	
MASE	0.897712804	

El MASE es menor que 1, por lo que es un mejor pronóstico que la que se obtendría de un pronóstico naïve (fijar todos los pronósticos para que correspondan los valores de la última observación). Véase Métricas de Precisión del Pronóstico. Los resultados se aproximan a aquellos que se obtuvieron en la demostración con el uso del Solver:

In-Sample Forecast Accuracy Metrics	
RMSE	0.316569323
MAE	0.247320195
MAPE	1.446469239
Scale (MR-bar)	0.275510204
MASE	0.897680708

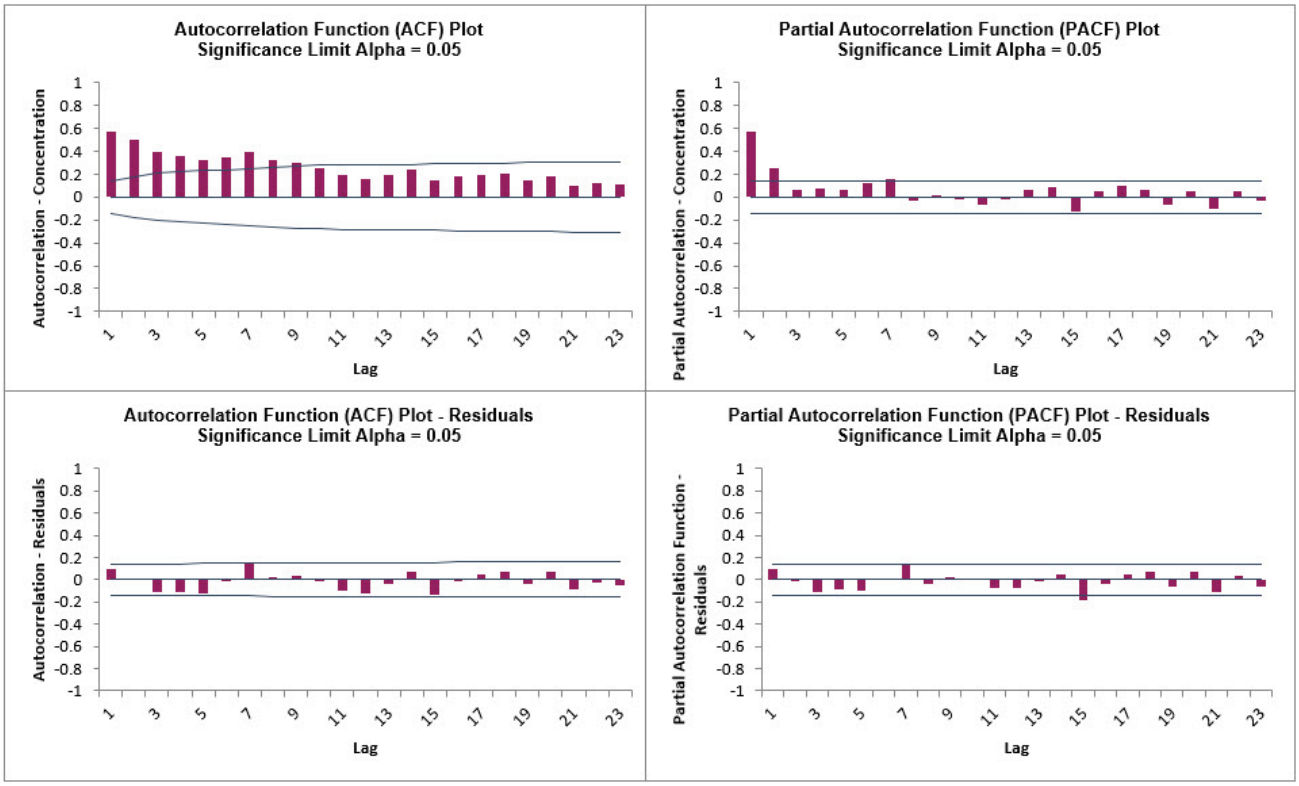
Se pueden ver las ecuaciones al hacer click en las celdas **P3** al **P7** del demo.

28. Se genera la Tabla de Pronóstico:

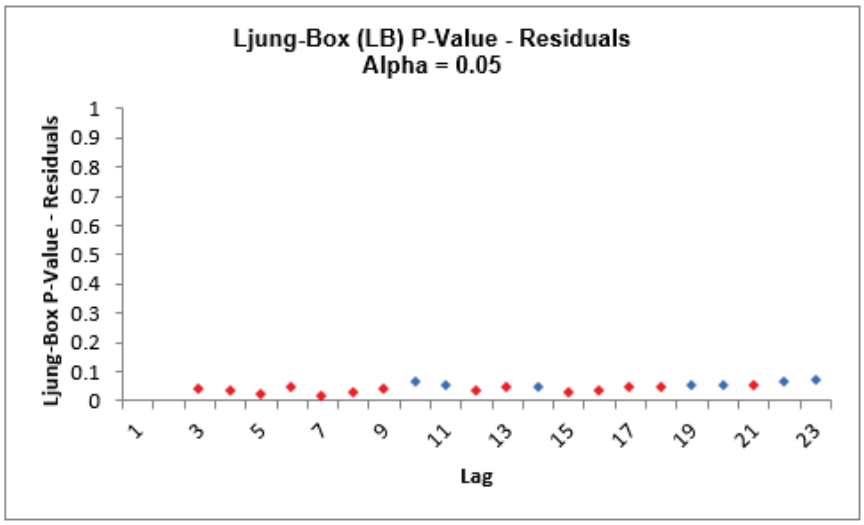
Forecast Table				
Period	Withhold Data	Forecast	Lower 95.0% PI	Upper 95.0% PI
198		17.50540234	16.8817641	18.12904059
199		17.50540234	16.8552318	18.15557289
200		17.50540234	16.82974058	18.1810641
201		17.50540234	16.80517674	18.20562795
202		17.50540234	16.78144587	18.22935881
203		17.50540234	16.75846858	18.25233611
204		17.50540234	16.73617733	18.27462736
205		17.50540234	16.7145141	18.29629058
206		17.50540234	16.69342865	18.31737604
207		17.50540234	16.67287705	18.33792764
208		17.50540234	16.65282071	18.35798398
209		17.50540234	16.63322546	18.37757923

Se trata de los mismos valores del pronóstico e intervalos de previsión que se muestran en los Gráficos de Pronóstico, pero se proporcionan para un análisis o un gráfico más detallado. Si se especifican los períodos de retención, también se muestran los **Datos de Retención**.

29. Hacer click en la hoja **Exp. Smooth. ACF PACF LB** para ver los Gráficos de la FAC/FACP/LB:



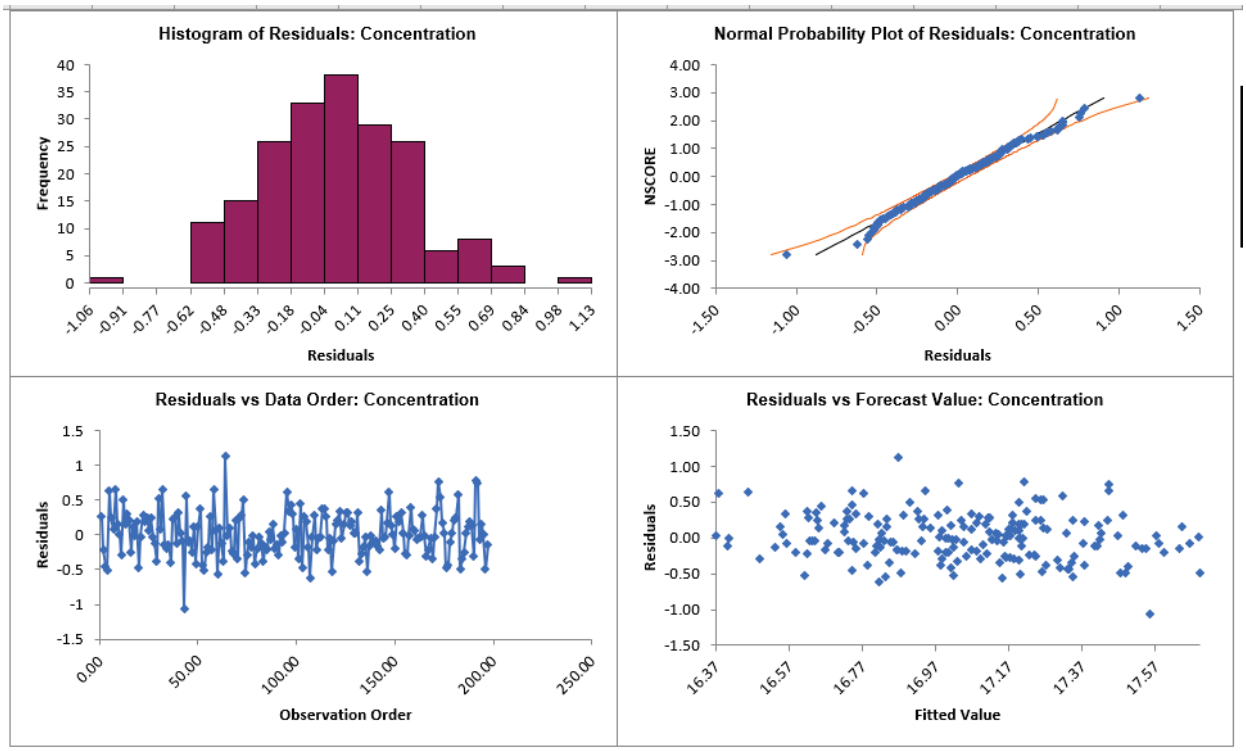
Estas gráficas coinciden con aquellas que se obtuvieron anteriormente de los residuos de las FAC/FACP. Se puede apreciar toda la autocorrelación que el modelo de Suavización Exponencial ha eliminado (con la excepción del rezago 15 en la FACP), por lo que este es un buen ajuste a los datos de la serie temporal.



La gráfica LB es un gráfico de la prueba Ljung-Box de los valores-P para varios rezagos y se emplea para determinar si un grupo de autocorrelaciones son significativas, (esto es, que las autocorrelaciones no provienen de series con ruido blanco). Para mayores detalles, véase **Test Ljung-Box**.

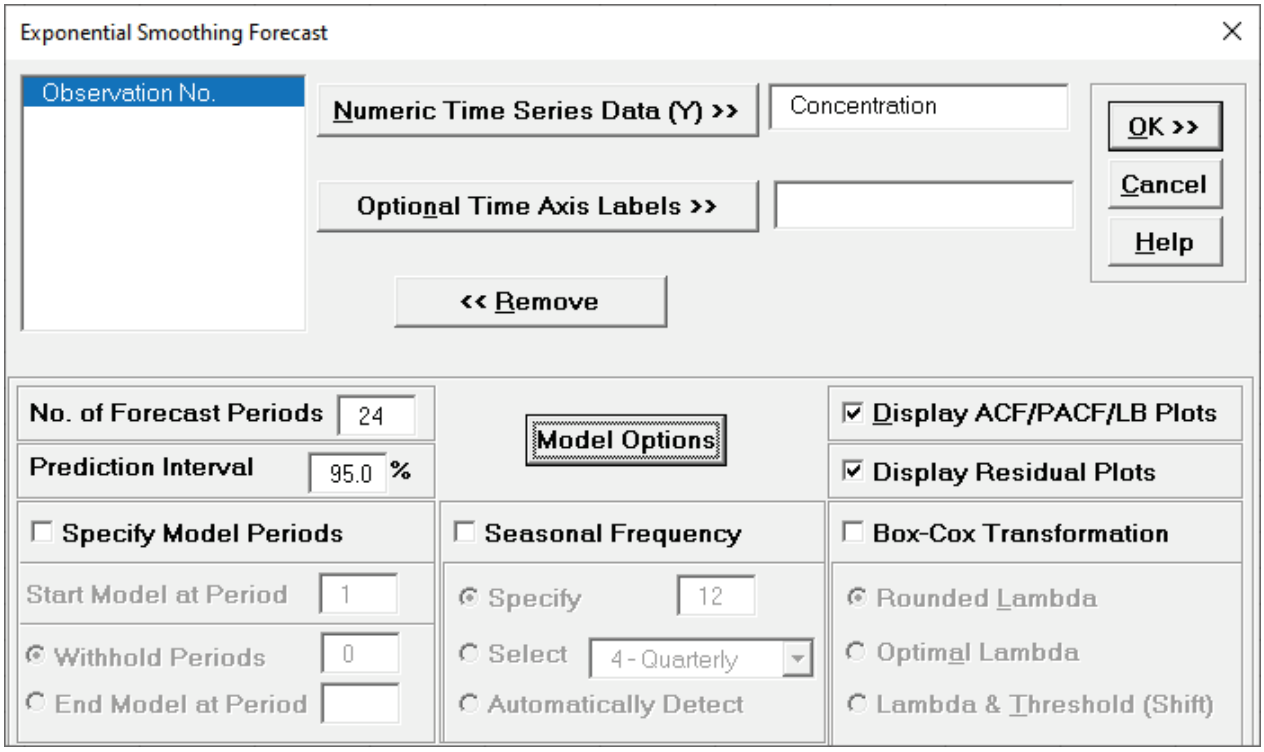
Los valores-P rojos son significativos (alfa=0,05) y los valores-P azules no son significativos. Es deseable que todos los valores-P sean azules. Los gráficos de las FAC/FACP indican que casi toda la correlación ha sido considerada por el modelo, pero el gráfico Ljung-Box muestra que aún queda algo de autocorrelación significativa, por lo que el modelo puede ser potencialmente mejorado. Esto no significa que el modelo sea malo, puede seguir siendo muy útil para fines de predicción, pero los intervalos de predicción pueden no proporcionar una cobertura precisa.

30. Hacer click en la hoja **Exp. Smoothing Residuals** para visualizar los Gráficos de los residuos:

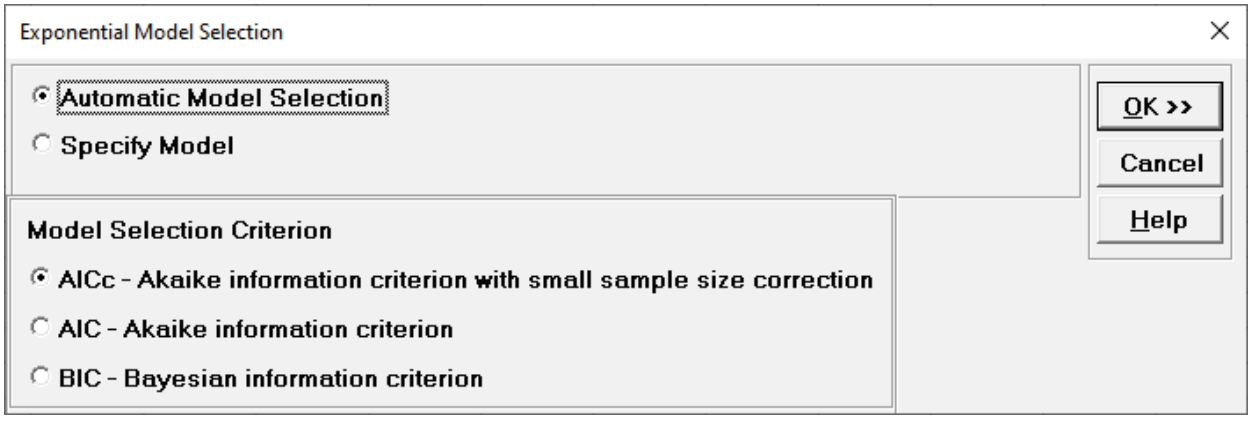


Estos gráficos de los residuos son los mismos que se utilizan en la Regresión Múltiple de SigmaXL. El histograma coincide con el que obtuvimos en la Demo (la curva normal no se aplica aquí). Los residuos están distribuidos de forma aproximadamente normal, con una línea aproximadamente recta en la gráfica de probabilidad normal. No hay valores extremos o patrones obvios en los gráficos. Más adelante, se aplicará un gráfico de control a los residuos para comprobar formalmente la existencia de valores atípicos significativos o causas asignables.

31. Se volverá a correr ahora la Suavización Exponencial con los datos de la Concentración, pero con el uso de Automatic Model Selection. Hacer click en el menú **Recall SigmaXL Dialog** o presione **F3** para repetir el último diálogo.



32. Hacer click en **Model Options**. Seleccionar **Automatic Model Selection**. Se hace uso por defecto de **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**.



33. Hacer click en **OK** para regresar al diálogo de Pronóstico de Suavización Exponencial. Hacer click en **OK**. Se genera el informe del pronóstico de suavización exponencial.

34. Desplazar hacia abajo el encabezado para ver el Modelo de Suavización Exponencial

Exponential Smoothing Forecast: Simple Exponential Smoothing with Multiplicative Errors (M, N, N) - Model Automatically Selected
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

Se selecciona el modelo de **Suavización Exponencial Simple con Errores Multiplicativos (M, N, N)** como el mejor ajuste para los datos de la Concentración con base en el criterio de AICc. Los pronósticos puntuales desarrollados por los modelos Multiplicativo y Aditivo son idénticos si utilizan los mismos valores de los parámetros de suavización. Sin embargo, el modelo Multiplicativo generará diferentes intervalos de predicción para acomodar el cambio en la varianza.

35. Se genera la Información del Modelo de Suavización Exponencial:

Exponential Smoothing Model Information	
Seasonal Frequency	1
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

Esto es un resumen de la información del modelo con Seasonal Frequency = 1 (nonseasonal); Model Selection Criterion = “AICc” y Box-Cox Transformation = “N/A” porque no se marcó la Transformación Box-Cox.

36. Las Estimaciones de los Parámetros son:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.295846234
l (initial level)	16.73860868

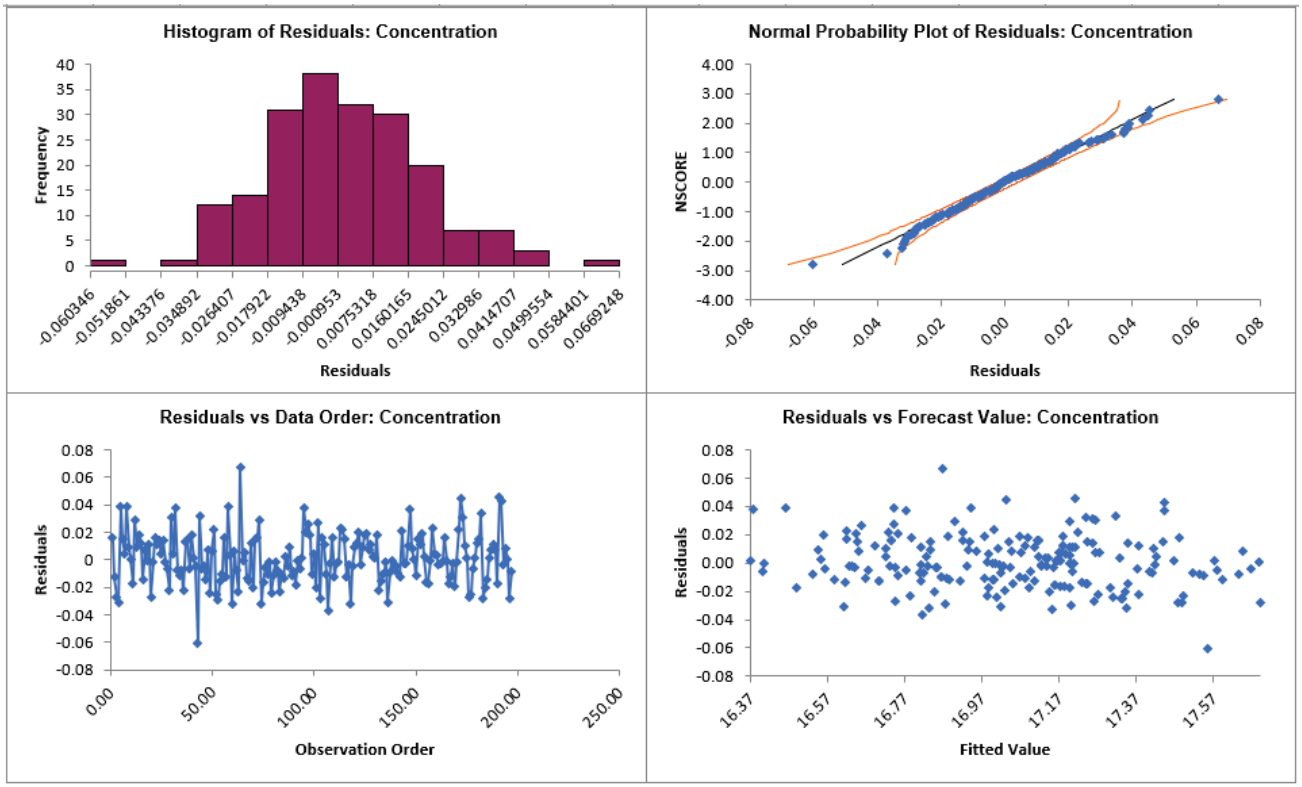
Estos valores se aproximan a las estimaciones de los parámetros que se obtuvieron arriba con el uso del modelo aditivo.

37. Las Estadísticas del Modelo de Suavización Exponencial son:

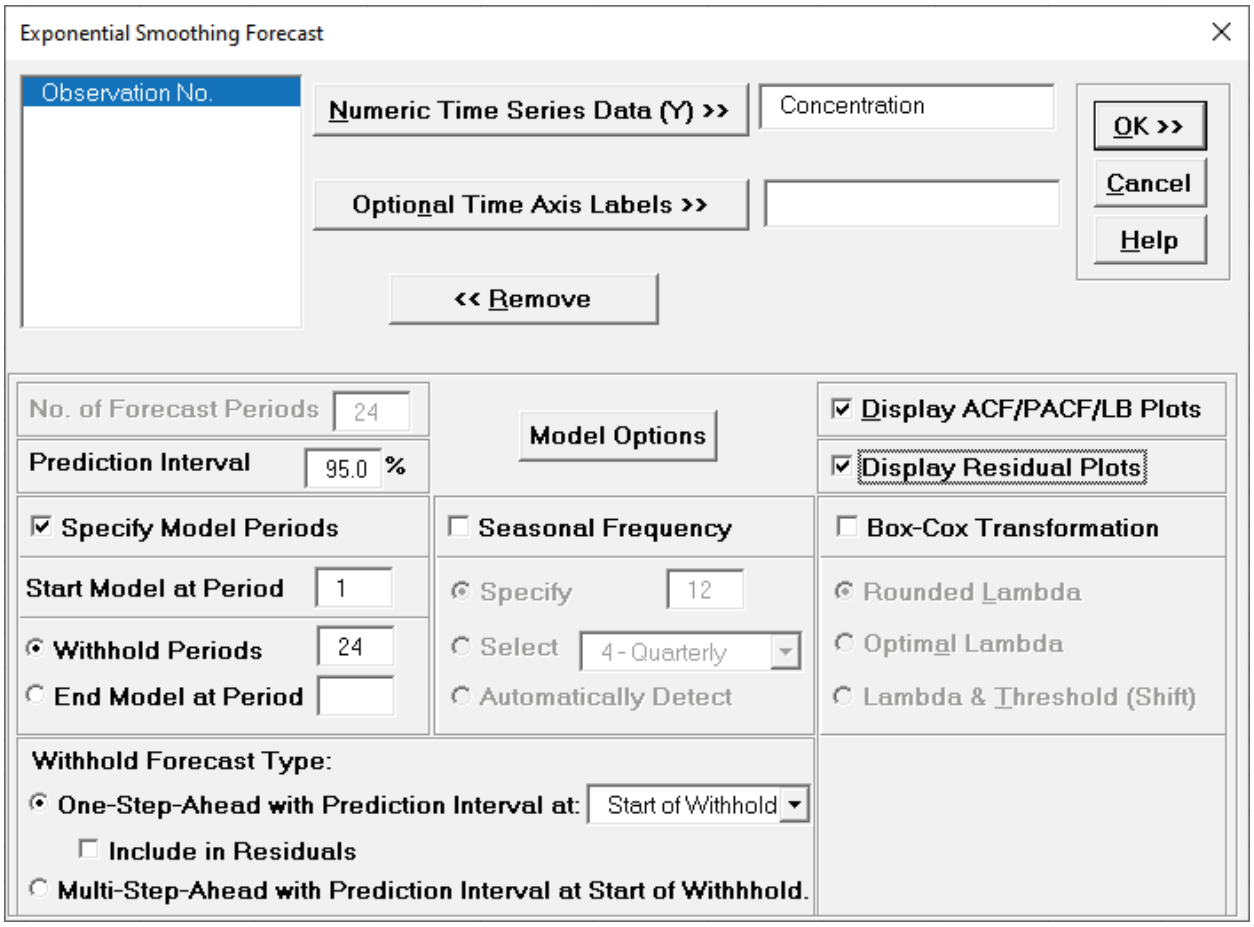
Exponential Smoothing Model Statistics	
No. of Observations	197
DF	195
StDev	0.01865
Variance	0.00035
Log-Likelihood	-293.601
AICc	593.326
AIC	593.201
BIC	603.051

El Log-Verosimilitud, AICc, AIC, y BIC se aproximan a los valores que se obtuvieron arriba con el uso del modelo aditivo, sin embargo el Log-Verosimilitud es ligeramente mayor obteniéndose un AICc menor, por lo que se seleccionó el modelo multiplicativo como el mejor. Tener en cuenta que la DesvEst y la Varianza son muy diferentes. Esta diferencia se debe a que los residuos multiplicativos son errores relativos: $e_t = \frac{y_t - \hat{y}_t}{\hat{y}_t}$.

38. Las métricas de precisión del pronóstico en la muestra, la Tabla del Pronóstico y los Gráficos FAC/FACP/LB para el multiplicativo se aproximan a los del modelo aditivo. Las gráficas de los residuos multiplicativos son iguales, no obstante observe la diferente escala debido a los errores relativos.



39. Se repetirá ahora la Suavización Exponencial con los datos de la Concentración con Automatic Model Selection, pero se emplea Specify Retención Periods. Hacer click en el menú **Recall SigmaXL Dialog** o presione **F3** para repetir el último diálogo. Marcar **Specify Model Periods**. Fijar **Withhold Periods** = 24 (esto es, se pronosticará 24 periodos y se los comparará contra la actual retención). Se hace uso por defecto de **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *Start of Withhold*.



- La opción **Specify Model Periods** permite especificar los periodos inicial y final que se emplean en la identificación automática del modelo y en la estimación de los parámetros. Generalmente, **Start Model at Period** se conserva = 1 y **Withhold Periods** especifica el número de periodos que se retienen para las pruebas fuera de la muestra. **End Model at Period** especifica el final del periodo, de manera que el tamaño de la muestra de retención sería: el número total de observaciones – fin del periodo.

- **Withhold Forecast Type: One-Step-Ahead** excluirá la muestra de la retención de la identificación automática del modelo y la estimación de los parámetros, sin embargo, hace uso de los datos de retención para actualizar la predicción del pronóstico de un paso hacia adelante. Esto es útil para evaluar el error del pronóstico cuando sólo existe interés en la predicción a corto plazo de un paso hacia adelante.

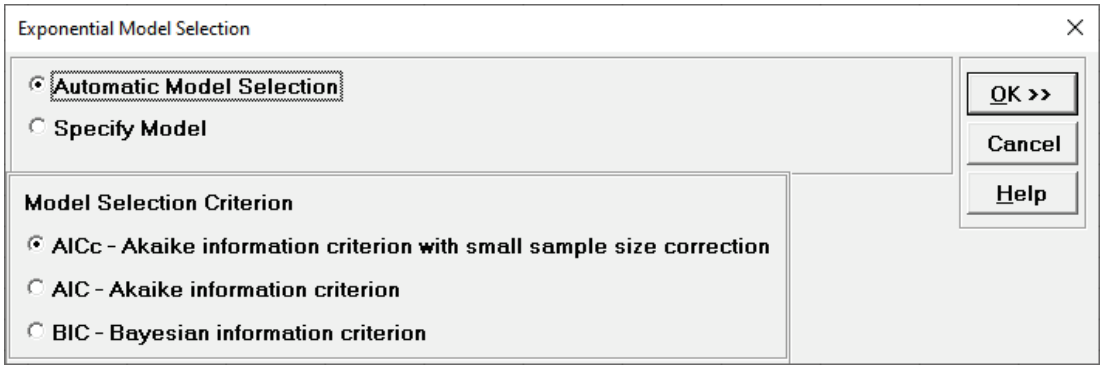
- **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *Start of Withhold* visualizará el intervalo de predicción para la duración de la muestra de retención. Tomar nota que la longitud del intervalo de predicción se lo determina mediante el número de periodos de retención, de tal forma que invalida el **número de Periodos del Pronóstico** especificados.

- **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *End of Withhold* visualizará el intervalo de predicción al final de la muestra de retención. La longitud del intervalo de predicción se lo determina mediante el número de periodos de retención, de tal forma que invalida el **número de Periodos del Pronóstico** especificados.

- **Include in Residuals** tratará los errores de pronósticos de un paso hacia adelante como residuos (aunque no formen parte del proceso de estimación del modelo) y se incluirán en las gráficas de los residuos de FAC/FACP/LB junto con el informe y los gráficos de los residuos. Por lo general permanece sin marcar.

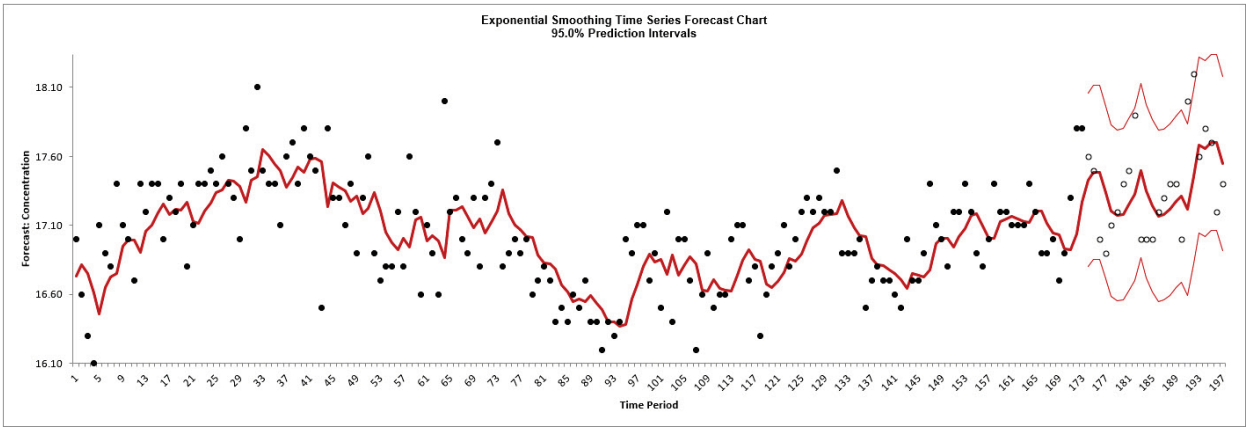
- **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold** excluirá la muestra de retención de la identificación automática del modelo y la estimación de parámetros y no utilizará los datos de retención para actualizar la predicción del pronóstico de un paso hacia adelante. Esto es útil para evaluar el error de pronóstico cuando existe interés en una ventana del pronóstico a largo plazo (horizonte). El intervalo de predicción se mostrará durante toda la duración de la muestra de retención. Tener en cuenta que la duración del intervalo de predicción se determinada por el número de períodos de retención, por lo que invalida el **número de Períodos de Pronósticos** especificados. Estos errores del pronóstico no se incluyen en los gráficos y el informe sobre los residuos de ACF/PACF/LB.

40. Hacer click en **Model Options**. Seleccionar **Automatic Model Selection**. Se hace uso por defecto **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**.



Consejo: Cuando emplea **Recall SigmaXL Dialog**, y no hay cambios en la configuración de **Model Option**, se utilizarán las configuraciones previas. No es necesario repetir este paso.

41. Hacer click en **OK** para regresar al diálogo del Pronóstico de Suavización Exponencial. Hacer click en **OK**. Se proporciona el informe del pronóstico de suavización exponencial:



Los puntos en blanco son los valores de los datos de la muestra de retención con un pronóstico de un paso hacia adelante e intervalos de predicción que se visualizan al comienzo de la muestra de retención.

42. Desplazar hacia abajo el encabezado para ver el Modelo de Suavización Exponencial

Exponential Smoothing Forecast: Simple Exponential Smoothing with Multiplicative Errors (M, N, N) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

De la misma forma que con los datos completos, el modelo de **Suavización Exponencial Simple** con **Errores Multiplicativos (M, N, N)** se ha seleccionado como el de mejor ajuste para los datos de la Concentración que se basa en el criterio AICc. El encabezado incluye también el número de periodos de retención especificados.

43. Las Estimaciones de los Parámetros son:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.303967286
l (initial level)	16.73554259

Estos valores se aproximan a los valores estimados de los parámetros que se han obtenido arriba (que utilizó todos los datos con el modelo multiplicativo).

44. Las Estadísticas del Modelo de Suavización Exponencial son:

Exponential Smoothing Model Statistics	
No. of Observations	173
DF	171
StDev	0.01841
Variance	0.00034
Log-Likelihood	-243.805
AICc	493.752
AIC	493.61
BIC	503.07

Estos valores se aproximan a las estadísticas del modelo que se ha obtenido arriba (que utilizó todos los datos con el modelo multiplicativo). Se hace uso aquí de sólo 173 de las 197 observaciones.

45. Las métricas de la Precisión del Pronóstico son:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	173	24
RMSE	0.31133921	0.35209899
MAE	0.243032918	0.273389284
MAPE	1.425751706	1.567334268
MASE	0.878186174	0.987877246

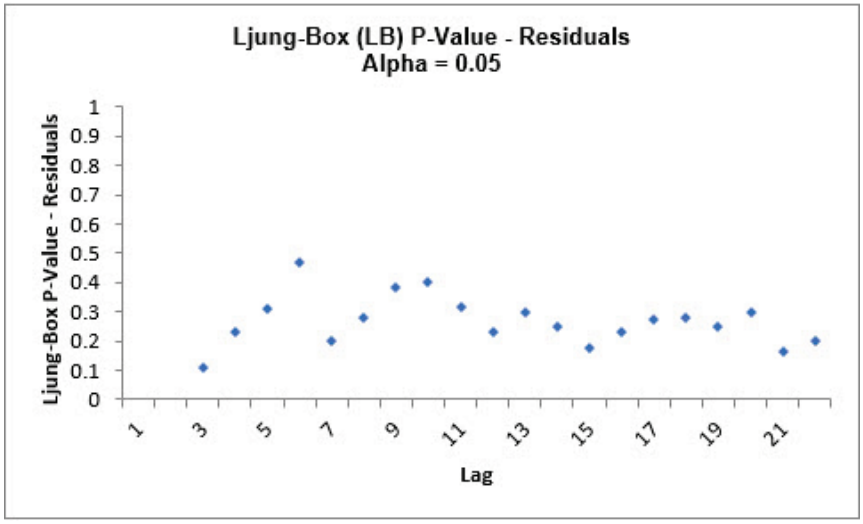
Como era esperarse, los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un Paso-Hacia-Adelante** son mayores que los errores del Pronóstico **Dentro-de-la-Muestra (Estimación) un Paso-Hacia-Adelante**.

46. Se genera la Tabla de Pronóstico:

Forecast Table				
Period	Withhold Data	One-Step-Ahead Forecast	Lower 95.0% PI	Upper 95.0% PI
174	17.6	17.43034504	16.80141557	18.05927452
175	17.5	17.4819146	16.85112437	18.11270483
176	17	17.48741197	16.85642338	18.11840056
177	16.9	17.33925468	16.71361197	17.96489738
178	17.1	17.20573562	16.58491061	17.82656064
179	17.2	17.17359545	16.55393014	17.79326077
180	17.4	17.18162157	16.56166665	17.80157649
181	17.5	17.24800147	16.6256514	17.87035154
182	17.9	17.32460078	16.69948682	17.94971474
183	17	17.49950332	16.86807845	18.13092819
184	17	17.34767065	16.72172427	17.97361703
185	17	17.24199015	16.61985698	17.86412331
186	17.2	17.16843306	16.54895401	17.78791211
187	17.3	17.17802838	16.55820311	17.79785365
188	17.4	17.21510376	16.59394072	17.8362668
189	17.4	17.27130617	16.64811521	17.89449713
190	17	17.31042488	16.68582242	17.93502734
191	18	17.21606587	16.59486812	17.83726363
192	18.2	17.4543562	16.82456035	18.08415206
193	17.6	17.68100752	17.04303354	18.31898151
194	17.8	17.65638389	17.01929838	18.29346939
195	17.7	17.70003849	17.06137782	18.33869916
196	17.2	17.70002679	17.06136654	18.33868704
197	17.4	17.548035	16.91485899	18.18121102

Se trata de los mismos valores de los pronósticos e intervalos de predicción que se muestran en el Gráfico de Pronóstico, pero se proporcionan para un análisis o un gráfico más detallado. También se muestran **los Datos de Retención**.

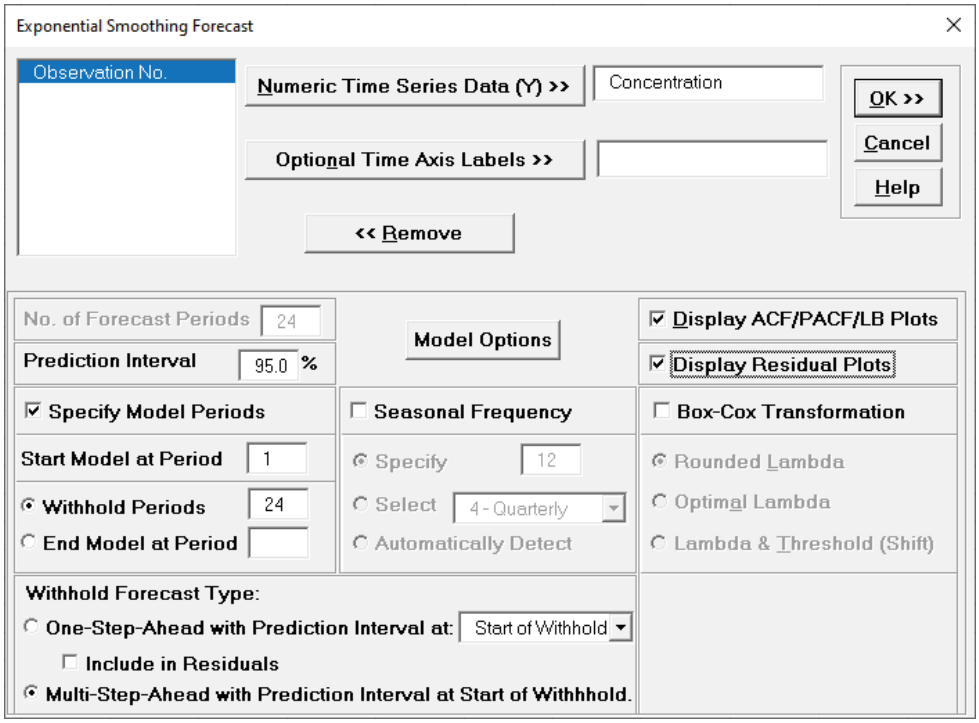
47. Las gráficas de los residuos de FAC/FACP/LB se basan en los datos dentro de la muestra. Las gráficas se parecen a las de los datos completos arriba, a excepción de los valores-P de Ljung-Box:



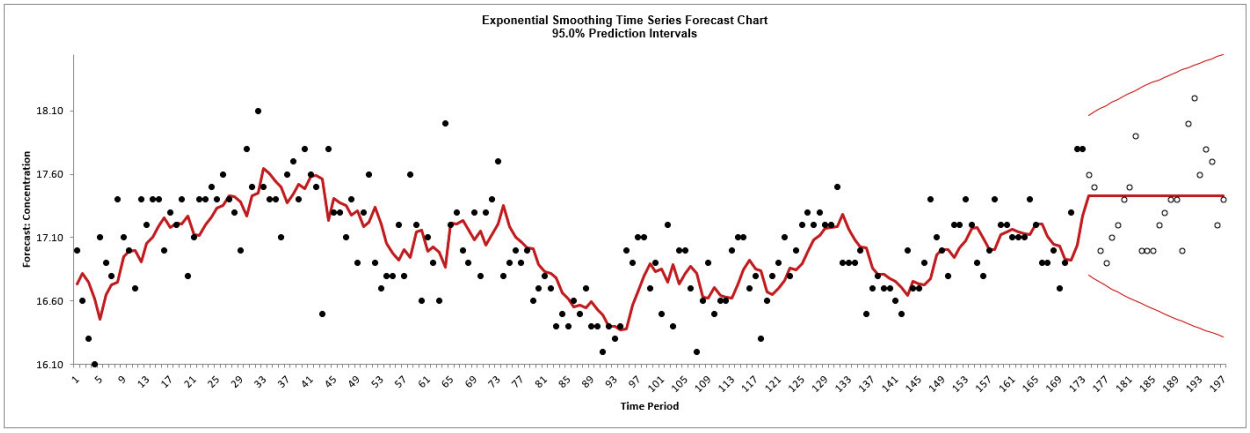
El modelo de **Suavización Exponencial Simple con Errores Multiplicativos (M, N, N)** tiene un mejor ajuste que el subconjunto de los datos completos, con todos los valores-p de color azul (>0.05).

48. Si **Include in Residuals** estuviese marcado entonces los residuos incluirían también los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un-Paso-Hacia-Adelante**.

49. Se repetirá ahora la Suavización Exponencial con los datos de la Concentración, pero con el empleo de Multi-Paso-Hacia-Adelante para el Pronóstico con Retención. Hacer click en el menú **Recall SigmaXL Dialog** o presione **F3** para repetir el último diálogo. Seleccionar **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**.



50. Hacer click en **OK**. Se proporciona el informe del pronóstico de suavización exponencial:



Los puntos en blanco son los valores de los datos en la muestra de retención con un pronóstico de varios pasos e intervalos de predicción mostrados al comienzo de la muestra de retención.

51. Las métricas de la Precisión del Pronóstico son:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Multi-Step-Ahead Forecast
N	173	24
RMSE	0.31133921	0.348228059
MAE	0.243032918	0.286752928
MAPE	1.425751706	1.647391419
MASE	0.878186174	1.036166041

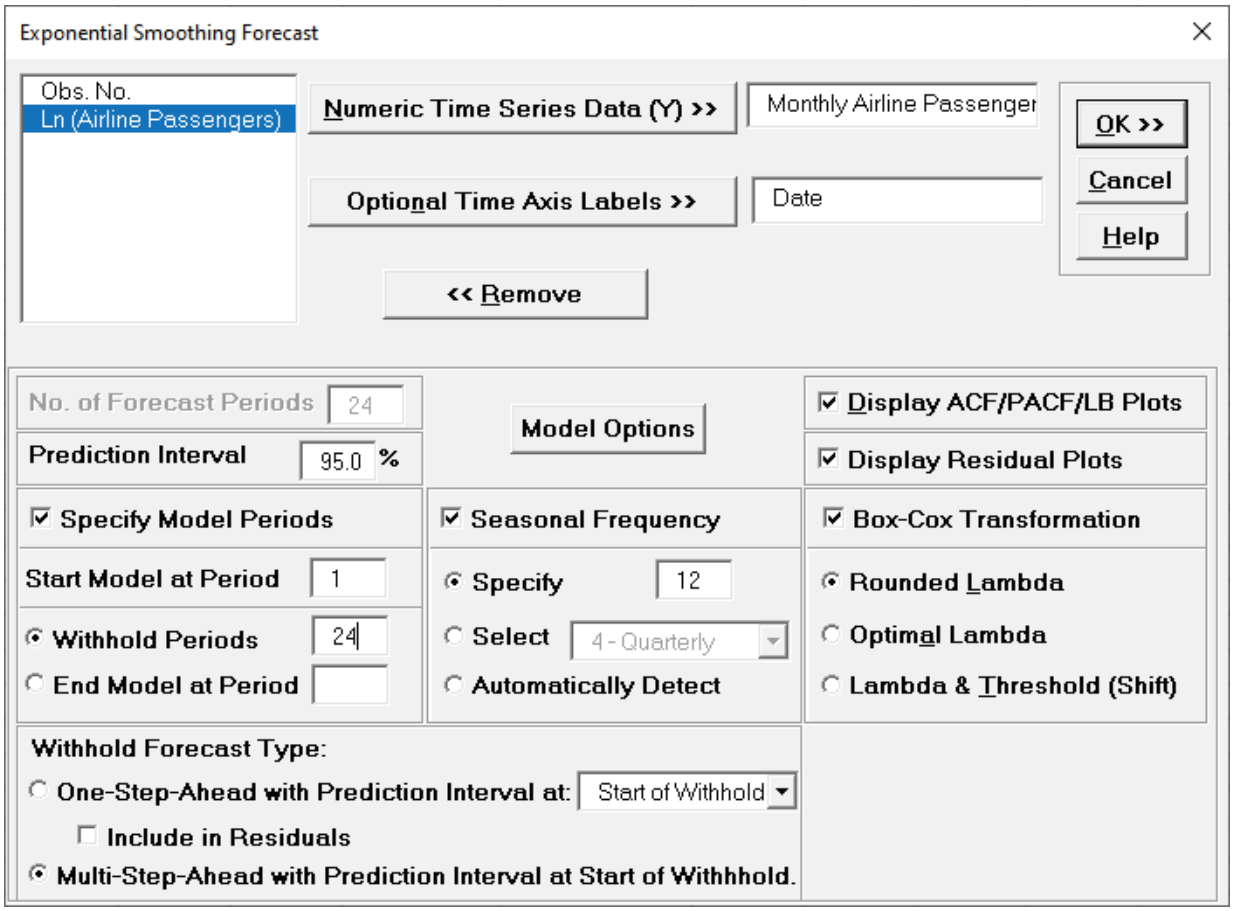
Tal como se esperaba, los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Multi-Paso-Hacia-Adelante** son mayores que los errores del **Pronóstico Dentro-de-la-Muestra (Estimación) Un-Paso-Hacia-Adelante** y los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un-Paso-Hacia-Adelante** de arriba.

Pasajeros Mensuales de Aerolínea - Serie G

52. Abrir **Monthly Airline Passengers - Series G.xlsx** (pestaña de la **Hoja 1**). Estos son los datos de una Serie G de Box y Jenkins, los pasajeros por mes de una aerolínea internacional desde enero 1949 a diciembre 1960.

53. Hacer click en **SigmaXL > Time Series Forecasting > Exponential Smoothing Forecast > Forecast.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

54. Seleccionar ***Monthly Airline Passengers***, hacer click en **Numeric Time Series Data (Y) >>**. Seleccionar Date, hacer click en **Optional Time Axis Labels >>**. Marcar **Display ACF/PACF/LB Plots** y **Display Residual Plots**. Marcar **Specify Model Periods**. Fijar **Withhold Periods** = 24 (esto es, se pronosticará 24 meses y compararán con la retención actual). Seleccionar **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Marcar **Seasonal Frequency** con **Specify** = 12. Marcar **Box-Cox Transformation** y seleccionar **Rounded Lambda** (seleccionado porque Run Chart mostró un incremento de la varianza estacional en el tiempo). Se utilizará por defecto **Prediction Interval** = 95.0 %.

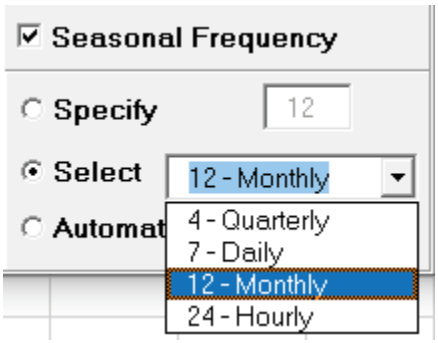


- **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold** excluirá la muestra que se retiene de la identificación automática del modelo y la estimación de parámetros y no utilizará los datos de la retención para actualizar el pronóstico previsto con un paso hacia adelante. Esto es útil para evaluar el error de pronóstico cuando existe

- interés en una ventana de pronóstico a largo plazo (horizonte). El intervalo de predicción se mostrará durante la duración de la muestra que se retiene. Obsérvese que la duración del intervalo de predicción está determinada por el número de períodos de retención, por lo que invalida el **número de Períodos de Pronósticos** especificados**.**

- Se emplea **Seasonal Frequency Specify** para especificar la frecuencia estacional. Tomar nota que la Suavización Exponencial se limita a una máxima frecuencia estacional de 24. Para mayores frecuencias utilizar **Suavización Exponencial – Descomposición Estacional Múltiple (DEM)**

- **Seasonal Frequency Select** proporciona un listado desplegable de las frecuencias estacionales de uso más común:



- Se debería usar **Seasonal Frequency Automatically Detect** si no se tiene la certeza del valor correspondiente a la frecuencia estacional (o realice una gráfica de Densidad Espectral antes de los gráficos de la Descomposición de Tendencia Estacional). Si se detecta una frecuencia > 24, arroja 1. Si ocurre aquello, por favor, emplear la Suavización Exponencial – Descomposición Estacional Múltiple (DEM).

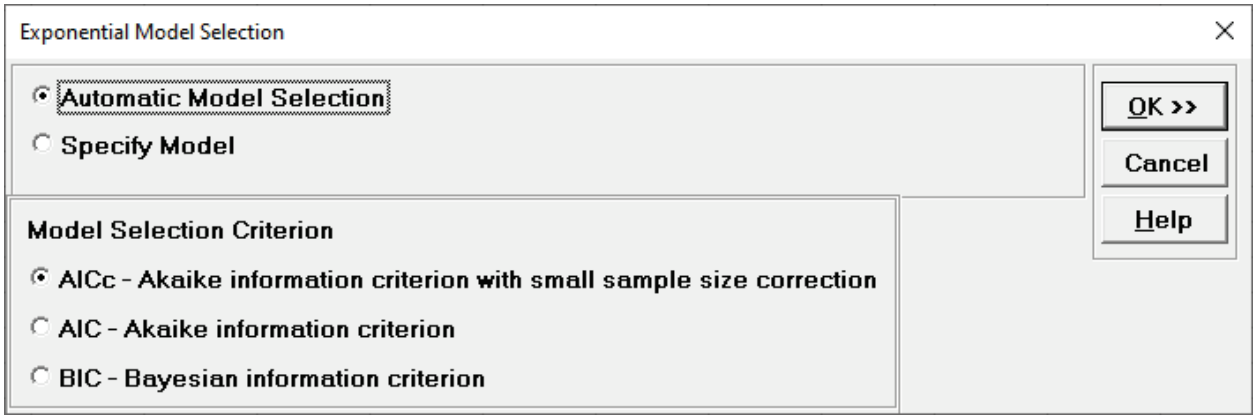
- **Box-Cox Transformation** with **Rounded Lambda s**eleccionará un Lambda = 0 (Ln), 0.5 (SQRT) o 1 (Sin transformación). El umbral (Shift) se calcula automáticamente si la serie temporal incluye el 0 o valores negativos, de lo contrario es 0.

- **Box-Cox Transformation** with **Optimal Lambda** utiliza el rango de 0 a 1 para Lambda. El umbral se calcula automáticamente si la serie temporal incluye el 0 o valores negativos.

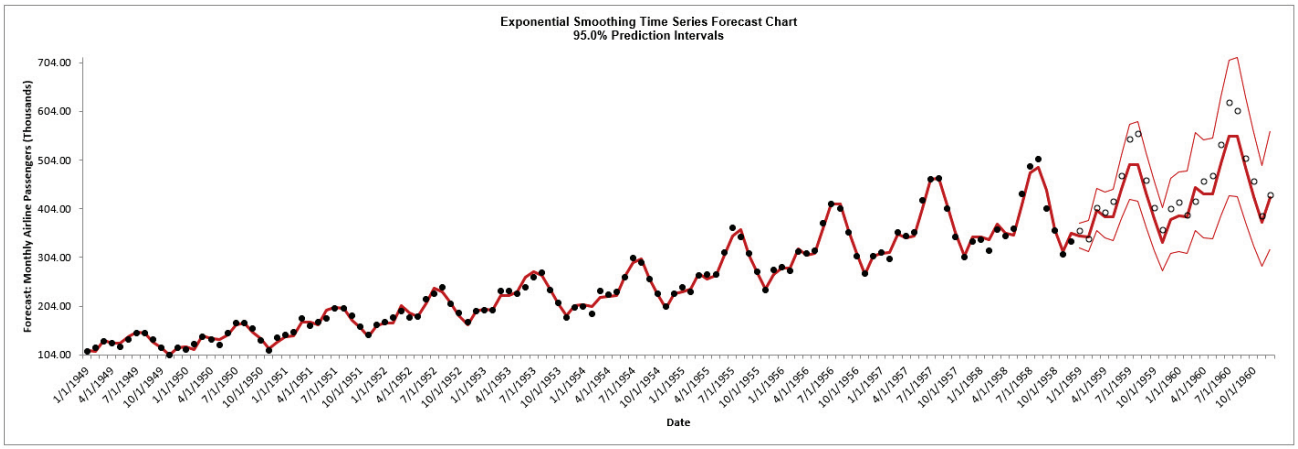
- **Box-Cox Transformation** con **Lambda & Threshold (Shift)** si se lo deja vacío, calculará el óptimo de lambda y el umbral. El usuario puede especificar también Lambda y el umbral. Se puede especificar también Lambda fuera del rango de 0 a 1, pero para el análisis de las series temporales prácticamente se debería limitar desde -1 hasta 2. Normalmente el umbral es 0, pero si la serie de tiempo incluye el 0 o valores negativos, un valor negativo del umbral

debería registrarse que sea más pequeño que el mínimo valor de los datos. Este valor se debería sustraer de los datos lo que resulta en valores positivos en la serie de tiempo.

55. Hacer click en **Model Options**.



56. Se hace uso por defecto la **Automatic Model Selection** con el **AICc** como el **Model Selection Criterion.** Hacer click en **OK** para regresar al diálogo del Pronóstico de Suavización Exponencial. Hacer click en **OK**. Se proporciona el informe del pronóstico de suavización exponencial:



Los puntos en blanco son los valores de los datos de la muestra de retención con un pronóstico de varios pasos e intervalos de predicción mostrados al comienzo de la muestra de retención. Obsérvese que los datos se han transformado utilizando Box-Cox, pero se aplica la transformación inversa para producir este gráfico con las unidades originales.

57. Desplazar hacia abajo el encabezado para ver el Modelo de Suavización Exponencial

Exponential Smoothing Model: Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters) (A, A, A) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

El modelo de **Tendencia Aditiva, Método Estacional Aditivo con Errores Aditivos (Holt-Winters) (A, A, A)** ha sido seleccionado automáticamente como el mejor ajuste para los datos de los Pasajeros de la Aerolínea transformados con Box-Cox con base al criterio AICc. Tener en

cuenta que no se consideran los modelos multiplicativos cuando se marca la **Transformación Box-Cox**. Si se opta por un modelo multiplicativo, entonces se podría transformar manualmente los datos y modelar Ln (Airline Passengers) sin marcar **Box-Cox Transformation**.

El encabezado incluye también el número de periodos de retención especificados.

58. Se genera la Información del Modelo de Suavización Exponencial:

Exponential Smoothing Model Information	
Seasonal Frequency	12
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

Esto es un resumen de la información del modelo con la Frecuencia Estacional = 12; Criterio de Selección del Modelo = "AICc" y Transformación Box-Cox = "Rounded Lambda" con Lambda = 0 (esto es, una transformación Ln).

59. Las Estimaciones de los Parámetros son:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.770415301
beta (trend smoothing)	0.0001
gamma (seasonal smoothing)	0.0001
l (initial level)	4.80925762
b (initial trend)	0.00935261
s1 (initial seasonal)	-0.099989196
s2 (initial seasonal)	-0.217929483
s3 (initial seasonal)	-0.076652841
s4 (initial seasonal)	0.063787774
s5 (initial seasonal)	0.198157807
s6 (initial seasonal)	0.205968013
s7 (initial seasonal)	0.108472035
s8 (initial seasonal)	-0.016634733
s9 (initial seasonal)	-0.008278553
s10 (initial seasonal)	0.033611999
s11 (initial seasonal)	-0.101347496
s12 (initial seasonal)	-0.089165324

- El error incluye el parámetro de suavización *alfa* y el valor del nivel inicial (*l*). El error es aditivo, pero en los datos transformados con Ln.
- La tendencia añade un parámetro de suavización (*beta*) y un valor de la tendencia inicial (*b*).

- El efecto estacional añade un parámetro de suavización (*gamma*) y los valores de las estaciones iniciales (*s1* hasta *s12*), restringido con suma cero parta el aditivo.

60. Las Estadísticas del Modelo de Suavización Exponencial son:

Exponential Smoothing Model Statistics	
No. of Observations	120
DF	104
StDev	0.037144
Variance	0.00138
Log-Likelihood	116.4916
AICc	-192.983
AIC	-198.983
BIC	-151.596

- El número de observaciones, *n* = 144 – 24 (*retención*) = 120
- Los grados de libertad (DF) = 120 (*n*) – 16 (los términos en el modelo, excluyendo *s12*) = 104
- Tener en cuenta que las estadísticas se basan en la transformación Ln de los datos, no en los datos originales.

61. Las métricas de la Precisión del Pronóstico son:

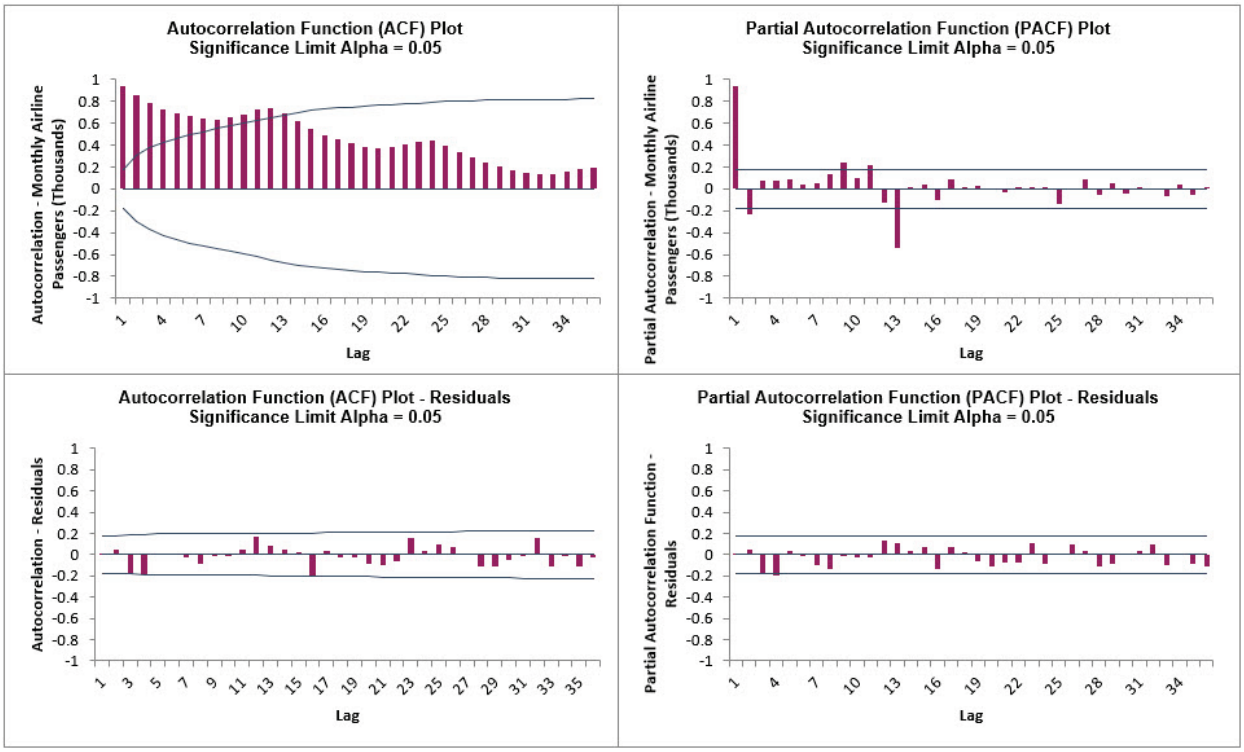
Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Multi-Step-Ahead Forecast
N	120	24
RMSE	8.67099231	33.06342473
MAE	6.481327147	27.83270693
MAPE	2.730421504	5.80549634
MASE	0.226825448	0.974054552

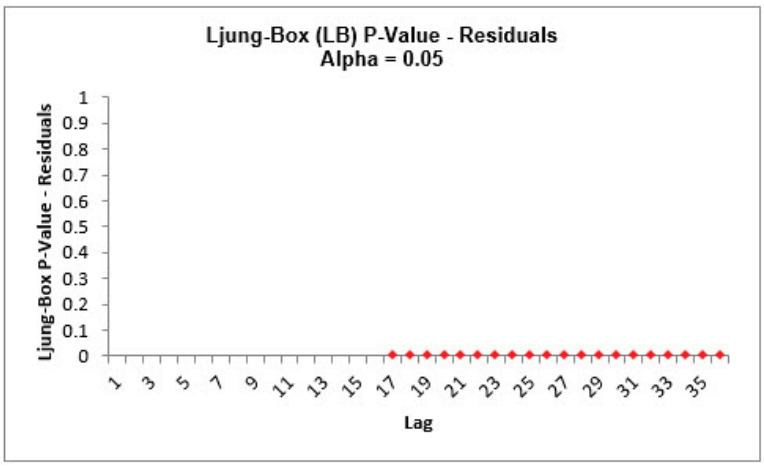
Como era esperarse, los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Multi- Pasos-Hacia-Adelante** son mayores que los errores del Pronóstico **Dentro-de-la-Muestra (Estimación) un Paso-Hacia-Adelante**. Tener en cuenta, si existiera interés en un pronóstico a corto plazo un paso hacia adelante, entonces se tendría que seleccionar **Withhold Forecast Type: One-Step-Ahead** y esta tabla mostraría los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un Paso-Hacia-Adelante**.

Las métricas de Precisión de los Pronósticos se calculan utilizando los datos reales sin procesar versus el pronóstico de la transformación inversa tal como se muestra en el Gráfico y la Tabla

de Pronósticos, lo que permite la comparación entre todos los tipos de modelos y transformaciones.

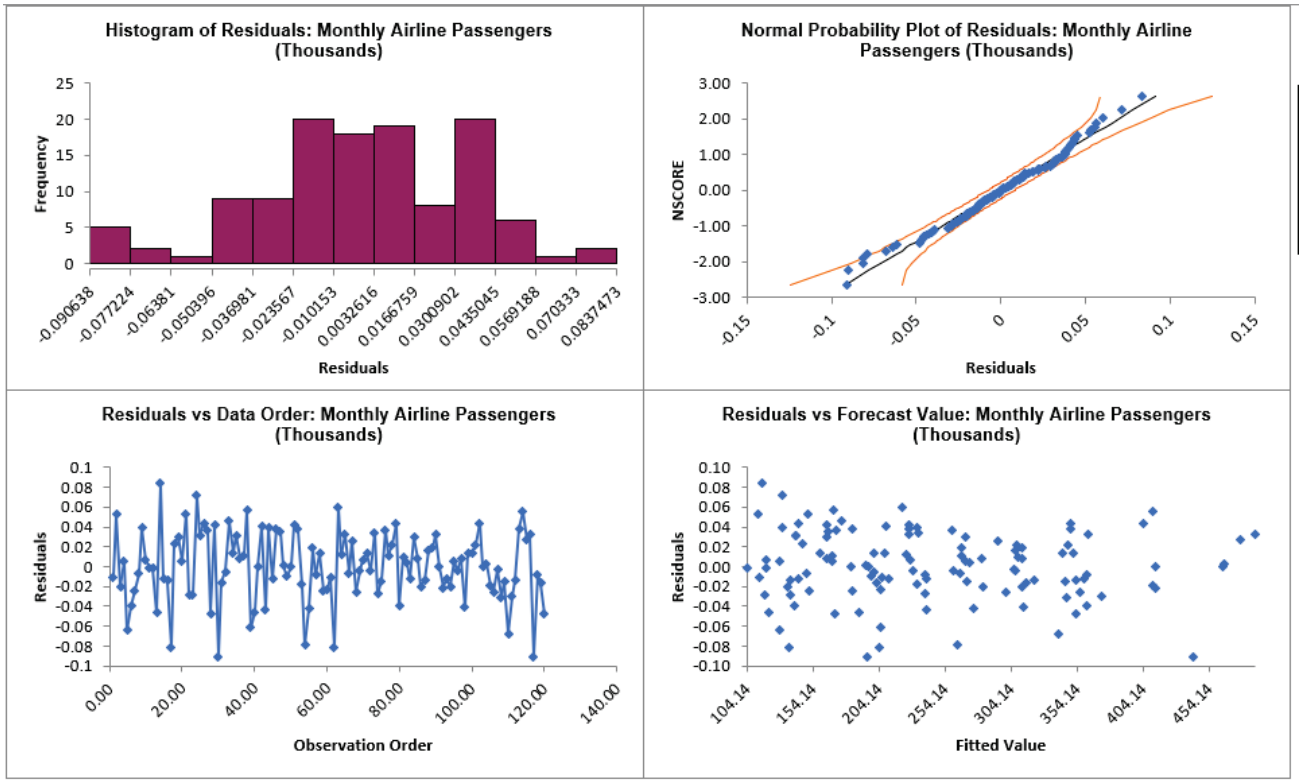
62. Hacer click en la hoja **Exp. Smooth. ACF PACF LB** para visualizar los gráficos FAC/FACP/LB:





Los gráficos de los residuos de FAC/FACP/LB indican que casi toda la autocorrelación es explicada por el modelo, sin embargo, el gráfico de Ljung-Box muestra que aún queda una autocorrelación significativa (los valores P rojos son significativos con un alfa=0.05) - por lo que el modelo puede ser potencialmente mejorado. Esto no significa que el modelo sea un mal modelo, todavía puede ser muy útil para fines de predicción, pero los intervalos de predicción pueden no proporcionar una cobertura precisa.

63. Hacer click en la hoja **Exp. Smoothing Residuals** para visualizar los gráficos de los Residuos:



Los residuos están aproximadamente normalmente distribuidos, con una línea casi recta en la gráfica de probabilidad normal. No hay valores extremos o patrones obvios en los gráficos.

Nótese que los residuos se basan en los datos transformados de Ln, no en los datos originales. La **Información del Modelo de Suavización Exponencial** a la derecha de los gráficos muestra la información sobre la Transformación Box-Cox

Exponential Smoothing Model Information	
Seasonal Frequency	12
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

Pronóstico de Suavización Exponencial – Descomposición Estacional Múltiple (DEM)

La suavización exponencial se limita a una frecuencia estacional máxima de 24. Para frecuencias más altas use la Suavización exponencial - Descomposición Estacional Múltiple (DEM). El componente estacional se elimina primero a través de la descomposición, un modelo con alisamiento exponencial no estacional ajustado al residual (+tendencia), y luego se vuelve a añadir el componente estacional. Para el pronóstico, se utiliza un pronóstico estacional naïve sobre el componente estacional. Obsérvese que los intervalos de predicción se derivan del modelo de suavización exponencial y no incluyen la incertidumbre en el componente estacional.

Como su nombre indica, la Descomposición Estacional Múltiple (DEM) también da cabida a la estacionalidad múltiple, por ejemplo, los datos de media hora con una frecuencia estacional de 48 observaciones diarias y 336 observaciones semanales. Al utilizar DEM, se recomienda limitar el período de pronóstico a 2* la frecuencia estacional dominante.

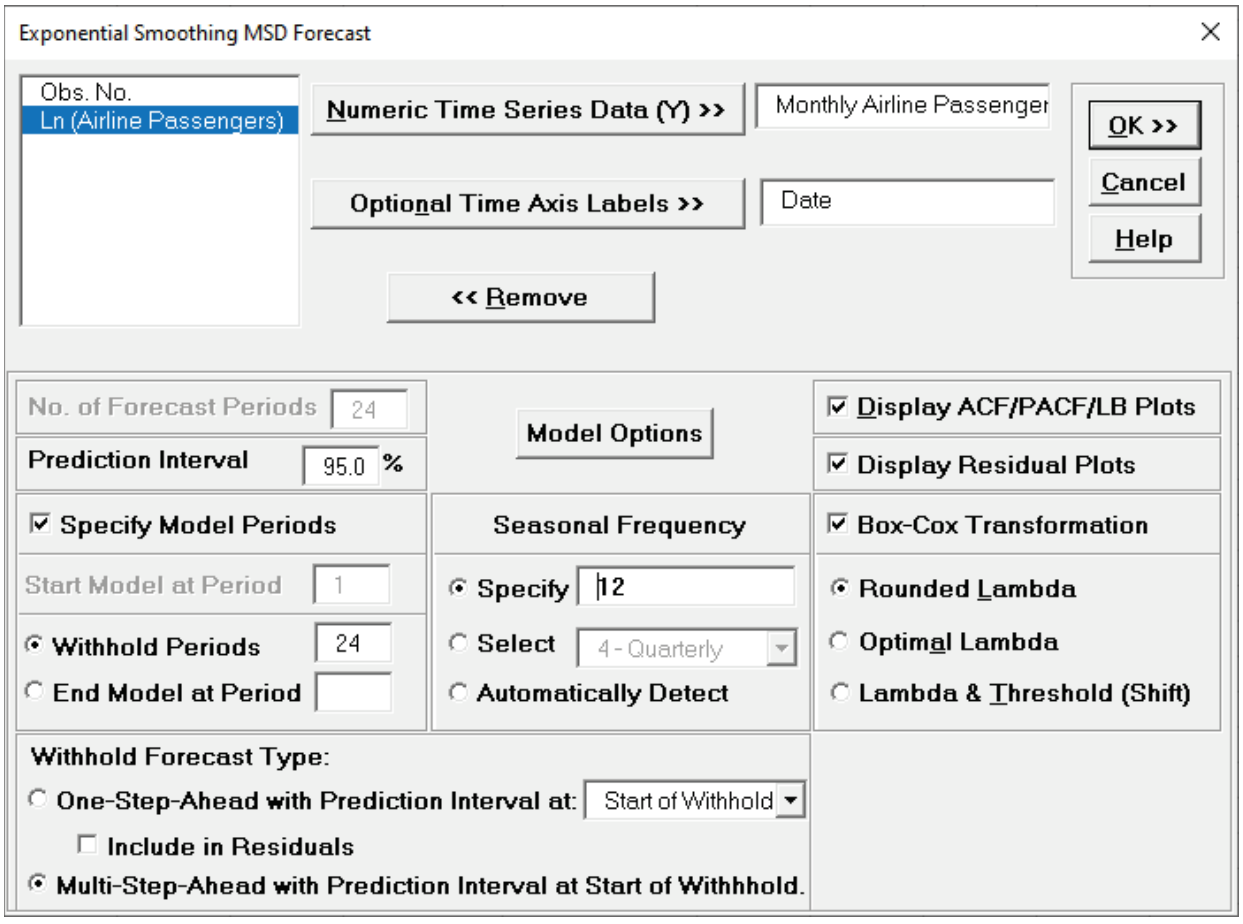
Para mayores detalles y fórmulas, véase el Apéndice: **Descomposición de Tendencia Estacional**

Pasajeros Mensuales de Aerolínea - Serie G

1. Abrir **Monthly Airline Passengers - Series G.xlsx** (pestaña de la Hoja 1). Estos datos son una Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional de enero 1949 a diciembre 1960. Véase **Cartas de Corrida, Gráficos FAC/FACP, Gráfico de Densidad Espectral, Gráfico de Descomposición de Tendencia Estacional** para estos datos. La opción de Descomposición Estacional Múltiple (DEM) no es necesaria para estos datos, sin embargo, como una forma de introducción, se empleará esto para compararlo con un previo análisis.

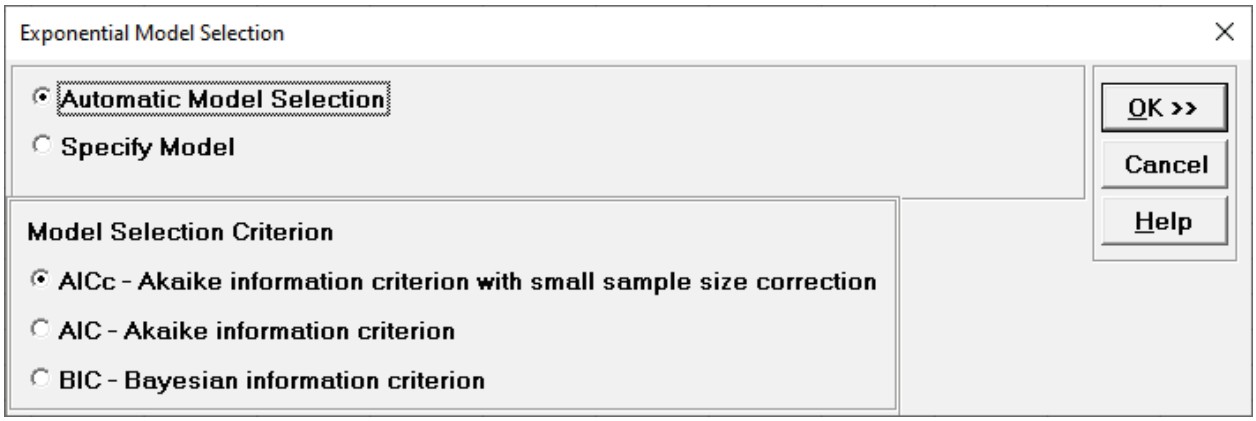
2. Hacer click en **SigmaXL > Time Series Forecasting > Suavización exponencial Forecast > Multiple Seasonal Decomposition Forecast**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar Use **Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Monthly Airline Passengers*, hacer click en **Numeric Time Series Data (Y) >>**. Seleccionar *Date*, hacer click en **Optional Time Axis Labels >>**. Marcar **Display FAC/FACP/LB Plots** y **Display Residual Plots**. Marcar **Specify Model Periods**. Fije **Withhold Periods** = 24. Seleccionar **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Seleccionar **Seasonal Frequency Specify** y registrar 12**.** Marcar **Box-Cox Transformation** y seleccionar **Rounded Lambda**. Se hace uso por defecto **Prediction Interval** = 95.0 %.

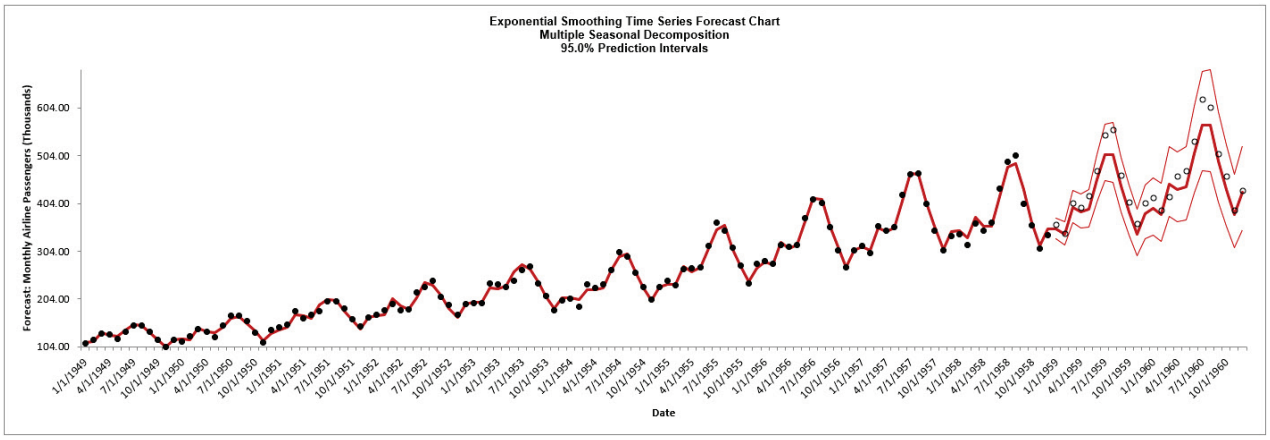


- **Seasonal Frequency** puede tener entradas múltiples, no obstante, se recomienda no más de 3 valores.

4. Hacer click en **Model Options**.



5. Se empleará por defecto **Automatic Model Selection** con **AICc** como el **Model Selection Criterion**. Hacer click en OK para regresar al diálogo Pronóstico de Suavización Exponencial DEM. Hacer click en **OK**. Se proporciona el informe del pronóstico de suavización exponencial:



6. Desplazar hacia abajo el encabezado para ver el Modelo de Suavización Exponencial

Exponential Smoothing Model (Multiple Seasonal Decomposition): Additive Trend Method with Additive Errors (Holt's Linear) (A, A, N) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

Posterior a la **Descomposición Estacional Múltiple**, el modelo corresponde al **Método de Tendencia Aditiva con Errores Aditivos (Lineal de Holt) (A, A, N)**.

Se selecciona **(A, A, N)** automáticamente como el de mejor ajuste para el de los datos desestacionalizados de los Pasajeros de la Aerolínea con base al criterio AICc.

El encabezado incluye también el número de periodos de retención especificados.

7. Se genera la Información del Modelo de Suavización Exponencial:

Exponential Smoothing Model Information	
Seasonal Frequency	12 Decomposition
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

Este es el resumen de la información sobre el modelo con Frecuencia Estacional = 12 aplicando Descomposición y el Criterio de Selección del Modelo = "AICc". La Transformación Box-Cox es "Rounded Lambda" con Lambda = 0 (Transformación Ln).

8. Las Estimaciones de Parámetros para los datos desestacionalizados de los Pasajeros de la Aerolínea son:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.674923407
beta (trend smoothing)	0.0001
l (initial level)	4.806932238
b (initial trend)	0.009425139

- El error incluye el parámetro de suavización *alfa* y el valor del nivel inicial (*l*). El error es aditivo, pero en los datos transformados con Ln.
- La tendencia agrega un parámetro de suavización (*beta*) y el valor de tendencia inicial (*b*).
- El parámetro de suavización (*gamma*) y los valores iniciales de estacionalidad no se procesan porque los datos han sido desestacionalizados.

9. Las Estadísticas del Modelo de Suavización Exponencial son:

Exponential Smoothing Model Statistics	
No. of Observations	120
DF	116
StDev	0.030549511
Variance	0.000933273
Log-Likelihood	133.3933791
AICc	-256.2604423
AIC	-256.7867581
BIC	-242.8492994

- El número de observaciones, *n* = 144 – 24 (*retención*) = 120
- Los grados de libertad (DF) = 120 (*n*) – 4 (los términos en el modelo) = 116
- Tener en cuenta que las estadísticas se basan en la transformación Ln de los datos, no en los datos originales.
- Las estadísticas del modelo son mejores que los correspondientes al análisis previo: menor DesvEst y Varianza, mayor Log-Verosimilitud y menor AICc, AIC y BIC. Esto era de esperarse porque los datos han sido desestacionalizados y de esta manera, el componente del error no se incluye.

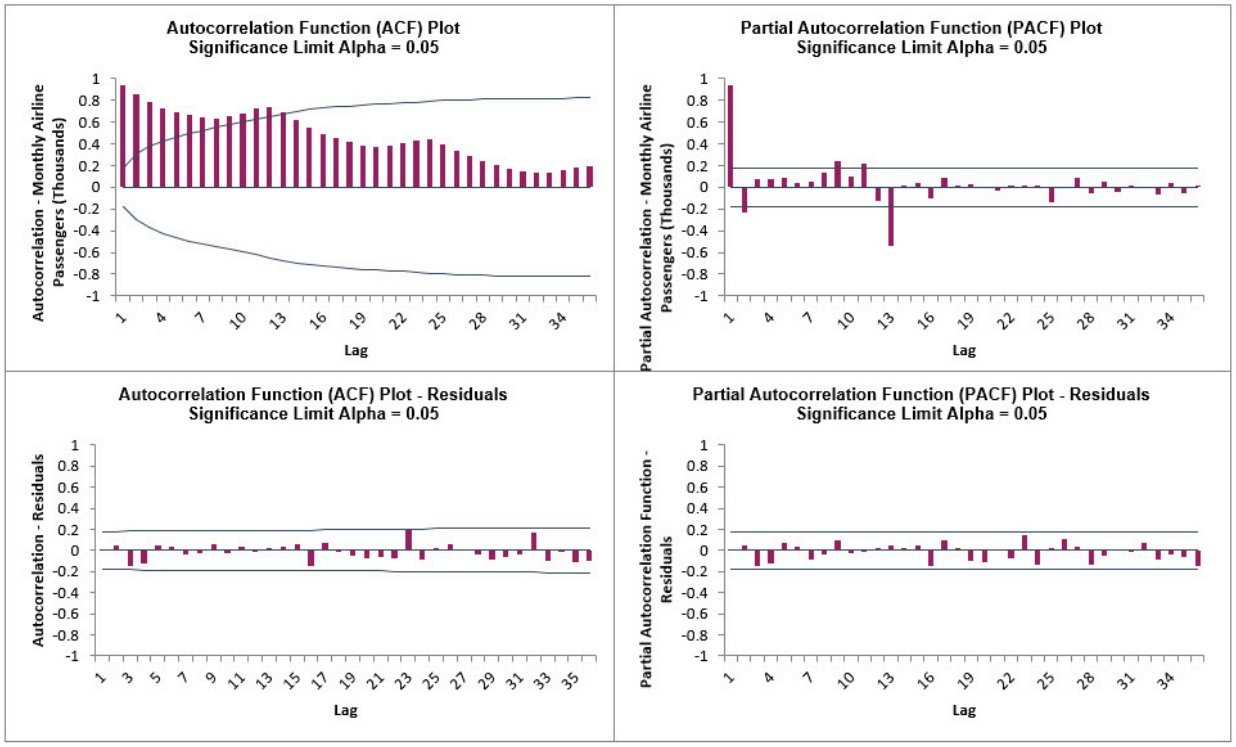
10. Las métricas de la Precisión del Pronóstico son:

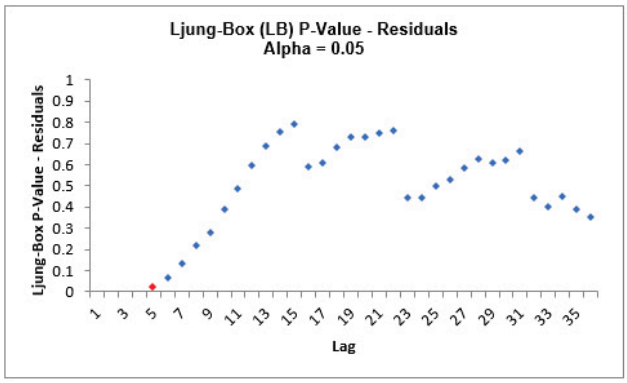
Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Multi-Step-Ahead Forecast
N	120	24
RMSE	7.115702971	26.70412899
MAE	5.455681009	22.99181598
MAPE	2.349324345	4.842245097
MASE	0.190931156	0.804639056

En comparación con el análisis previo, tanto los errores del Pronóstico Dentro de la Muestra (Estimación), Un-Paso-Hacia-Adelante y los errores del Pronóstico Fuera de la Muestra (Retención), Multi-Pasos-Hacia-Adelante son ligeramente más pequeños. (Esto no se esperaba, por lo general, un modelo de Suavización exponencial estacional o ARIMA daría un pronóstico más exacto, basado en la comparación de métodos que usan datos de competencia de pronósticos). Sin embargo, dado que se está pronosticando por dos años, ambos modelos se ven muy bien.

Las métricas de la precisión de la previsión se calculan utilizando los datos reales sin procesar versus la transformada inversa del pronóstico que se visualiza en la Gráfica y la tabla de Pronóstico, de modo que permiten la comparación entre todos los tipos y transformaciones de los modelos.

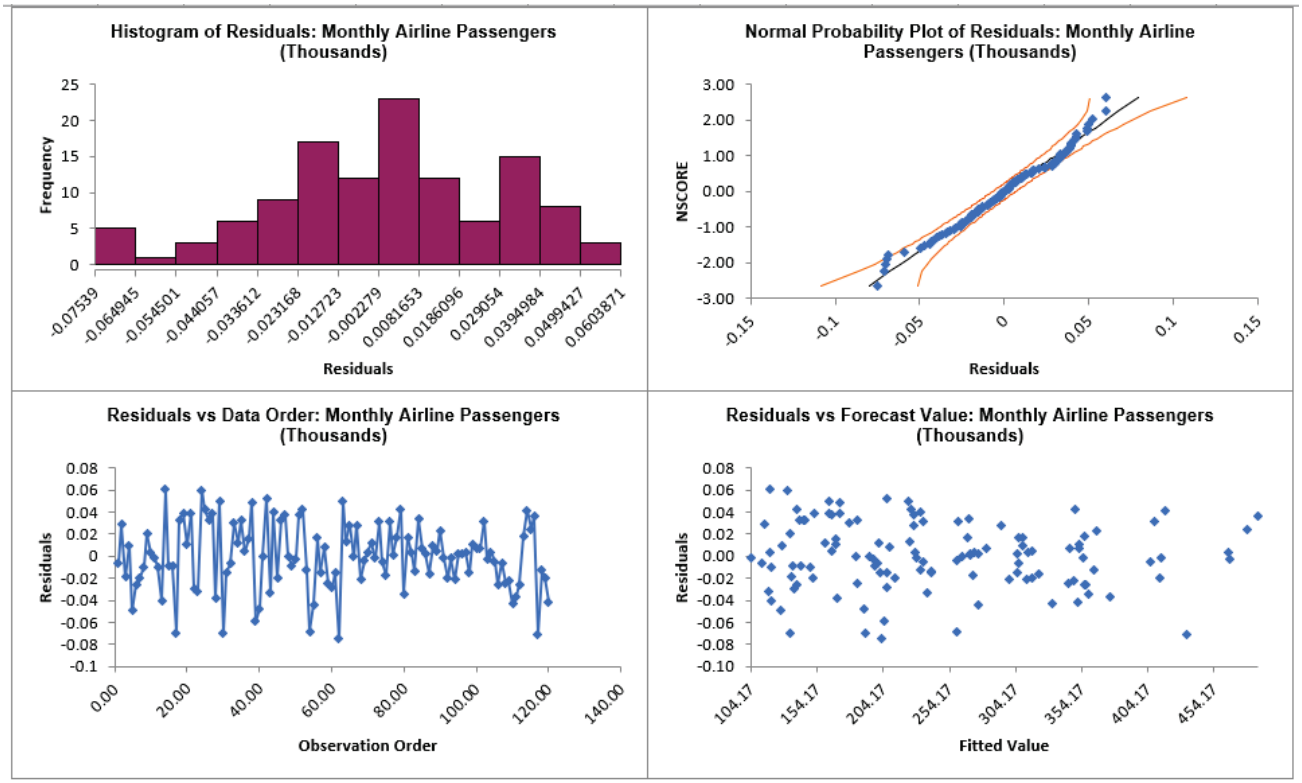
11. Hacer click en la hoja **Exp Smooth MSD ACF PACF LB** para ver los Gráficos FAC/FACP/LB:





Los Gráficos de los Residuos FAC/FACP son similares a los del análisis previo e indican que casi toda la autocorrelación ha sido considerada en el modelo, sin embargo, la gráfica Ljung-Box confirma que este es un mejor ajuste, con la mayoría de valores-P de color azul (> 0.05).

12. Hacer click en la hoja **Exp Smooth MSD Residuals** para visualizar los Gráficos de los Residuos:



Los residuos están distribuidos de forma aproximadamente normal, con una línea más o menos recta en la gráfica de probabilidad normal. No hay valores extremos o patrones obvios en los gráficos.

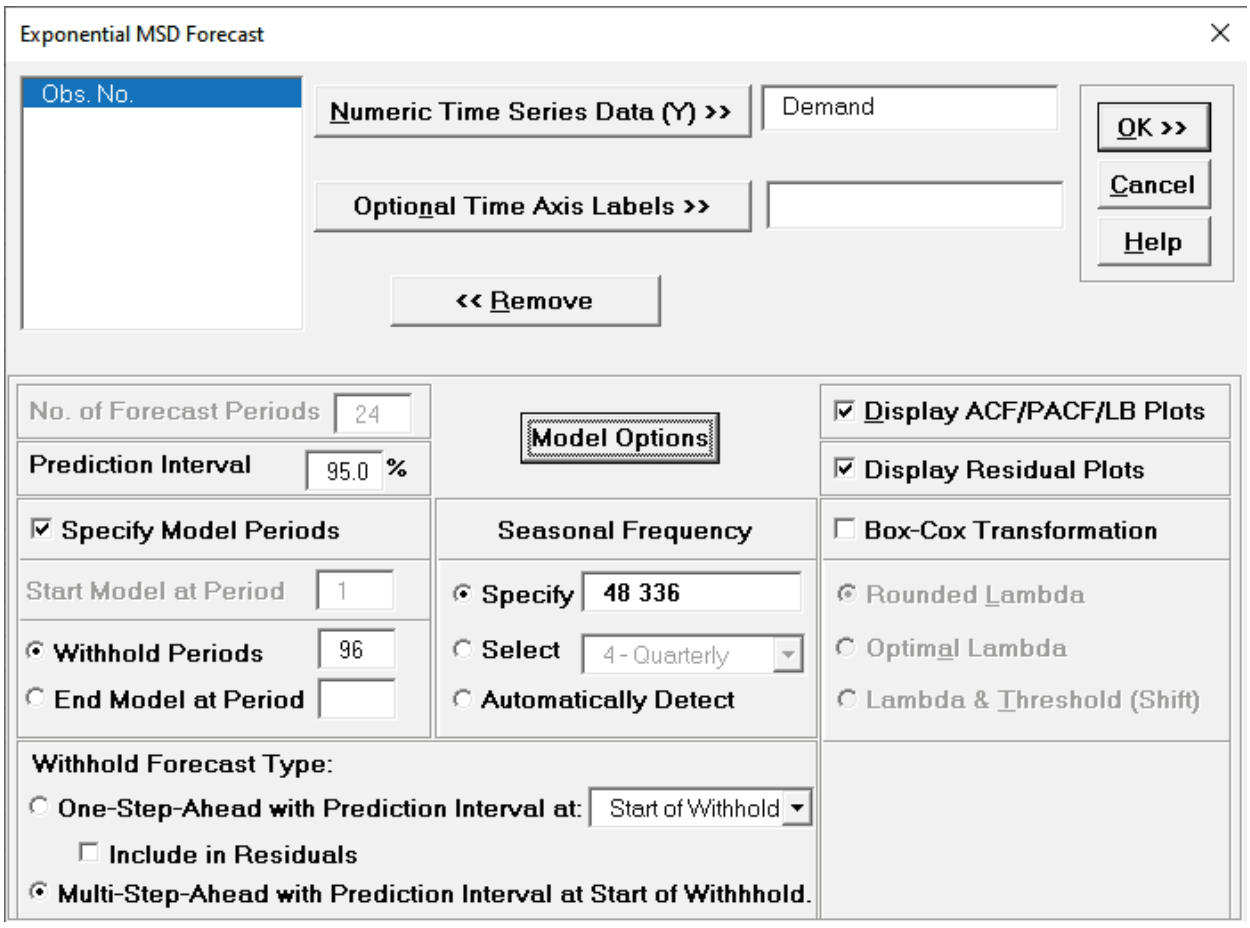
Nótese que los residuos de la DEM son los valores finales observados – los valores pronosticados, por lo que no hay diferencias de escala si el modelo utiliza un error aditivo o multiplicativo. Como se utilizó una transformación Box-Cox, los residuos están en unidades transformadas de Ln.

Demanda de Media-Hora de Electricidad Estacional Múltiple – Taylor

13. Abrir **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (pestaña de la Hoja 1). Esta es la demanda de electricidad de media hora (MW) en Inglaterra y Wales desde el lunes, 5 de junio, 2000 hasta el Domingo, 27 de agosto, 2000 (taylor, R forecast). Estos datos tienen una múltiple estacionalidad con frecuencia = 48 (observaciones diarias) y 336 (observaciones semanales), con un total de 4032 observaciones. Véase para estos datos Cartas de Corrida, Gráficos FAC/FACP, Gráfico de Densidad Espectral, Gráficos de Descomposición de Tendencia Estacional,

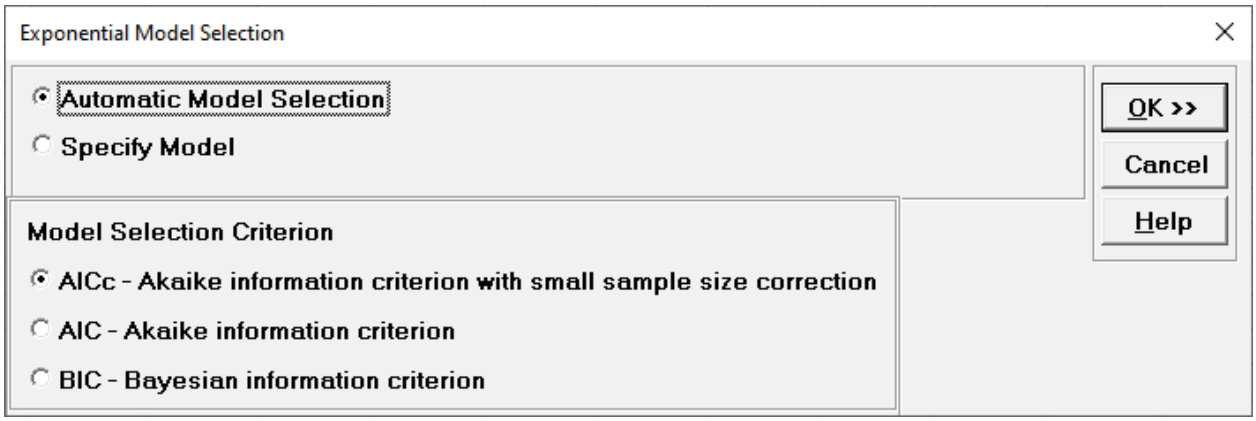
14. Hacer click en **SigmaXL > Time Series Forecasting > Suavización exponencial Forecast > Multiple Seasonal Decomposition Forecast**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

15. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**. Marcar **Display FAC/FACP/LB Plots** y **Display Residual Plots**. Marcar **Specify Model Periods**. Set **Retention Periods** = 96. Seleccionar **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Marcar **Seasonal Frequency** with **Specify** = 48 336**.** Desmarcar **Box-Cox Transformation**. Se hace uso por defecto **Prediction Interval** = 95.0 %.

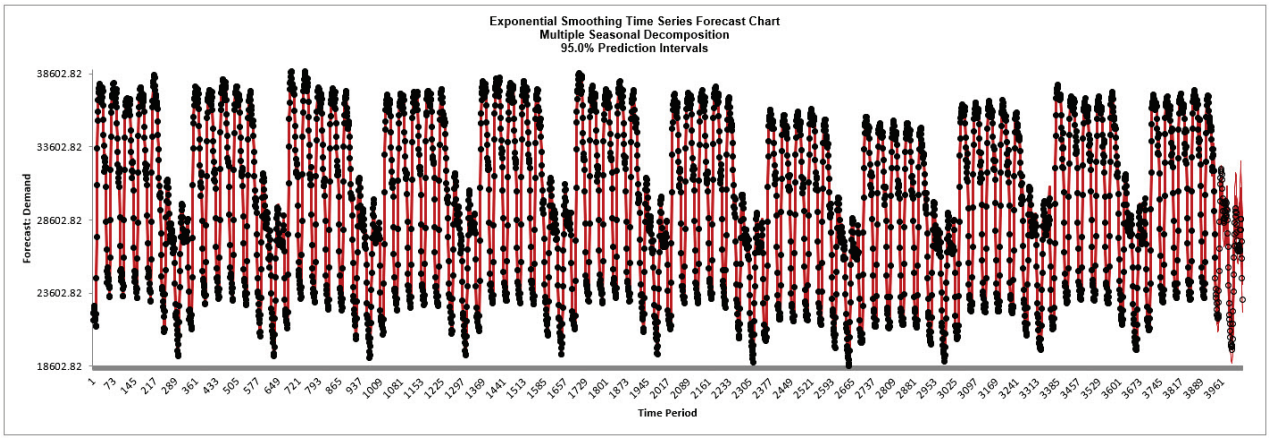


Withhold Periods es 2*la frecuencia estacional dominante (48). La frecuencia dominante se obtiene de los Gráficos de Densidad Espectral. Start Model at Period = 1 se encuentra siempre inhabilitada para la DEM.

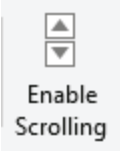
16. Hacer click en **Model Options**.



17. Se empleará por defecto **Automatic Model Selection** con **AICc** como el **Model Selection Criterion**. Hacer click en **OK** para regresar al diálogo Pronóstico de Suavización Exponencial DEM. Hacer click en **OK**. Se proporciona el informe del pronóstico de suavización exponencial:

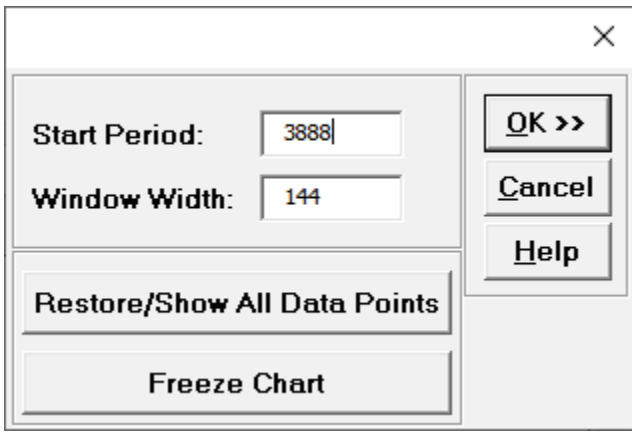


18. Se puede ampliar en los últimos 3 días, esto es, en el periodo de tiempo de las 144 media-horas, con el uso del desplazamiento de la carta. Hacer click en **SigmaXL Chart Tools > Enable Scrolling**



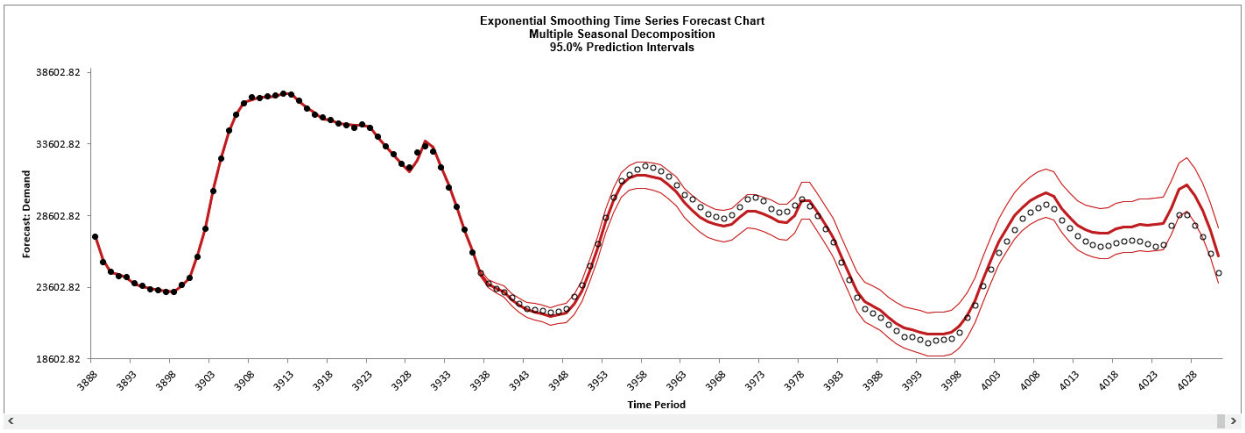
Puede que se le indique con un mensaje de advertencia que el formato personalizado del gráfico se borrará. Puede evitar ver esta advertencia marcar **Save this choice as default and do not show this form again**.

19. Hacer click en **OK.** Aparece el diálogo de desplazamiento lo que permite especificar el **Start Period** y **Window Width**. Registrar **Start Period** = 3888 and **Window Width** = 144:



Puede hacer click en cualquier instante en **Restore/Show All Data Points** o **Freeze Chart**. Al congelar el gráfico se removerá el desplazamiento y se descargará el diálogo. El diálogo del desplazamiento se cargará también si cambia la hoja. Hacer click en **SigmaXL Chart Tools > Enable Scrolling** para reestablecer el diálogo.

20. Hacer click en **OK.** Aparece una barra del desplazamiento debajo del gráfico del pronóstico. Puede cambiar también el **Start Subgroup** y **Window Width** y **Update**.



Puede desplazarse haciendo click a la derecha o a la izquierda, con el ancho de la ventana especificado en 144.

Los puntos blancos son los valores de los datos de la muestra que se retiene con un pronóstico de varios pasos e intervalos de predicción mostrados al inicio de la muestra que se retiene. El modelo funciona bastante bien para predecir los valores de demanda de las retenciones de 96 medias horas. Obsérvese que el error de predicción aumenta cuanto más lejos se pronostica.

Hacer click en **Cancel** para salir del diálogo del desplazamiento.

21. Desplazarse hacia abajo para observar el título del Modelo de Suavización Exponencial:

Exponential Smoothing Model (Multiple Seasonal Decomposition): Simple Exponential Smoothing with Multiplicative Errors (M, N, N) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 96 withhold periods.

Después de la **Descomposición Estacional Múltiple**, el modelo es una **Suavización Exponencial Simple con Errores Multiplicativos (M, N, N).**

Se seleccionó automáticamente el modelo **(M, N, N)** como el mejor ajuste para los datos desestacionalizados de Demanda que se basan en el criterio AICc.

El encabezado incluye también el número de periodos de retención especificados.

22. Se genera la Información del Modelo de Suavización Exponencial:

Exponential Smoothing Model Information	
Seasonal Frequency	48, 336 Decomposition
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

Este es el resumen de la información del modelo con Seasonal Frequency = 48, 336 utilizando la descomposición y el Model Selection Criterion = “AICc”.

23. La Estimación de los Parámetros para los datos de la Demanda desestacionalizada son:

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.9999
l (initial level)	29798.08498

- El error incluye el parámetro de suavización *alfa* y el valor del nivel inicial (*l*). Con *alfa* = 0.9999, esto es efectivamente un pronóstico naïve, con errores multiplicativos.
- No se calcula el parámetro estacional de suavización (*gamma*) porque se han desestacionalizado los datos.

24. Las Estadísticas del Modelo de Suavización Exponencial son:

Exponential Smoothing Model Statistics	
No. of Observations	3936
DF	3934
StDev	0.003336007
Variance	1.11289E-05
Log-Likelihood	-34366.34649
AICc	68738.69908
AIC	68738.69298
BIC	68757.52674

- El número de observaciones, *n* = 4032 – 96 (*retención*) = 3936
- Los grados de libertad (DF) = 3936 (*n*) – 2 (los términos en el modelo) = 3934
- Tener en cuenta que las estadísticas del modelo se calculan utilizando los datos desestacionalizados y los residuos son errores multiplicativos (relativos): $e_t = \frac{y_t - \hat{y}_t}{\hat{y}_t}$.

25. Las métricas de la Precisión del Pronóstico son:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Multi-Step-Ahead Forecast
N	3936	96
RMSE	101.5029509	776.7232991
MAE	73.73426784	649.5237662
MAPE	0.253138484	2.45436482
MASE	0.040157098	0.353743116

Como era esperarse, los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un Paso-Hacia-Adelante** son mayores que los errores del Pronóstico **Dentro-de-la-Muestra (Estimación) Un Paso-Hacia-Adelante**. Tener en cuenta que si se estuviera interesado principalmente en un pronóstico Un-paso-Hacia-Adelante a corto plazo, entonces se hubiera seleccionado **Pronóstico de Retención Tipo: Un Paso-Hacia-Adelante** y la tabla de arriba mostraría los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un Paso-Hacia-Adelante**.

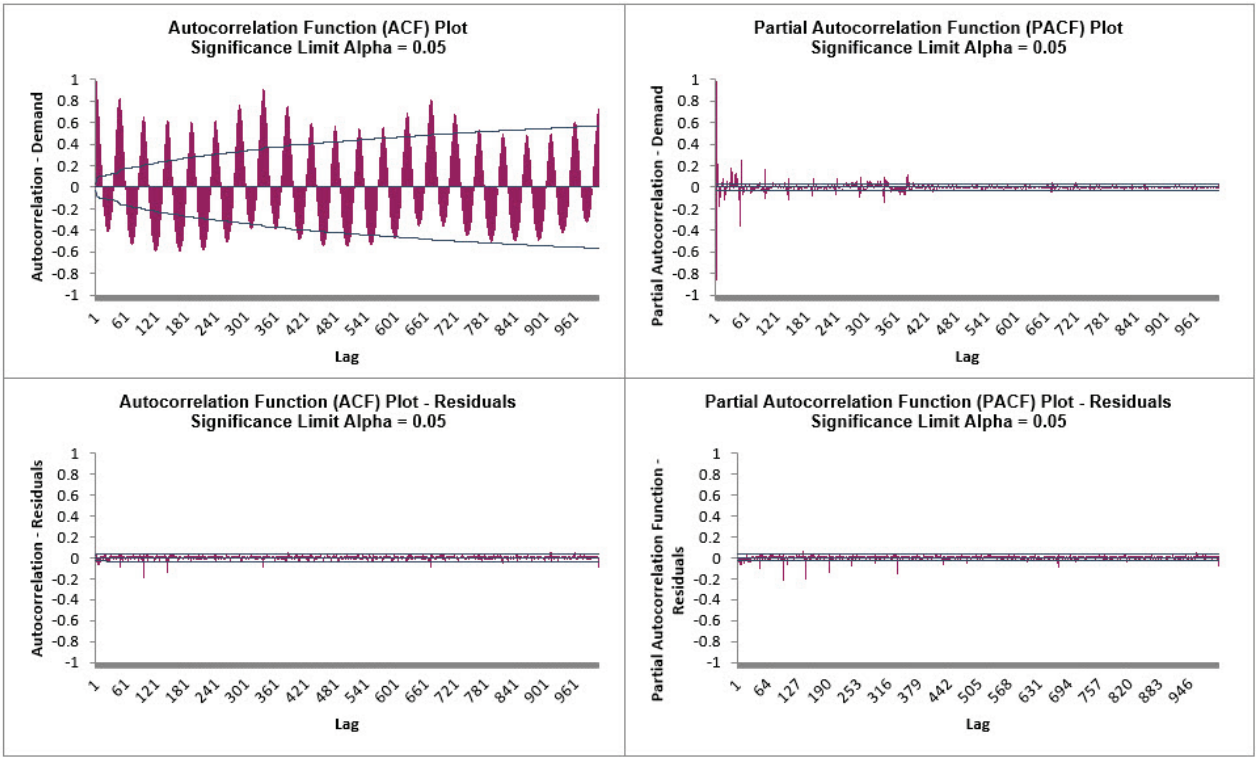
Las métricas de la Precisión del Pronóstico se calculan utilizando los datos reales sin procesar versus el pronóstico tal como se visualizan en el Gráfico y la Tabla de Pronóstico, de modo que, a diferencia de las estadísticas del modelo anterior, permite la comparación entre todos los tipos y transformaciones del modelo de pronóstico.

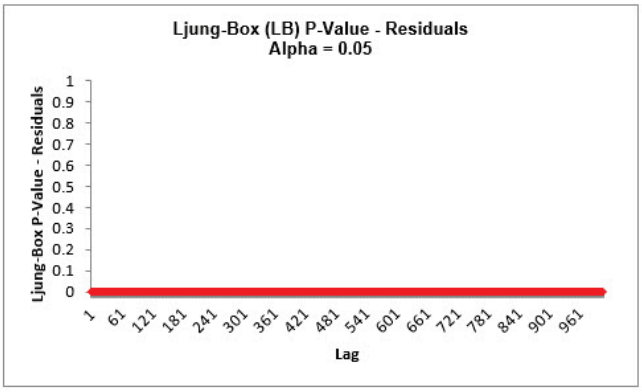
26. Se genera la Tabla de Pronóstico:

Forecast Table				
Period	Withhold Data	Multi-Step-Ahead Forecast	Lower 95.0% PI	Upper 95.0% PI
3937	24653	24595.68006	24399.19061	24792.1695
3938	23879	23839.01162	23561.1467	24116.87654
3939	23508	23546.05076	23205.74185	23886.35968
3940	23275	23300.43291	22907.48021	23693.38561
3941	22890	22735.89622	22296.56272	23175.22972
3942	22462	22349.79507	21868.5296	22831.06054

Estos son los mismos valores del pronóstico e intervalos de predicción que se visualizan en el Gráfico del Pronóstico, pero al que se le ha agregado análisis y gráficos extras (por ejemplo: una Carta de Corrida para los errores del pronóstico). Se visualiza también los **Datos de la Retención**.

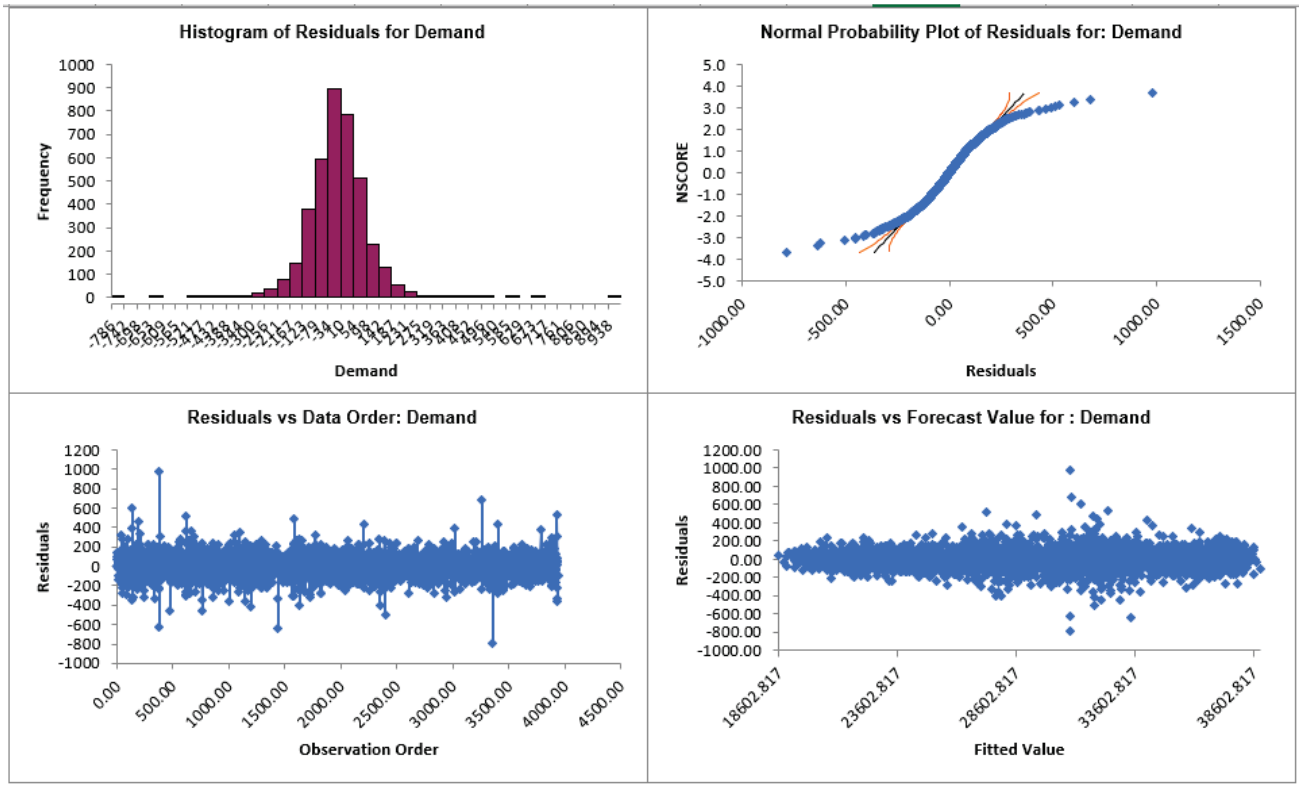
27. Hacer click en la hoja **Exp Smooth MSD ACF PACF LB** para visualizar los Gráficos de FAC/FACP/LB:





Los Gráficos de FAC/FACF de los Residuos indican que gran parte de la autocorrelación se ha considerado en el modelo, pero el gráfico de Ljung-Box muestra que aún queda alguna autocorrelación significativa (los valores P rojos son significativos con alfa=0,05) - por lo que el modelo puede ser potencialmente mejorado. Esto no significa que el modelo sea un mal modelo, todavía puede ser muy útil para fines de predicción, pero los intervalos de predicción pueden no proporcionar una cobertura precisa.

28. Hacer click en la hoja **Exp Smooth MSD Residuals** para visualizar los Gráficos de los Residuos:



Los residuos no se distribuyen normalmente y hay valores extremos atípicos. Estos deben ser investigados con una carta de control sobre los residuos. Los valores atípicos en la demanda de electricidad se explican a menudo por la temperatura y esto es algo que se considerará más

adelante con un conjunto de datos diferentes (Demanda Diaria de Electricidad con Predictoras – ElecDaily).

29. Tener en cuenta que los Residuos de la DEM son los valores finales observados – los pronosticados, por lo que no hay diferencias de escala si el modelo utiliza el error aditivo o multiplicativo. Si se utiliza una transformación Box-Cox, entonces los residuos están en unidades transformadas.

Carta de Control de Suavización Exponencial

Control Estadístico de Procesos (CEP) para Datos Auto correlacionados

Se crea una carta de control de Individuales utilizando los residuos del modelo de pronósticos de Suavización Exponencial.

La carta de Límites Móviles utiliza la predicción de un paso como la línea central, por lo que los límites de control se moverán con la línea central. Si se utiliza una transformación Box-Cox, se aplica una transformación inversa para calcular los límites de control. Si los residuos son de suavización exponencial multiplicativa, los límites de control son aproximados y las señales fuera de control pueden no coincidir exactamente con la Carta de Individuales. Si eso ocurre, se debe utilizar la Carta de Individuales para determinar qué puntos están fuera de control.

Las Características conocidas "Add Data", "Show Last 30" y "Scroll" en SigmaXL Chart Tools están disponibles para estas cartas de control. Para “Add Data”, los modelos de series de tiempo no son reajustados, sino que se usan para calcular los valores residuos de los nuevos datos.

Para mayores detalles y referencias, véase el Apéndice: **Cartas de Control para Datos Autocorrelacionados**.

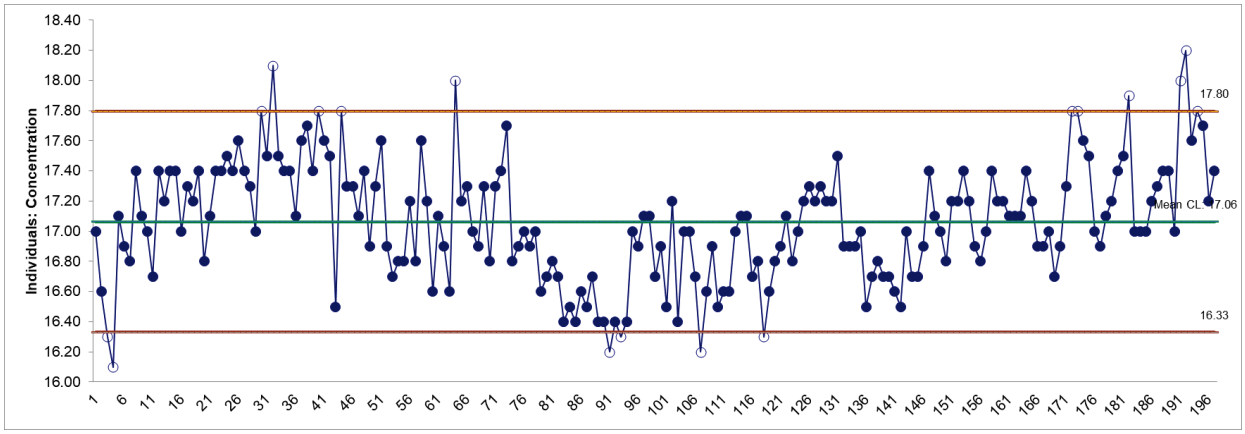
Tener en cuenta que la Carta de Rangos Móviles y las Pruebas de Causas Especiales no están disponibles aquí, pero el usuario puede almacenar y Seleccionar Residuos, entonces crearlos con **SigmaXL > Control Charts > Individuals & Moving Range**.

Concentración de Proceso Químico - Serie A

1. Abrir **Chemical Process Concentration – Series A.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie A de Box y Jenkins, un conjunto de 197 valores de la concentración de un proceso químico tomados a intervalos de 2 horas. Véase para estos datos Cartas de Corrida, Gráficos FAC/FACP, Gráfico de Densidad Espectral, y Gráfico de Descomposición de Tendencia Estacional.

2. Previamente se observó que este proceso tiene una autocorrelación significativa. Con el fin de apreciar el impacto sobre la carta de control, se construirá cartas de Individuales con los datos sin procesar. Hacer click en **SigmaXL > Control Charts > Individuals**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

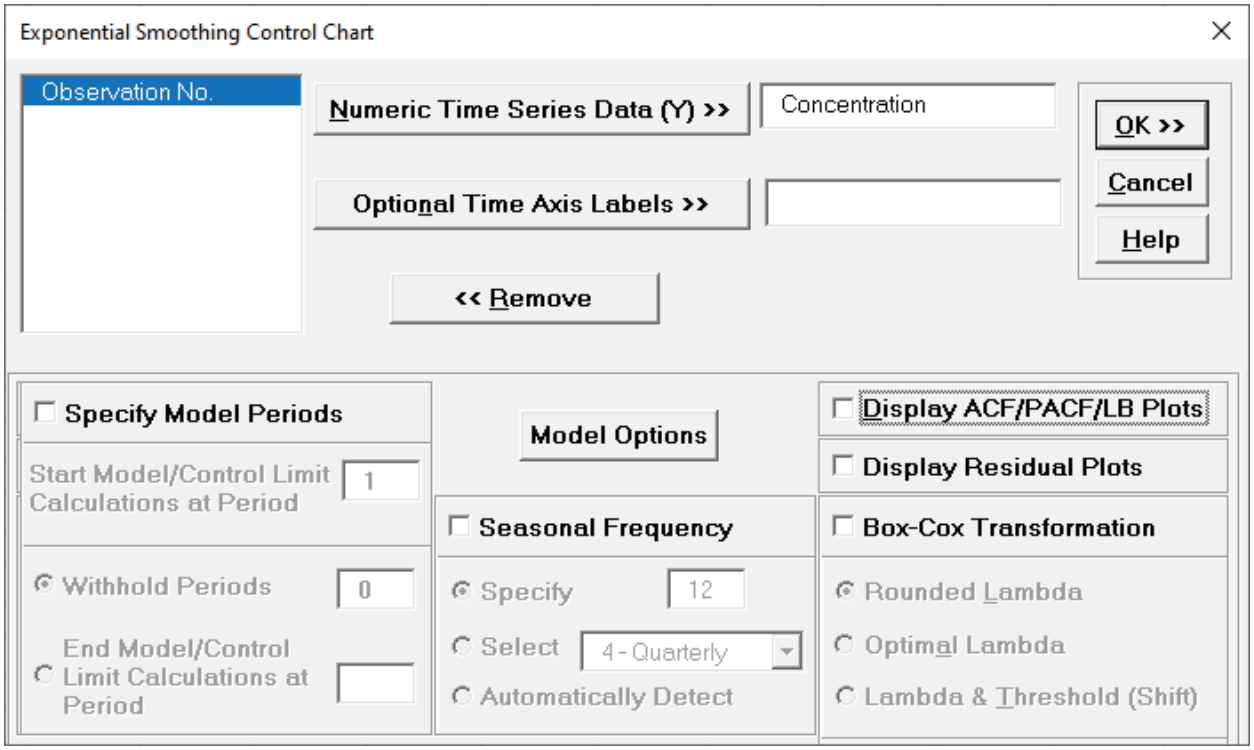
3. Seleccionar *Concentration*, hacer click en **Numeric Data Variable (Y) >>**. Hacer click en **OK**. Se genera una Carta de Control de Individuales:



Existen 17 puntos de datos fuera de control, principalmente debido a la autocorrelación. La búsqueda de causas asignables con el uso de esta carta parecería inútil.

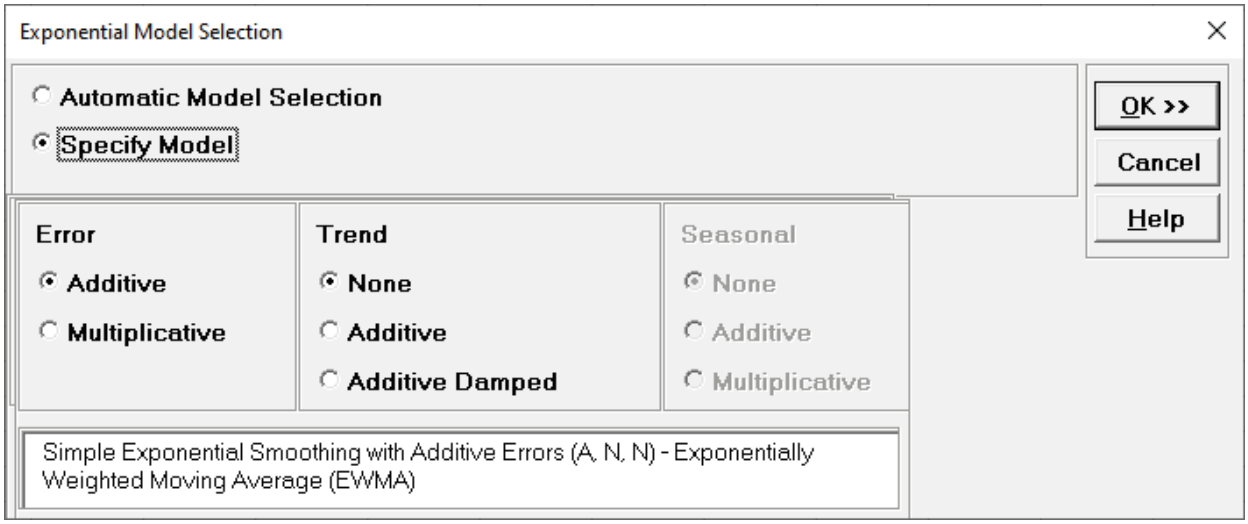
4. Hacer ahora click en la pestaña **Hoja 1** y **SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Control Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

5. Seleccionar *Concentration*, hacer click en **Numeric Time Series Data (Y) >>**. Desmarcar **Display ACF/PACF/LB Plots.** Desmarcar **Display Residual Plots**, **Specify Model Periods, Seasonal Frequency** y **Box-Cox Transformation**.

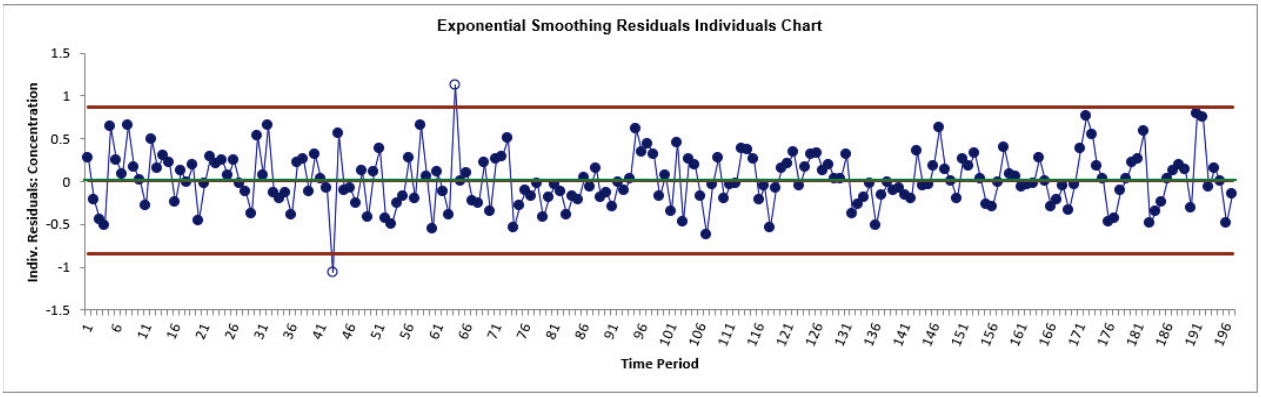


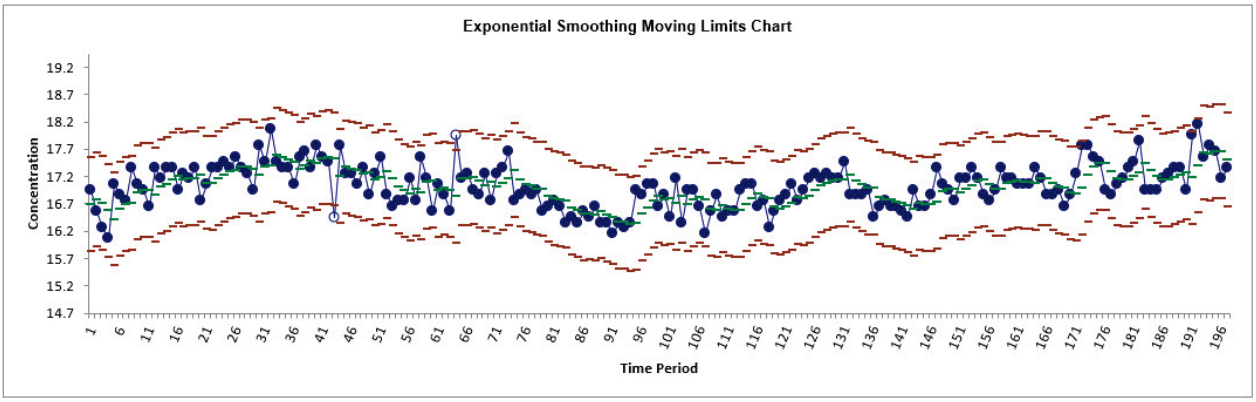
En vista que se correrá el mismo modelo (A, N, N) que se usó antes, no se necesitarán los gráficos de los residuos ni los de FAC/FAC/LB.

6. Hacer click en **Model Options**. Seleccionar **Specify Model**.



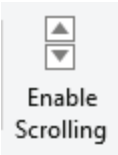
7. Se hace uso por defecto **Error: Additive** y **Trend: None**, que es un modelo de suavización exponencial simple, o Promedio Móvil Ponderado Exponencialmente (EWMA). Hacer click en **OK** para regresar al diálogo de la Carta de Control de Suavización Exponencial. Hacer click en **OK**. Se genera las cartas de control de suavización exponencial:





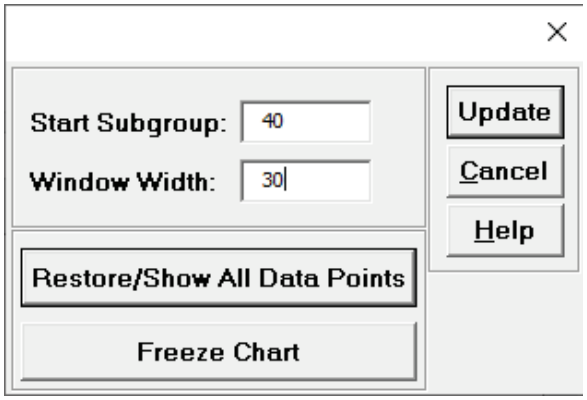
Se dispone sólo de dos puntos fuera de control en la carta de Individuales que se debe investigar. La carta de Límites Móviles emplea la predicción de un paso como línea central, de manera que los límites de control se mueven con la línea central.

8. Se puede desplazar a través de los puntos de la carta de datos. Hacer click en **SigmaXL Chart Tools > Enable Scrolling**

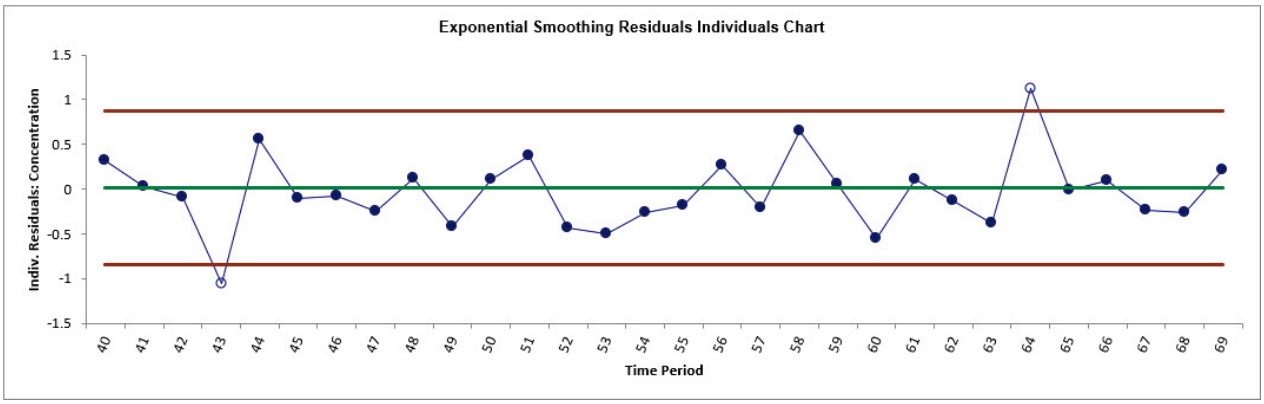


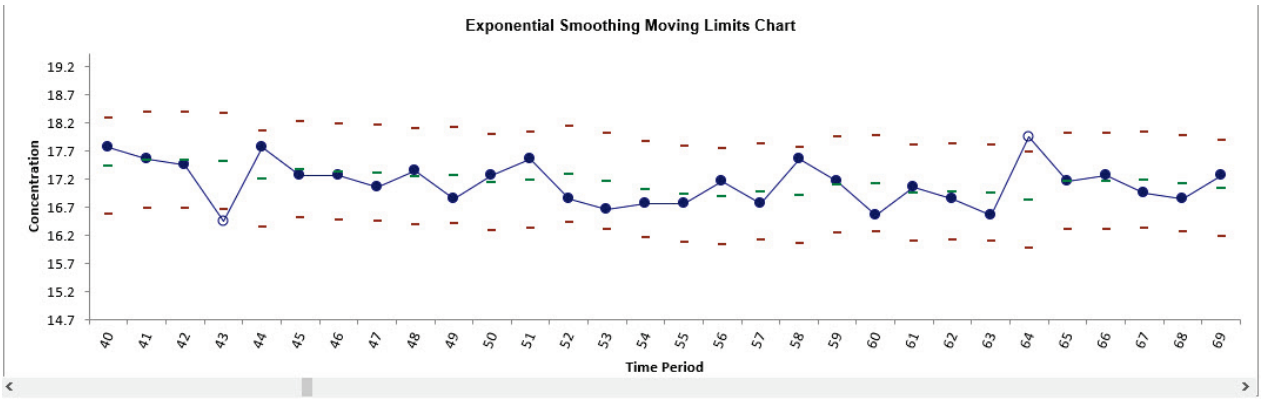
Puede que se le indique con un mensaje de advertencia que el formato personalizado del gráfico se borrará. Se puede evitar ver esta advertencia al marcar **Save this choice as default and do not show this form again**.

9. Hacer click en **OK.** El diálogo del desplazamiento, al parecer, permite especificar al **Start Subgroup** y al **Window Width**. Registrar **Start Subgroup** = 40 y el **Window Width** = 30 para visualizar los dos puntos de datos fuera de control.



10. Hacer click en **OK.** Esto permite ampliar los puntos fuera de control 43 y 64.





La observación **43** del modelo de pronóstico de suavización exponencial es menor de lo esperado. La observación **64** es mayor de lo esperado.

11. Hacer click en **Cancel** para salir del diálogo de desplazamiento.

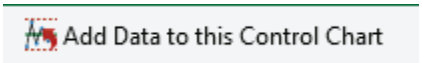
12. Se agregará un nuevo punto de dato a la Serie A de datos de la Concentración. Se calcularán los residuos utilizando el mismo modelo que el anterior sin reestimar los parámetros del modelo ni recalcular los límites de control. Esto se conoce también como la aplicación de la "Fase II" de

un gráfico de control, en la que una señal fuera de control debe dar lugar a una investigación de la causa asignable y a la aplicación de una acción correctiva o un ajuste del proceso. Hacer click en la Hoja 1, registrar el valor 19 tal como se muestra en la celda **B199** (y de forma opcional el número de la observación 198 en la celda **A199**).

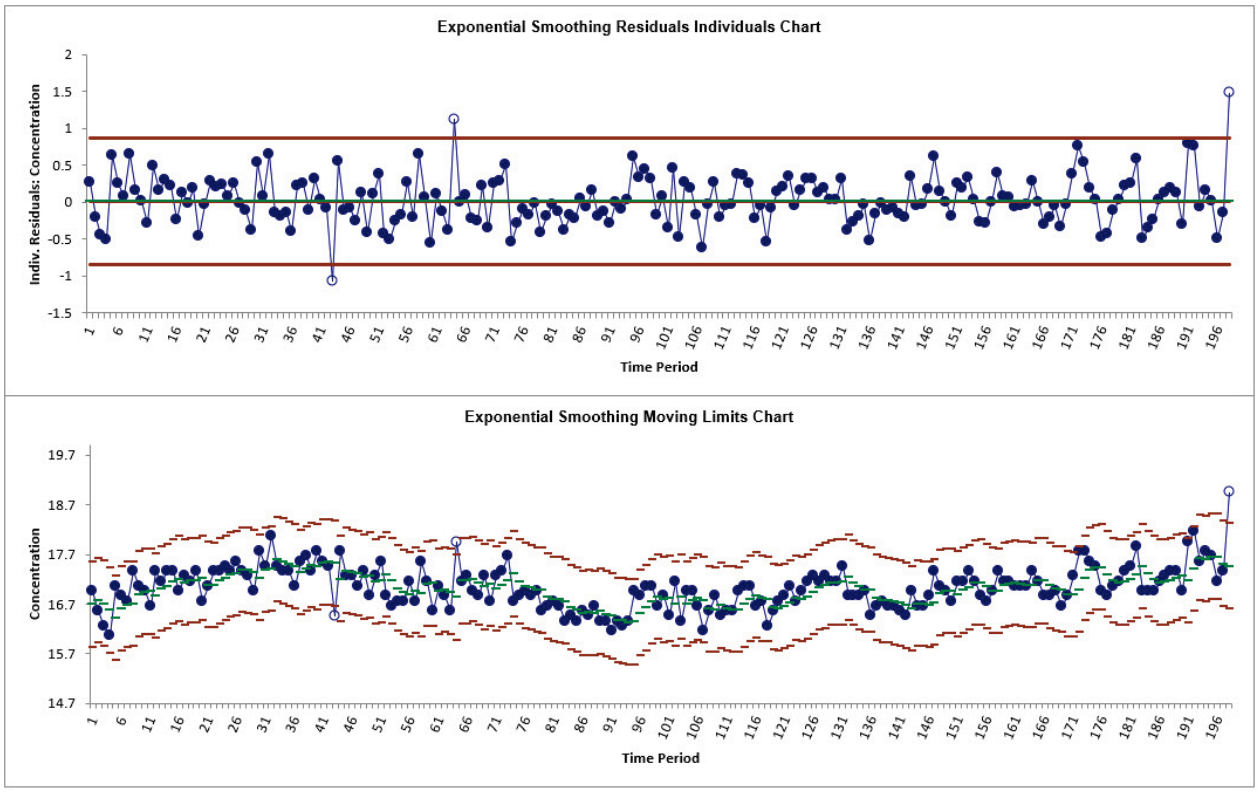
196	195	17.7
197	196	17.2
198	197	17.4
199	198	19

13. Hacer click en la pestaña **Exp. Smoothing Control Charts** (si existe más de una carta de control en la hoja de ejercicio, seleccionar, por favor, la gráfica donde se han de agregar los datos).

14. Hacer click en **SigmaXL Chart Tools > Add Data to this Control Chart**



15. Se actualizan ahora con nuevos datos los residuos de la Carta de Control y los Límites Móviles de las Cartas, ilustrando con esto un punto de dato fuera de control:



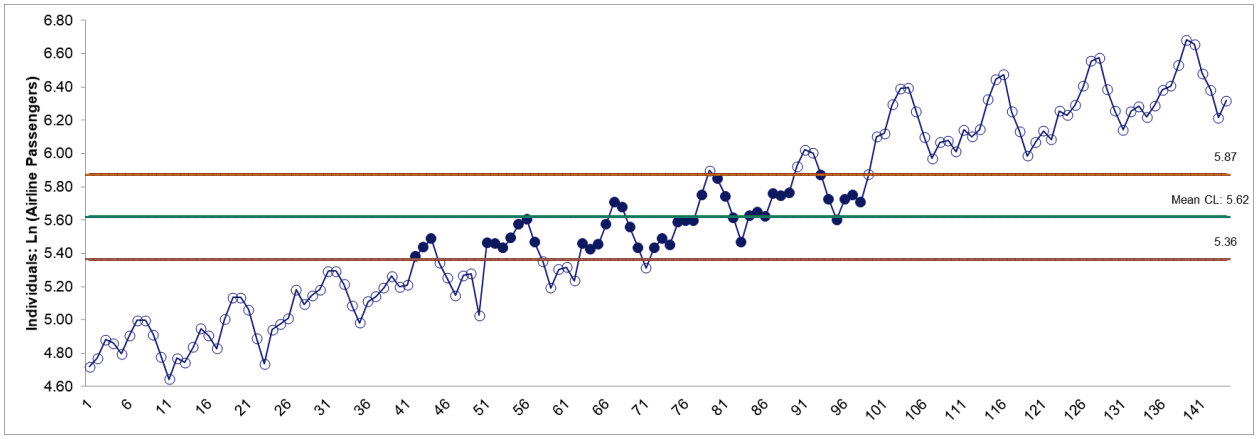
16. Se recomienda renombrar el libro de ejercicios como **Chemical Process Concentration – Series A_AddData.xlsx**, de tal manera que el uso posterior de los datos de la Concentración no incluya el punto de dato añadido.

Pasajeros Mensuales de la Aerolínea - Modificado para Cartas de Control

17. Abrir **Monthly Airline Passengers – Modified for Control Charts.xlsx** (pestaña de la Hoja 1). Estos datos son una Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional desde enero 1949 hasta diciembre 1960. Se aplica una transformación Ln (evitando el uso de una transformación Box-Cox), se añade un valor atípico en el dato 50 (-0.25), y se aplica un cambio de nivel en el dato 100 (+0.25). Se agregan variables codificadas para ayudar a distinguir un valor atípico versus un cambio de nivel, pero se analizarán posteriormente con el método de pronósticos ARIMA con Predictoras. La Suavización Exponencial no permite el uso de predictoras.

18. Anteriormente se pudo apreciar que este proceso tiene una autocorrelación significativa con una fuerte tendencia y estacionalidad. Para ver el impacto en una carta de control, se construirá una carta de individuales con los datos sin procesar. Hacer click en **SigmaXL > Control Charts > Individuals**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

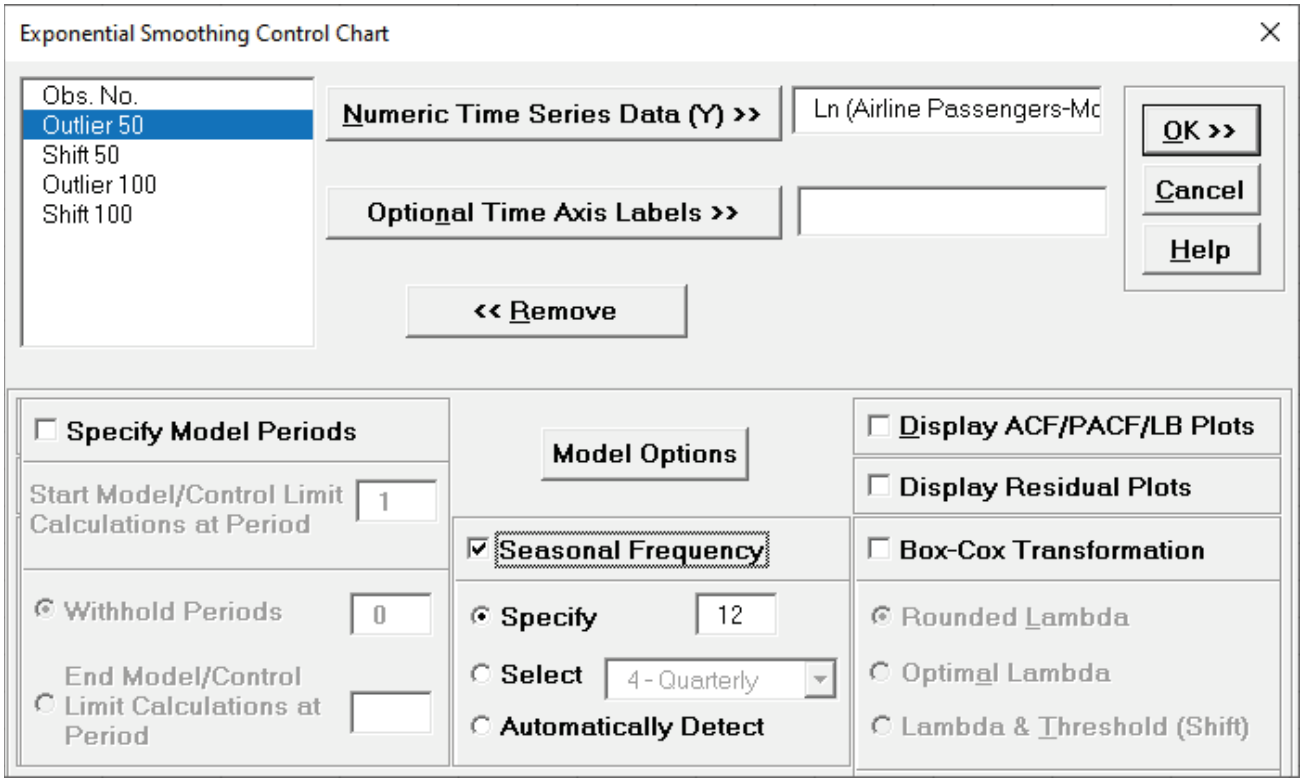
19. Seleccionar *Ln (Airline Passengers-Modified)*, hacer click en **Numeric Data Variable (Y) >>**. Hacer click en **OK**. Se genera una Carta de Control de Individuales:



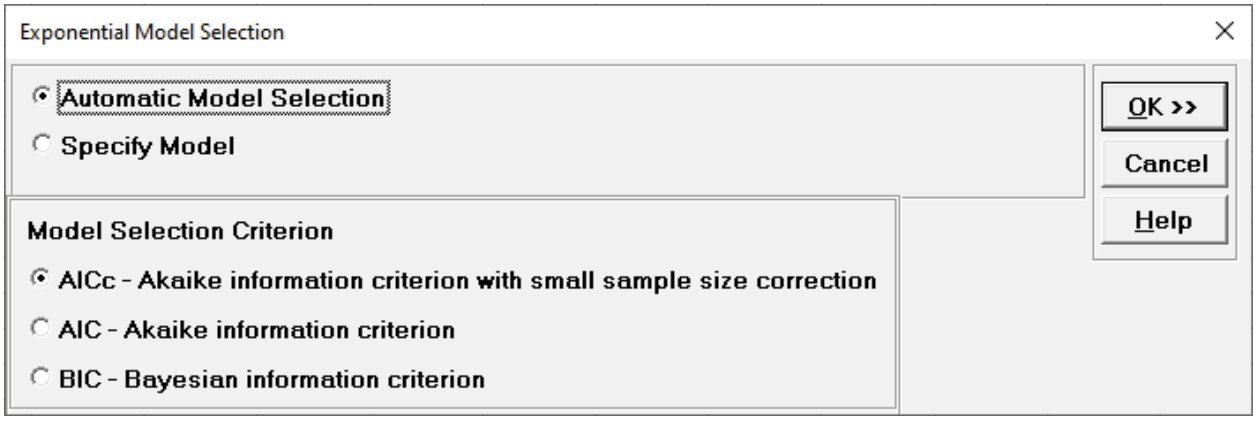
Con una fuerte tendencia, estacionalidad y autocorrelación positiva, esta carta de control no tiene sentido.

20. Hacer ahora click en la pestaña de la **Hoja 1** y **SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Control Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

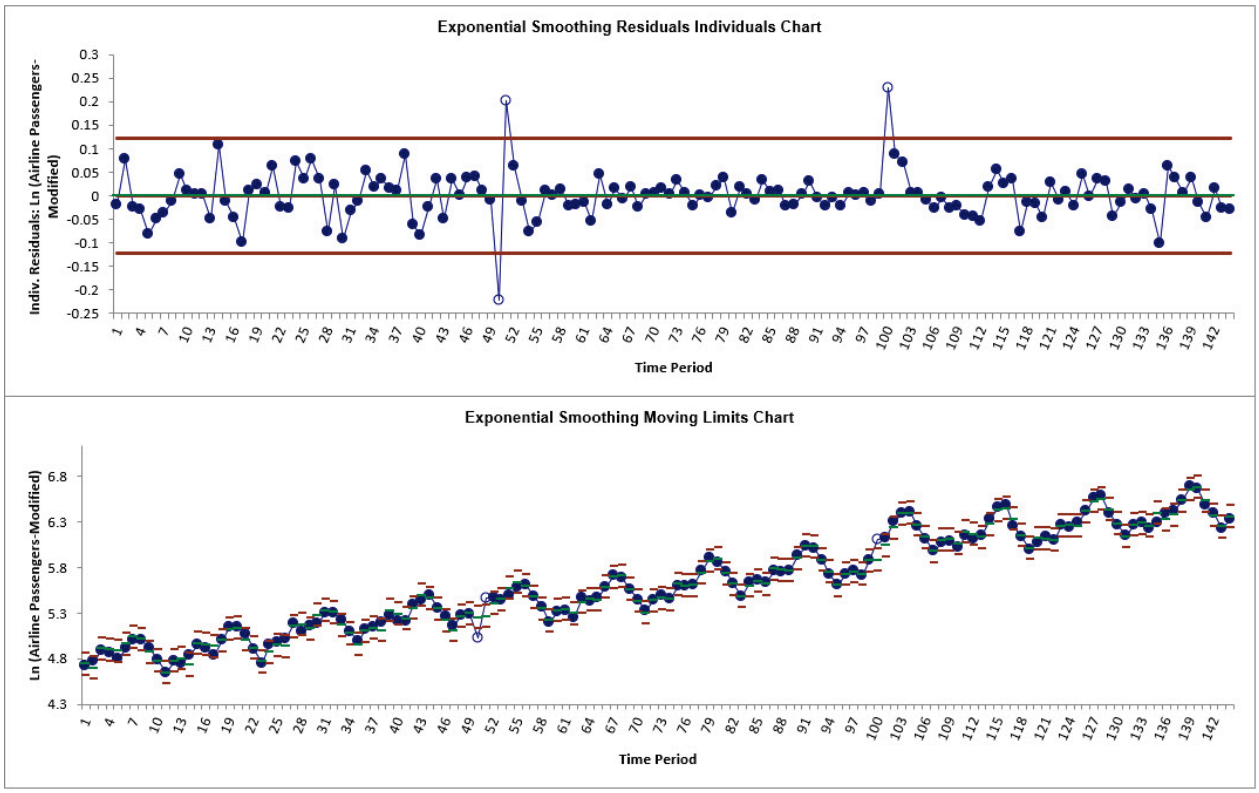
21. Seleccionar *Ln (Airline Passengers-Modified)*, hacer click en **Numeric Time Series Data (Y) >>**. Desmarcar **Display ACF/PACF/LB Plots** y **Display Residual Plots.** Marcar **Seasonal Frequency** con **Specify** = 12. Desmarcar **Model Periods** y **Box-Cox Transformation**.



22. Hacer click en **Model Options**.

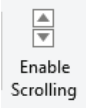


23. Se empleará por defecto **Automatic Model Selection** con **AICc** como el **Model Selection Criterion**. Hacer click en **OK** para regresar al diálogo de la Carta de Control de Suavización Exponencial. Hacer click en **OK**. Se generan Cartas de Control de Individuales:



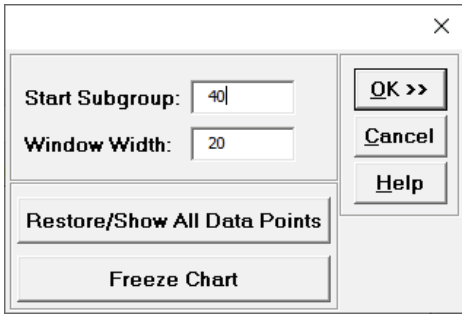
Ahora se puede observar en forma clara los puntos de datos fuera de control en los periodos 50, 51 y 100 en la carta de Individuales. Para visualizar los puntos de la carta de Límites Móviles se hará uso del desplazamiento.

24. Hacer click en **SigmaXL Chart Tools > Enable Scrolling**

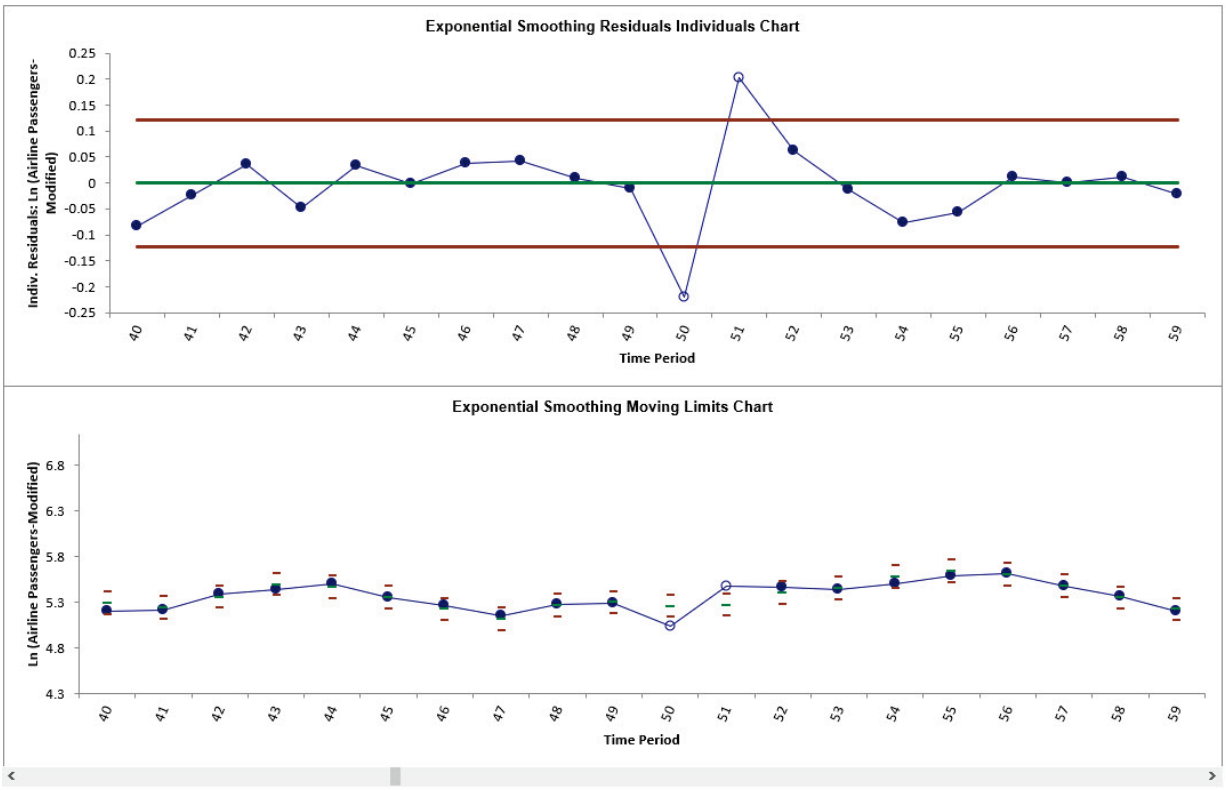


Puede que se le indique con un mensaje de advertencia que se borrará el formato personalizado de la carta. Puede evitar ver esta advertencia al marcar **Save this choice as default and do not show this form again**.

25. Hacer click en **OK.** Aparece el diálogo de desplazamiento lo que permite especificar el **Start Subgroup** y **Window Width**. Registrar **Start Subgroup** = 40 and **Window Width** = 20 para visualizar los primeros dos puntos de datos fuera de control.



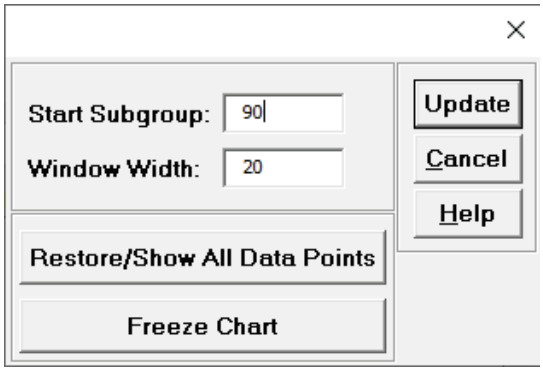
26. Hacer click en **OK.** Esto permite la ampliación en los puntos 50 y 51 fuera de control.



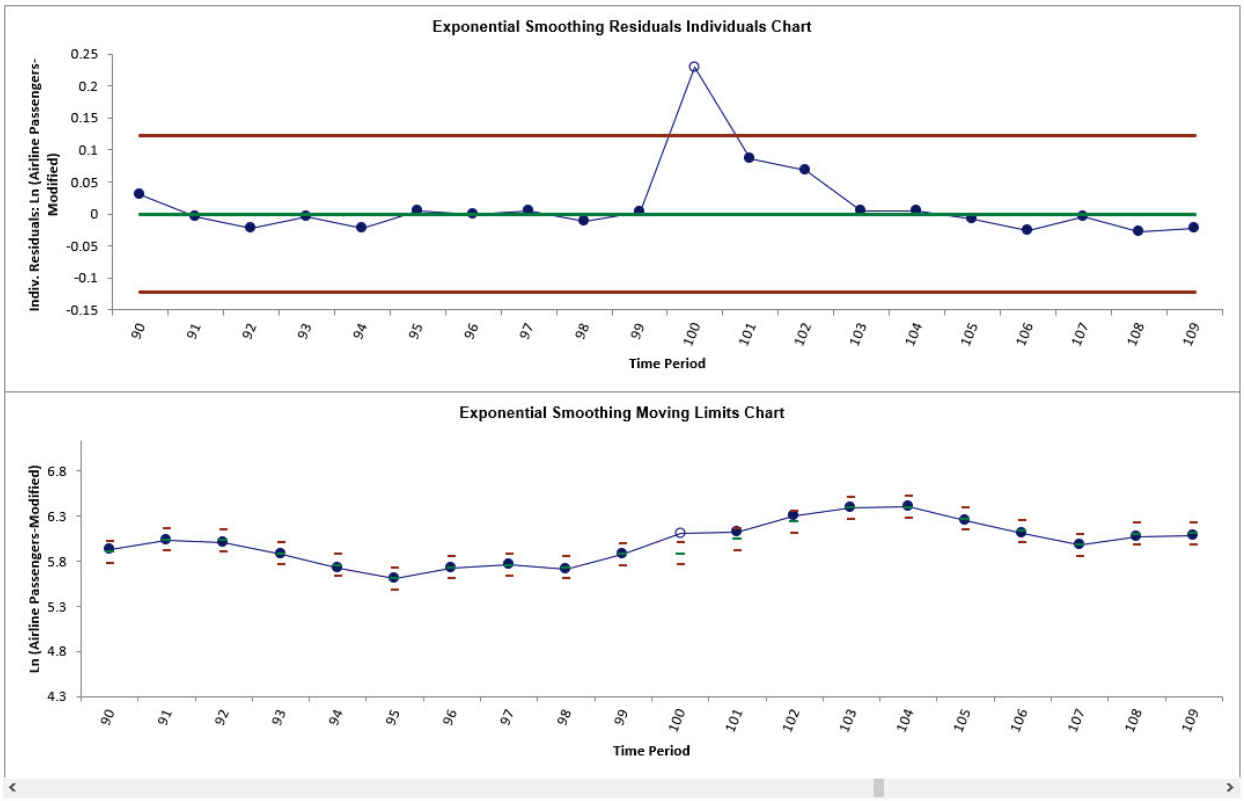
La observación **50** es menor que la que se espera del modelo de pronóstico de suavización exponencial. La observación **51** es mayor de la esperada. Una posterior investigación revelará que esto es debido a un solo valor atípico negativo.

Consejo: El desplazamiento mantiene el ajuste mínimo y máximo del eje Y original. Puede que desee cambiar esto a automático al hacer click en el eje Y, hacer click derecho en **Format Axis**, hacer click en Bounds Minimum **Reset** y Bounds Maximum **Reset.** Esto cambia los ajustes del eje a Auto para que cuando se desplace o actualice el eje Y se ajuste también automáticamente.

27. Registrar ahora **Start Subgroup** = 90 y **Window Width** = 20 para visualizar el tercer punto de datos fuera de control.



28. Hacer click en **Update**.



La observación **100** es mayor que la que se espera del modelo de pronóstico de suavización exponencial. Una posterior investigación revelará que esto es debido a un cambio en la media.

29. Hacer click en **Cancel** para salir del diálogo de desplazamiento.

30. Desplazar hacia abajo el encabezado para ver el Modelo de Suavización Exponencial

Exponential Smoothing Model: Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters) (A, A, A) - Model Automatically Selected
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

Se seleccionó automáticamente el modelo del Método de **Tendencia Aditiva Estacional Aditivo con Errores Aditivos (Holt-Winters) (A, A, A)** como el mejor ajuste para los datos Ln modificados de Pasajeros de la Aerolínea con base al criterio AICc.

31. Las Estimaciones de los Parámetros y las Estadísticas del Modelo de Suavización Exponencial son un poco diferentes que las correspondientes al anterior análisis debido a que se introdujo un valor atípico y un cambio de nivel, así también se están utilizando todos los datos, es decir, no existen periodos de retención. Tomar en cuenta que anteriormente se aplicó la Transformación Box-Cox con Lambda=0 y aquí, en cambio, se está aplicando Ln de los datos.

32. Las métricas de la Precisión del Pronóstico son:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	144	
RMSE	0.049508351	
MAE	0.033506668	
MAPE	0.609538124	
MASE	0.230072112	

Tomar en cuenta que estos errores del pronóstico son muy diferentes a los correspondientes al anterior análisis en donde se calculó el error del pronóstico en los datos sin procesar versus los valores finales pronosticados, pero con los datos Ln de Pasajeros de la Aerolínea.

Cartas de Control Suavización Exponencial Descomposición Estacional Múltiple (DEM)

La Suavización Exponencial se limita a una frecuencia estacional máxima de 24. Para mayores frecuencias aplique la Suavización Exponencial – Descomposición Estacional Múltiple (DEM). Se remueve al componente estacional a través de la descomposición, un modelo exponencial no estacional se ajusta a los residuos (+tendencia), y entonces se vuelve a agregar el componente estacional.

Como su nombre indica, la Descomposición Estacional Múltiple (DSM) también da cabida a la estacionalidad múltiple, por ejemplo, los datos de media hora con una frecuencia estacional de 48 observaciones por día y 336 observaciones por semana.

Para este método de pronóstico se crea una carta de control de individuales de los residuos. La carta de Límites Móviles utiliza la predicción de un paso como la línea central, por lo que los límites de control se moverán con la línea central. Si se utiliza una transformación Box-Cox, entonces se aplica una transformación inversa para calcular los límites de control. Si los residuos son de suavización exponencial multiplicativa, los límites móviles de control son aproximados y las señales fuera de control pueden no coincidir exactamente con la Carta de Individuales. Si eso ocurre, se debe usar la Carta de Individuales para determinar qué puntos están fuera de control.

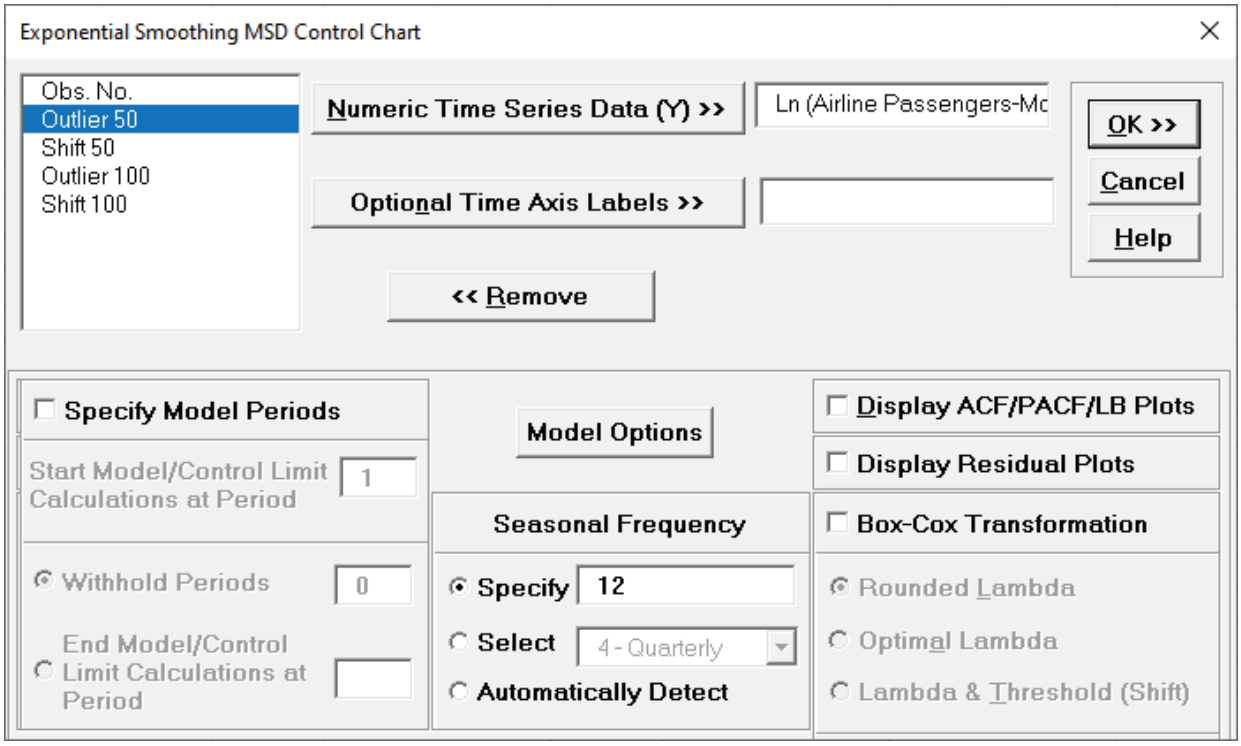
Las características conocidas “Add Data”, “Show Last 30” y “Scroll” de SigmaXL Chart Tools están disponibles para estas cartas de control. En el caso de "Add Data", los modelos de series de tiempo no se reajustan, sino que se utilizan para calcular los valores residuos de los nuevos datos.

Pasajeros Mensuales de la Aerolínea - Modificado para Cartas de Control

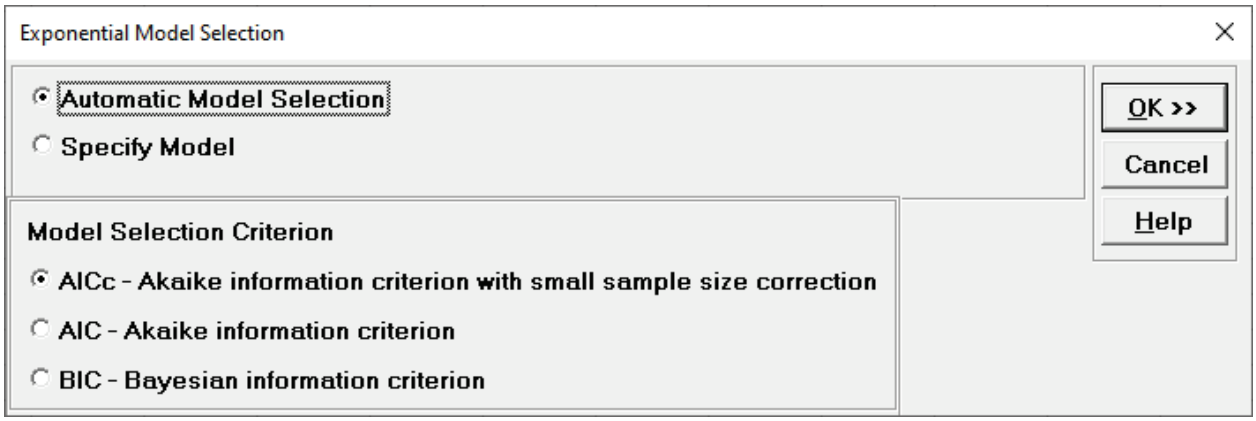
1. Abrir **Monthly Passengers Airline – Modified for Control Charts.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional desde enero 1949 hasta diciembre 1960. Se aplica una transformación Ln (evitando el uso de una transformación Box-Cox), se agrega un valor atípico en el dato 50 (-0.25) y se aplica un cambio de nivel (+0.25), con el inicio en 100. La opción de Descomposición Estacional Múltiple (MSD) no es necesaria para estos datos, pero como una forma de introducción será útil para compararlo con el análisis previo.

2. Hacer click en **SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Multiple Seasonal Decomposition Control Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

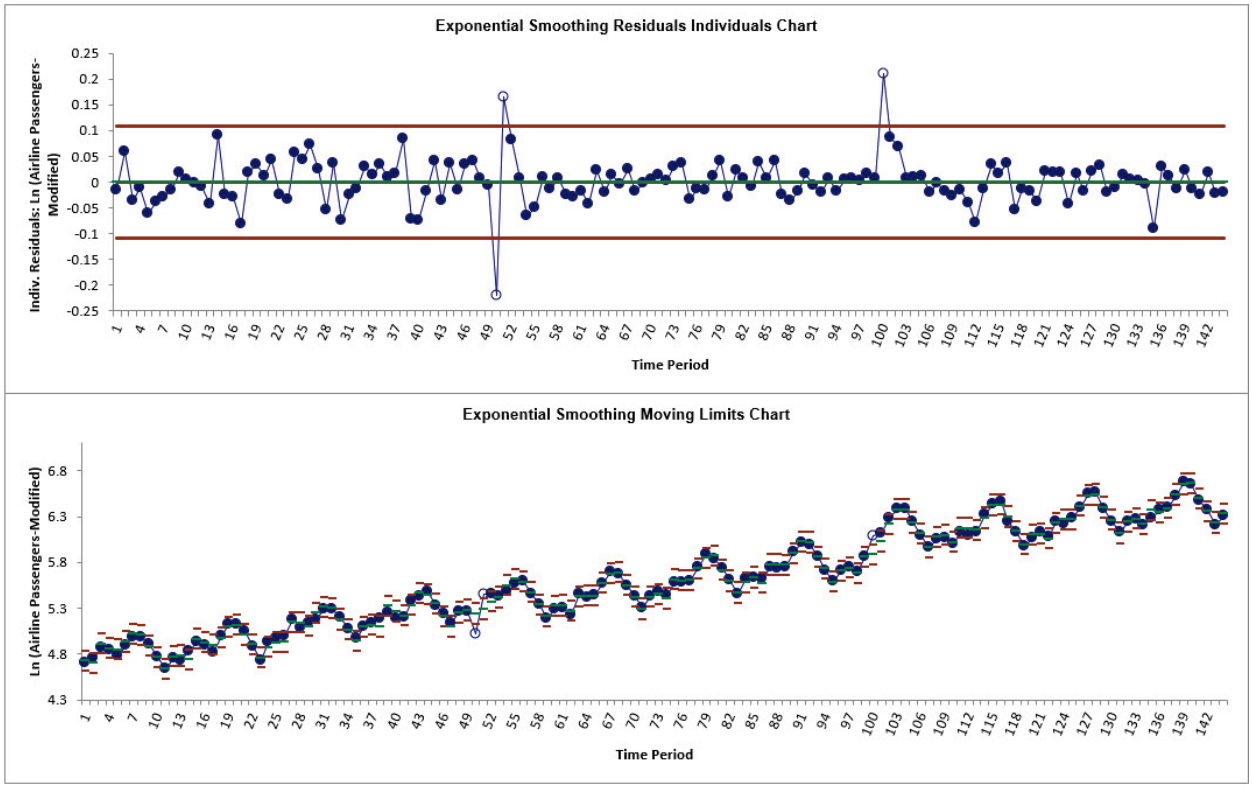
3. Seleccionar *Ln (Airline Passengers-Modified)*, hacer click en **Numeric Time Series Data (Y) >>**. Desmarcar **Display ACF/PACF/LB Plots** y **Display Residual Plots.** Marcar **Seasonal Frequency** con **Specify** = 12**.** Desmarcar **Specify Model Periods** y **Box-Cox Transformation**.



4. Hacer click en **Model Options**.



5. Se empleará por defecto **Automatic Model Selection** con **AICc** como el **Model Selection Criterion**. Hacer click en **OK** para regresar al diálogo de Carta de Control de Suavización Exponencial. Hacer click en **OK**. Se generan Cartas de Control de Suavización Exponencial (DEM):



Se puede apreciar ahora claramente los puntos de datos fuera de control en los periodos 50, 51 y 100 de los residuos en la carta de Individuales. Esto coincide con lo observado previamente con las Cartas de Control regulares de Suavización Exponencial.

6. Desplazar hacia abajo el encabezado para ver el Modelo de Suavización Exponencial

Exponential Smoothing Model (Multiple Seasonal Decomposition): Additive Trend Method with Additive Errors (Holt's Linear) (A, A, N) - Model Automatically Selected
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

Se seleccionó automáticamente el modelo del **Método de Tendencia Aditiva con Errores Aditivos (Lineal de Holt) (A, A, N)** como el mejor ajuste para los datos desestacionalizados Ln modificados de los Pasajeros de la Aerolínea con base del criterio AICc. No existen periodos de retención.

7. Se genera la Información del Modelo de Suavización Exponencial:

Exponential Smoothing Model Information	
Seasonal Frequency	12 Decomposition
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

Este es el resumen de la información sobre el modelo con Frecuencia Estacional = 12 aplicando Descomposición y el Criterio de Selección del Modelo = “AICc”. La Transformación Box-Cox es “N/A”.

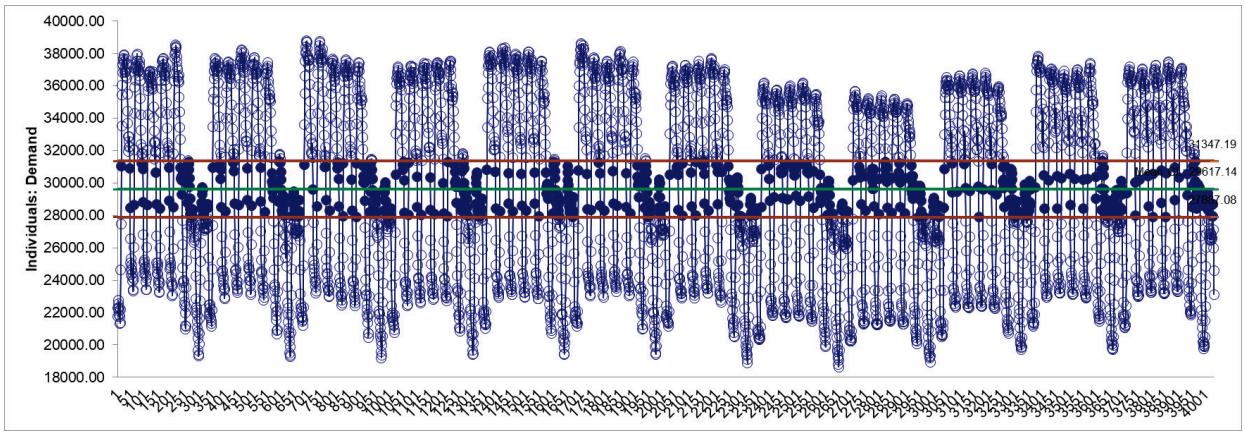
8. No se revisarán las Estimaciones de los Parámetros, las Estadísticas del Modelo y la Precisión del Pronóstico porque son casi similares a los valores del DEM que se dieron anteriormente, aunque hay que tener en cuenta que las ligeras diferencias se deben a la introducción de un valor atípico y un cambio de nivel, así también porque ahora se están utilizando todos los datos, es decir, no hay períodos de retención. Anteriormente se utilizó una Transformación Box-Cox con Lambda=0 y aquí se está utilizando los Ln de los datos.

Demanda de Media Hora de Electricidad Estacional Múltiple – Taylor

9. Abrir **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (pestaña de la **Hoja 1**). Esta es la demanda de electricidad de media hora (MW) en Inglaterra y Wales desde el lunes, 5 de junio, 2000 hasta el Domingo, 27 de agosto, 2000 (Taylor, R forecast). Estos datos tienen una múltiple estacionalidad con frecuencia = 48 (observaciones diarias) y 336 (observaciones semanales), con un total de 4032 observaciones. Véase Cartas de Corrida, Gráficos FAC/FACP, Gráfico de Densidad Espectral, y Gráficos de Descomposición de Tendencia Estacional para estos datos

10. Se construirá primero una clásica Carta de Control de Individuales con los datos sin procesar. Hacer click en **SigmaXL > Control Charts > Individuals**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

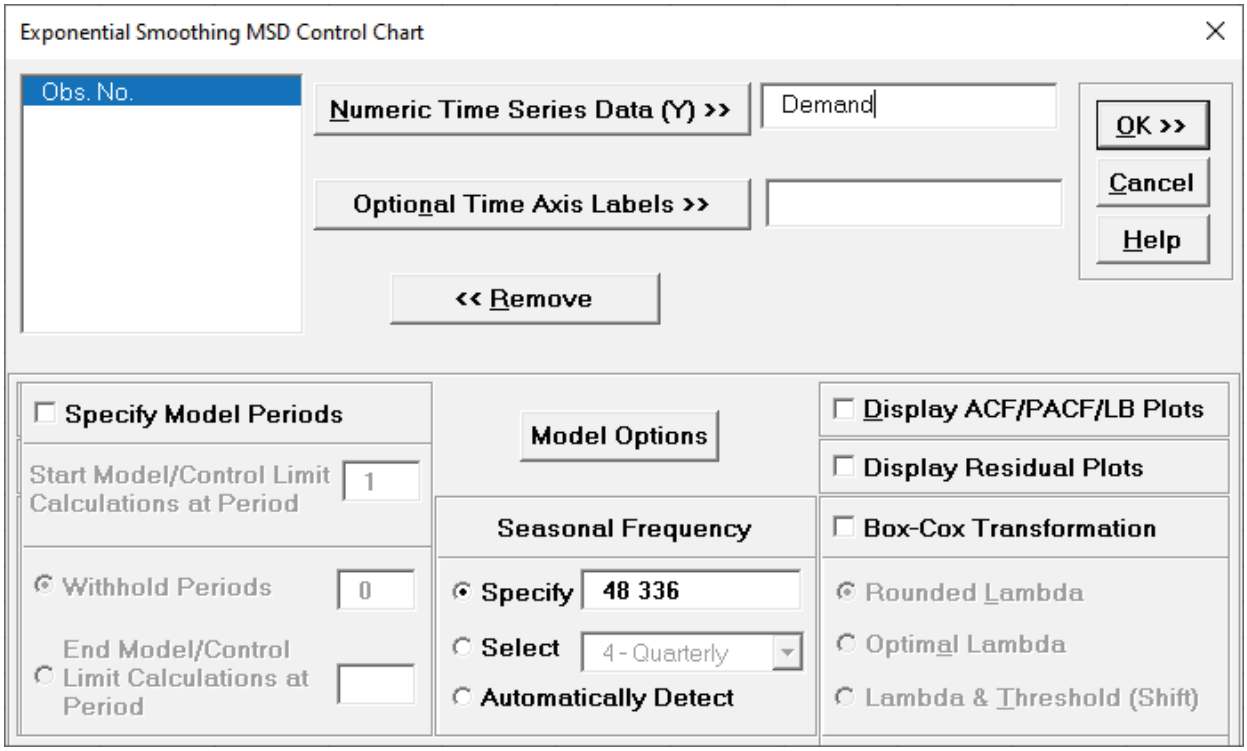
11. Seleccionar *Demand*, hacer click en **Numeric Data Variable (Y) >>**. Hacer click en **OK**. Se genera una Carta de Control de Individuales:



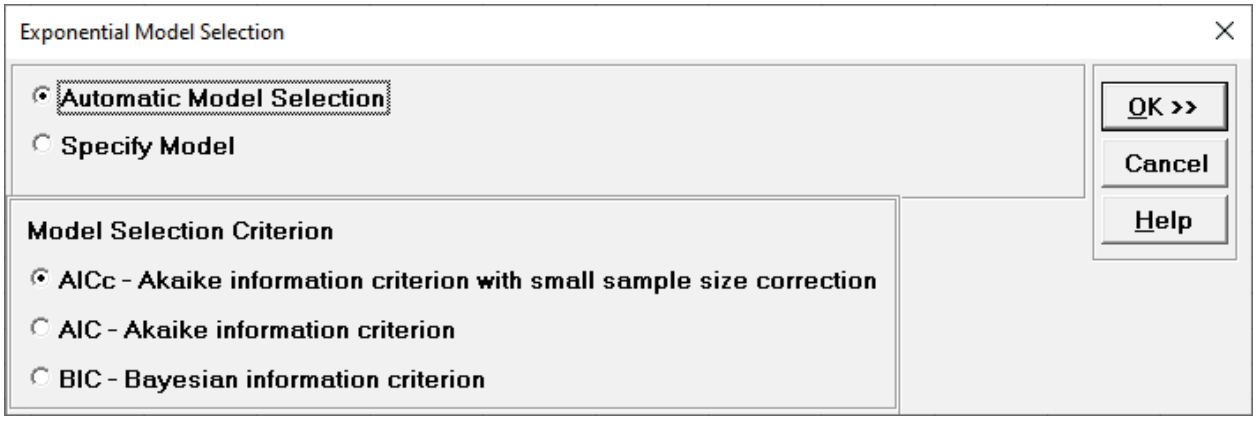
Con la estacionalidad de alta frecuencia, esta carta de control carece de sentido.

12. Hacer click en **SigmaXL > Time Series Forecasting > Smoothing Exponential Control Chart > Multiple Seasonal Decomposition Control Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

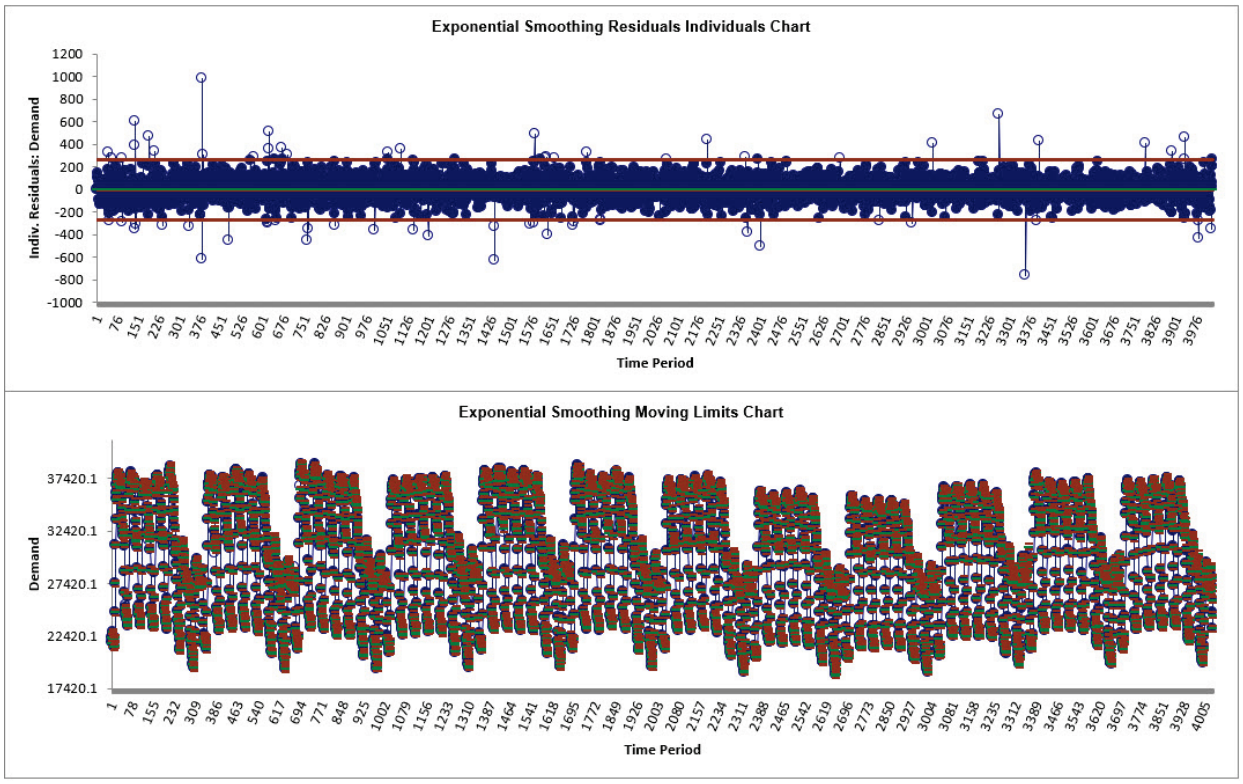
13. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y)** >>. Desmarcar **Display ACF/PACF/LB Plots** y **Display Residual Plots**. Marcar **Seasonal Frequency** con **Specify** = 48 336. Desmarcar **Specify Model Periods** y **Box-Cox Transformation**.



14. Hacer click en **Model Options**.

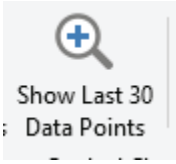


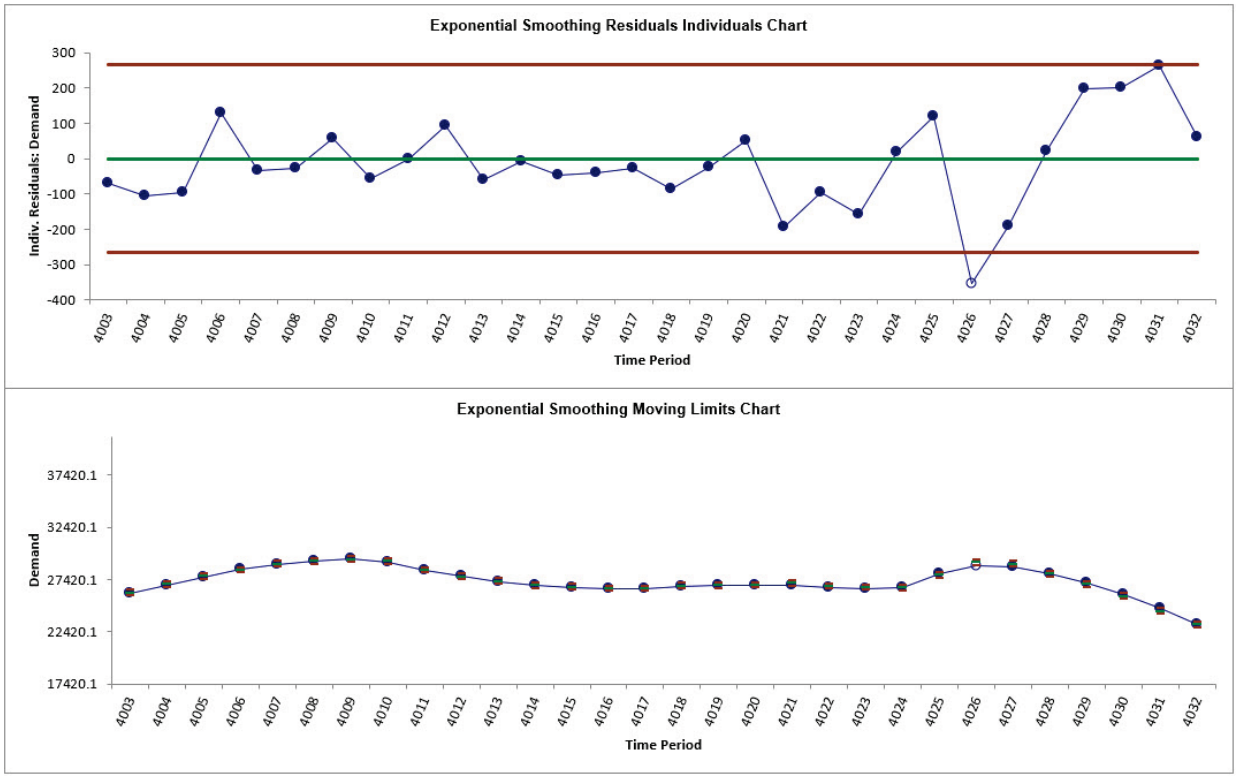
15. Se empleará por defecto **Automatic Model Selection** con **AICc** como el **Model Selection Criterion**. Hacer click en **OK** para regresar al diálogo de la Carta de Control de Suavización Exponencial DEM. Hacer click en **OK**. Se generan las cartas de control de suavización exponencial DEM:



Se tiene ahora una carta que se puede utilizar para identificar las causas asignables. El número de señales fuera de control se ha reducido dramáticamente.

16. Ud. puede habilitar el desplazamiento o la ampliación para visualizar los 30 puntos de las Cartas de Control. Se realizará lo segundo aquí. Hacer click en **SigmaXL Chart Tools > Show Last 30 Data Points**.

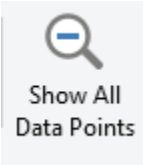




El punto del dato fuera de control en el periodo 4026 es claramente visible ahora.

Consejo: La escala del eje Y para la carta de Límites Móviles hace difícil que el mismo punto se vea. Ud. puede desear cambiar los valores máximos y mínimos, hacer click derecho en **Format Axis**, y cambiar los valores Bounds Minimum Maximum.

17. Para reestablecer la carta, hacer click en **SigmaXL Chart Tools > Show All Data Points**.



18. Desplazar hacia abajo el encabezado para ver el Modelo de Suavización Exponencial

Exponential Smoothing Model (Multiple Seasonal Decomposition): Simple Exponential Smoothing with Multiplicative Errors (M, N, N) - Model Automatically Selected
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

Se seleccionó automáticamente el modelo de **Suavización Exponencial con Errores Multiplicativos (M, N, N)** como el mejor ajuste para los datos de la Demanda desestacionalizada con base al criterio AICc. No existen periodos de retención.

Este es el mismo modelo usado anteriormente con el Pronóstico de Suavización Exponencial DEM, pero que empleó una retención de la muestra de 96.

Pronóstico ARIMA

El modelo de Promedio Móvil Integrado Autorregresivo (ARIMA) fue desarrollado por Box y Jenkins. La determinación automática por defecto del mejor orden del modelo en SigmaXL utiliza el método escalonado de Hyndman y Khandakar (véase fpp2).

Estacionariedad, Diferencias y Constante

El ARIMA asume que la serie de tiempo es estacionaria, es decir, tiene la propiedad de que la media, la varianza y la estructura de autocorrelación no cambian con el tiempo. Si la media de una serie temporal no es estacionaria (por ejemplo, la tendencia), esto puede corregirse mediante la diferenciación, calculando las diferencias entre las observaciones consecutivas para los datos no estacionales y entre los períodos consecutivos para los datos estacionales (por ejemplo, enero de 2019 - enero de 2018, etc.). En el caso de los datos no estacionales, esto implicará típicamente 1 o 2 órdenes de diferenciación. Este orden es el término integrado d. Para los datos estacionales, esto implicará típicamente 1 orden de diferenciación. Este orden es el término integrado estacional D. Un término constante c es opcional:

- Si d+D=0, un término constante en el modelo es la media.
- Si d+D=1, un término constante en el modelo es una tendencia (deriva).
- Si d+D>1, un término constante sería una tendencia cuadrática o superior, por lo que la constante no debe incluirse.
- Se recomienda que d+D no sea > 3.
- Si la varianza no es estacionaria, utilice una transformación Box-Cox.

Modelo autorregresivo (AR)

En un modelo autorregresivo, se pronostica la variable de interés utilizando una combinación lineal de valores de la variable que ocurrieron en el pasado. El término autorregresivo indica que es una regresión de la variable contra sí misma:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

donde ε_t es ruido blanco. Esto es casi como una regresión múltiple, pero con valores de rezago de y_t como predictoras. Esto es a lo que se refiere a un modelo AR(p), un modelo autorregresivo de orden p (fpp2).

Modelo de Promedio Móvil (MA)

En lugar de utilizar en una regresión los valores del pasado de una variable de pronóstico, un modelo de promedio móvil utiliza errores de pronósticos del pasado en un modelo similar a una regresión:

$$y_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

Esto es a lo que se refiere a un modelo MA(q), un modelo de promedio móvil de orden q (fpp2).

Modelo Autorregresivo Integrado Promedio Móvil (ARIMA)

Si se combina las diferencias con un modelo de auto regresión con uno de promedio móvil, se obtiene un modelo ARIMA no estacional:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} + \varepsilon_t$$

de donde y'_t es la serie en diferencias. Este es un modelo ARIMA (p, d, q) , de donde p es el número de términos autorregresivos, d es el grado de diferenciación y q es el número de términos del promedio móvil.

ARIMA Estacional

Para el caso estacional, el modelo consiste en términos similares a los componentes no estacionales del modelo, pero que implican retrocesos del período estacional. El modelo estacional es ARIMA (P, D, Q), y el combinado es ARIMA (p, d, q) (P, D, Q).

Orden del Modelo ARIMA

El orden del modelo puede determinarse automáticamente o el usuario lo especifica. El **Procedimiento Escalonado** emplea el método paso a paso de Hyndman y Khandakar (véase el Apéndice: **Selección Automática del Modelo (Identificación**):

- El D Estacional (0 o 1) se determina con el uso de una prueba Estacional de Potencia o lo especifica el usuario.
- La d no estacionaria (0, 1, o 2) se determina con el uso de una prueba KPSS (Kwiatkowski-Phillips-Schmidt-Shin) modificada de raíz unitaria o el usuario la especifica.
- El orden de AR (p) y MA (q) va desde 0 hasta un máximo de 5.
- El SAR (P) y SMA (Q) estacional va desde 0 hasta un máximo de 2.
- La constante (se incluye o no se incluye) si $d + D \leq 2$; no se incluye si $d + D > 2$.

Extended Model Search realizará una búsqueda sobre el mismo rango de valores del orden como Stepwise pero lo lleva a cabo usando todas las combinaciones, sujetas a las siguientes restricciones para una consistencia y eficiencia computacional:

- Se determinan d y D utilizando los mismos métodos como el Stepwise, o especificado por el usuario.
- Constante (incluida o no incluida) si $d + D \leq 2$; no incluida si $d + D > 2$.
- $p + q + P + Q \leq 7$.
- Los límites de tiempo de la Computación los especifica el usuario.

En general, por defecto se recomienda el **Stepwise Procedure** sobre la **Extended Model Search**, ya que es mucho más rápido y por lo general encuentra el mejor modelo ARIMA, o uno más simple que se acerca al mejor modelo ARIMA**.**

Estimación del Parámetro del Modelo y Valores Ausentes

Los parámetros del modelo se resuelven mediante la maximización no lineal de la función Log-Verosimilitud. Hay dos modelos disponibles: la suma condicional de cuadrados (CSS) y la máxima verosimilitud del espacio de estado Kalman. La CSS se utiliza siempre para las estimaciones iniciales y se usa si n>500 o la frecuencia estacional >12 para la eficiencia de los cálculos. Los filtros de Kalman permiten cálculos exactos y pueden lidiar con los valores ausentes. Para el CSS, si se encuentran valores faltantes, se utiliza el rango contiguo más grande. Para mayores detalles y las fórmulas de los parámetros, Véase el Apéndice: **Promedio Móvil Autorregresivo Integrado-ARIMA.**

Estadísticas Modelo ARIMA y Criterio de Información para Comparación de Modelo

Las estadísticas del modelo ARIMA son similares a las utilizadas en Suavización exponencial. Se relaciona el Log-Verosimilitud con - Ln (Error Cuadrado Medio), por lo que se maximiza. Los criterios de información AICc, AIC y BIC se calculan usando -2*Log-Verosimilitud e incorporan una penalización por el número de los términos en el modelo, por lo que cuanto más pequeño es mejor. Se utilizan en la Selección Automática de Modelos. El AICc es el Criterio de Información por defecto, el cual se basa en el performance del error de pronóstico con los datos de la competencia. Para mayores detalles, véase **Criterio de Información para Comparación de Modelo**.

Métricas de Precisión del Pronóstico

Las Métricas de Precisión del ARIMA son las mismas que aquellas empleadas en las Métricas de Precisión del Pronóstico:

Raíz cuadrada del error cuadrático medio: las siglas en inglés, RMSE $= \sqrt{\text{media}(e_t^2)}$

Error medio absoluto: MAE $= \text{media}(|e_t|)$

Error porcentual medio absoluto: MAPE $= \text{media}\left(\left|\frac{100e_t}{y_t}\right|\right)$

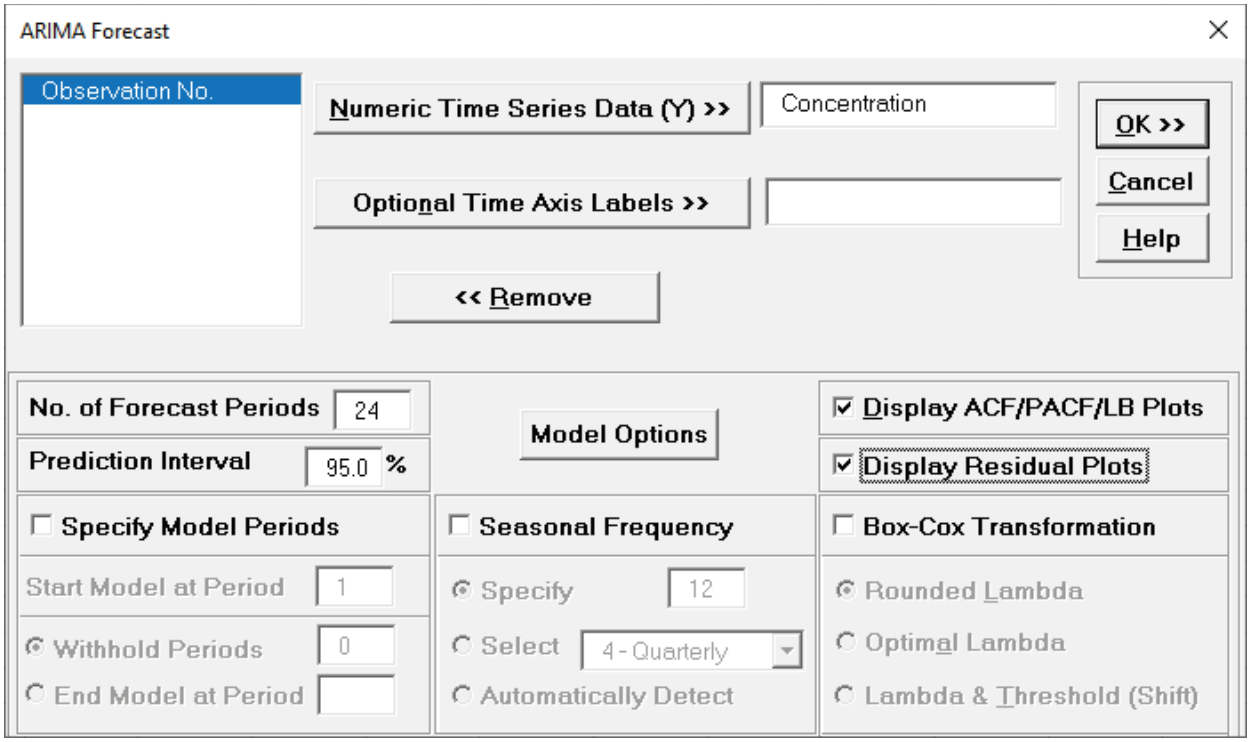
Error en escala medio absoluto: MASE $= \text{media}(|e_t|)/escala$

Chemical Process Concentration - Series A

1. Abrir **Chemical Process Concentration – Series A.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie A de Box y Jenkins, un conjunto de 197 valores de la concentración de un proceso químico tomados a intervalos de 2 horas. Véase Cartas de Corrida, Gráficos FAC/FACP, Gráfico de Densidad Espectral, y Gráficos de Descomposición de Tendencia Estacional para estos datos .

2. Hacer click en **SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Concentration*, hacer click en **Numeric Time Series Data (Y) >>**. Marcar **Display ACF/PACF/LB Plots** y **Display Residual Plots**. Desmarcar **Specify Model Periods, Seasonal Frequency** y **Box-Cox Transformation**. Se hace uso por defecto **No. of Forecast Periods** = 24 y **Prediction Interval** = 95.0 %.



- **Optional Time Axis Labels** se mostrará en el eje de tiempo de la carta de pronósticos. Si se utiliza, también se deben incluir las fechas de los períodos del pronóstico, de lo contrario el eje de tiempo quedará en blanco para los períodos del pronóstico.

- **No. of Forecast Periods** son el número de valores de las series temporales que se deben pronosticar (horizonte de pronóstico). El pronóstico más preciso será el primer valor pronosticado (un paso hacia adelante).

- **Prediction Interval** % es el nivel de confianza para las predicciones individuales. Por ejemplo, un intervalo de predicción del 95% contiene un rango de valores que debería incluir el valor futuro real con una probabilidad del 95%. El intervalo se hará más grande cuanto más distante sea la predicción.

- **Model Options** abre otro diálogo que permite fijar las opciones automáticas o especificar un modelo.

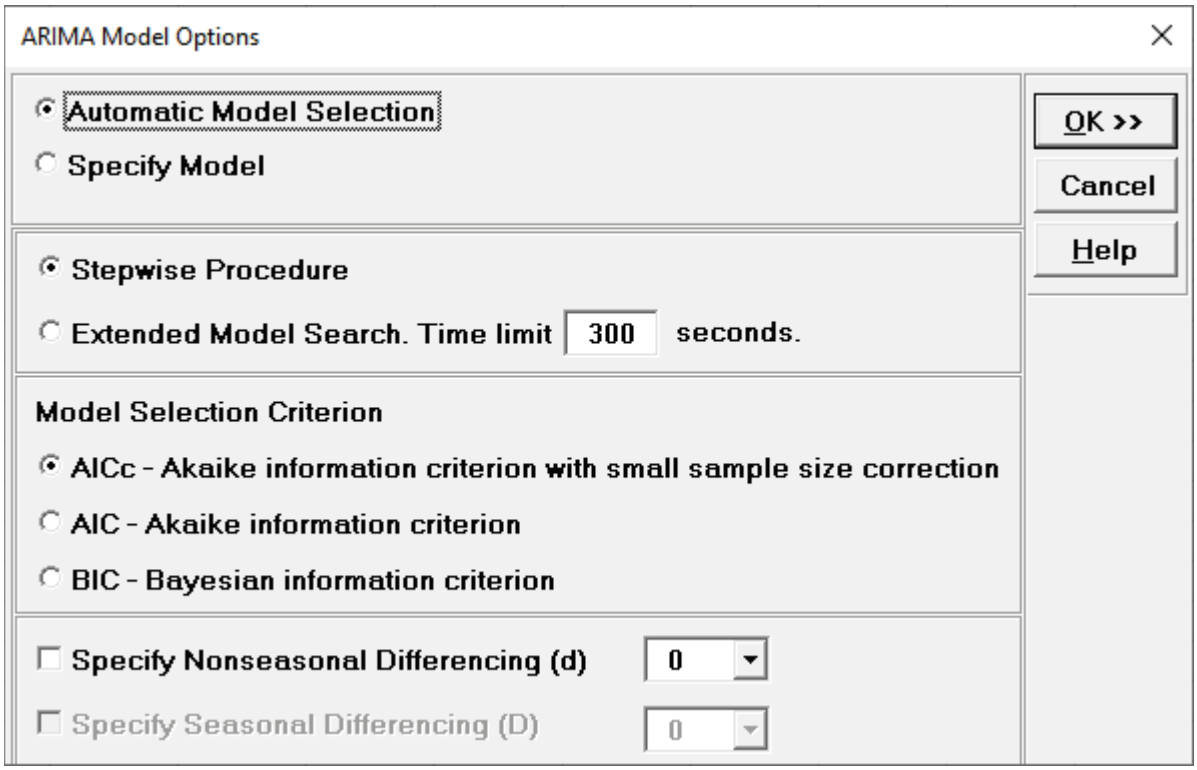
- **Display ACF/PACF/LB option** generará los gráficos de la FAC y FACP para los datos sin procesar, así como para los residuos del modelo. El gráfico LB es una gráfica de los valores p de prueba de Ljung-Box para varios rezagos y se utiliza para determinar si un grupo de autocorrelaciones son significativas, (esto es, las autocorrelaciones no provienen de una serie de ruido blanco). Para mayores detalles, véase **Ljung-Box Test**.

- **Display Residual Plots** generará una tabla de residuos del modelo y las habituales gráficas de residuos del modelo: histograma, gráfica de probabilidad normal, residuos frente al orden de los datos, y valores de los residuos vs aquellos de los del pronóstico. Nótese que si se aplica una transformación de Box-Cox, los residuos que se transforman no serán iguales a los actuales del pronóstico.

- **Specify Model Periods** se utilizan para especificar un período de inicio, un período final o para retener la muestra. Los datos que se retienen no se utilizan en la estimación del modelo, por lo que son muy útiles para la validación y comparación del modelo. Esto se utilizará en un ejemplo posterior.

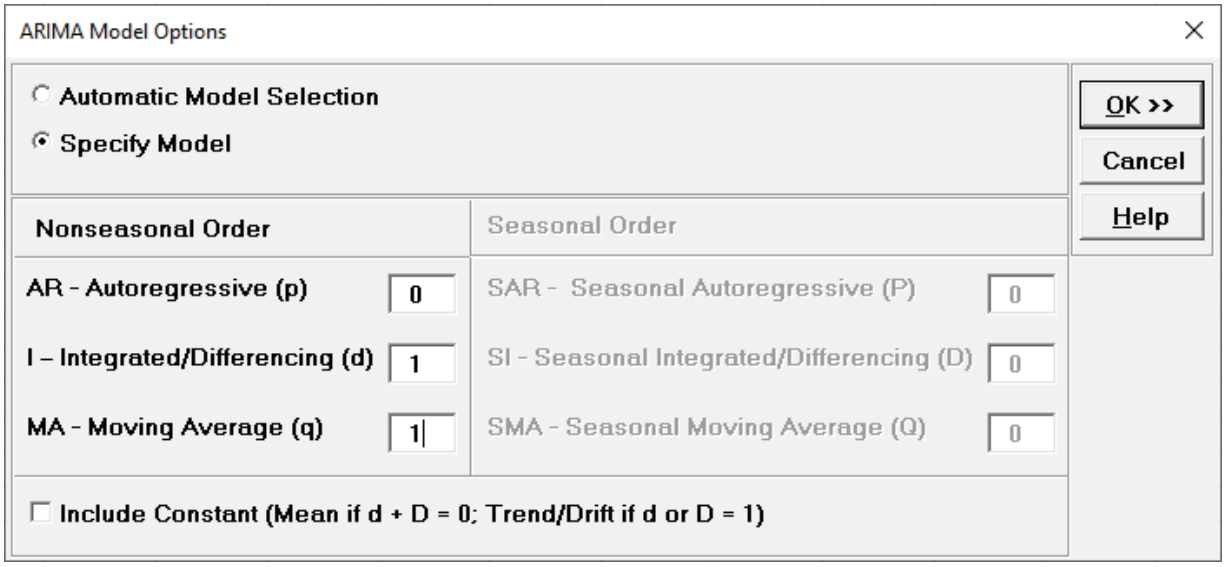
- **Seasonal Frequency** y **Box-Cox Transformation** se usarán en un posterior ejemplo.

4. Hacer click en **Model Options**.



- **Automatic Model Selection** se empleará posteriormente. Es la Selección por defecto.
- **Stepwise Procedure** Selecciona el uso del Stepwise Procedure.
- **Extended Model Search** está descrito arriba. El **Time limit** puede incrementarse para los modelos estacionales con una alta frecuencia estacional y/o un número grande de observaciones.
- **Model Selection Criterion** es la métrica del criterio de información que se emplea en Automatic Model Selection. AICc es la Selección por defecto.
- **Specify Nonseasonal Differencing (d)** = 0, 1, o 2, invalida la diferenciación no estacional automática. Es útil para comparar los modelos para los casos que son "casi no estacionarios" (véase Box y Jenkins).
- **Specify Seasonal Differencing (D)** = 0 o 1, invalida la diferenciación estacional automática. Se inhabilita cuando se desmarca **Seasonal Frequency** en el previo diálogo.
- Al hacer click en **OK** se acepta la configuración y retorna al previo diálogo. Al hacer click en **Cancel** se cancelará cualquier cambio y se regresa al previo diálogo.

5. Seleccionar **Specify Model**. Specify **Nonseasonal Order I – Integrated/Differencing (d)** = 1 y **MA – Moving Average (q)** = 1. Desmarcar **Include Constant**.

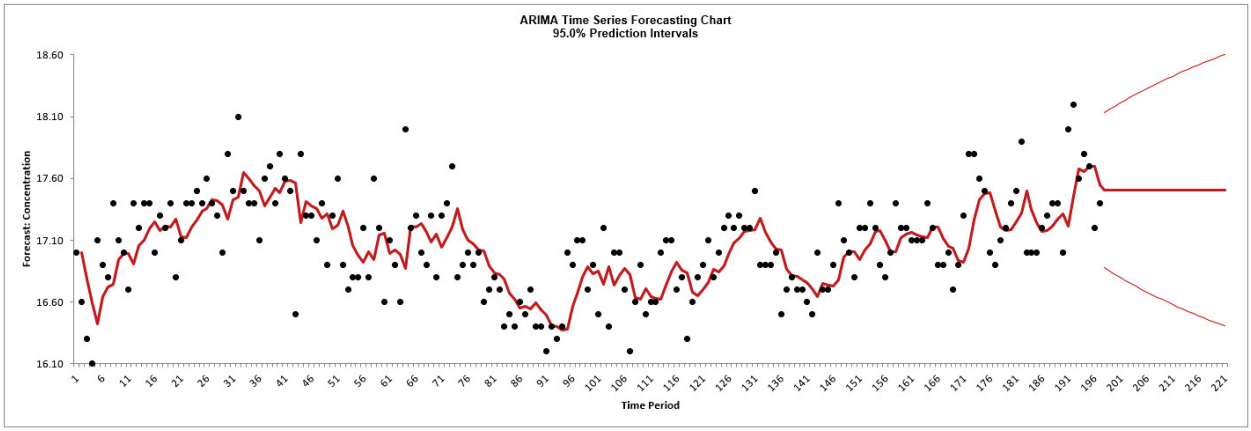


- **Specify Model** permite especificar en forma manual los valores **Nonseasonal Order** y **Seasonal Order** y la opción de **Include Constant (Mean if d + D = 0; Trend/Drift if d or D = 1)**. Si d + D = 0, entonces el término constante es la media; si d o D = 1, entonces el término constante es una Tendencia; Si d + D > 1, entonces el término constante es cuadrático o una mayor potencia – no recomendable.

- **Seasonal Order** está inhabilitado porque se ha desmarcado **Seasonal Frequency** en el diálogo previo.

El especificado ARIMA (0,1,1) es equivalente al modelo de suavización exponencial simple (con ligeras diferencias debido a la estimación del valor inicial).

6. Hacer click en **OK** para regresar al diálogo del Pronóstico ARIMA. Hacer click en **OK**. Se genera el informe del pronóstico ARIMA:



Como era de esperarse, esto es muy similar a la suavización exponencial como resultado del uso del modelo de **Suavización Exponencial Simple con Errores Aditivos (A, N, N) – Promedio Móvil Ponderado Exponencial (EWMA)**. El valor pronosticado dentro de la muestra del ARIMA es ligeramente diferente y empieza en el segundo periodo debido a la diferenciación.

7. Desplazarse hacia abajo para ver el encabezado del Modelo ARIMA:

ARIMA Model: Concentration - User Specified Model
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

De haberse marcado Specify Model Periods en el diálogo principal, el inicio, el final o la Selección de la Retención estaría resumida aquí también.

8. Se genera el Resumen del Modelo ARIMA:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	1
Include Constant	0
No. of Predictors	0
Model Selection Criterion	Specified
Box-Cox Transformation	N/A
Lambda	
Threshold	

Esto es el resumen de la información del modelo: ARIMA (0,1,1) sin constante ni predictoras. Seasonal Frequency = 1 (nonseasonal); Model Selection Criterion = “Specified” porque el modelo ha sido especificado por el usuario; y Box-Cox Transformation = “N/A” porque se ha desmarcado Box-Cox Transformation.

9. Las Estimaciones de los Parámetros son:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.699386992	0.064505199	10.8423	**0.0000**

El valor del coeficiente del parámetro MA_1 es aproximadamente igual a 1 – alfa = 1 - 0.2948 = 0.7052 en Estimaciones de Parámetros. La ligera diferencia se debe a la estimación del valor inicial.

Las Estimaciones de los Parámetros del ARIMA incluyen pruebas de significancia; valores-P < 0.05 son significativos y se resaltan en color rojo. Esto puede ser útil para el refinamiento del modelo con múltiples predictoras (y se demostrará más adelante). Tenga en cuenta que para la selección del orden de los modelos AR/MA, se debe utilizar el AICc mínimo, en lugar de las pruebas de significancia (Véase Kostenko, A.V. and Hyndman, R.J.).

10. Las Estadísticas del Modelo ARIMA son las siguientes:

ARIMA Model Statistics	
No. of Observations	197
DF	195
StDev	0.317609332
Variance	0.100875688
Log-Likelihood	-53.5087905
AICc	111.0797573
AIC	111.0175811
BIC	117.5738104

Los grados de libertad (DF) = $n - 2$ (1 término en el modelo, el orden de diferenciación es 1). Véase Estadísticas del Modelo ARIMA y Criterio de Información para Comparación del Modelo .

Al comparar con el Modelo Estadístico de Suavización exponencial, se aprecia que la DevEst y la Varianza son aproximadamente iguales, pero el Log-Verosimilitud, AICc, AIC y BIC son muy diferentes. Esto se debe a las diferentes fórmulas que se utilizan en la función de verosimilitud. No se puede utilizar el Criterio de Información para comparar los modelos ARIMA y Suavización Exponencial para determinar qué modelo tiene el mejor ajuste.

11. Las métricas de Precisión del Pronóstico Dentro de la Muestra son:

Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) One-Step-Ahead Forecast**
N	196	
RMSE	0.318324153	
MAE	0.248181659	
MAPE	1.45163544	
MASE	0.900807504	

MASE es menos de uno, por lo que es un mejor pronóstico que lo que se obtendría de un pronóstico naïve (establezca todos los pronósticos como el valor de la última observación). Véase Métricas de Precisión del Pronóstico.

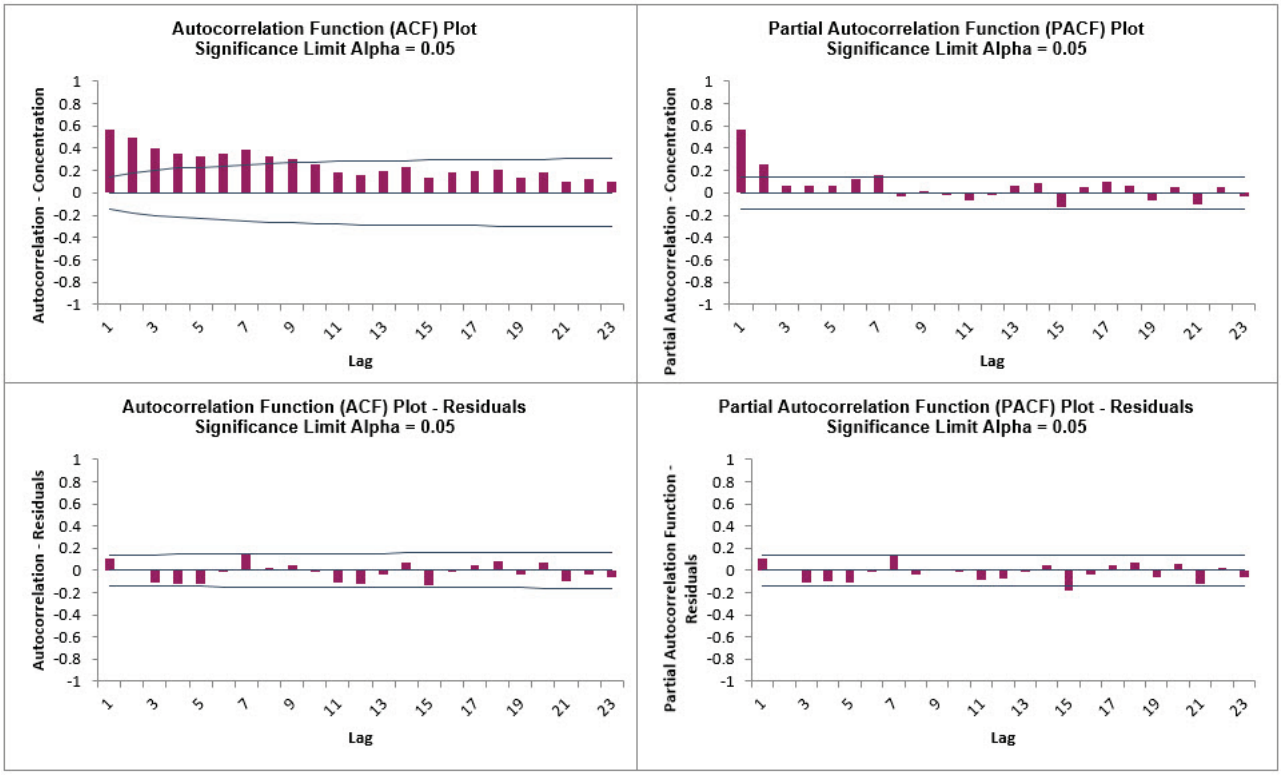
Al comparar con las métricas de Precisión del Pronóstico de Suavización Exponencial Dentro de la Muestra, como era de esperarse, se aprecia que los errores de pronóstico son aproximadamente los mismos. Nótese que se pierde una observación (N=196) ya que no se dispone de un valor pronosticado en el período de tiempo = 1 debido a la diferenciación.

12. Se genera la Tabla de Pronóstico:

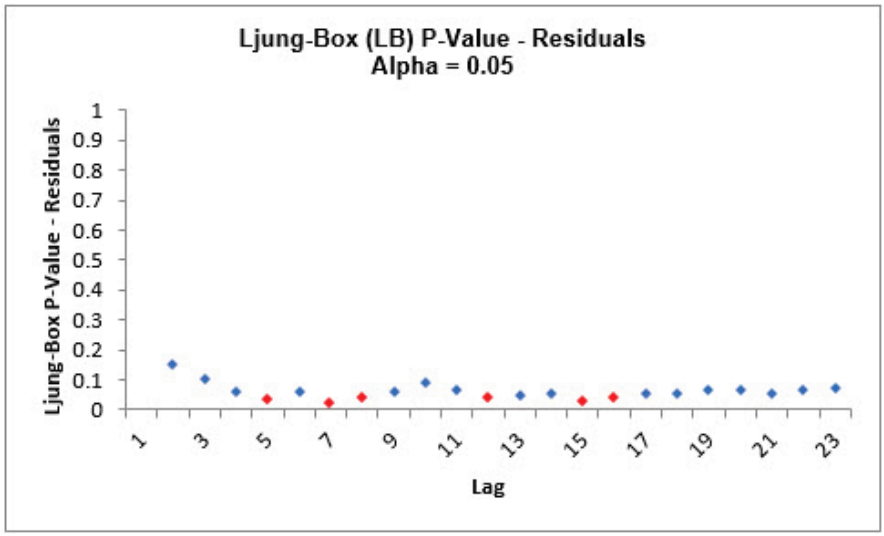
Forecast Table				
Period	Withhold Data	Forecast	Lower 95.0% PI	Upper 95.0% PI
198		17.50392441	16.87755395	18.13029487
199		17.50392441	16.84986402	18.1579848
200		17.50392441	16.82329967	18.18454915
201		17.50392441	16.79773386	18.21011496
202		17.50392441	16.77306181	18.23478701
203		17.50392441	16.74919586	18.25865296
204		17.50392441	16.72606181	18.28178701
205		17.50392441	16.70359618	18.30425264
206		17.50392441	16.68174418	18.32610464
207		17.50392441	16.66045813	18.34739069
208		17.50392441	16.63969619	18.36815263
209		17.50392441	16.61942146	18.38842736

Estos son los mismos valores del pronóstico e intervalos de previsión que se muestran en los Gráficos de Pronósticos, pero que se proporcionan para un análisis o un gráfico más detallado. Si se especifican los Periodos de Retención, los Datos de Retención se mostrarán también.

13. Hacer click en la hoja **ARIMA ACF PACF LB** para visualizar los Gráficos FAC/FACP/LB Plots:



Estos son aproximadamente los mismos que se obtuvieron anteriormente con los gráficos de FACF/FACP/LB. Se puede apreciar que toda la autocorrelación ha sido eliminada por el modelo de suavización exponencial (con la excepción del rezago 15 en la FACP), por lo que esto es un buen ajuste a los datos de las series temporales.

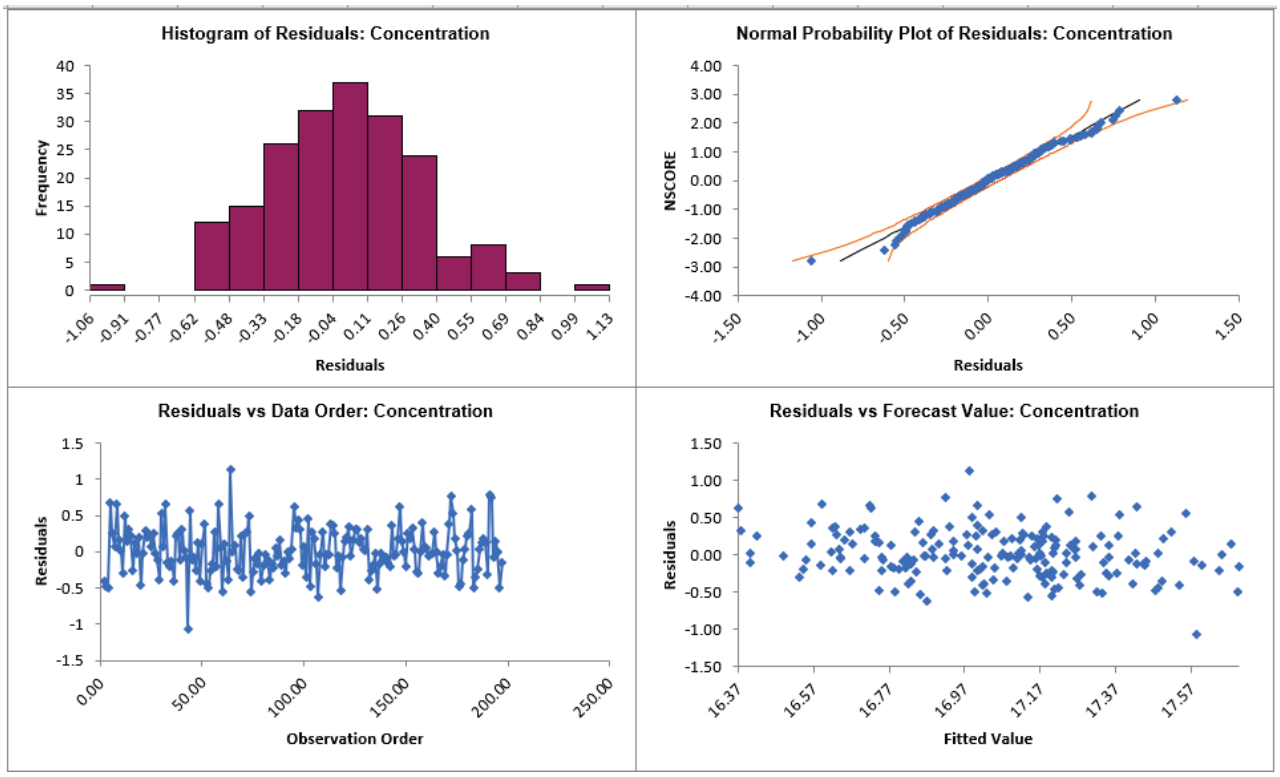


El gráfico LB es un gráfico de los valores-P de la prueba de Ljung-Box para varios rezagos y se utiliza para determinar si un grupo de autocorrelaciones es significativo, (es decir, las autocorrelaciones no provienen de una serie con ruido blanco). Para mayores detalles, véase **Test Ljung-Box** .

Los valores-P rojos son significativos (alfa=0.05) y los valores-P azules no son significativos. Es deseable que todos los valores-P sean azules. Los gráficos de la FAC/FACP indicaron que casi toda la correlación ha sido explicada por el modelo, pero el gráfico de Ljung-Box muestra que aún queda alguna autocorrelación significativa, por lo que el modelo puede ser potencialmente mejorado. Esto no significa que el modelo sea malo, puede seguir siendo muy útil para fines de predicción, pero los intervalos de predicción pueden no proporcionar una cobertura precisa.

Parece haber menos valores-P significativos que los que se obtuvo anteriormente con el gráfico LB, pero puede que no sea una diferencia práctica, dada la similitud de todas las demás estadísticas.

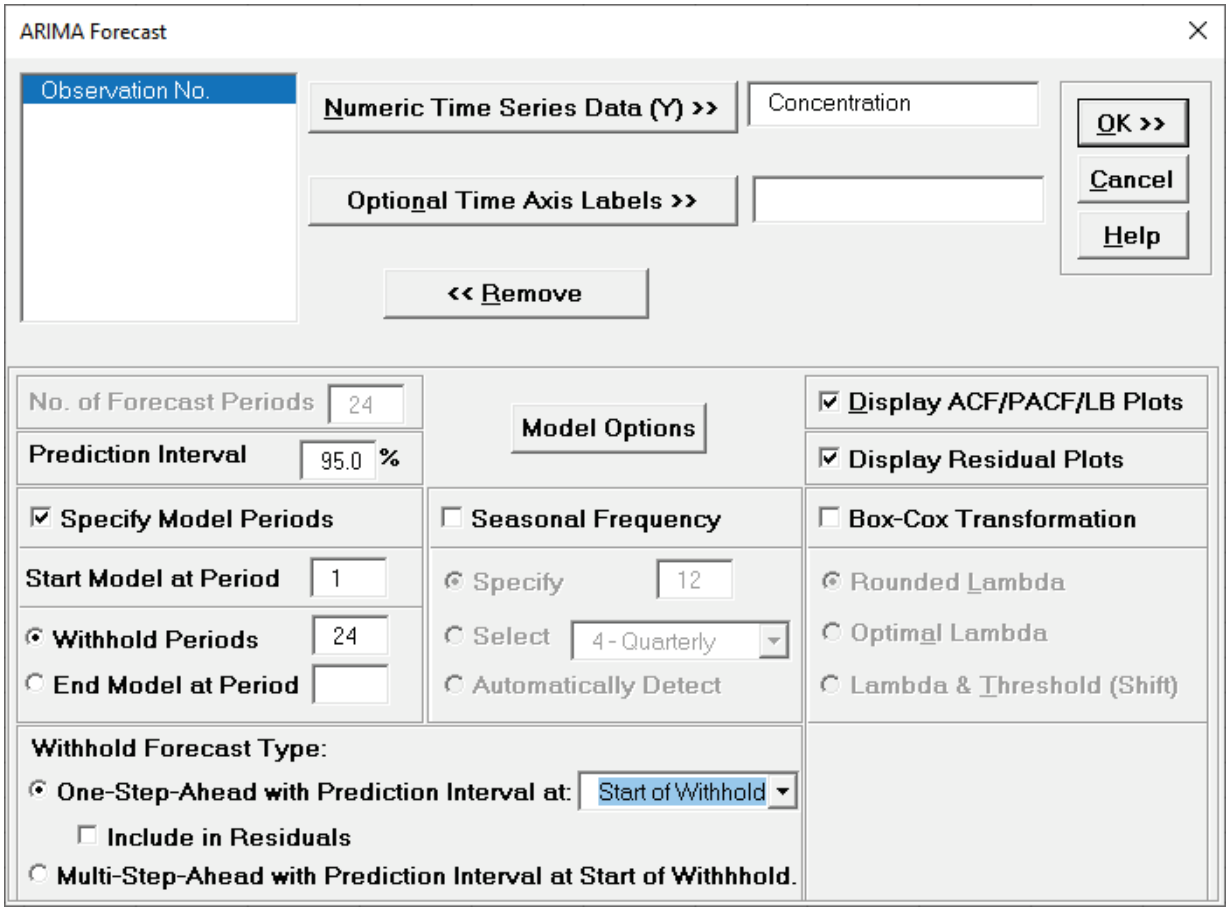
14. Hacer click en la hoja **ARIMA Residuals** para visualizar los Gráficos de los Residuos:



Los residuos están distribuidos de forma casi normal, con una línea aproximadamente recta en la gráfica de probabilidad normal. No hay valores extremos atípicos o patrones obvios en los gráficos. Posteriormente, se aplicará una carta de control a los residuos para comprobar formalmente la existencia de valores atípicos significativos o causas asignables.

Estos son aproximadamente los mismos que se obtuvieron anteriormente con los gráficos de los Residuos.

15. Se repetirá ahora el Pronóstico ARIMA sobre los datos de la Concentración con Automatic Model Selection y Specify Withhold Periods. Hacer click en el menú **Recall SigmaXL Dialog** o presione **F3** para recordar el último diálogo. Marcar **Specify Model Periods**. Fijar **Retention Periods** = 24. Se hace uso por defecto del **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *Start of Withhold*.



- La opción **Specify Model Periods** permite especificar los periodos inicial y final que se emplean en la identificación automática del modelo y en la estimación de los parámetros. Generalmente, **Start Model at Period** se conserva = 1 y **Withhold Periods** especifica el número de periodos que se retienen para las pruebas fuera de la muestra. **End Model at Period** especifica el final del periodo, de manera que el tamaño de la muestra de retención sería: el número total de observaciones – fin del periodo.

- **Withhold Forecast Type: One-Step-Ahead** excluirá la muestra de la retención de la identificación automática del modelo y la estimación de los parámetros, sin embargo, hace uso de los datos de retención para actualizar la predicción del pronóstico de un paso hacia

adelante. Esto es útil para evaluar el error del pronóstico cuando sólo existe interés en la predicción a corto plazo de un paso hacia adelante.

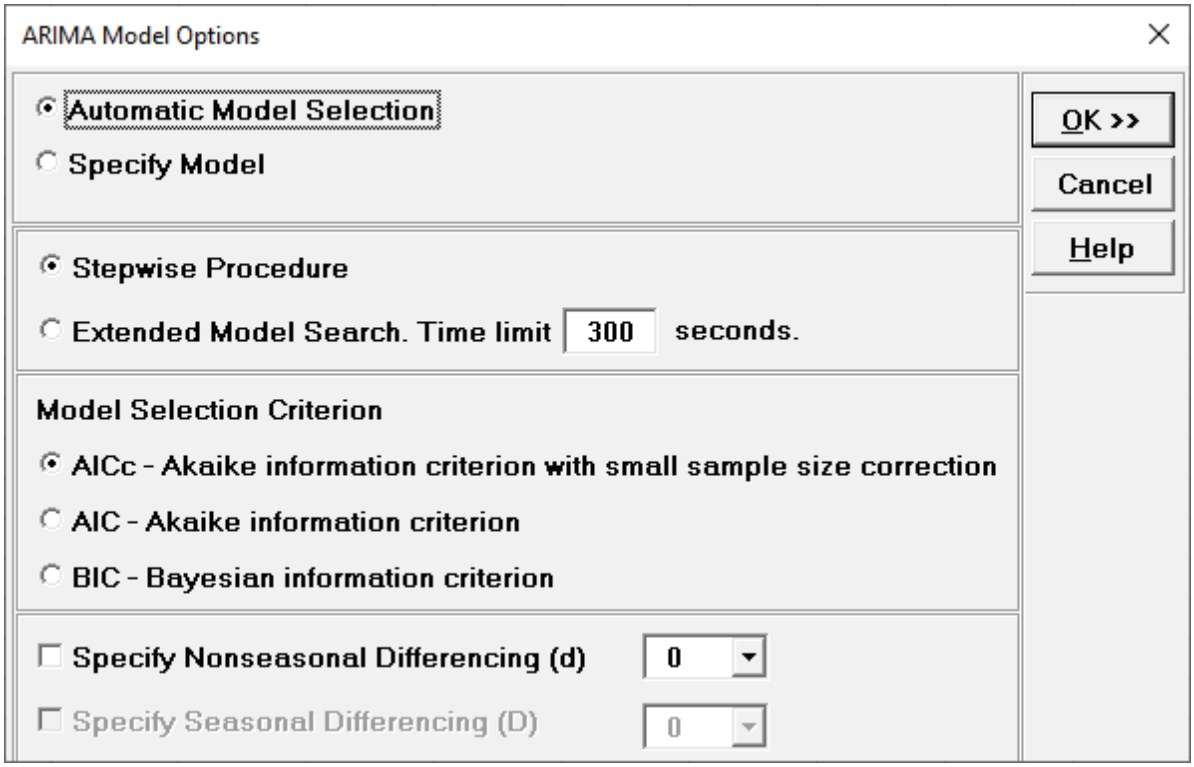
- **Withhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *Start of Withhold* visualizará el intervalo de predicción para la duración de la muestra de retención. Tomar nota que la longitud del intervalo de predicción se lo determina mediante el número de periodos de retención, de tal forma que invalida el **número de Periodos del Pronóstico** especificados.

- **Wirhhold Forecast Type: One-Step-Ahead with Prediction Interval at:** *End of Withhold* visualizará el intervalo de predicción al final de la muestra de retención. La longitud del intervalo de predicción se lo determina mediante el número de periodos de retención, de tal forma que invalida el **número de Periodos del Pronóstico** especificados.

- **Include in Residuals** tratará los errores de pronósticos de un paso hacia adelante como residuos (aunque no formen parte del proceso de estimación del modelo) y se incluirán en las gráficas de los residuos de ACF/PACF/LB junto con el informe y los gráficos de los residuos. Por lo general permanece sin marcar.

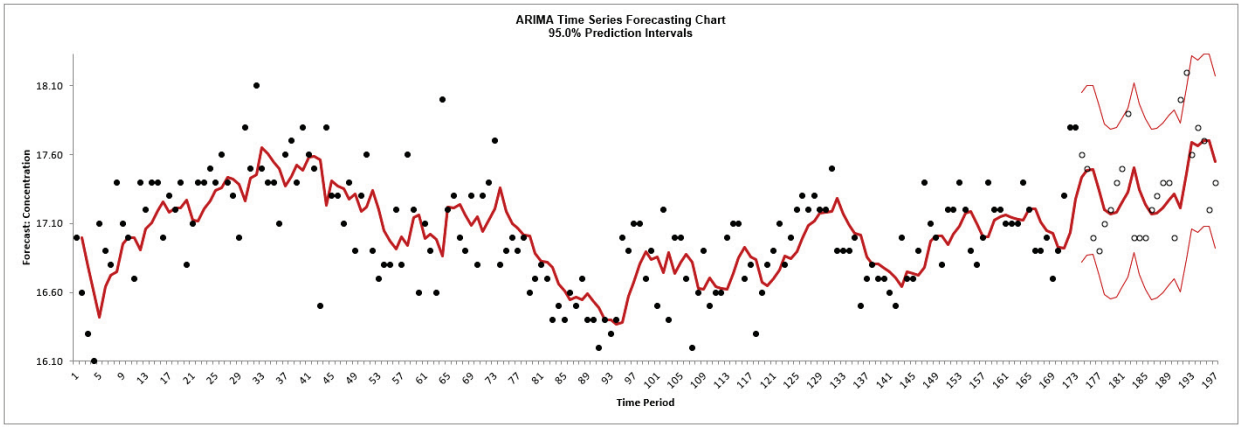
- **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold** excluirá la muestra de retención de la identificación automática del modelo y la estimación de parámetros y no utilizará los datos de retención para actualizar la predicción del pronóstico de un paso hacia adelante. Esto es útil para evaluar el error de pronóstico cuando existe interés en una ventana del pronóstico a largo plazo (horizonte). El intervalo de predicción se mostrará durante toda la duración de la muestra de retención. Tener en cuenta que la duración del intervalo de predicción se determinada por el número de períodos de retención, por lo que invalida el **número de Períodos de Pronósticos** especificados. Estos errores del pronóstico no se incluyen en los gráficos y el informe sobre los residuos de ACF/PACF/LB.

16. Hacer click en **Model Options**. Seleccionar **Automatic Model Selection**. Se hace uso por defecto: **Stepwise Procedure** and **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**, desmarcar **Specify Nonseasonal Differencing (d)**.



Consejo: Cuando se utiliza el **diálogo Recall SigmaXL** y si no hay cambios en la configuración de **Model Option,** se utilizarán los ajustes anteriores. No es necesario repetir este paso**.**

17. Hacer click en **OK** para retornar al diálogo del Pronóstico ARIMA. Hacer click en **OK**. Se genera el informe del pronóstico ARIMA:



Los puntos blancos son los valores de los datos en la muestra de retención con un pronóstico de un paso hacia adelante e intervalos de predicción que se visualizan al comienzo de la muestra de retención.

Esto es muy similar a la Carta de Pronóstico de la Suavización Exponencial que se desarrolló con el uso del modelo de **Suavización Exponencial Simple con Errores Multiplicativos (M, N, N)** modelo. El valor inicial del pronóstico ARIMA dentro de la muestra es ligeramente diferente y empieza en el segundo periodo debido a la diferenciación.

18. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA:

ARIMA Model: Concentration - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

19. Se genera el Resumen del Modelo ARIMA:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	1
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

El modelo ARIMA (0,1,1) que se especificó arriba en forma manual, se ha seleccionado automáticamente con base al criterio AICc.

20. Las Estimaciones de los Parámetros son:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.690028724	0.065305747	10.56613	**0.0000**

Esto es muy cercano a la estimación de los parámetros que se obtuvo arriba (con el uso de todos los datos).

21. Las Estadísticas del Modelo ARIMA son:

ARIMA Model Statistics	
No. of Observations	173
DF	171
StDev	0.312472686
Variance	0.09763918
Log-Likelihood	-44.18071519
AICc	92.4324363
AIC	92.36143039
BIC	98.65641934

Estos valores se aproximan a las estadísticas del modelo que se han obtenido arriba empleando todos los datos. Aquí se han usado sólo 173 de las 197 observaciones.

22. Las métricas de la Precisión del Pronóstico son:

Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) One-Step-Ahead Forecast**
N	172	24
RMSE	0.313295746	0.352257753
MAE	0.243952039	0.273585123
MAPE	1.431266547	1.568577996
MASE	0.881507368	0.988584899

Como era esperarse, los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un Paso-Hacia-Adelante** son mayores que los errores del Pronóstico **Dentro-de-la-Muestra (Estimación) un Paso-Hacia-Adelante**.

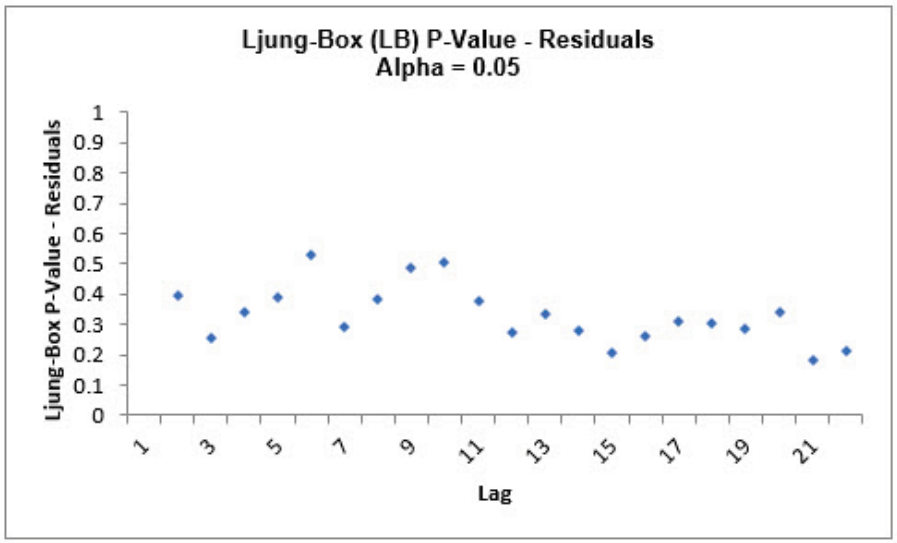
Estas son muy similares a las Métricas de Precisión del Pronóstico de Suavización Exponencial que se generaron con el uso del modelo de **Suavización Exponencial Simple con Errores Multiplicativos (M, N, N)**. Tener en cuenta que se pierde una observación dentro de la muestra (N=172) porque no se dispone del valor pronosticado en el periodo de tiempo = 1 debido a la diferenciación.

23. Se genera la Tabla de Pronóstico:

Forecast Table				
Period	Withhold Data	One-Step-Ahead Forecast	Lower 95.0% PI	Upper 95.0% PI
174	17.6	17.43699636	16.82022147	18.05377125
175	17.5	17.48752281	16.87207955	18.10296607
176	17	17.49139038	16.87775042	18.10503034
177	16.9	17.33907348	16.72282244	17.95532451
178	17.1	17.20297331	16.58502829	17.82091833
179	17.2	17.17105454	16.55470128	17.78740781
180	17.4	17.1800268	16.56542588	17.79462772
181	17.5	17.24821218	16.63450051	17.86192384
182	17.9	17.32625917	16.71316165	17.93935668
183	17	17.50410235	16.88693756	18.12126714
184	17	17.3478451	16.72798565	17.96770455
185	17	17.24002311	16.61980167	17.86024455
186	17.2	17.16562284	16.54613085	17.78511483
187	17.3	17.17627877	16.55847435	17.79408319
188	17.4	17.2146288	16.59825692	17.83100068
189	17.4	17.27208855	16.6568128	17.8873643
190	17	17.31173742	16.69785289	17.92562196
191	18	17.21510778	16.60123859	17.82897696
192	18.2	17.45840182	16.83590796	18.08089568
193	17.6	17.68827595	17.05843876	18.31811315
194	17.8	17.66091294	17.03262123	18.28920466
195	17.7	17.70402594	17.07708046	18.33097141
196	17.2	17.70277801	17.07747894	18.32807709
197	17.4	17.54693127	16.91921585	18.17464669

Estos son los mismos valores del pronóstico y de los intervalos de predicción que se visualizaron en la Carta del pronóstico, el cual ha sido proporcionado para un posterior análisis y gráficos. También se visualizan los **Datos de Retención**.

24. Los gráficos de los residuos de FAC/FACP/LB se basan en los datos dentro dela muestra. Los gráficos se parecen a aquellos a los correspondientes a los datos completos de arriba, excepto por los valores-P de la prueba Ljung-Box:



Como en el caso de la suavización exponencial, el modelo ARIMA (0,1,1) es un mejor ajuste al subconjunto que el correspondiente a los datos completos, con todos los valores-p de color azul (> 0.05).

25. Al marcar **Include in Residuals** entonces los residuos también incluirían los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un-Paso-Hacia-Adelante**.

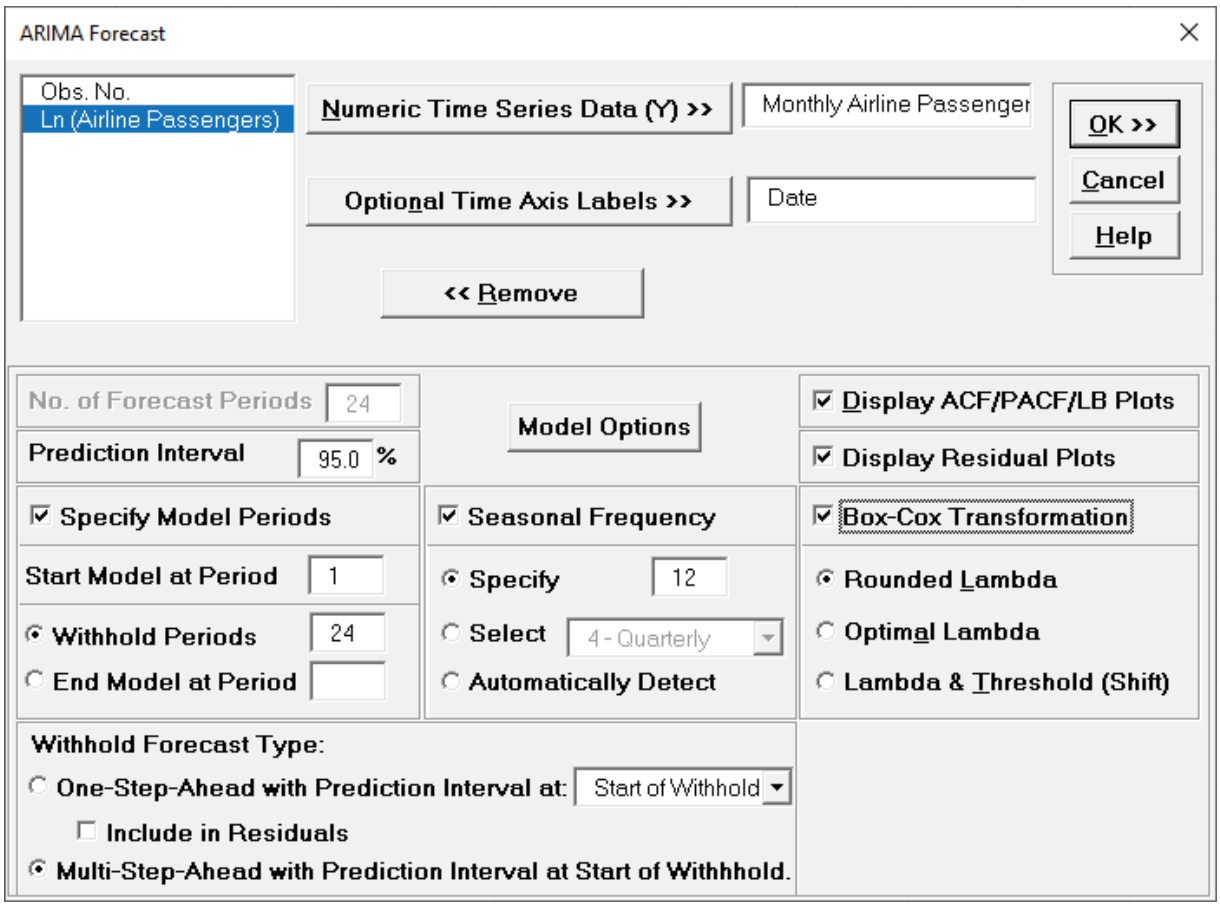
26. Se puede repetir el análisis de arriba con el uso de **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold** (Hacer click en el menú **Recall SigmaXL Dialog** o presione **F3** para repetir el último diálogo), pero no se lo realizará aquí. Los resultados serían muy similares a aquellos que se obtuvieron con la **suavización exponencial.**

Monthly Airline Passengers - Series G

27. Abrir **Monthly Airline Passengers - Series G.xlsx** (pestaña de la Hoja 1). Estos datos son una Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional de enero 1949 a diciembre 1960. Véase para estos datos **Cartas de Corrida, Gráficos FAC/FACP/LB, Gráficos de Densidad Espectral, Gráficos de Descomposición de Tendencia Estacional**

28. Hacer click en **SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

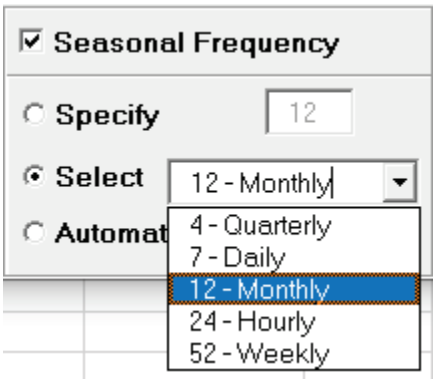
29. Seleccionar *Monthly Airline Passengers*, hacer click en **Numeric Time Series Data (Y) >>**. Seleccionar *Date*, hacer click en **Optional Time Axis Labels >>**. Marcar **Display ACF/PACF/LB Plots** y **Display Residual Plots**. Marcar **Specify Model Periods**. Fijar **Withhold Periods** = 24. Seleccionar **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Marcar **Seasonal Frequency** with **Specify** = 12**.** Marcar **Box-Cox Transformation** y Seleccionar **Rounded Lambda** (se ha seleccionado porque la **Carta de Corrida** mostró un incremento en la varianza estacional a lo largo del tiempo). Se hace uso por defecto del **Prediction Interval** = 95.0 %.



- **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold** excluirá la muestra de retención de la identificación automática del modelo y la estimación de parámetros y no utiliza los datos de retención para actualizar el pronóstico de un paso hacia adelante. Esto es útil para evaluar el error de pronóstico cuando se está interesado en una ventana de pronóstico a largo plazo (horizonte). El intervalo de predicción se visualizará durante toda la duración de la muestra de retención. Tener en cuenta que la duración del intervalo de predicción está determinada por el número de períodos de retención, por lo que anula el **No. of Forecast Periods** especificados.

- **Seasonal Frequency Specify** se usa para especificar la frecuencia estacional.

- **Seasonal Frequency Select** proporciona una lista desplegable que se emplea usualmente con las frecuencias estacionales:



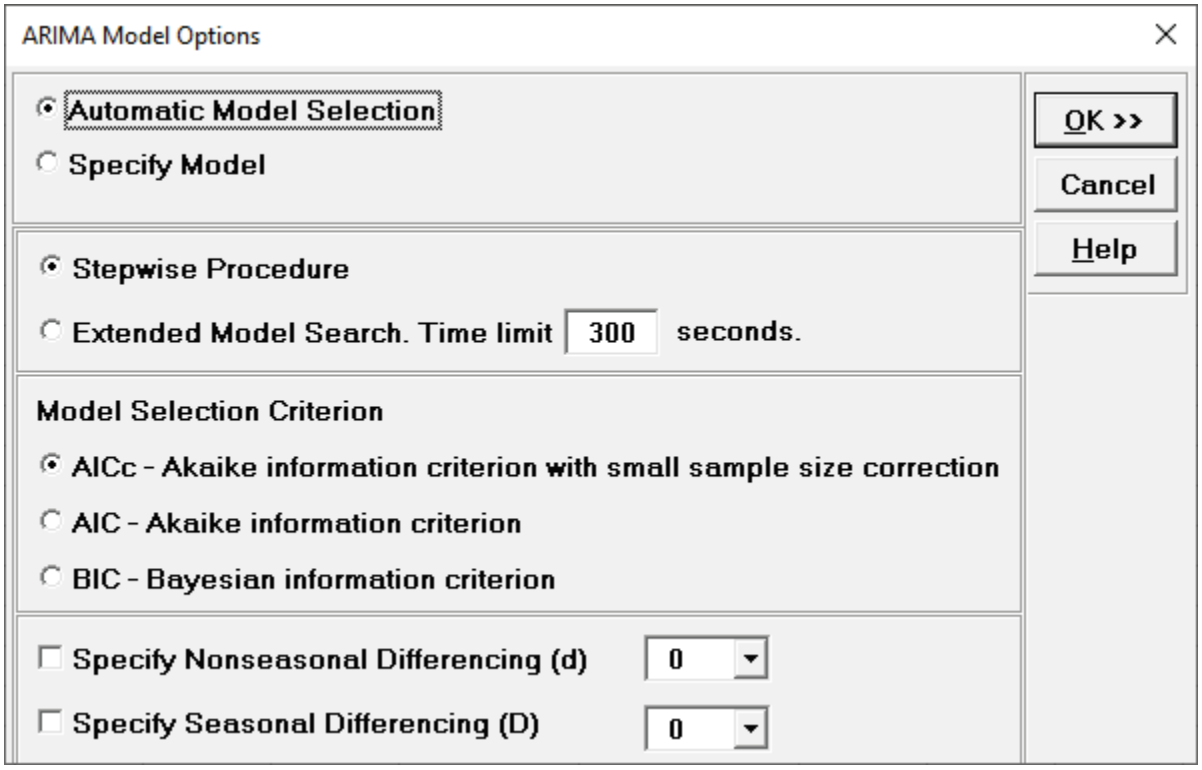
- **Seasonal Frequency Automatically Detect** debería usarse si no está seguro cuál es el valor de la frecuencia estacional (o realice un Diagrama de Densidad Espectral antes del Gráfico de Descomposición de la Tendencia Estacional).

- **Box-Cox Transformation** con **Rounded Lambda** seleccionará Lambda = 0 (Ln), 0.5 (SQRT) o 1 (Sin transformar). Threshold (Shift) se procesa en forma automática si la serie de tiempo incluye 0 o valores negativos, de lo contrario es 0.

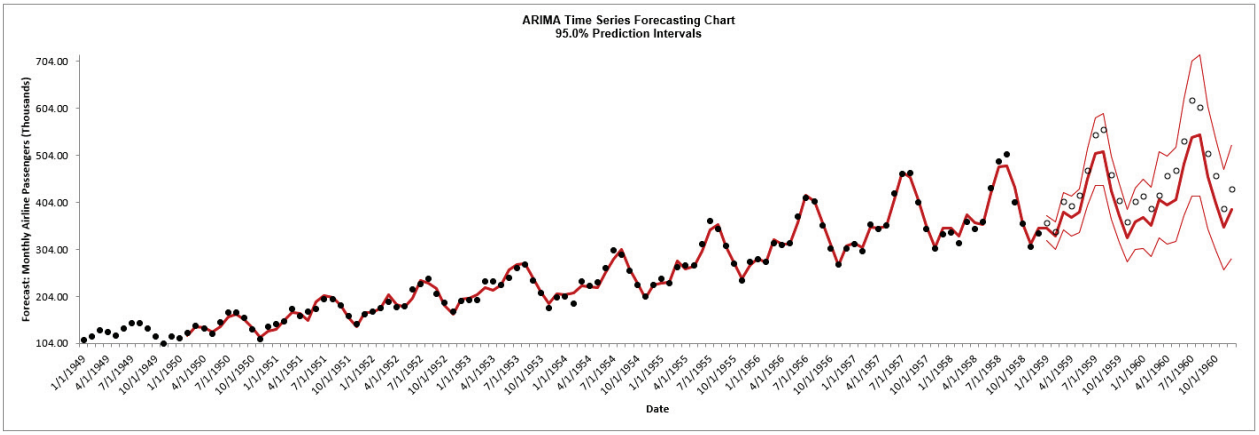
- **Box-Cox Transformation** con **Optimal Lambda** usa el rango de 0 a 1 para Lambda. El umbral (threshold) se procesa en forma automática si la serie de tiempo incluye 0 o valores negativos.

- **Box-Cox Transformation** con **Lambda & Threshold (Shift)** si se lo deja vacío, procesará el óptimo de lambda y el umbral. El usuario puede especificar también Lambda y el Threshold. Se puede especificar Lambda fuera del rango de 0 a 1, pero para el análisis de las series de tiempo, se debería limitar entre -1 y 2. El umbral, por lo general, es 0, no obstante, si los datos de la serie temporal incluyen 0 o valores negativos, se debería registrar un valor negativo para el umbral que sea más pequeño que el valor mínimo. Se restará este valor de los datos que resultaren de los valores positivos de la serie temporal.

30. Hacer click en **Model Options**. Seleccionar **Automatic Model Selection**. Se hace uso por defectos de: **Stepwise Procedure** y **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**, desmarcar **Specify Nonseasonal Differencing (d)** y **Specify Seasonal Differencing (D)**.



31. Hacer click en **OK** para retornar al diálogo del Pronóstico ARIMA. Hacer click en **OK**. Se genera el informe del pronóstico ARIMA:



Los puntos en blanco son los valores de los datos de la muestra de retención con un pronóstico de varios pasos e intervalos de predicción mostrados al comienzo de la muestra de retención. Obsérvese que los datos se han transformado utilizando Box-Cox, pero se aplica la transformación inversa para producir este gráfico con las unidades originales.

Esto es muy similar a la Carta de Pronóstico de Suavización Exponencial que se elaboró con el uso del modelo del método de **Tendencia Aditiva, Estacional Aditivo con Errores Aditivos (Holt-Winters) (A, A, A)**. El valor inicial dentro de la muestra pronosticado con ARIMA comienza en el 14º período de tiempo debido a las diferencias no estacionales y estacionales.

32. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA:

ARIMA Model: Monthly Airline Passengers (Thousands) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

33. Se genera el Resumen del Modelo ARIMA:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	1
SMA Order (Q)	1
Seasonal Frequency	12
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

El modelo ARIMA (0,1,1) (0,1,1) (sin constante) se seleccionó automáticamente con base al criterio AICc. Esto concuerda con los resultados que se mostraron en la Carta de Corrida de los datos diferenciados en forma manual y es el modelo empleado en Box y Jenkins, Capítulo 9., "Análisis de Series de Tiempo Estacionales".

No existen predictoras. Seasonal Frequency = 12; Model Selection Criterion = "AICc" y Box-Cox Transformation = "Rounded Lambda" con Lambda = 0 (esto es, una transformación Ln).

34. Las Estimaciones de los Parámetros son:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.342313249	0.100902427	3.392517	**0.0010**
SMA_1	0.540469465	0.087677292	6.164304	**0.0000**

A diferencia del ejemplo anterior con los datos de Concentración no estacionales y sin tendencia, no se puede comparar estos parámetros con el modelo de suavización exponencial.

35. Las Estadísticas del Modelo ARIMA son:

ARIMA Model Statistics	
No. of Observations	120
DF	105
StDev	0.037414431
Variance	0.00139984
Log-Likelihood	197.5047754
AICc	-388.776541
AIC	-389.009551
BIC	-380.991064

- El número de observaciones, *n* = 144 – 24 (*retención*) = 120
- Los grados de libertad (DF) = 120 (*n*) – 15 (2 los términos en el modelo, 1 diferencia no estacional, 12 diferencia estacional) = 105
- Tener en cuenta que las estadísticas se basan en la transformación Ln de los datos, no en los datos originales.

Comparando con el Modelo Estadístico de Suavización exponencial, vemos que la DesvEst y la Varianza son aproximadamente iguales, pero el Log-Verosimilitud, AICc, AIC y BIC son muy diferentes. Esto se debe a las diferentes fórmulas que se utilizan en la función de verosimilitud. No se puede utilizar el Criterio de Información para comparar los modelos ARIMA y Suavización Exponencial para determinar qué modelo tiene el mejor ajuste.

36. Las métricas de la Precisión del Pronóstico son:

Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) Multi-Step-Ahead Forecast**
N	107	24
RMSE	9.44337388	43.18833723
MAE	7.384036891	39.45185993
MAPE	3.001604734	8.51734062
MASE	0.258417364	1.380687256

Al comparar la Precisión del Pronóstico, se tiene errores Fuera-de-la-Muestra mayores. El MASE es mayor que uno, de esta manera es un peor pronóstico que aquel que se obtendría a partir de un pronóstico naïve estacional. Sin embargo, es importante tener presente que se está pronosticando para 24 meses, lo cual es muy razonable. Al seleccionar una Muestra de Retención Multi-Pasos Hacia-Adelante de 12 meses (o una Muestra de Un-Paso Hacia-Adelante) mejoraría sustancialmente el pronóstico.

Como era esperarse, los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Multi- Pasos-Hacia-Adelante** son mayores que los errores del Pronóstico **Dentro-de-la-Muestra (Estimación) un Paso-Hacia-Adelante**. Tener en cuenta, si existiera interés en un pronóstico a corto plazo un paso hacia adelante, entonces se tendría que seleccionar **Withhold Forecast Type: One-Step-Ahead** y esta tabla mostraría los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un Paso-Hacia-Adelante**.

Obsérvese que se pierden 13 observaciones dentro de la Muestra (N=107) dado que no se dispone de un valor pronosticado en el periodo = 1 al 13, debido a la diferenciación.

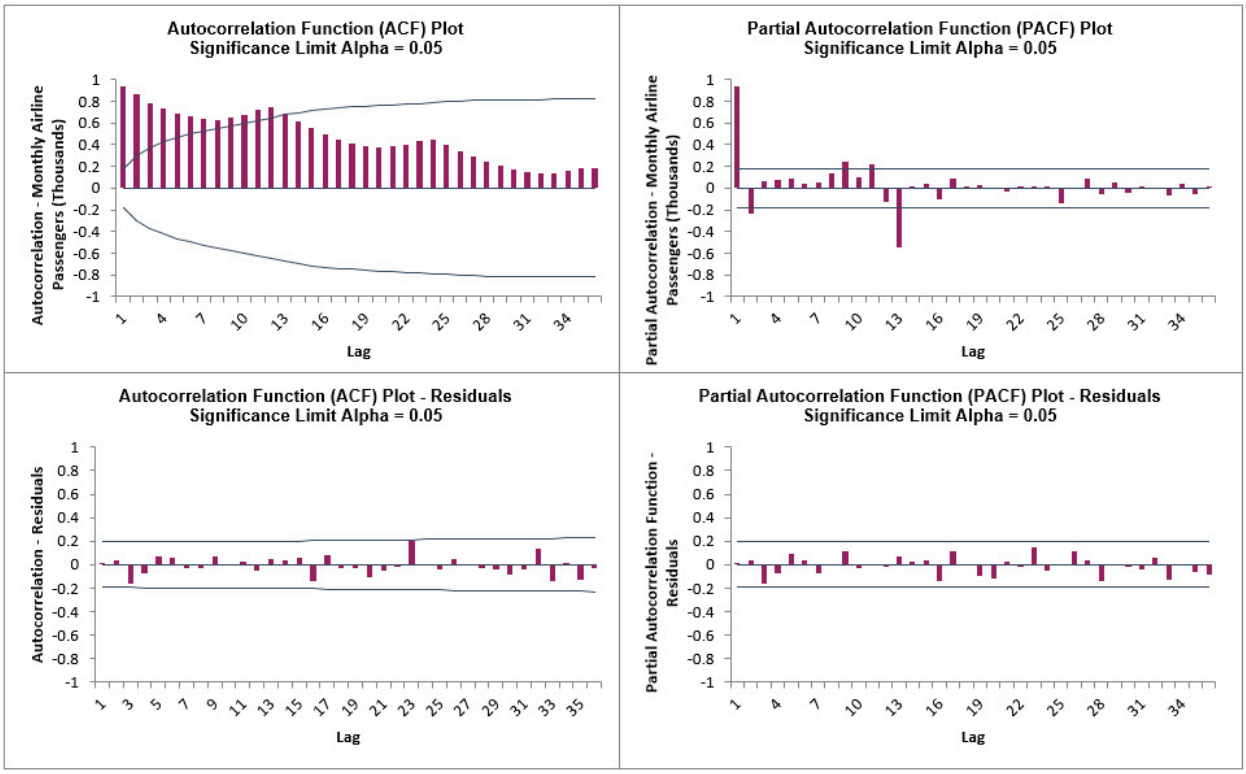
Las métricas de Precisión se calculan utilizando los datos reales sin procesar frente a la transformada inversa del pronóstico tal y como se muestra en el Gráfico y la Tabla de Pronóstico, lo que permite la comparación entre todos los tipos de modelos y las transformaciones.

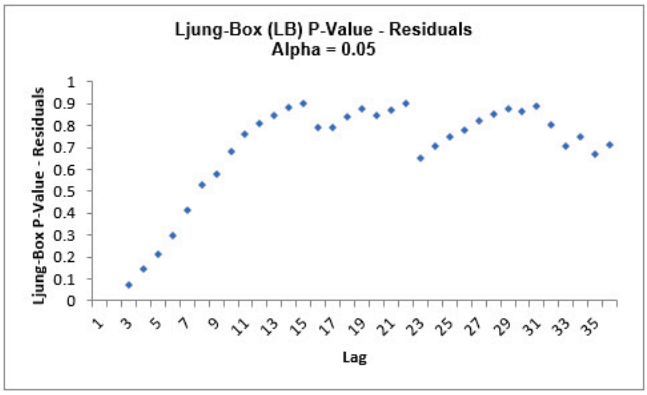
37. Se genera la Tabla de Pronóstico:

Forecast Table				
Period	Withhold Data	Multi-Step Ahead Forecast	Lower 95.0% PI	Upper 95.0% PI
121	360	348.5828062	323.6916935	375.3879856
122	342	331.5740082	303.4388404	362.3178982
123	406	383.1608225	346.2907533	423.9565003
124	396	372.606915	333.0281012	416.8894835
125	420	382.9486611	338.8267505	432.8161127
126	472	453.1350644	397.199074	516.9483013
127	548	507.198003	440.7280155	583.6929018
128	559	511.2341986	440.6029023	593.1881166
129	463	429.3989001	367.2058264	502.1255171
130	407	376.442152	319.5422535	443.4740391
131	362	328.1517697	276.5830729	389.3354095
132	405	362.9392672	303.8283644	433.5504089
133	417	372.7865814	306.4056918	453.5484783
134	391	354.596781	287.8353252	436.8431047
135	419	409.7655151	328.7189301	510.7943655
136	461	398.4787992	316.1075385	502.3143522
137	472	409.5386223	321.4319641	521.7959068
138	535	484.5984041	376.4753418	623.7742214
139	622	542.4151917	417.2724015	705.0891434
140	606	546.7316396	416.6289379	717.4621313
141	508	459.2141241	346.749965	608.1546736
142	461	402.5803351	301.3041689	537.8980544
143	390	350.9369201	260.4037852	472.9452062
144	432	388.1398803	285.6122259	527.4724013

Se trata de los mismos valores del pronóstico e intervalos de previsión que se visualizan en la Carta de Pronóstico, pero que se proporcionan para un análisis o un gráfico más detallado. Obsérvese que esta tabla indica el número del Período de Pronóstico mientras que el Gráfico muestra la Fecha.

38. Hacer click en la hoja **ARIMA ACF PACF LB** para visualizar los Gráficos de FAC/FACP/LB:





Los Gráficos de las FAC/FACP de los Residuos indican que la autocorrelación ha sido explicada por el modelo. El gráfico de Ljung-Box confirma que no queda ninguna autocorrelación significativa (todos los valores-P azules > 0.05), por lo que esto es un buen ajuste a los datos de la serie de tiempo (en comparación con los Gráficos de FAC/FACP/LB).

39. Hacer click en la hoja **ARIMA Residuals** para visualizar los Gráficos de los Residuos:

Los residuos están distribuidos de forma casi normal, con una línea aproximadamente recta en la gráfica de probabilidad normal. No hay valores extremos atípicos o patrones obvios en los gráficos.

Nótese que los residuos se basan en los datos transformados de Ln, no en los datos originales. La **información del modelo ARIMA** a la derecha de los gráficos muestra la información de la Transformación de Box-Cox.

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	1
SMA Order (Q)	1
Seasonal Frequency	12
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

El Pronóstico ARIMA con Predictoras

El modelo ARIMA permite predictoras continuas o categóricas, similar a la regresión múltiple con SigmaXL. Para mayores detalles, véase el apéndice: ARIMA con Covariables (Predictoras).

A fin de proporcionar un pronóstico, se debe agregar los valores adicionales de la predictora (X) al conjunto de datos antes de realizar el análisis. El número de períodos de pronóstico será igual al número de filas de las predictoras adicionales. Alternativamente, se pueden utilizar los valores de la predictora de una muestra de retención. Si no se proporciona ninguno de los dos, SigmaXL utilizará la última fila los de valores de predicción y calculará un período de pronóstico.

Las predictoras continuas son numéricas, las predictoras categóricas pueden ser de texto o numéricas. Todas las predictoras deben tener el mismo número de filas. Una predictora no puede tener todos los valores iguales.

Al igual que con la regresión lineal múltiple, las predictoras no deben estar fuertemente correlacionadas. SigmaXL eliminará automáticamente los términos con factores de inflación de muy alta varianza (VIF > 100) y proporcionará un mensaje de advertencia "Multicolinealidad detectada en las predictoras". “Se han eliminado los siguientes predictoras...".

De la misma forma que se realiza con la regresión lineal múltiple, se codifican automáticamente como “ficticias” las variables predictoras categóricas. A las predictoras se les agregará un nivel a los nombres de los términos en la tabla de Estimaciones de los Parámetros. El primer nivel (ordenado en forma alfanumérica) es un nivel de referencia oculto. De existir tres niveles, sólo dos aparecerán en la tabla de Estimaciones de los Parámetros.

Los valores ausentes, aunque se permiten en los datos de las series temporales (Y), no se permiten con los predictoras. Aparecerá un mensaje de advertencia: "Se han detectado valores ausentes con los pronósticos. Las siguientes predictoras fueron eliminadas...". Si todos las predictoras tienen valores ausentes aparece un mensaje de error. Se hace una excepción para los valores ausentes en las primeras filas para acomodar los indicadores con rezagos.

A veces el impacto de una predictora no será simple e inmediato. Por ejemplo, una campaña publicitaria puede tener un impacto en las ventas durante algún tiempo después de haber culminado la campaña, y las ventas en un mes dependerán de los gastos de publicidad en cada uno de los meses anteriores. En esta situación, se debe tener en cuenta los efectos rezagados de la predictora (Hyndman, fpp2, 9.6 Predictoras Rezagadas). Un gráfico CCF preblanqueado mostrará qué desfases deben incluirse en el modelo.

SigmaXL puede acomodar las predictoras rezagadas, pero deben crearse manualmente mediante **SigmaXL > Utilities > Lag Data** (o simplemente: copiar, desplazar una fila hacia abajo, pegar una nueva columna, borrar una fila extra en la parte inferior, añadir el encabezado de la columna para los datos de la predictora, e incluirlo como una predictora en el modelo ARIMA). Obsérvese en Box y Jenkins, Cap. 11, en el que se describe un modelo del método más complejo llamado Función de Transferencia, pero esto no se proporciona en SigmaXL.

Demanda Diaria de Electricidad con Predictoras – ElecDaily

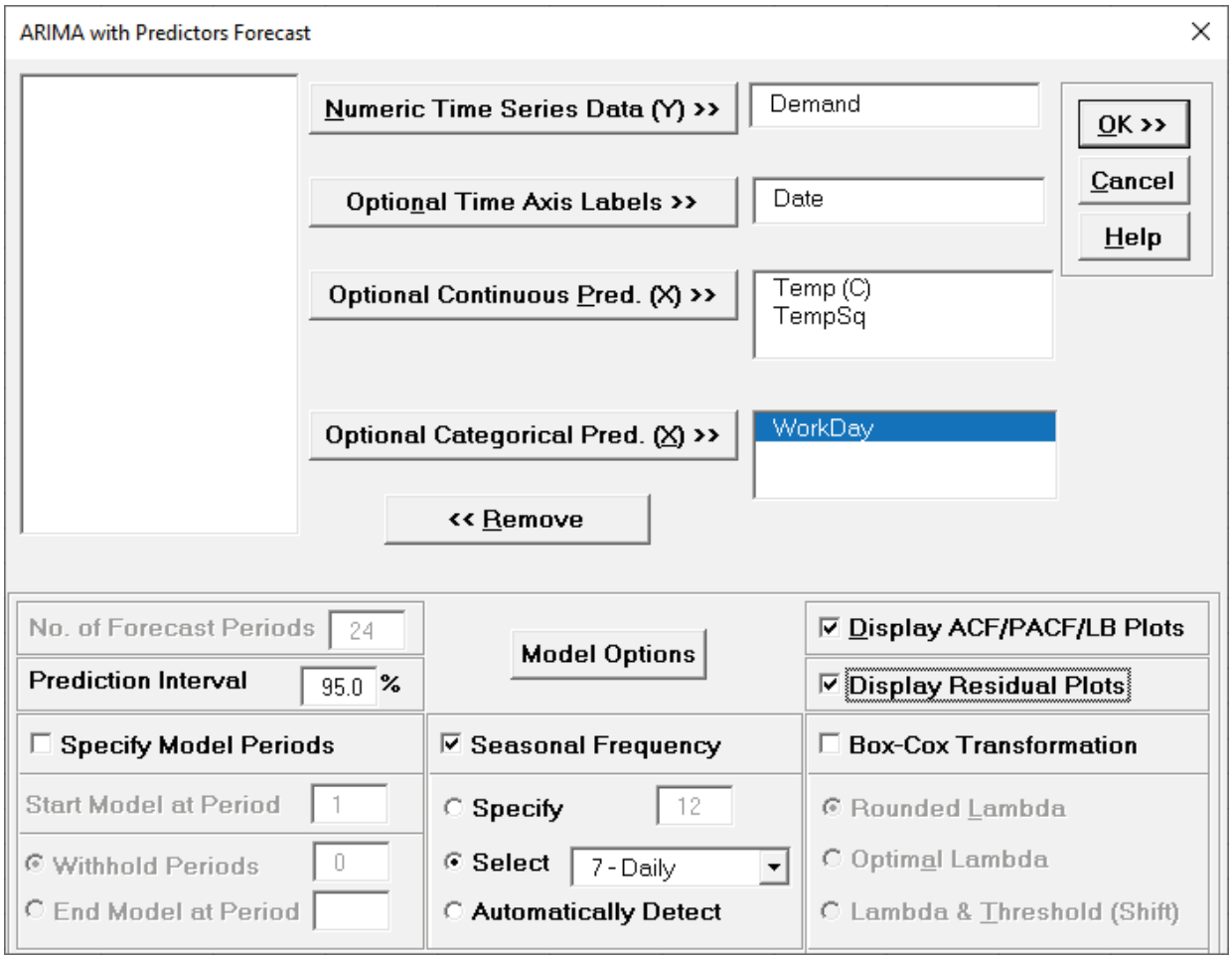
1. Abrir **Demanda Diaria de Electricidad con Predictoras – ElecDaily.xlsx** (pestaña de la Hoja 1). Esta es la demanda diaria de electricidad (GW) para el estado de Victoria, Australia, todos los días durante 2014 (Hyndman, fpp2, Section 9.3, **Ejemplo: Forecasting electricity demand**, https://otexts.com/fpp2/forecasting.html)). La temperatura (C) es la temperatura máxima diaria en grados centígrados para la ciudad de Melbourne. TempSq es la temperatura al cuadrado. WorkDay toma el valor 1 en los días laborables y 0 en los demás. Estos datos tienen una frecuencia estacional = 7 (observaciones por semana). Véase para estos datos Cartas de Corrida, Gráficos de FAC/FACP, Gráfico de Densidad Espectral y Gráficos de Descomposición de Tendencia Estacional

2. Hacer click en la pestaña de la hoja **Forecast 2 Weeks**. En el seguimiento al ejemplo del cual se hace referencia, se empleará el modelo ARIMA para pronosticar 14 días adelante empezando en Enero 1ro, 2015 (un día de asueto por ser el primer día del año). Se podría obtener pronóstico del clima para esos 14 días, pero con fines de ilustración, se establecerá que la temperatura para los 14 días permanezca constante en 26 grados y TempSq sea 676. Desplazarse hacia abajo para visualizar los datos agregados. Tener en cuenta que la Fecha, Temp (C), TempSq y WorkDay se han agregado para esos 14 días, pero los valores de la Demanda están vacíos:

1	Date	Demand	Temp (C)	TempSq	WorkDay
366	12/31/2014	186.370181	25.5	650.25	1
367	1/1/2015		26	676	0
368	1/2/2015		26	676	1
369	1/3/2015		26	676	0
370	1/4/2015		26	676	0
371	1/5/2015		26	676	1
372	1/6/2015		26	676	1
373	1/7/2015		26	676	1
374	1/8/2015		26	676	1
375	1/9/2015		26	676	1
376	1/10/2015		26	676	0
377	1/11/2015		26	676	0
378	1/12/2015		26	676	1
379	1/13/2015		26	676	1
380	1/14/2015		26	676	1

3. Hacer click en **SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast with Predictors**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

4. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**; seleccionar *Date*, hacer click en **Optional X-Axis Labels >>**; seleccionar *Temp (C)* and *TempSq*, hacer click en **Optional Continuous Pred. >>**; seleccionar *WorkDay*, hacer click en **Optional Categorical Pred. >>**. Marcar **Display ACF/PACF/LB Plots** y **Display Residual Plots**. Marcar **Seasonal Frequency** con **Seleccionar** = 7 - Daily (or **Specify** = 7). Desmarcar **Specify Model Periods y Box-Cox Transformation**. Se hace uso por defecto de **Prediction Interval** = 95.0 %.

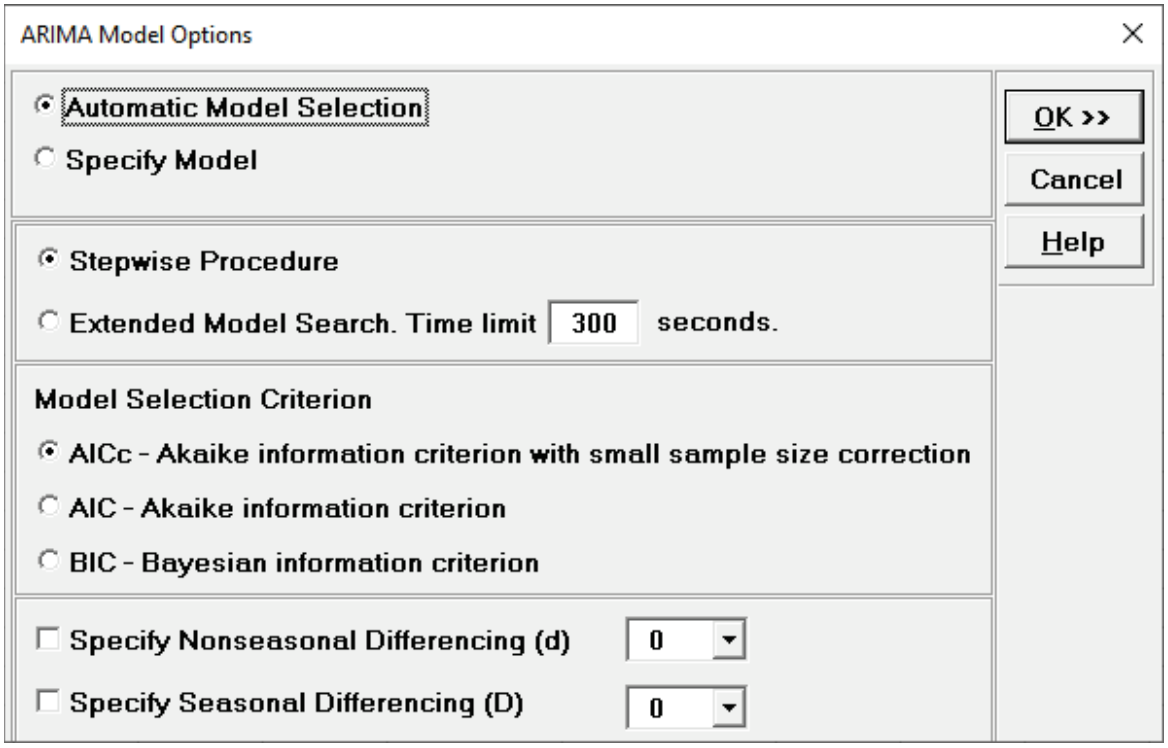


- **No. of Forecast Periods** está inhabilitado porque estos se determinan en función del número adicional de filas de predictoras que se proporcionan, en este ejemplo, 14.

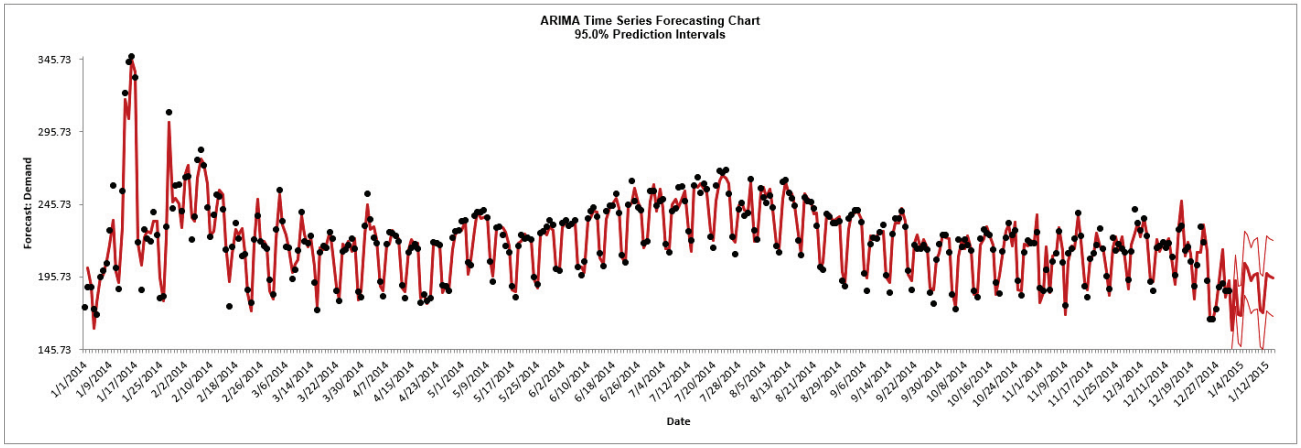
- **Optional Continuous Pred. (X)** son predictoras continuos.

- **Optional Categorical Pred. (X)** son predictoras categóricos. En este ejemplo WorkDay podría ser continuo o categórico porque está codificado como 0, 1.

5. Hacer click en **Model Options**. Seleccionar **Automatic Model Selection**. Se hace uso por defecto: **Stepwise Procedure** and **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**, desmarcar **Specify Nonseasonal Differencing (d)** y **Specify Seasonal Differencing (D)**.



6. Hacer click en **OK** para regresar al diálogo del Pronóstico ARIMA. Hacer click en **OK**. Este es un modelo complejo, de tal forma que los cálculos tomarán aproximadamente uno a dos minutos. Se genera el informe del pronóstico ARIMA:



Esto concuerda con la Figure 9.8 que aparece en Hyndman, fpp2, Sección 9.3, **Ejemplo: Forecasting electricity demand**, https://otexts.com/fpp2/forecasting.html. Existen pequeñas diferencias en los valores iniciales pronosticados.

7. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA:

ARIMA Model: Demand - Model Automatically Selected
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

Si se hubiera marcado en el diálogo principal Specify Model Periods, al inicio, final o en la Withhod Selection se resumiría aquí también.

8. Se genera el Resumen del Modelo ARIMA:

ARIMA Model Summary	
AR Order (p)	2
I Order (d)	1
MA Order (q)	2
SAR Order (P)	2
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	7
Include Constant	0
No. of Predictors	3
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

Este es un resumen de la información del modelo: ARIMA (2,1,2) (2,0,0) sin constante y con 3 predictoras. Seasonal Frequency = 7; Model Selection Criterion = “AICc” y Box-Cox Transformation = “N/A” porque se desmarcó Box-Cox Transformation.

Esto concuerda con el modelo proporcionado en Hyndman, fpp2, Sección 9.3, **Ejemplo: Forecasting electricity demand**, https://otexts.com/fpp2/forecasting.html.

9. Las Estimaciones de los Parámetros son:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
AR_1	-0.063223451	0.075658448	0.83564	0.4039
AR_2	0.673128346	0.067270503	10.0063	**0.0000**
MA_1	0.022660844	0.043288704	0.52348	0.6010
MA_2	0.929862871	0.039474102	23.5563	**0.0000**
SAR_1	0.200902989	0.053912363	3.72647	**0.0002**
SAR_2	0.402632085	0.05676416	7.09307	**0.0000**
Temp (C)	-7.501559029	0.446098708	16.8159	**0.0000**
TempSq	0.17890261	0.008530253	20.9727	**0.0000**
WorkDay_1	30.56943168	1.295720007	23.5926	**0.0000**

Las Estimaciones de los Parámetros de ARIMA incluyen tests de significancia; valores-P <0 .05 son significativos y se resaltan en color rojo. Todos las predictoras son significativos. Los términos AR_1 and MA_1 no son significativos, pero deben conservarse en el modelo por cuestiones de jerarquía. Tener en cuenta que la Selección del orden del modelo AR/MA implica que el AICc mínimo debería usarse, en lugar de los tests de significancia (véase Kostenko, A.V. and Hyndman, R.J.).

La predictora categórica *WorkDay_1* tiene el 0 como nivel de referencia oculto. Si existen tres niveles de categoría, solo dos aparecerían en la tabla.

10. Las Estadísticas del Modelo ARIMA son:

ARIMA Model Statistics	
No. of Observations	365
DF	355
StDev	6.565744215
Variance	43.1089971
Log-Likelihood	-1200.73256
AICc	2422.088353
AIC	2421.465123
BIC	2460.436662

Los grados de libertad (DF) = *n* – 10 (9 los términos en el modelo, 1 orden de diferenciación). Véase Estadísticas del Modelo ARIMA y Criterio de Información para Comparación del Modelo.

11. Las Métricas de Precisión del Pronóstico Dentro de la Muestra son:

Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) One-Step-Ahead Forecast**
N	364	
RMSE	6.615605184	
MAE	4.768089714	
MAPE	2.162224989	
MASE	0.327382095	

El MASE es menor que 1, por lo que es un mejor pronóstico que aquel que se obtendría de un pronóstico naïve (fijar todos los pronósticos para que sean el valor de la última observación). Véase Métricas de Precisión del Pronóstico.

Se puede repetir el análisis (usar Recall SigmaXL Dialog) con una muestra de retención para obtener los errores del Pronóstico Fuera-de-la-Muestra Un-Paso-Hacia-Adelante o Multi-Pasos-Hacia -Adelante, pero no se lo realizará aquí.

12. Se genera la Tabla de Pronóstico:

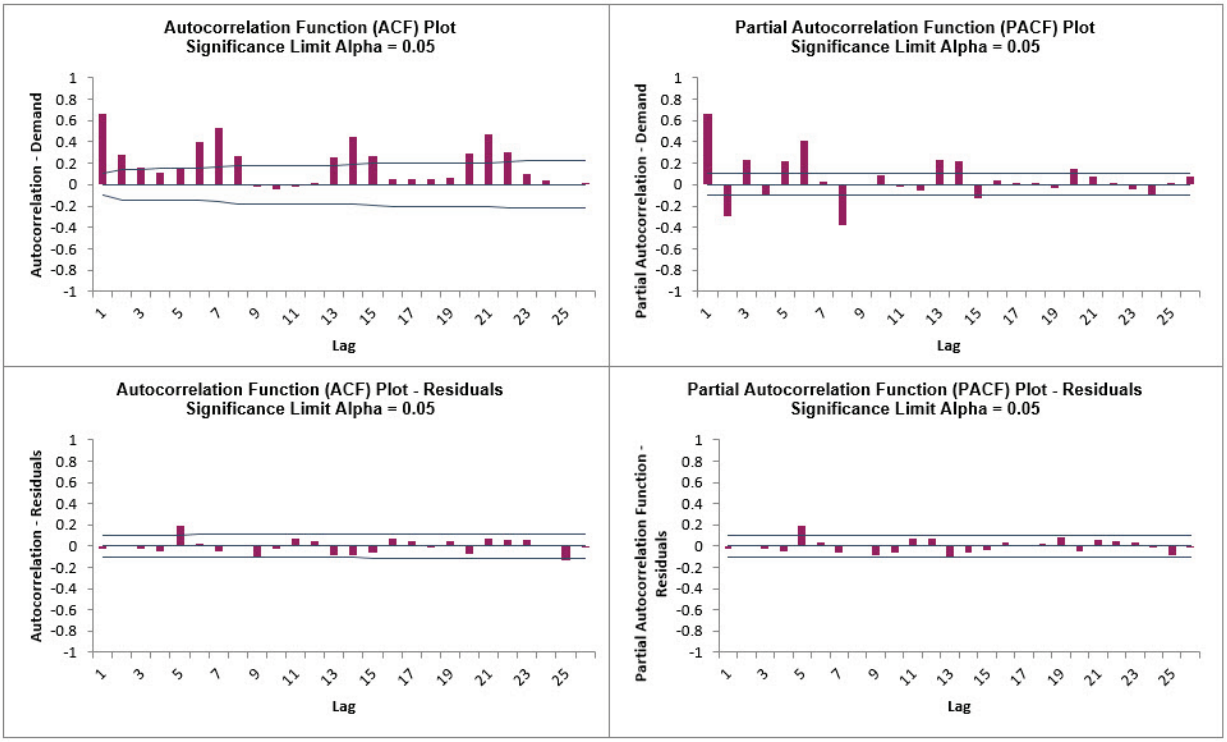
Forecast Table				
Period	**Withhold Data**	**Forecast**	**Lower 95.0% PI**	**Upper 95.0% PI**
366		159.2517864	146.3402335	172.1633392
367		193.2139395	175.7207745	210.7071045
368		169.8531097	150.3787979	189.3274216
369		168.9969534	147.9371527	190.0567541
370		205.1573872	183.2956912	227.0190833
371		201.9616978	179.3846259	224.5387697
372		193.1293236	170.1638907	216.0947565
373		197.1190445	173.1860174	221.0520716
374		198.0338842	173.4650501	222.6027184
375		173.0122062	147.9556892	198.0687231
376		171.1330507	145.7330385	196.5330629
377		198.1491778	172.4687919	223.8295637
378		196.3934701	170.5020251	222.2849151
379		194.6921179	168.6187184	220.7655175

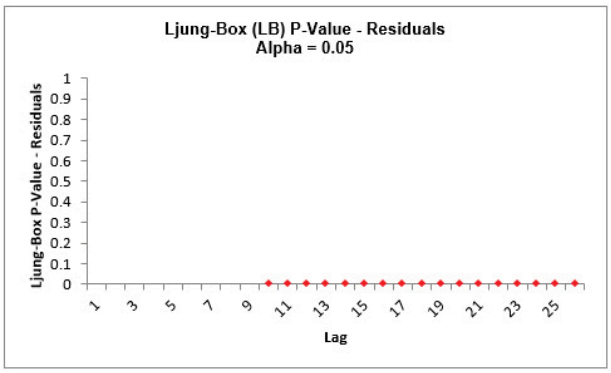
Estos son los mismos valores del pronóstico e intervalos de predicción que se visualizaron en la Carta del Pronóstico pero que se han proporcionado para un posterior análisis y gráficos. Si se especifican los Periodos de Retención, se visualizarán también los Datos de Retención.

13. Los Valores de la predictora del Pronóstico son los valores adicionales de la predictora (X) valores usados para obtener el pronóstico de 14 periodos en la Tabla de Pronósticos:

Predictor Values for Forecast		
Temp (C)	TempSq	WorkDay_1
26	676	0
26	676	1
26	676	0
26	676	0
26	676	1
26	676	1
26	676	1
26	676	1
26	676	1
26	676	0
26	676	0
26	676	1
26	676	1
26	676	1

14. Hacer click en la hoja **ARIMA ACF PACF LB** para visualizar los Gráficos FAC/FACP/LB:





Se puede apreciar que gran parte de la autocorrelación ha sido eliminada por el modelo ARIMA con Predictoras (con la excepción de los rezagos 5 y 25 en la FAC, rezago 5 en la FACP).

El gráfico de Ljung-Box muestra que todavía queda alguna autocorrelación significativa (los valores-P en color rojo son significativos con un alfa=0.05) - por lo que el modelo puede ser potencialmente mejorado. Esto no significa que el modelo sea malo, todavía puede ser muy útil para fines de predicción, pero los intervalos de predicción pueden no proporcionar una cobertura precisa.

15. Hacer click en la hoja **ARIMA Residuals** para visualizar los Gráficos de los Residuos:

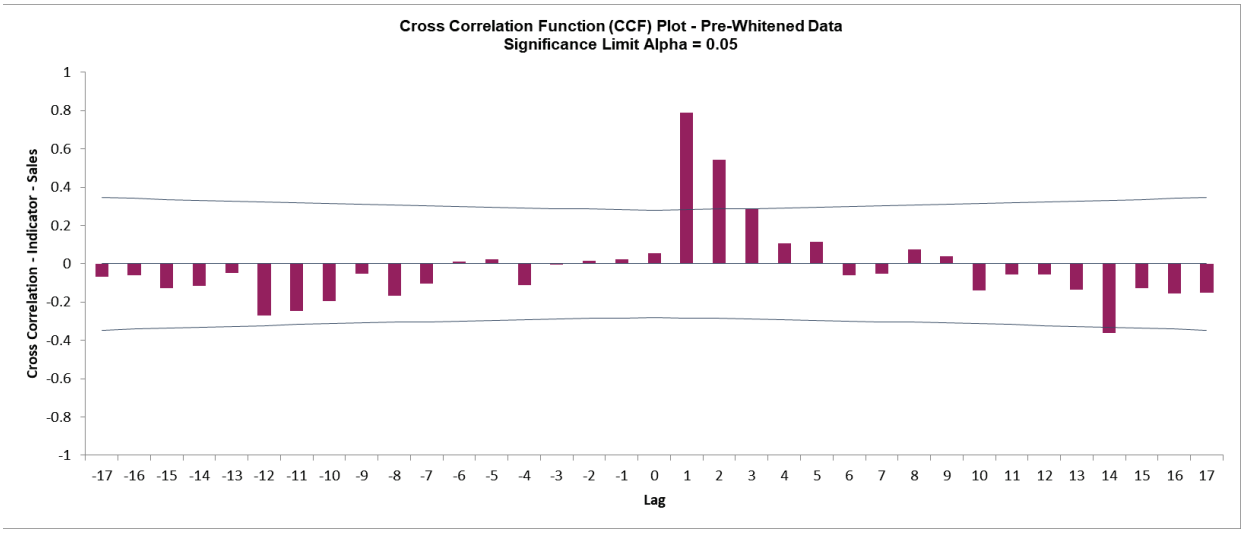
Al observar el histograma y la gráfica de probabilidad normal, hay algunos valores atípicos (pero más pequeños de lo que habrían sido si no se hubiera incluido los predictoras). Posteriormente, estos serán investigados con una carta de control de los residuos.

Obsérvese que se proporcionan gráficos de residuos adicionales para cada predictor.

Ventas con un Indicador - Serie M Modificada

16. Abrir **Sales with Indicator - Modified Series M.xlsx**. (pestaña de la **Hoja 1**). Esta es una Serie M de datos modificados de Box y Jenkins, con los valores de 150 ventas mensuales corporativas con un indicador principal, los datos fueron modificados al convertirlos a valores trimestrales al promediarlo cada tres meses, es decir 50 trimestres, etiquetados como Q1-Y1, Q-Y1, etc. Se realizó esto para simplifica el análisis del indicador principal. Aunque los datos fueron mensuales y se los resumió a trimestrales, serán considerados como no estacionales, como se llevó a cabo con Box y Jenkins. Véase para estos datos Cartas de Corrida, Gráficos de FAC/FACP, Gráfico de CCF, Gráfico de Densidad Espectral, Gráficos de Descomposición de Tendencia.

17. Recordar que el Gráfico de la CCF del preblanqueo para estos datos es:

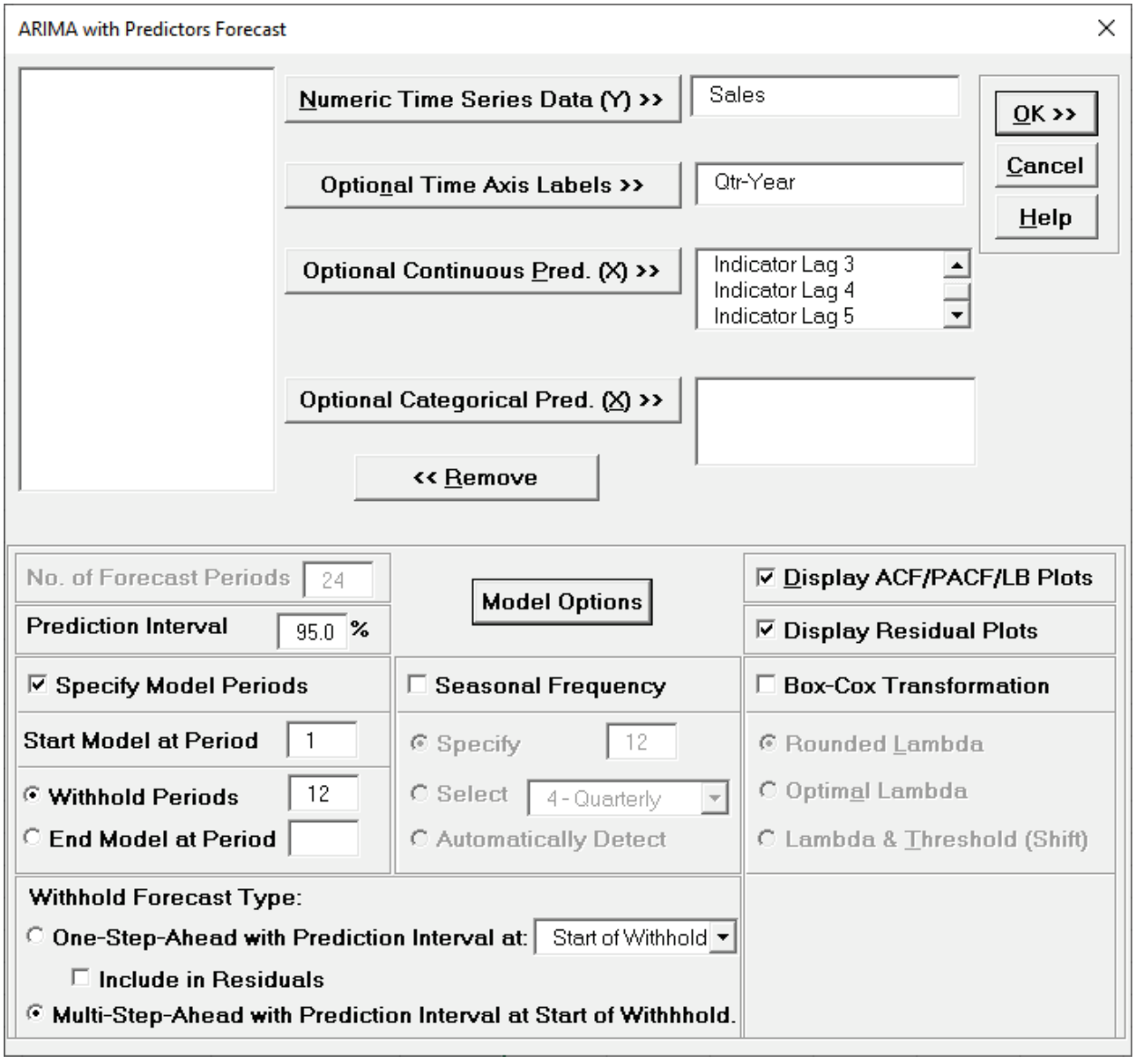


El preblanqueo de los datos ha alterado dramáticamente el gráfico de la CCF, lo que permite observar el patrón de correlación cruzada subyacente. Los rezagos 1 y 2 son significativamente positivos, y el rezago 3 está justo en la línea de significancia. Usar esto como una guía para determinar qué rezagos se pueden incluir en el modelo, sin embargo, es posible que algunos rezagos adicionales puedan ser significativos. En este ejemplo, se incluirá inicialmente hasta el rezago 5.

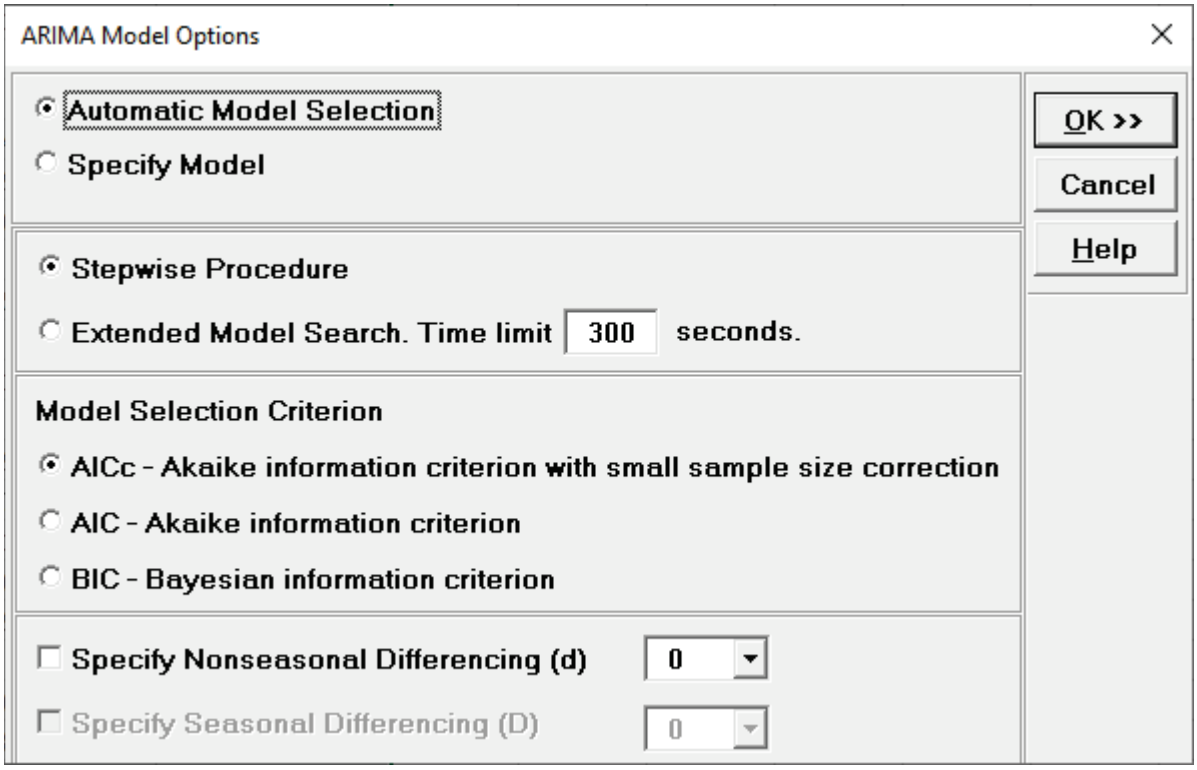
Obsérvese que mientras que X sea llamado un indicador destacado, es decir, X antecede a Y en el tiempo, el rezago positivo significa que la variable X está rezagando a la variable Y en términos de la estructura de correlación. SigmaXL utiliza esta convención como se indica en Box y Jenkins (2016, pp. 437-440).

18. Hacer click en la pestaña de la hoja **Indicator with Lags> SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast with Predictors**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

19. Seleccionar *Sales*, hacer click en **Numeric Time Series Data (Y) >>**; seleccionar *Qtr-Year*, hacer click en **Optional Time Axis Labels >>**; seleccionar *Indicator* to *Indicator Lag 5*, hacer click en **Optional Continuous Pred. >>**. Marcar **Display ACF/PACF/LB Plots** y **Display Residual Plots**. Marcar **Especificar Model Periods**. Fijar **Withhold Periods** = 12 (i.e., 3 years). Seleccionar **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Desmarcar **Specify Model Periods**, **Seasonal Frequency** y **Box-Cox Transformation**. Se hace uso por defecto de **Prediction Interval** = 95.0 %.

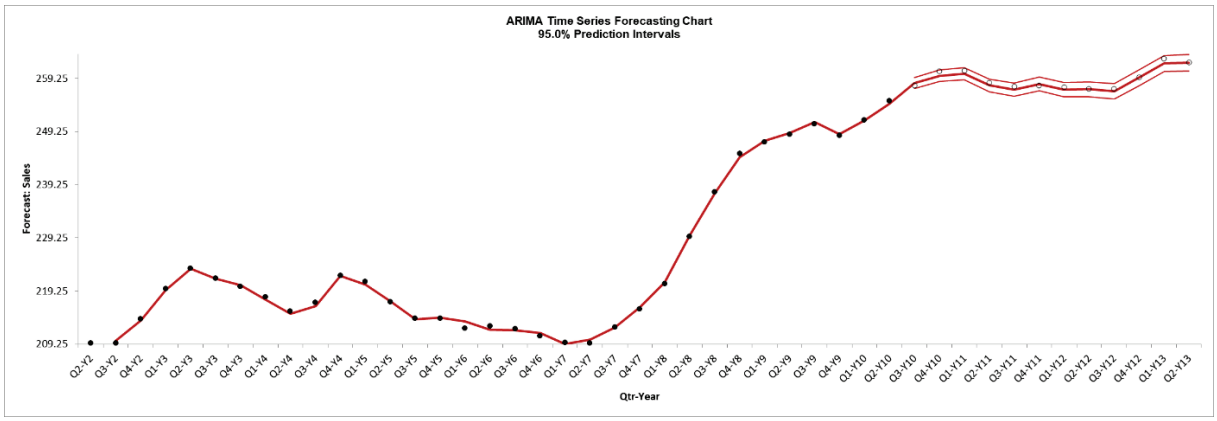


20. Hacer click en **Model Options**. Seleccionar **Automatic Model Selection**. Se hace uso por defecto de: **Stepwise Procedure** y **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**, desmarcar **Specify Nonseasonal Differencing (d)** y **Specify Seasonal Differencing (D)**.



Consejo: Cuando se utiliza el diálogo **Recall SigmaXL** y si no hay cambios en la configuración de **Model Option**, se utilizarán los ajustes anteriores. No es necesario repetir este paso.

21. Hacer click en **OK** para retornar al diálogo del Pronóstico ARIMA. Hacer click en **OK**. Se genera el informe del Pronóstico ARIMA con Predictoras:



Los puntos en blanco son los valores de los datos de la muestra de retención con un pronóstico de varios pasos e intervalos de predicción que se visualizaron al comienzo de la muestra de retención. El modelo utiliza los valores del indicador de rezago en la muestra de retención y lo hace bastante bien al pronosticar los 12 trimestres de los valores de las Ventas.

Obsérvese que el gráfico comienza en el Q2-Y2. Esto se debe a que se han eliminado las primeras 5 filas de datos X e Y, ya que hay valores que faltan en los intervalos del indicador 1 al 5.

22. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA:

ARIMA Model: Sales - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 12 withhold periods. Missing values in initial rows of predictors. Deleted first 5 rows of X and Y data.

23. Se genera el Resumen del Modelo ARIMA:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	1
Include Constant	1
No. of Predictors	6
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

Este es el resumen de la información del modelo: ARIMA (0,1,1) sin constante y 6 predictoras. Seasonal Frequency = 1 (nonseasonal); Model Selection Criterion = “AICc” y Box-Cox Transformation = “N/A” porque se desmarcó Box-Cox Transformation.

24. Las Estimaciones de los Parámetros son:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.661826535	0.171947916	3.84899	**0.0008**
Const:Trend	0.107385184	0.033507981	3.20476	**0.0038**
Indicator	-0.193804762	0.279261631	0.69399	0.4944
Indicator Lag 1	8.199265436	0.304329926	26.942	**0.0000**
Indicator Lag 2	5.392904003	0.2952757	18.264	**0.0000**
Indicator Lag 3	2.481788513	0.301027513	8.24439	**0.0000**
Indicator Lag 4	0.885376557	0.324714746	2.72663	**0.0118**
Indicator Lag 5	0.154053202	0.264999213	0.58133	0.5664

Las Estimaciones de los Parámetros de ARIMA incluyen tests de significancia; valores-P <0 .05 son significativos y se resaltan en color rojo. El Gráfico de la CCF Plot sugiere que sólo los rezagos 1 y 2, posiblemente el 3, serían significativos, pero se puede apreciar aquí que el rezago 4 es también significativo (de allí por qué la CCF es solo una guía aproximada). Posteriormente, se repetirá el modelo y se removerá el Indicador & Indicador de rezago 5 y se comparará en forma manual los valores AICc.

25. Las Estadísticas del Modelo ARIMA con Predictoras son:

ARIMA Model Statistics	
No. of Observations	33
DF	24
StDev	0.505833251
Variance	0.255867277
Log-Likelihood	-19.8991901
AICc	65.98019828
AIC	57.7983801
BIC	70.99000323

- El número de observaciones, *n* = 50 – 12 (*retención*) – 5 (*filas eliminadas*) = 33
- Los grados de libertad (DF) = 33 (*n*) – 9 (8 los términos en el modelo, 1 diferencia no estacional) = 24

26. Las métricas de la Precisión del Pronóstico son:

Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Multi-Step-Ahead Forecast**
N	32	12
RMSE	0.441579854	0.519299269
MAE	0.355486014	0.433503476
MAPE	0.160638285	0.167034359
MASE	0.123200929	0.150239472

Los errores del pronóstico Fuera-de-la-Muestra son sólo apenas mayores que los de Dentro-de-la-Muestra, así que esto es una buena predicción.

El MASE es menor que 1, por lo que es un mejor pronóstico que aquel que se obtendría de un pronóstico naïve (fijar todos los pronósticos para que sean el valor de la última observación). Véase Métricas de Precisión del Pronóstico.

27. Se genera la Tabla de Pronóstico:

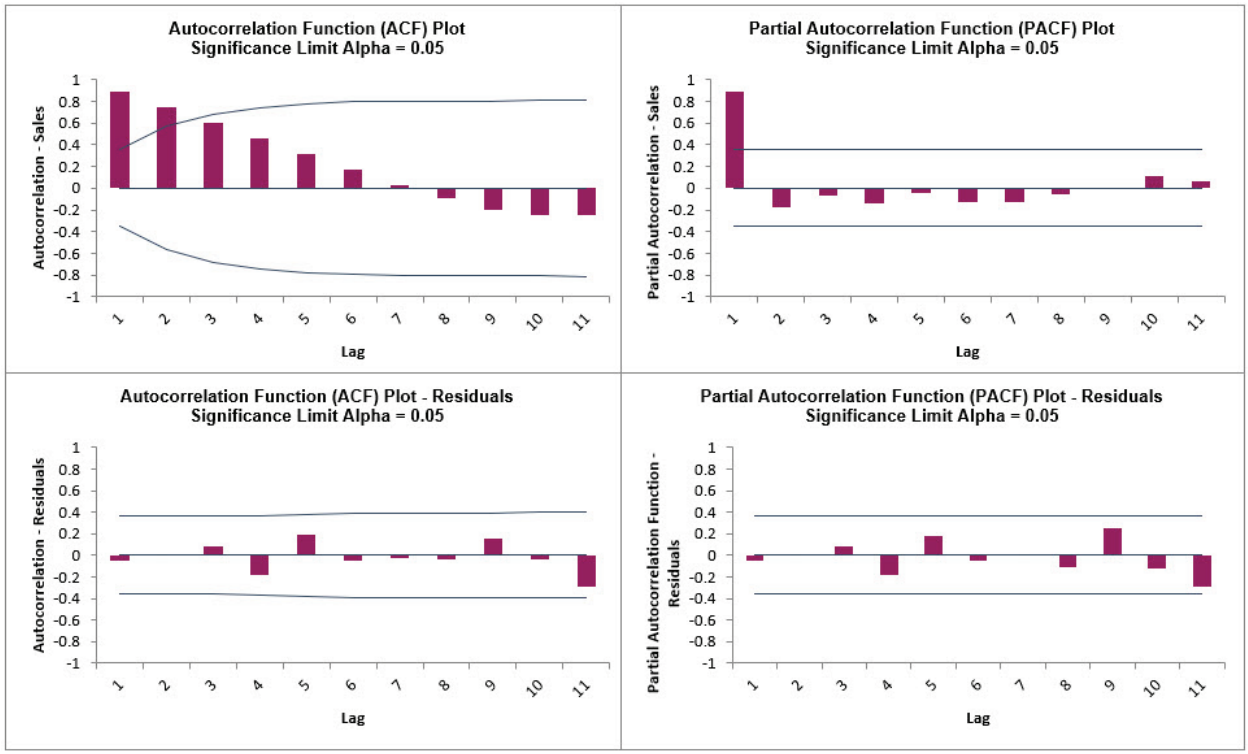
Forecast Table				
Period	**Withhold Data**	**Multi-Step Ahead Forecast**	**Lower 95.0% PI**	**Upper 95.0% PI**
34	257.9333333	258.3217379	257.2913893	259.3520865
35	260.5666667	259.7168718	258.6292016	260.8045419
36	260.7333333	260.1163414	258.974223	261.2584597
37	258.4333333	257.9037045	256.7096181	259.097791
38	257.6666667	257.0834025	255.8395172	258.3272877
39	257.9333333	258.1603883	256.8686227	259.4521539
40	257.5666667	257.1341925	255.7962588	258.4721261
41	257.2666667	257.1889296	255.8063687	258.5714904
42	257.3	256.7947588	255.3689669	258.2205507
43	259.4	259.3538829	257.8861328	260.8216331
44	263	262.0730127	260.5644708	263.5815546
45	262.2333333	262.2149879	260.6667286	263.7632471

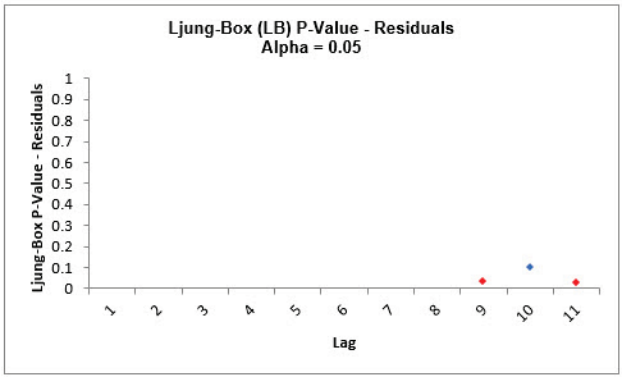
Se trata de los mismos valores del pronóstico e intervalos de predicción que se visualizan en la Carta de Pronóstico, que se proporcionan para un análisis o un gráfico más detallado. Observe que esta tabla genera el número del Período de Pronóstico mientras que el Gráfico muestra el Trimestre-Año.

28. Los valores de predicción para el pronóstico son los valores de predicción adicionales (X) utilizados para obtener el pronóstico de 12 períodos en la tabla de pronóstico:

Predictor Values for Forecast					
Indicator	**Indicator Lag 1**	**Indicator Lag 2**	**Indicator Lag 3**	**Indicator Lag 4**	**Indicator Lag 5**
13.4166667	13.5	13.27	13.13333333	12.62666667	13.01
13.4066667	13.41666667	13.5	13.27	13.13333333	12.62666667
13.13	13.40666667	13.41666667	13.5	13.27	13.13333333
13.21	13.13	13.40666667	13.41666667	13.5	13.27
13.3566667	13.21	13.13	13.40666667	13.41666667	13.5
13.13	13.35666667	13.21	13.13	13.40666667	13.41666667
13.2233333	13.13	13.35666667	13.21	13.13	13.40666667
13.16	13.22333333	13.13	13.35666667	13.21	13.13
13.4966667	13.16	13.22333333	13.13	13.35666667	13.21
13.6033333	13.49666667	13.16	13.22333333	13.13	13.35666667
13.4433333	13.60333333	13.49666667	13.16	13.22333333	13.13
13.56	13.44333333	13.60333333	13.49666667	13.16	13.22333333

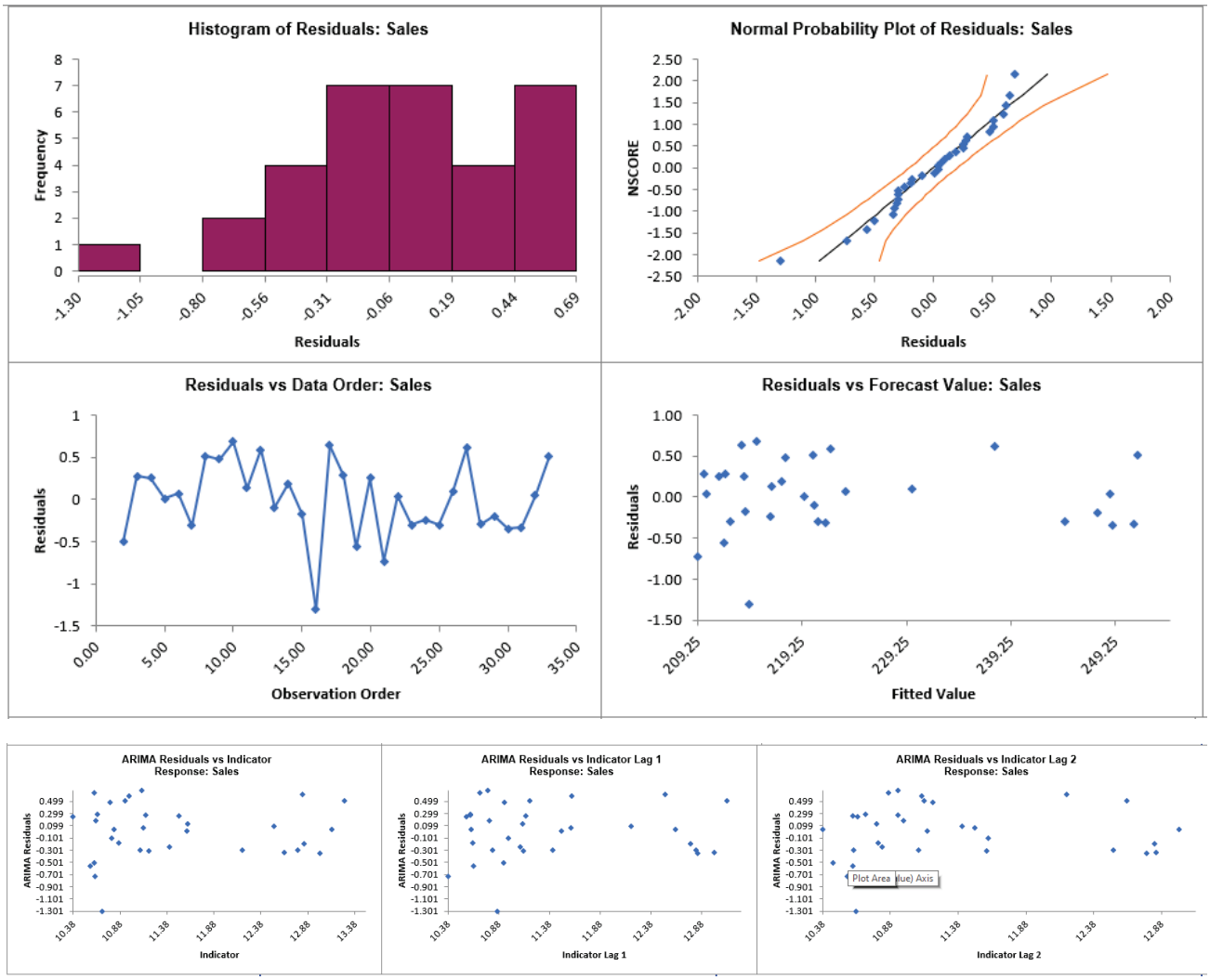
29. Hacer click en la hoja **ARIMA ACF PACF LB** para visualizar los Gráficos de FAC/FACFP/LB:





Se puede apreciar que gran parte de la autocorrelación se ha eliminado por el modelo ARIMA con Predictoras. El gráfico de Ljung-Box muestra que todavía queda alguna autocorrelación significativa (los valores-P de color rojo son significativos con un alfa = 0,05) - por lo que el modelo puede ser potencialmente mejorado. Esto no significa que el modelo sea un mal modelo, todavía puede ser muy útil para fines de predicción, pero los intervalos de predicción pueden no proporcionar una cobertura precisa.

30. Hacer click en la hoja **ARIMA Residuals** para visualizar los Gráficos de los Residuos:

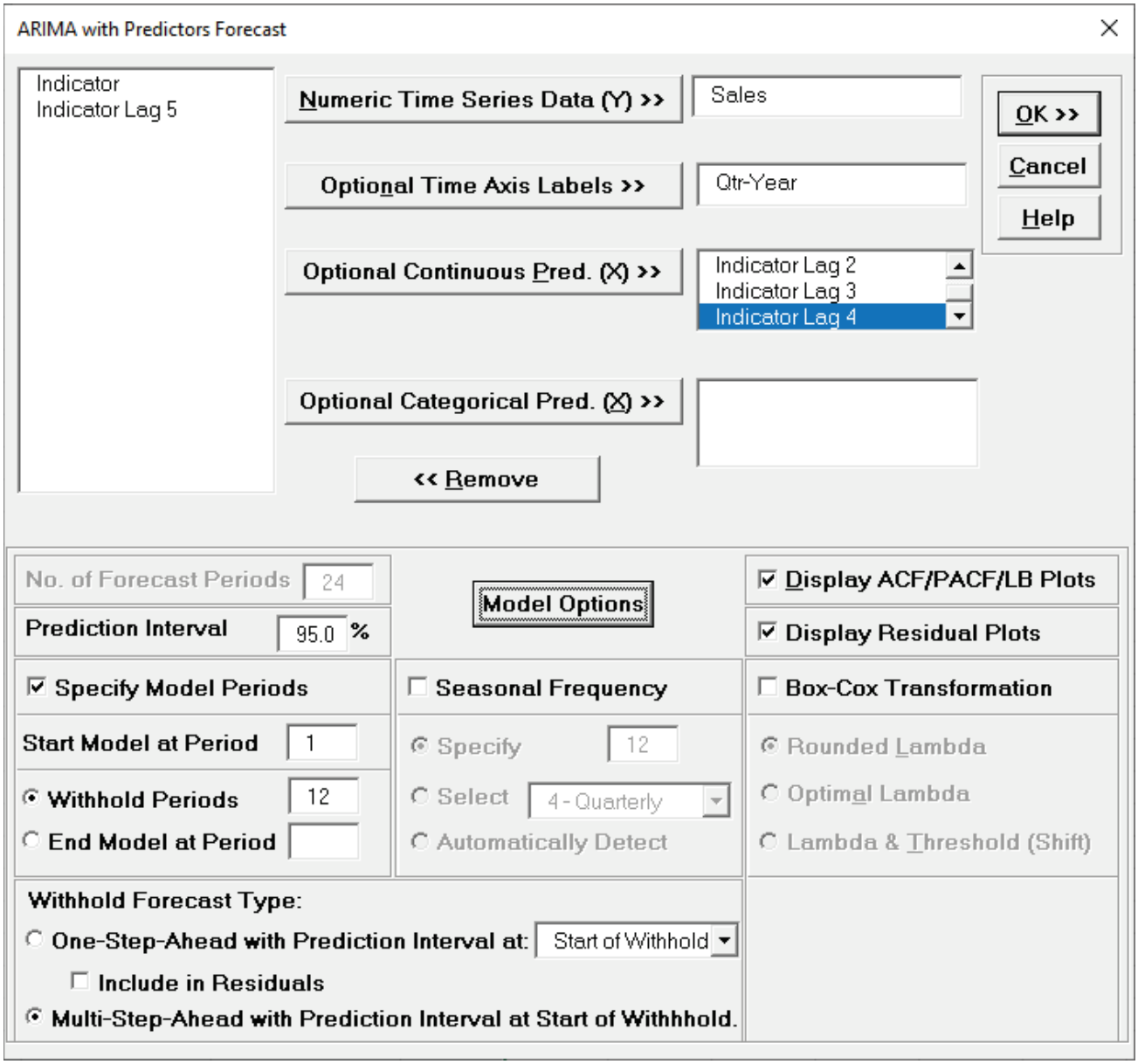


Los residuos están distribuidos de forma casi normal, con una línea aproximadamente recta en la gráfica de probabilidad normal. No hay valores extremos atípicos ni patrones obvios en los gráficos. El histograma puede sugerir un sesgo a la izquierda, pero el tamaño de la muestra es bastante pequeño. (También se pueden aplicar sobre los residuos con el uso de SigmaXL > Statistical Tools > Descriptive Statistics: Options, Additional Normality Tests, lo que mostraría que los residuos son normales).

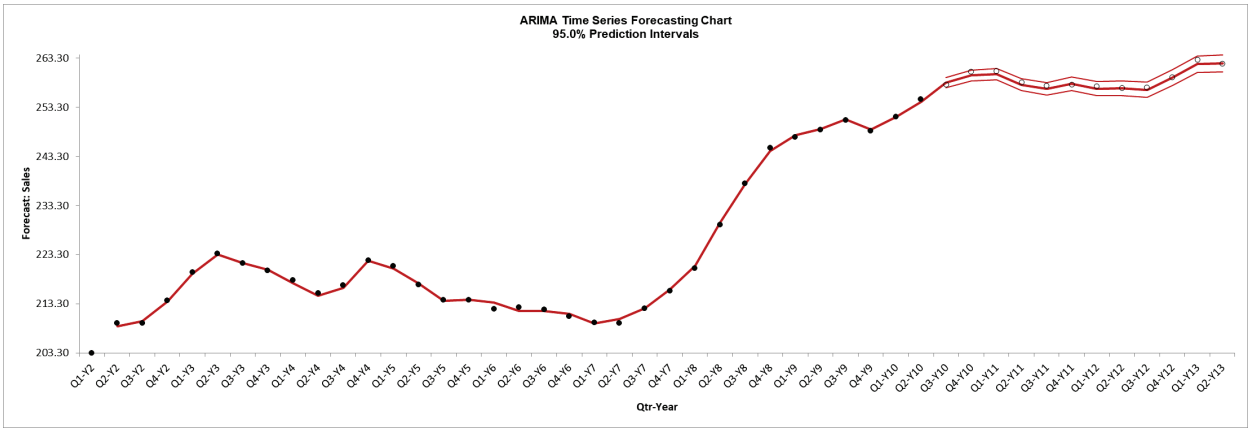
Obsérvese que los gráficos adicionales de los residuos se proporcionan para cada predictor.

31. Se repetirá el análisis y se removerá los términos que no sean significativos, *Indicador* y *Rezago del Indicador 5*. Hacer click en el menú **Recall SigmaXL Dialog** o presione **F3** para repetir el diálogo.

32. Seleccionar *Indicator,* hacer click en Remove; seleccionar *Indicator Lag 5*, hacer click en Remove:



33. No se realizarán cambios a **Model Options**. Hacer click en **OK**. Se genera el informe del Pronóstico ARIMA con Predictoras:



Tener en cuenta ahora comienza en Q1-Y2. Esto se debe a que se han eliminado las primeras 4 filas de datos de X e Y, debido a que faltan valores en los intervalos de los indicadores 1 a 4.

34. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA:

ARIMA Model: Sales - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 12 withhold periods. Missing values in initial rows of predictors. Deleted first 4 rows of X and Y data.

35. Se genera el Resumen del Modelo ARIMA:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	1
Include Constant	1
No. of Predictors	4
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

Este es un resumen de la información del modelo ARIMA (0,1,1) sin ninguna constante que no ha cambiado. Ahora existen 4 predictoras. Seasonal Frequency = 1 (nonseasonal); Model Selection Criterion = “AICc” y Box-Cox Transformation = “N/A” debido a que Box-Cox Transformation no estaba marcada.

36. Las Estimaciones de los Parámetros son:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.579458707	0.137596707	4.21128	**0.0003**
Const:Trend	0.110416608	0.039494253	2.79576	**0.0094**
Indicator Lag 1	8.14204032	0.260590425	31.2446	**0.0000**
Indicator Lag 2	5.381655285	0.288605873	18.6471	**0.0000**
Indicator Lag 3	2.521912726	0.293140527	8.60308	**0.0000**
Indicator Lag 4	0.932902504	0.262059255	3.55989	**0.0014**

Las Estimaciones del Parámetro ARIMA incluyen los tests de significancia; los Valores-P <0.05 son significativos y están resaltados en color rojo. El Gráfico de la CCF sugiere que sólo los Rezagos 1 y 2, posiblemente el 3, serían significativos, pero se puede apreciar que el Rezago 4 también es significativo (de allí por qué la CCF es simplemente una guía).

37. Las Estadísticas del Modelo ARIMA con Predictoras son:

ARIMA Model Statistics	
No. of Observations	34
DF	27
StDev	0.491053186
Variance	0.241133231
Log-Likelihood	-20.6849129
AICc	59.84982585
AIC	55.36982585
BIC	65.84537878

- El número de observaciones, *n* = 50 – 12 (*retención*) – 4 (*filas eliminadas*) = 34
- Los grados de libertad (DF) = 34 (*n*) – 7 (6 los términos en el modelo, 1 diferencia no estacional) = 27
- El AICc es menor que las Estadísticas del Modelo ARIMA con Predictoras que incluyen los *indicadores* de los términos no significativos.

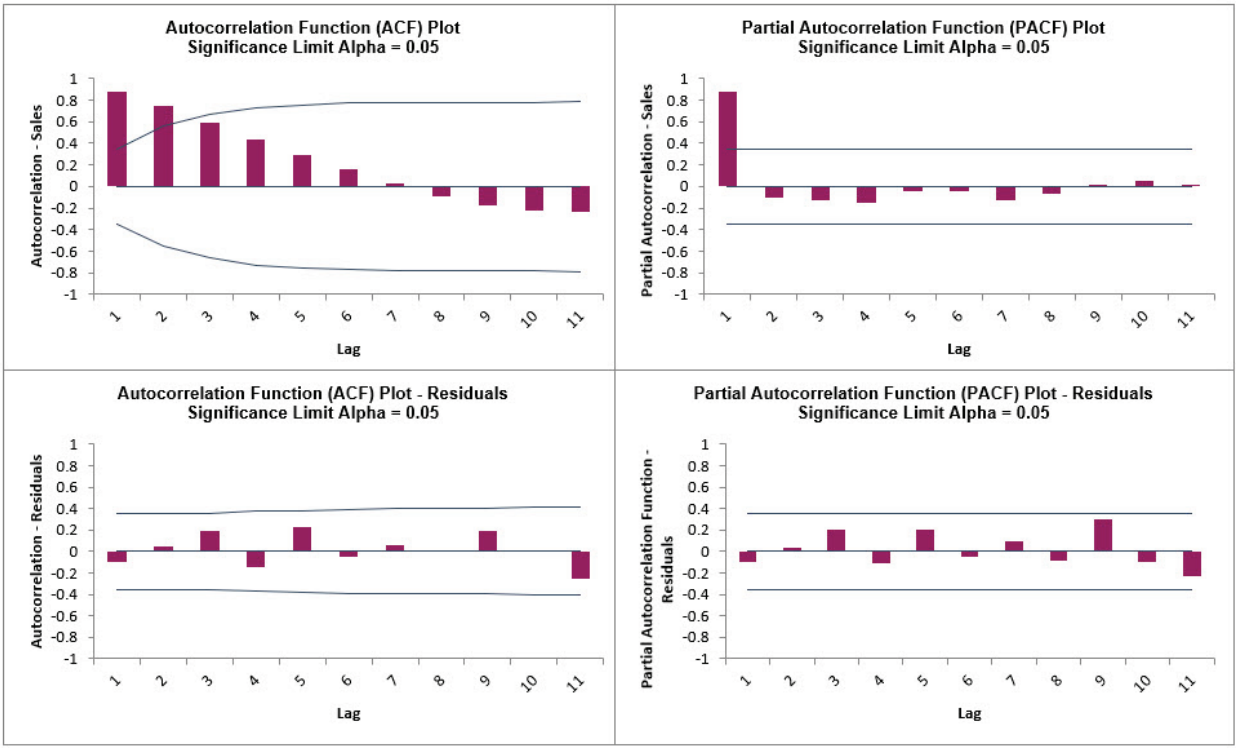
38. Las métricas de la Precisión del Pronóstico son:

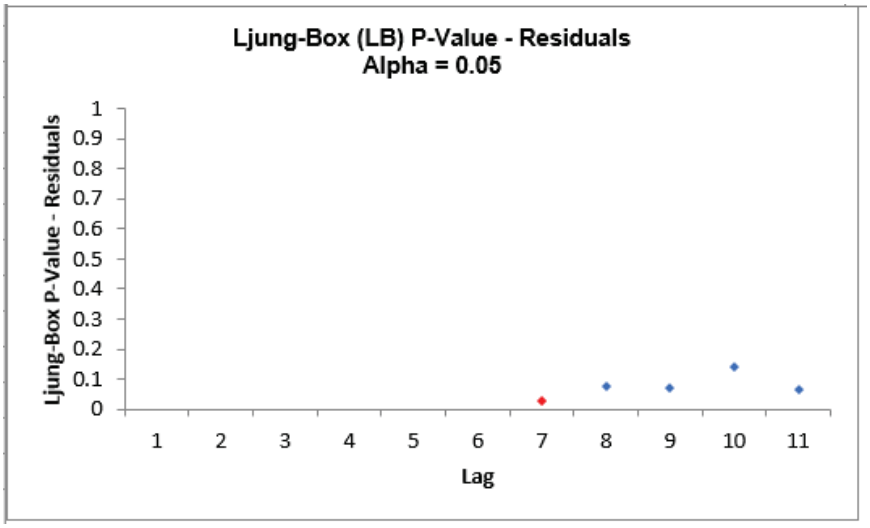
Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) Multi-Step Ahead Forecast**
N	33	12
RMSE	0.44993435	0.505510545
MAE	0.363143972	0.420790326
MAPE	0.164705945	0.162132467
MASE	0.121786088	0.141118707

En comparación con las métricas de Precisión del Pronóstico, las métricas de Precisión de los errores del Pronóstico Fuera-de-la-Muestra son ligeramente menores, por lo que esta es una buena predicción con un modelo más simple.

El MASE es menor que 1, por lo que es un mejor pronóstico que la que se obtendría a partir de un pronóstico naïve (establecer que todos los pronósticos sean el valor de la última observación). Véase Métricas de Precisión del Pronóstico.

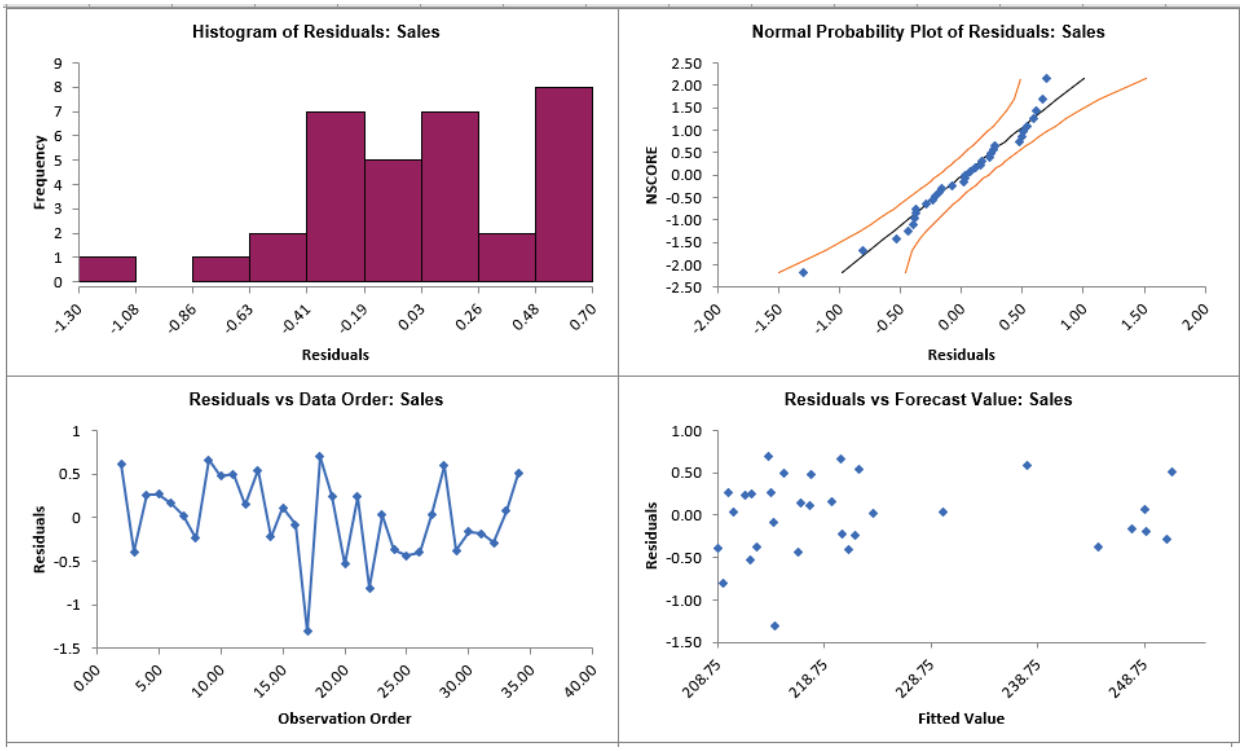
39. Hacer click en la hoja **ARIMA ACF PACF LB** para visualizar los Gráficos FAC/FACP/LB:

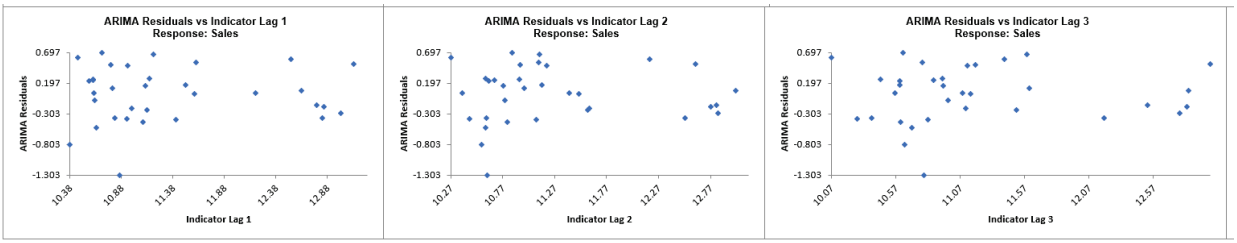




Se puede observar que gran parte de la autocorrelación se ha eliminado por el modelo ARIMA con Predictoras. El gráfico de Ljung-Box muestra menores valores-P significativos que los de los gráficos de FAC/FACP/LB del modelo.

40. Hacer click en la hoja **ARIMA Residuals** para visualizar los Gráficos de los Residuos:





Los residuos tienen un aspecto similar a los Gráficos de los Residuos del modelo casi de una distribución normal, con una línea aproximadamente recta en el gráfico de probabilidad normal. No hay valores extremos o patrones obvios de valores atípicos en los gráficos. El histograma podría sugerir un sesgo a la izquierda, pero el tamaño de la muestra es bastante pequeño. (Se pueden aplicar también tests de Normalidad para los residuos con el uso de SigmaXL > Statistical Tools > Descriptive Statistics: Options, Additional Normality Tests, y estos mostrarían todos los residuos que son normales).

Obsérvese que los gráficos adicionales de los residuos se proporcionan para cada predictor, *Indicador de Rezagos 1 al 4*.

Pronóstico Descomposición Estacional Múltiple ARIMA – (DEM)

ARIMA no tiene un límite de frecuencia teórico, pero por cuestión de la eficiencia computacional y para minimizar la pérdida potencial de observaciones por diferenciación, se recomienda usar ARIMA - Descomposición Estacional Múltiple (DEM) para una frecuencia estacional mayor de 52 (o con frecuencias múltiples).

El componente estacional se elimina primero a través de la descomposición, un modelo ARIMA no estacional se ajusta a los residuos (tendencia +), y luego se vuelve a añadir el componente estacional. Para el pronóstico, se utiliza un pronóstico estacional naïve sobre el componente estacional. Obsérvese que los intervalos de predicción se derivan del modelo ARIMA y no incluyen la incertidumbre en el componente estacional.

Como su nombre indica, la Descomposición Estacional Múltiple (DEM) también da cabida a la estacionalidad múltiple, por ejemplo: los datos de media hora con una frecuencia estacional de 48 observaciones por día y 336 observaciones por semana. Al utilizar DEM, se recomienda limitar el período de pronóstico a 2* una frecuencia estacional dominante.

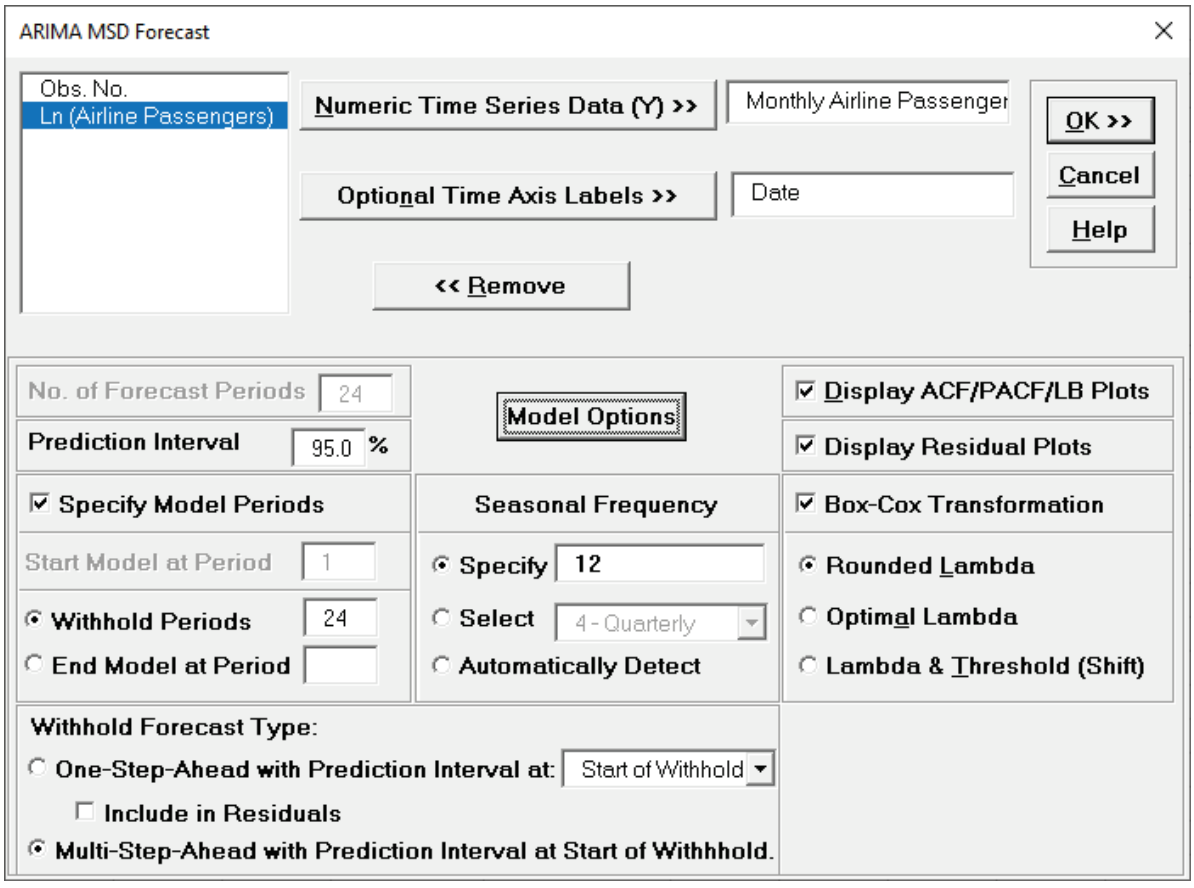
Para mayores detalles y las fórmulas, véase el Apéndice: **Descomposición de Tendencia Estacional.**

Monthly Airline Passengers - Series G

1. Abrir **Monthly Airline Passengers - Series G.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional de enero 1949 a diciembre 1960. Véase para estos datos las Cartas de Corrida, los Gráficos de FAC/FACP, el Gráfico de Densidad Espectral y los Gráficos de Descomposición de Tendencia Estacional. No es necesaria para estos datos la opción de Descomposición Estacional Múltiple (DEM), pero como una forma de introducción, se lo utilizará para compararla con el previo análisis ARIMA.

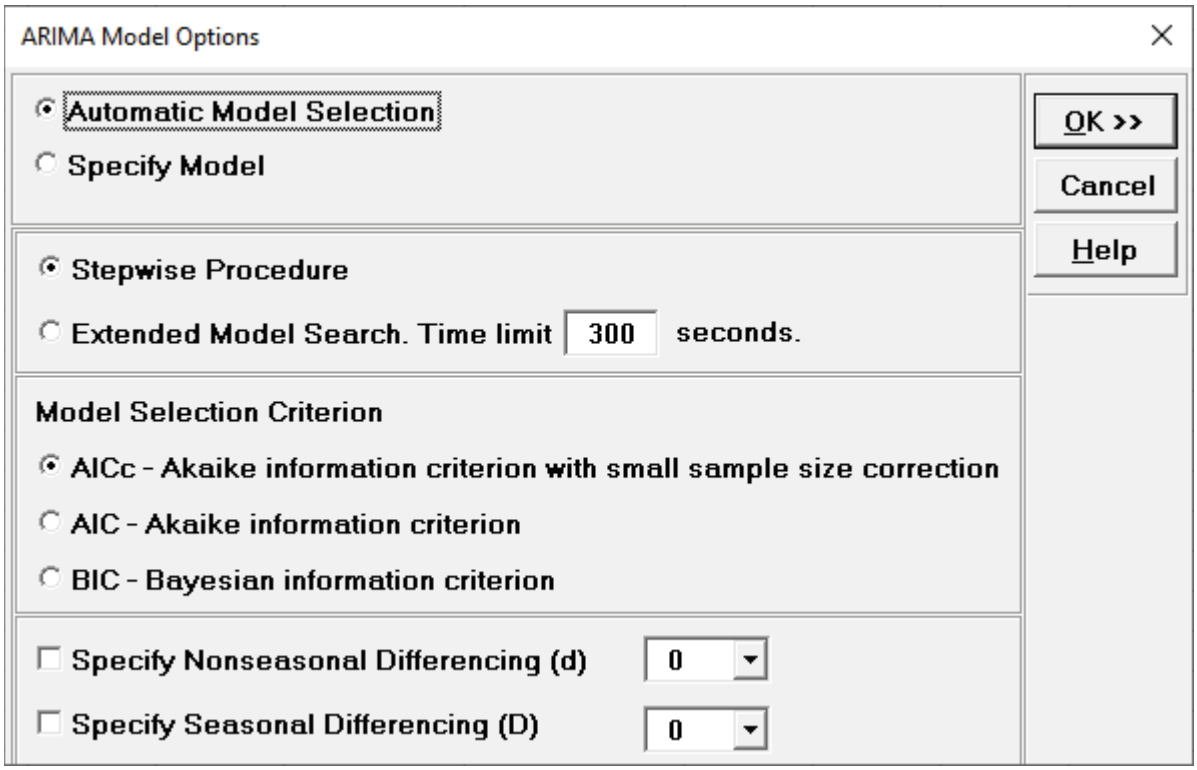
2. Hacer click en **SigmaXL > Time Series Forecasting > ARIMA Forecast > Multiple Seasonal Decomposition Forecast**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Monthly Airline Passengers*, hacer click en **Numeric Time Series Data (Y) >>**. Seleccionar *Date*, hacer click en **Optional Time Axis Labels >>**. Marcar **Display ACF/PACF/LB Plots** y **Display Residual Plots**. Marcar **Specify Model Periods**. Establecer **Withhold Periods** = 24. Seleccionar **Withhod Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold**. Seleccionar **Seasonal Frequency Specify** y Registrar 12**.** Marcar **Box-Cox Transformation** and Seleccionar **Rounded Lambda**. Se hace uso por defecto de **Prediction Interval** = 95.0 %.

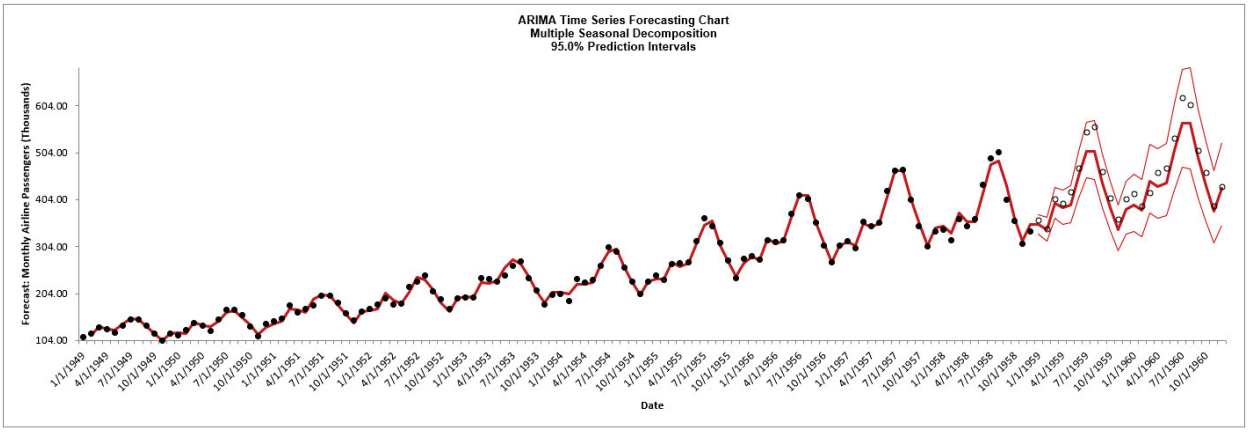


- **Seasonal Frequency** puede tener múltiples entradas, sin embargo, no se recomienda más de 3 valores.

4. Hacer click en **Model Options**.



5. Se empleará por defecto **Automatic Model Selection** con **AICc** como el **Model Selection Criterion**. Hacer click en **OK** para retornar al diálogo del Pronóstico ARIMA DEM. Hacer click en **OK**. Se genera el informe del pronóstico ARIMA:



6. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA:

ARIMA Model (Multiple Seasonal Decomposition): Monthly Airline Passengers (Thousands) - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 24 withhold periods.

7. Se genera el Resumen del Modelo ARIMA:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	12 Decomposition
Include Constant	1
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	Rounded Lambda
Lambda	0
Threshold	0

Este es un resumen de la información del modelo para los datos desestacionalizados: ARIMA (0,1,1) con una constante. La Frecuencia estacional = 12 usando el Criterio de Descomposición y Selección de Modelos = "AICc". No hay términos estacionales en el modelo. La transformación Box-Cox es "Lambda redondeada" con Lambda = 0 (transformación Ln).

8. Las Estimaciones de los Parámetros desestacionalizados para los Pasajeros de la Aerolínea son:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.322113898	0.095098865	3.38715	**0.0010**
Const:Trend	0.009427003	0.001881534	5.01027	**0.0000**

9. Las Estadísticas del Modelo ARIMA son:

ARIMA Model Statistics	
No. of Observations	120
DF	117
StDev	0.030409792
Variance	0.000924755
Log-Likelihood	247.7313206
AICc	-489.253946
AIC	-489.462641
BIC	-481.125271

- El número de observaciones, *n* = 144 – 24 (*retención*) = 120
- Los grados de libertad (DF) = 120 (*n*) – 3 (2 los términos en el modelo, 1 diferencia no estacional) = 117
- Tener en cuenta que las estadísticas se basan en la transformación Ln de los datos, no en los datos originales.
- Al comparar las Estadísticas del Modelo de Suavización Exponencial, se puede apreciar que la DesvEst y la Varianza son aproximadamente las mismas, pero el Log-Verosimilitud, el AICc, el AIC y el BIC son muy diferentes. Esto se debe al uso de fórmulas distintas que se emplean en la función de Verosimilitud. No puede usar el Criterio de Información para comparar los modelos ARIMA y Suavización Exponencial para determinar cuál modelo tiene el mejor ajuste.

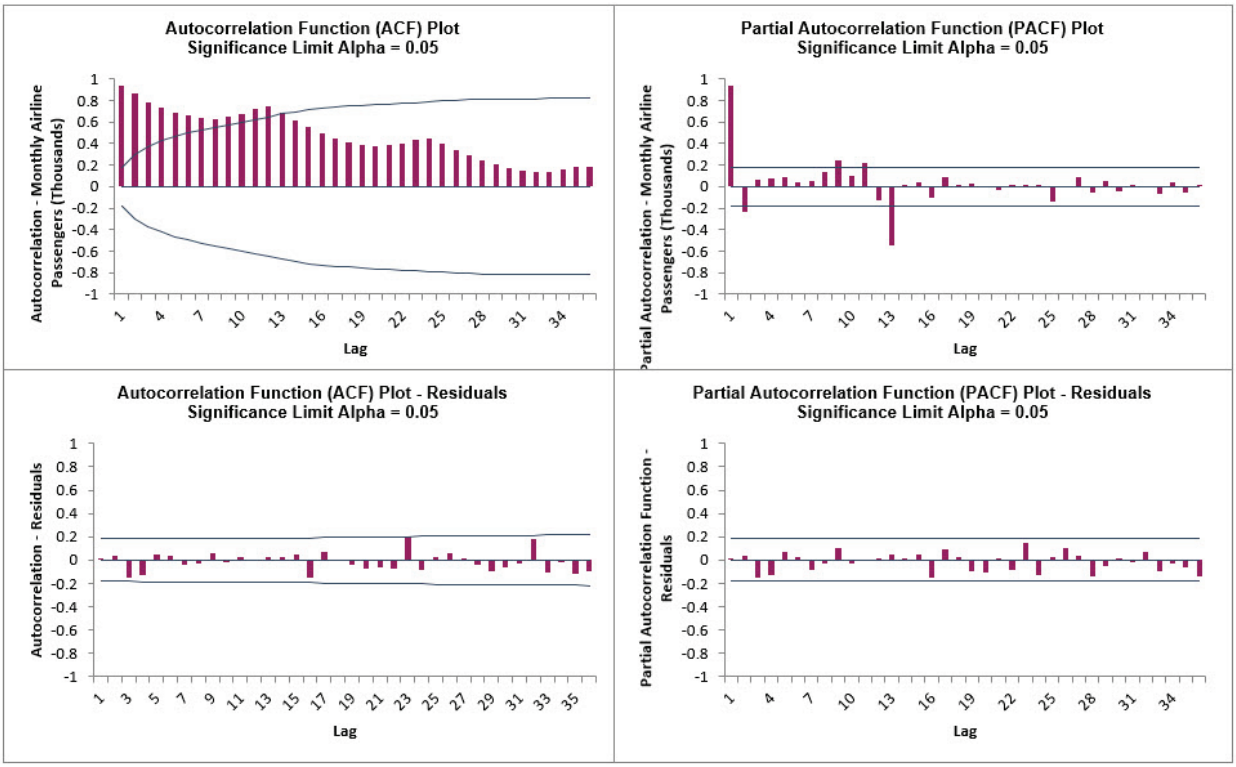
10. Las métricas de la Precisión del Pronóstico son:

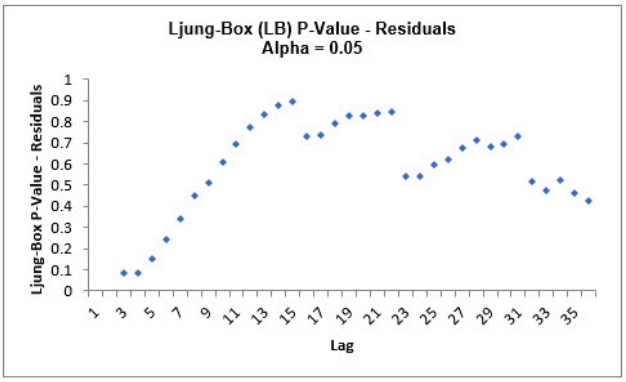
Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) Multi-Step Ahead Forecast**
N	119	24
RMSE	7.145744373	26.75002158
MAE	5.494721757	23.0424413
MAPE	2.364132406	4.85347701
MASE	0.192297456	0.806410778

En comparación con las métricas de Precisión del Pronóstico de Suavización Exponencial DEM, se puede apreciar que los resultados del ARIMA DEM son casi los mismos.

Las métricas de la Precisión del Pronóstico se calculan utilizando los datos actuales sin procesar versus el pronóstico de la transformada inversa tal y como se muestra en el Gráfico y la Tabla de Pronóstico, lo que permite la comparación entre todos los tipos de modelos y transformaciones.

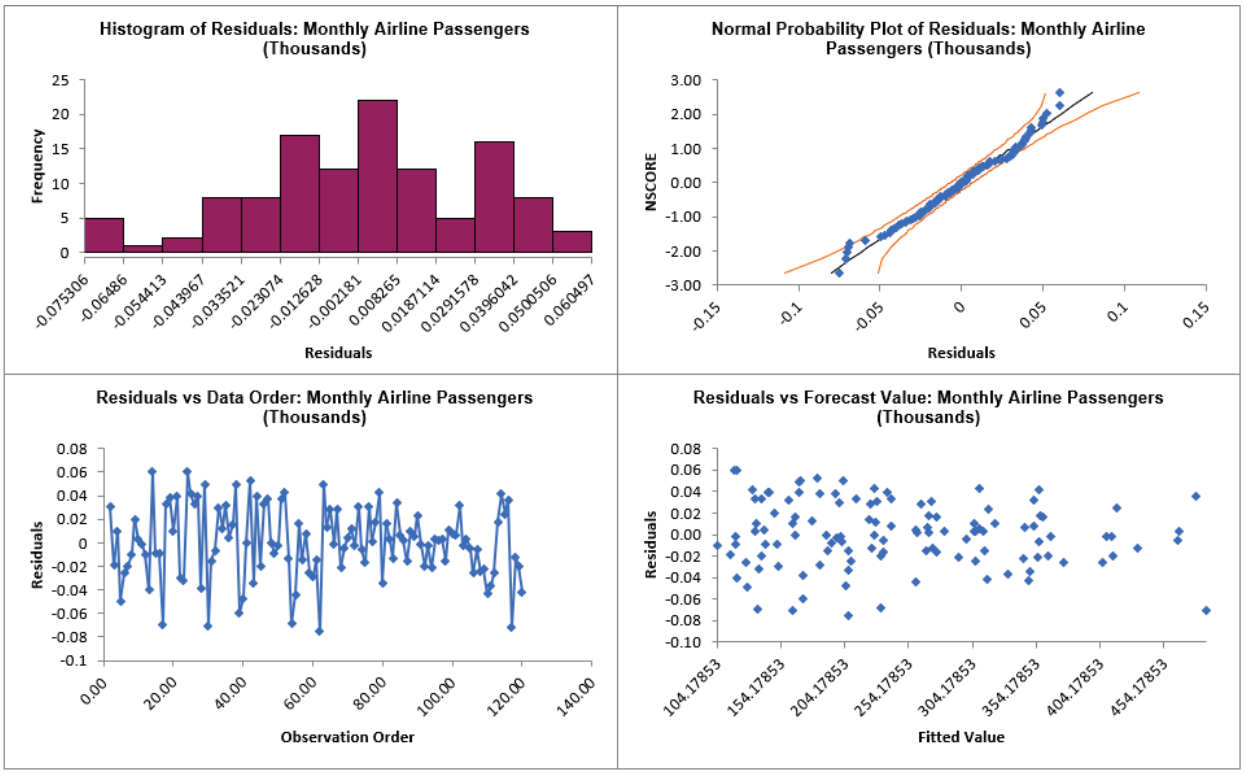
11. Hacer click en **ARIMA MSD ACF PACF LB** para visualizar los Gráficos FAC/FACP/LB:





Los Gráficos de FAC/FACP ARIMA DEM de los Residuos son similares a los Gráficos FAC/FACP/LB e indican que la autocorrelación ha sido considerada en el modelo. Todos los valores-P de Ljung-Box son azules (es decir, no significativos), con Valores-P > 0.05.

12. Hacer click en la hoja **ARIMA MSD Residuals** para visualizar los Gráficos de los Residuos:



Los residuos están distribuidos de forma casi normal, con una línea aproximadamente recta en la gráfica de probabilidad normal. No hay valores extremos atípicos o patrones obvios en las cartas. Las gráficas de residuos son aproximadamente las mismas que en los Gráficos de Residuos.

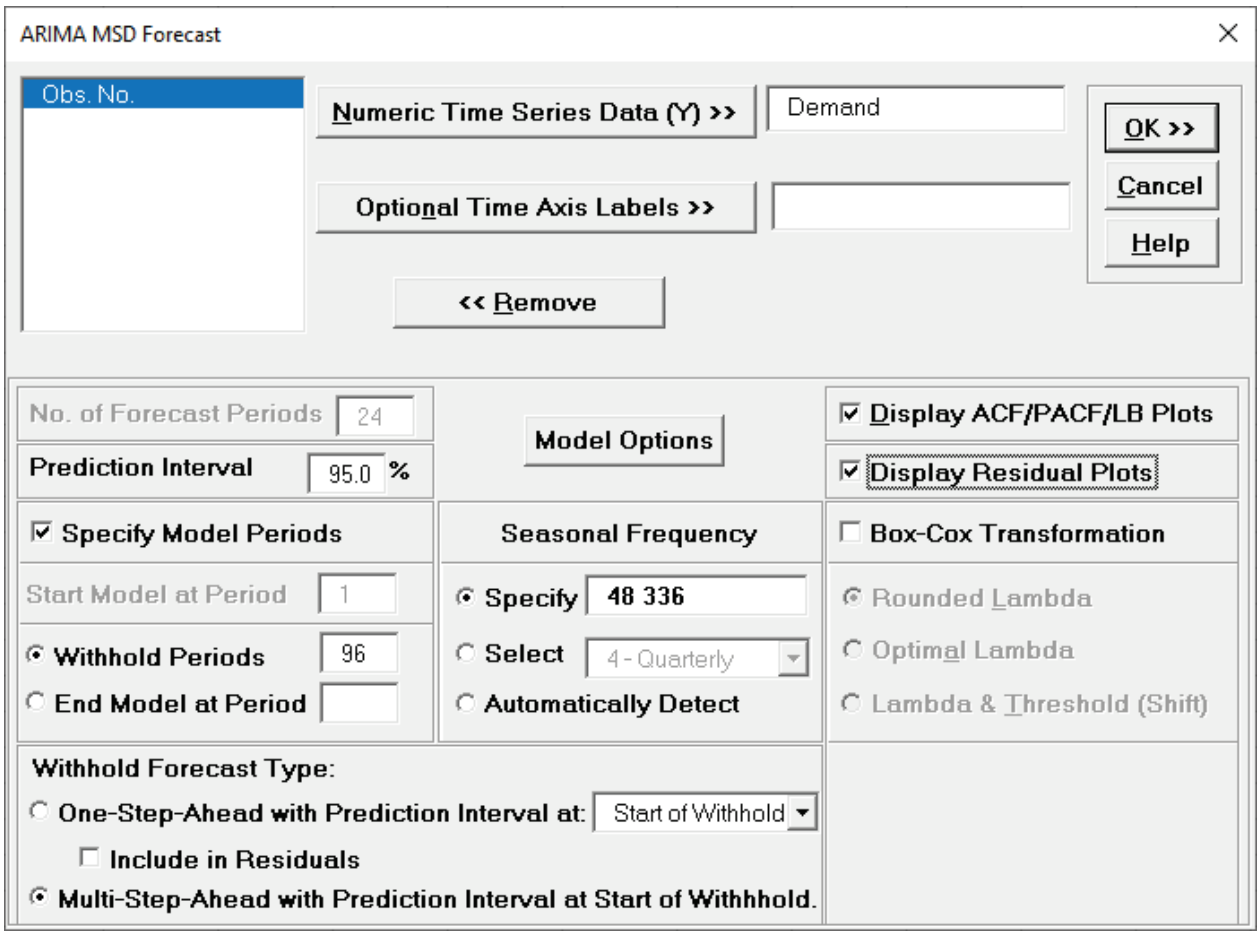
Como se realizó una transformación Box-Cox, los residuos están en unidades transformadas de Ln.

Demanda de Media Hora de Electricidad Estacional Múltiple – Taylor

13. Abrir **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (pestaña de la **Hoja 1**). Esta es la demanda de electricidad de media hora (MW) en Inglaterra y Wales desde el lunes, 5 de junio, 2000 hasta el Domingo, 27 de agosto, 2000 (taylor, R forecast). Estos datos tienen una múltiple estacionalidad con frecuencia = 48 (observaciones diarias) y 336 (observaciones semanales), con un total de 4032 observaciones. Véase para estos datos Cartas de Corrida, Gráficos de FAC/FACP, Gráfico de Densidad Espectral, Gráficos de Descomposición de Tendencia Estacional

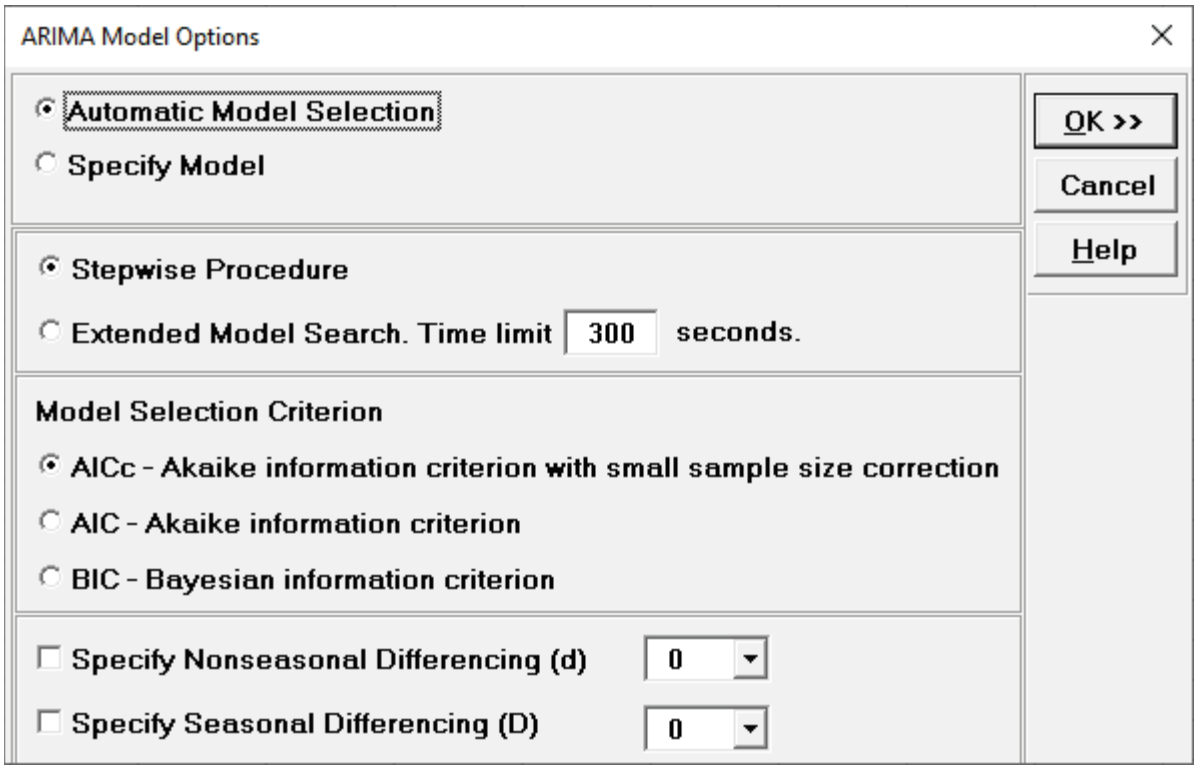
14. Hacer click en **SigmaXL > Time Series Forecasting > ARIMA Forecast > Multiple Seasonal Decomposition Forecast**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

15. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**. Marcar **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Marcar **Specificy Model Periods**. Fijar **Withhold Periods** = 96. Seleccionar **Withhold Forecast Type: Multi-Step-Ahead with Prediction Interval at Start of Withhold.** Marcar **Seasonal Frequency** with **Specify** = 48 336**.** Desmarcar **Box-Cox Transformation**. Se hace uso por defecto **Prediction Interval** = 95.0 %.

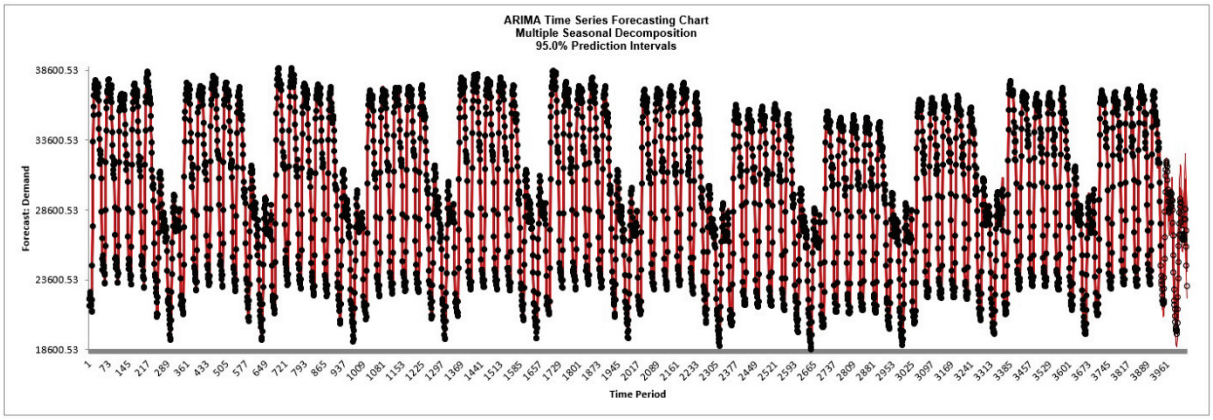


Withhold Periods es 2* frecuencia estacional dominante (48). La frecuencia dominante se obtiene de la gráfica de Densidad Espectral. Inicie el Modelo con un Period = 1 siempre se encuentra sin activar para el DSM.

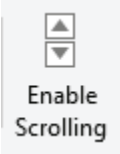
16. Hacer click en **Model Options**.



17. Se empleará por defecto **Automatic Model Selection** con **AICc** como el **Model Selection Criterion**. Hacer click en **OK** para regresar al diálogo del Pronóstico ARIMA (DEM). Hacer click en **OK**. Se genera el informe del pronóstico ARIMA:

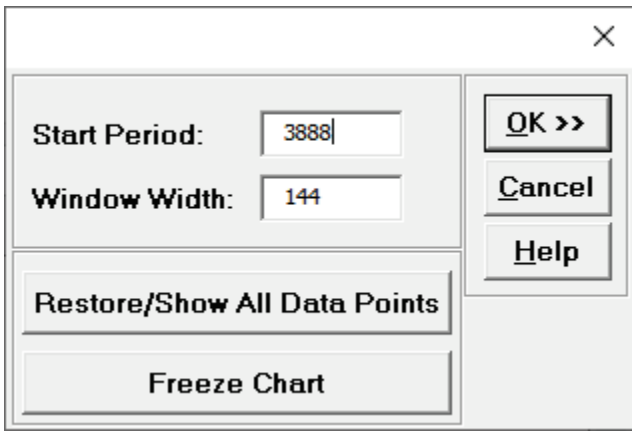


18. Se desea acercarse a los últimos 3 días, es decir, 144 periodos de tiempo de media hora, usando el desplazamiento de las cartas. Hacer click en **SigmaXL Chart Tools > Enable Scrolling**



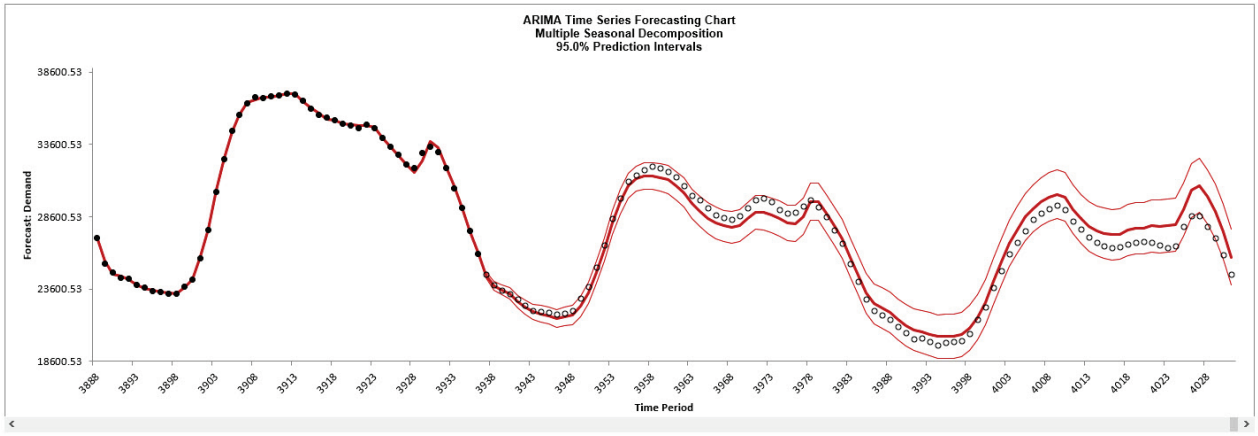
Es probable que se le indique con un mensaje de advertencia que el formato personalizado de la carta se borrará. Puede evitar ver esta advertencia al marcar **grabar esta opción como predeterminada y no mostrarla de nuevo**.

19. Hacer click en **OK.** Aparece el diálogo de desplazamiento lo que permite especificar el **Start Period** and **Window Width**. Registrar **Start Period** = 3888 y **Window Width** = 144:



En cualquier momento, puede hacer click en **Restore/Show All Data Points** o **Freeze Chart**. Al congelar la carta se eliminará el desplazamiento y se descargará el diálogo. El diálogo de desplazamiento también se descargará si cambia las hojas de trabajo. Para reestablecer el diálogo, hacer click en **SigmaXL Chart Tools > Enable Scrolling**.

20. Hacer click en **OK.** Una barra de desplazamiento aparecerá debajo de la carta del pronóstico**.** Se puede cambiar también el **Start Subgroup** y **Window Width** y **Update**.



Se puede desplazar al hacer click a la derecha o a la izquierda, con una anchura de ventana especificada de 144.

Los puntos en blanco son los valores de los datos de la muestra de retención con un pronóstico de varios pasos e intervalos de predicción que se visualizaron al comienzo de la muestra de retención. El modelo funciona bastante bien en la predicción de los 96 valores de la demanda de media hora de retención. Obsérvese que el error de predicción aumenta cuanto más distante es la predicción. La Carta de Pronóstico ARIMA DEM tiene aproximadamente el mismo aspecto que el anterior.

Hacer click en **Cancel** para dejar el diálogo de desplazamiento.

21. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA:

ARIMA Model (Multiple Seasonal Decomposition): Demand - Model Automatically Selected
Model Periods: Model parameter estimates calculated excluding 96 withhold periods.

22. Se genera el Resumen del Informe del Modelo ARIMA:

ARIMA Model Summary	
AR Order (p)	2
I Order (d)	1
MA Order (q)	0
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	48, 336 Decomposition
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

Este es el resumen de la información del modelo para los datos desestacionalizados: ARIMA (2,1,0) sin constante. La Frecuencia Estacional = 48, 336 con el uso del Decomposition y Model Selection Criterion = "AICc". No existen términos estacionales en el modelo.

23. Las Estimaciones de los Parámetros para los datos de la Demanda desestacionalizada son:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
AR_1	0.033677486	0.010694482	3.14905	**0.0017**
AR_2	-0.022798133	0.0160335	1.42191	0.1551

Las Estimaciones de los Parámetros ARIMA incluyen los tests de significancia; Valores-P < 0.05 son significativos y se resaltan con color rojo. Esto puede ser útil para la mejora del modelo con múltiple predictoras pero para la Selección del orden del modelo AR/MA, se debe emplear el mínimo AICc, en lugar de los tests de significancia (Véase Kostenko, A.V. and Hyndman, R.J.).

24. Las Estadísticas del Modelo ARIMA son:

ARIMA Model Statistics	
No. of Observations	3936
DF	3933
StDev	98.8439103
Variance	9770.11861
Log-Likelihood	-23647.086
AICc	47315.0387
AIC	47315.0326
BIC	47333.8656

- El número de observaciones, *n* = 4032 – 96 (*retención*) = 3936
- Los grados de libertad (DF) = 3936 (*n*) – 3 (2 los términos en el modelo, 1 diferencia no estacional) = 3933
- Tener en cuenta que las estadísticas del modelo se calcularon con el uso de datos desestacionalizados.

25. Las métricas de la Precisión del Pronóstico son:

Forecast Accuracy		
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) Multi-Step Ahead Forecast**
N	3935	96
RMSE	101.4595732	775.9894896
MAE	73.7919471	649.1824792
MAPE	0.253330267	2.452663125
MASE	0.040188511	0.353557244

Como era esperarse, los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un Paso-Hacia-Adelante** son mayores que los errores del Pronóstico **Dentro-de-la-Muestra (Estimación) Un Paso-Hacia-Adelante**. Observar que si hubiera existido el interés en el pronóstico a corto plazo de Un Paso-Hacia-Adelante, entonces se hubiera seleccionado el Pronóstico de Tipo Retención Un Paso-Hacia-Adelante y la tabla de arriba mostraría los errores del **Pronóstico Fuera-de-la-Muestra (Retención) Un-Paso-Hacia-Adelante**.

Al comparar las métricas de Precisión del Pronóstico, anterior, se puede apreciar que los resultados del ARIMA DEM son casi los mismos.

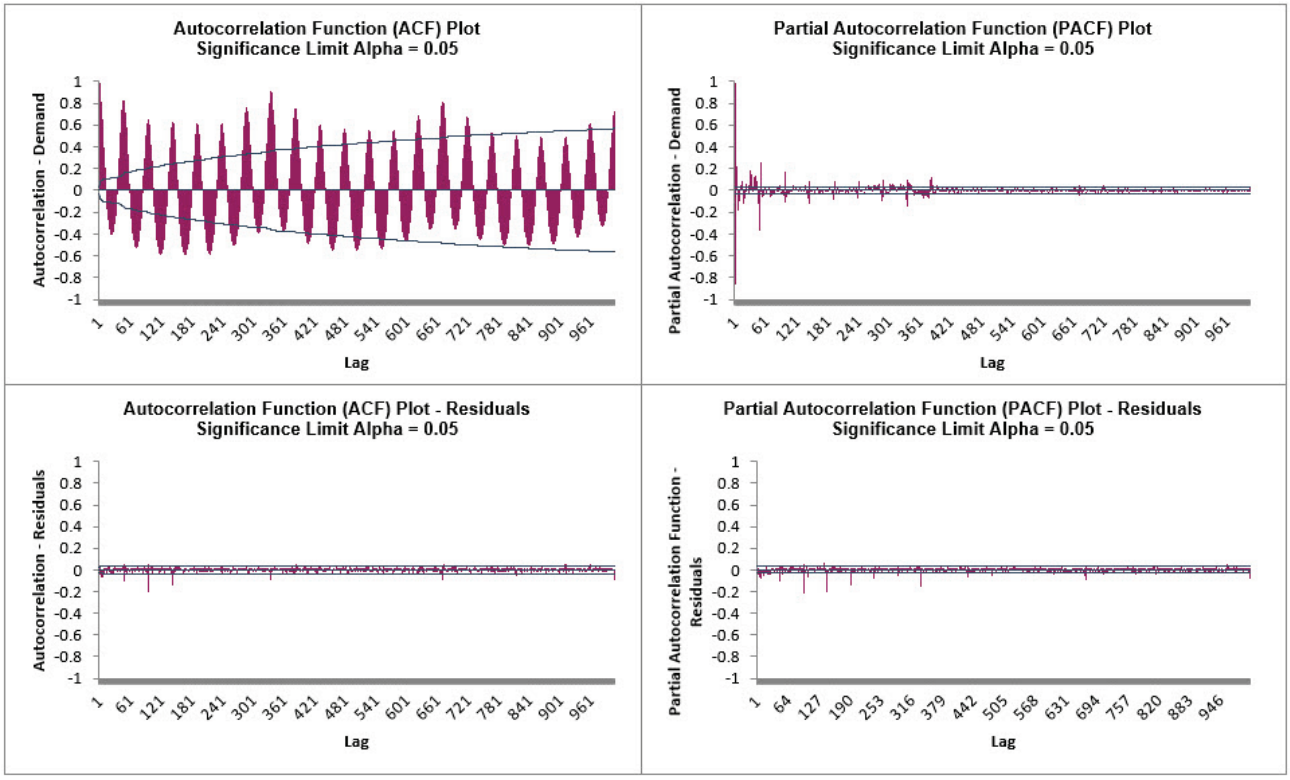
Las métricas de Precisión del Pronóstico se calculan utilizando los datos reales sin procesar versus los del pronóstico que se visualizaron en la Carta de Pronóstico y en la Tabla, y a diferencia de las estadísticas del modelo anterior, permite la comparación entre todos los tipos y transformaciones del modelo de pronóstico.

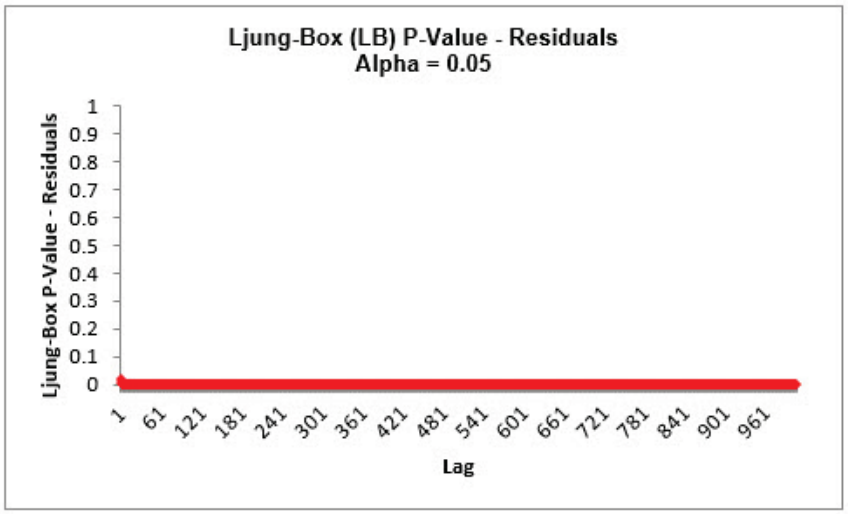
26. Se genera la Tabla de Pronóstico:

Forecast Table				
Period	**Withhold Data**	**Multi-Step Ahead Forecast**	**Lower 95.0% PI**	**Upper 95.0% PI**
3937	24653	24591.46494	24397.67483	24785.25505
3938	23879	23836.7087	23557.99505	24115.42235
3939	23508	23543.90813	23203.10958	23884.70668
3940	23275	23298.25208	22905.19703	23691.30713
3941	22890	22733.71045	22294.53559	23172.88531
3942	22462	22347.61	21866.71441	22828.5056

Se trata de los mismos valores del pronóstico e intervalos de predicción que se visualizan en la Carta de Pronósticos, pero que se proporcionan para un análisis o un gráfico más detallado (por ejemplo: Se genera una Carta de Corrida de los errores del pronóstico). Se visualizan también los **Datos de Retención**.

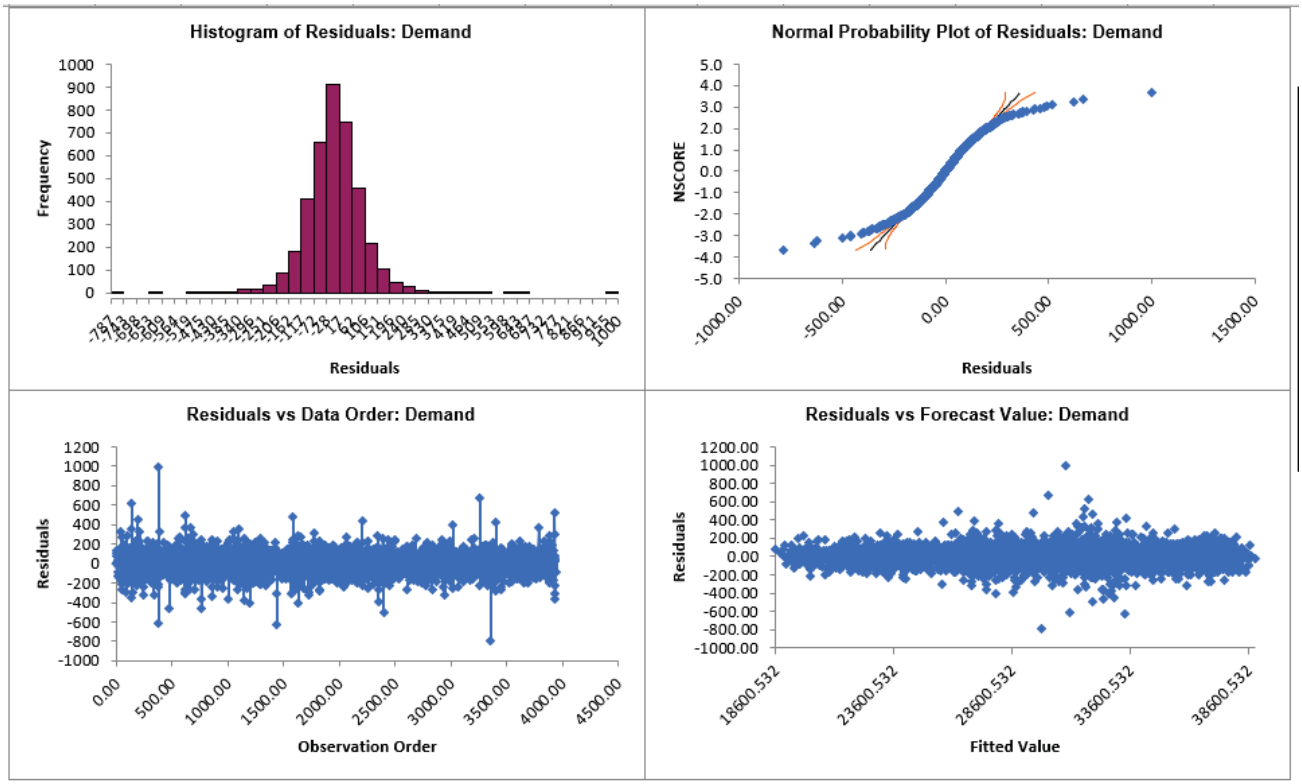
27. Hacer click en la hoja **ARIMA MSD ACF PACF LB** para visualizar los Gráficos FAC/FACP/LB:





Los Gráficos de FAC/FACP de los Residuos indican que gran parte de la autocorrelación ha sido considerada por el modelo, pero el gráfico de Ljung-Box muestra que todavía queda alguna autocorrelación significativa (los valores-P rojos son significativos con un alfa=0,05) - por lo que el modelo puede ser potencialmente mejorado. Esto no significa que el modelo sea un mal modelo, todavía puede ser muy útil para fines de predicción, pero los intervalos de predicción pueden no proporcionar una cobertura precisa.

28. Hacer click en la hoja **ARIMA MSD Residuals** para visualizar los Gráficos de los Residuos:



Al igual que con los Gráficos de los Residuos, los residuos no se distribuyen normalmente y hay valores extremos atípicos. Estos deben ser investigados con una carta de control de residuos. Los valores atípicos en la demanda de electricidad se explican a menudo por la Temperatura.

Cartas de Control ARIMA

Control Estadístico de Proceso (CEP) para Datos Autocorrelacionados

Se crea una carta de control de individuales utilizando los residuos del modelo de pronóstico ARIMA.

La carta de Límites Móviles utiliza la predicción de un paso como la línea central, por lo que los límites de control se moverán con la línea central. Si se utiliza una transformación Box-Cox, se aplica una transformación inversa para calcular los límites de control.

Las populares funciones "Añadir datos", "Mostrar los últimos 30" y "Desplazamiento" de las herramientas de las Cartas en SigmaXL están disponibles para estas cartas de control. Los modelos de series de tiempo no se reajustan, sino que se utilizan para calcular los valores de los residuos de los nuevos datos.

Para mayores detalles y referencias, véase el Apéndice: **Cartas de Control para Datos Autocorrelacionados**.

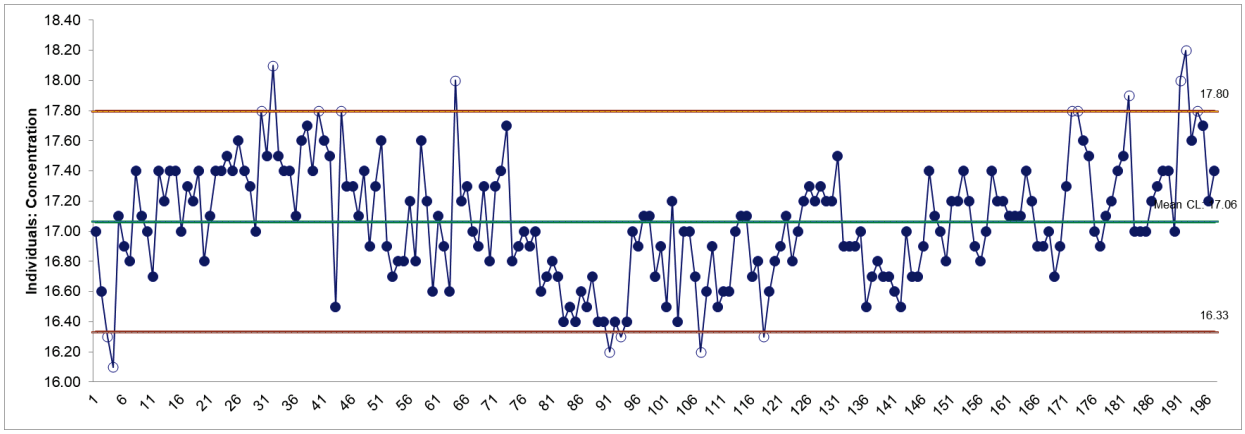
Tener en cuenta que una Carta de Rangos Móviles y las Pruebas de Causas Especiales no están disponibles aquí, pero el usuario puede almacenar y seleccionar los residuos, y entonces crearlos con **SigmaXL > Control Charts > Individuals & Moving Range**.

Concentración del Proceso Químico - Serie A

1. Abrir **Chemical Process Concentration – Series A.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie A de Box y Jenkins, un conjunto de 197 valores de la concentración de un proceso químico tomados a intervalos de 2 horas. Véase Cartas de Corrida, Gráficos FAC/FACP, Gráfico de Densidad Espectral, y Gráficos de Descomposición de Tendencia Estacional para estos datos .

2. Hacer click en **SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

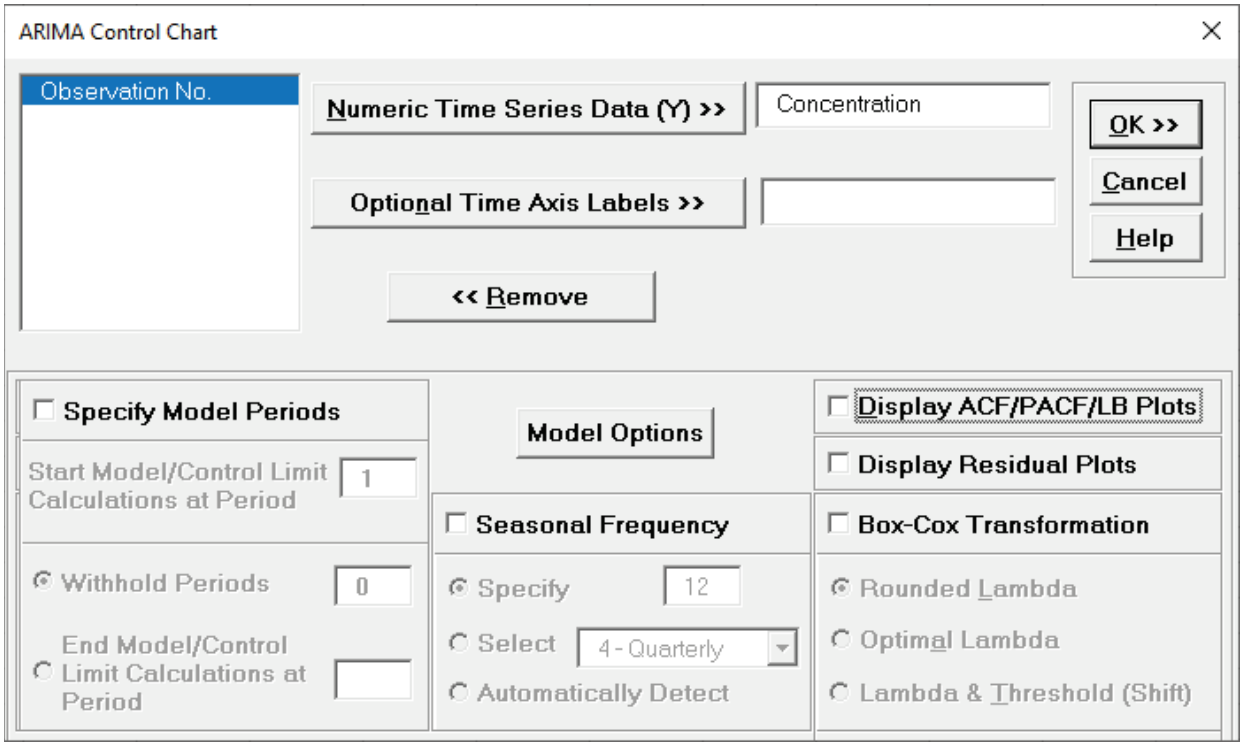
3. Seleccionar *Concentration*, hacer click en **Numeric Data Variable (Y) >>**. Hacer click en **OK**. Se genera una Carta de Control de Individuales:



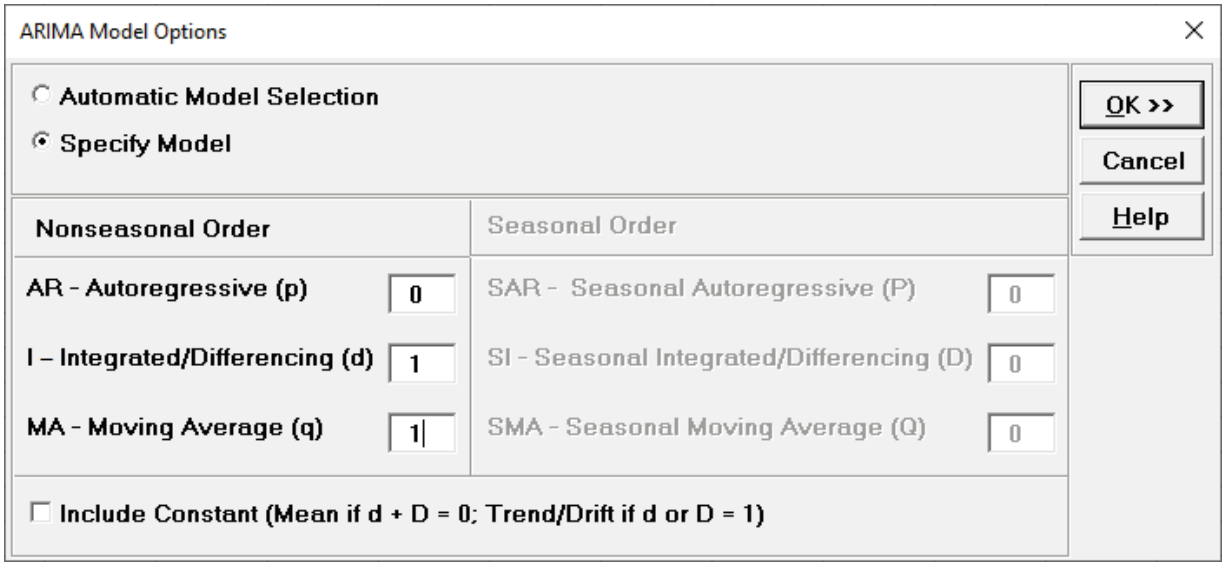
Existen 17 puntos de datos fuera de control, en gran parte debido a la autocorrelación. Buscar las causas asignables usando este gráfico tal como está, sería inútil.

4. Hacer click ahora en la pestaña de la **Hoja 1** y **SigmaXL > Time Series Forecasting > ARIMA Control Chart > Control Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

5. Seleccionar *Concentration*, hacer click en **Numeric Time Series Data (Y) >>**. Desmarcar **Display ACF/PACF/LB Plots.** Desmarcar **Display Residual Plots**, **Specify Model Periods, Seasonal Frequency** y **Box-Cox Transformation**.

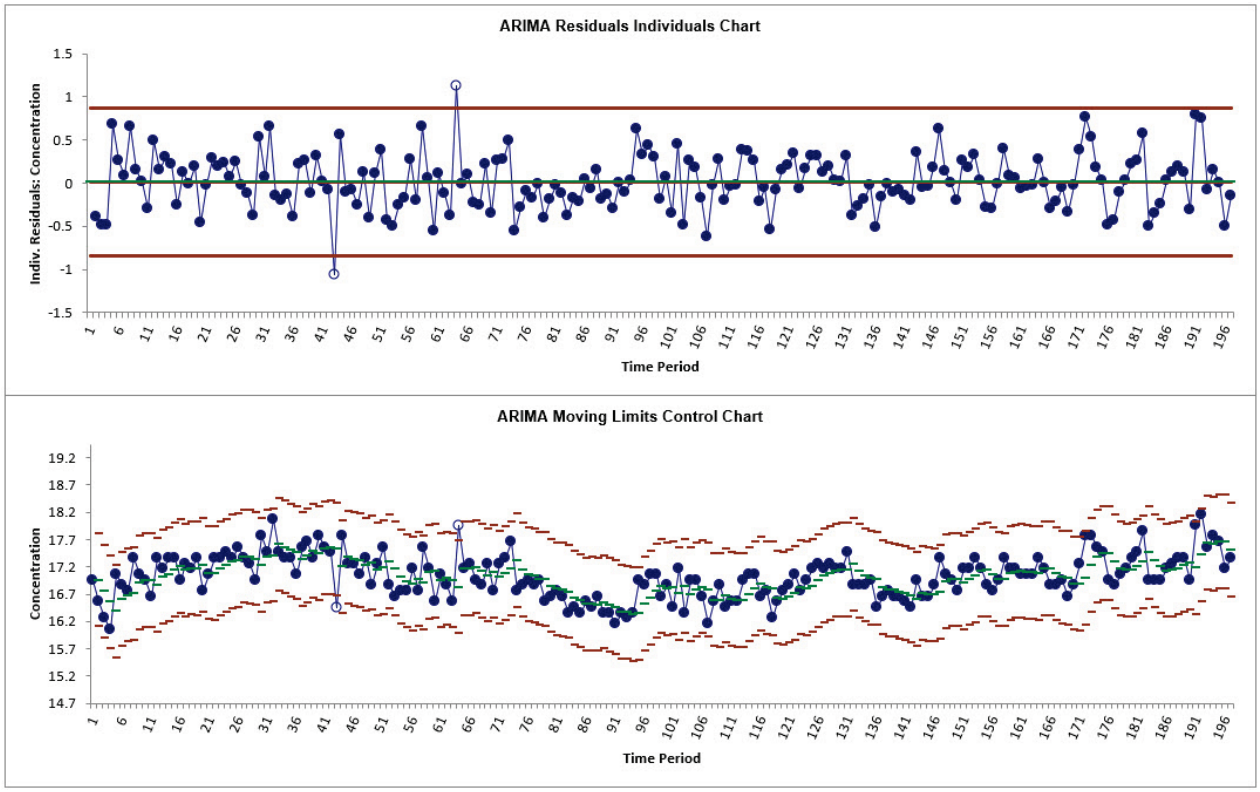


6. Hacer click en **Model Options**. Seleccionar **Specify Model**. Especificar **Nonseasonal Order I – Integrated/Differencing (d)** = 1 and **MA – Moving Average (q)** = 1. Desmarcar **Include Constant**.



En general, se usaría **Automatic Model Selection**, pero se desea demostrar la carta de control ARIMA utilizando el modelo ARIMA (0,1,1) como se usó anteriormente, y compararlo con las cartas de control de suavización exponencial equivalentes.

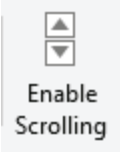
7. Hacer click en **OK** para retornar al diálogo de la Carta de Control ARIMA. Hacer click en **OK**. Se genera la carta de control ARIMA:



Ahora sólo se dispone de puntos de datos fuera de control en la tabla de Individuales para investigar. La Carta de Límites Móviles utiliza la predicción de un paso como la línea central, por lo que los límites de control se mueven con la línea central.

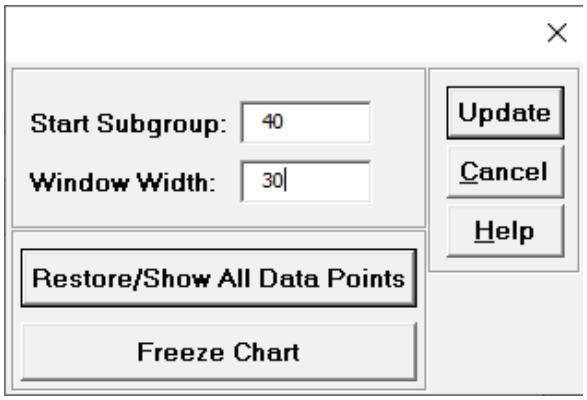
Como era de esperarse, estas son aproximadamente las mismas que las Cartas de Control de Suavización Exponencial, con ligeras diferencias en los valores iniciales. La Carta de Individuales de los Residuos de ARIMA no muestra un punto de datos en el período de tiempo = 1 debido a las diferencias.

8. Se desplazará a través de los datos de puntos de la Carta como se lo hace en las Cartas de Control de Suavización Exponencial. Hacer click en **SigmaXL Chart Tools > Enable Scrolling**

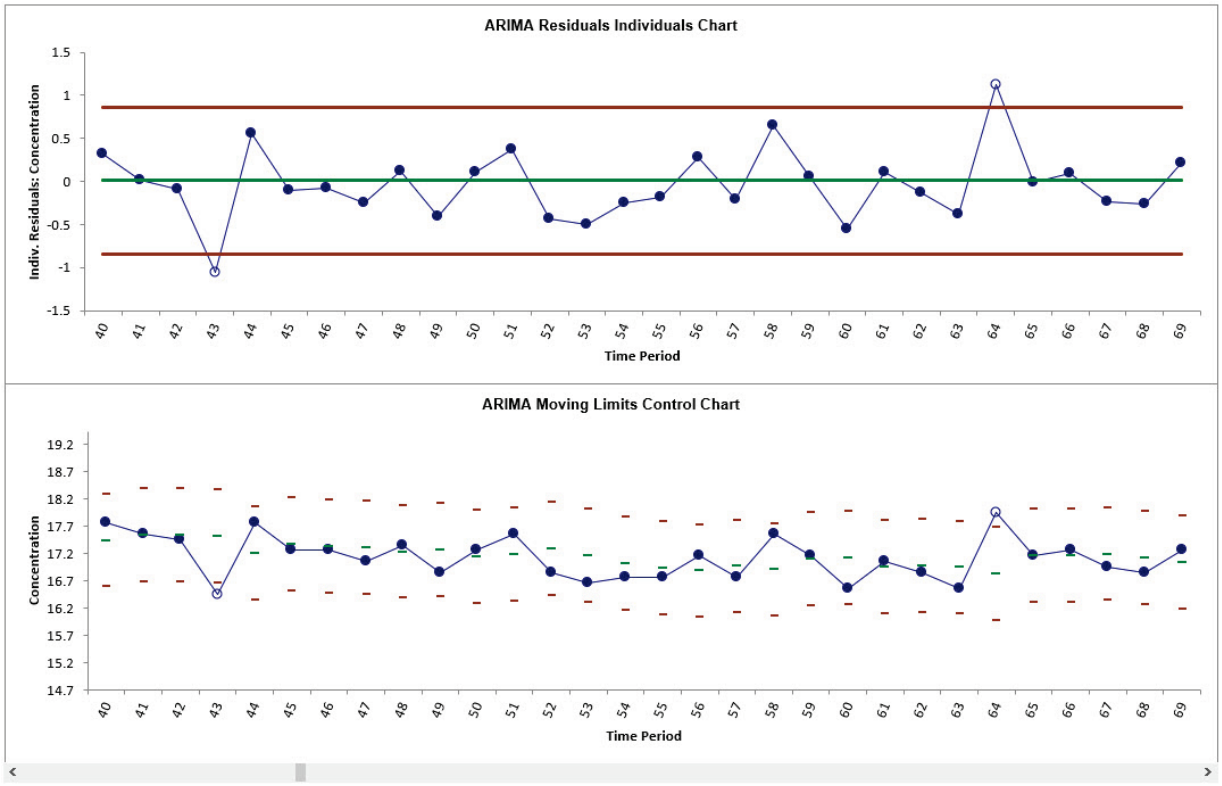


Es probable que se le indique con un mensaje de advertencia que el formato personalizado de la carta se borrará. Puede evitar ver esta advertencia al marcar **grabar esta opción como predeterminada y no volver a mostrar esta advertencia**.

9. Hacer click en **OK.** Aparece el diálogo de desplazamiento lo que permite especificar el **Start Subgroup** y **Window Width**. Registrar **Start Subgroup** = 40 y **Window Width** = 30 para visualizar los dos puntos de los datos fuera-de-control.



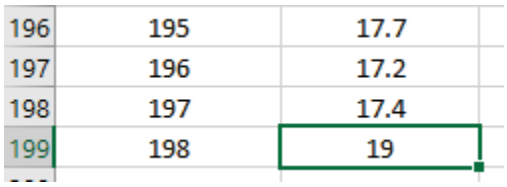
10. Hacer click en **OK.** Esto permite amplificar la imagen de los puntos fuera-de-control 43 y 64.



La observación **43** es menor de lo que se esperaba del modelo de pronóstico ARIMA. La observación **64** es mayor de lo que se esperaba.

11. Hacer click en **Cancel** para salir del diálogo de desplazamiento.

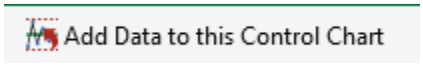
12. Se añadirá ahora un nuevo punto de datos a los datos de la Concentración de la serie A. Se calcularán los residuos utilizando el mismo modelo que el anterior sin reestimar los parámetros del modelo o sin volver a calcular los límites de control. Esto se lo conoce también como la aplicación de la "Fase II" de una Carta de Control, en la que una señal de fuera de control debe conducir a una investigación de la causa asignable y a la aplicación de una acción correctiva o un ajuste del proceso. Hacer click en la Hoja1, registrar el valor 19 como se muestra en la celda **B199** (y opcionalmente el número de observación 198 en la celda **A199**).



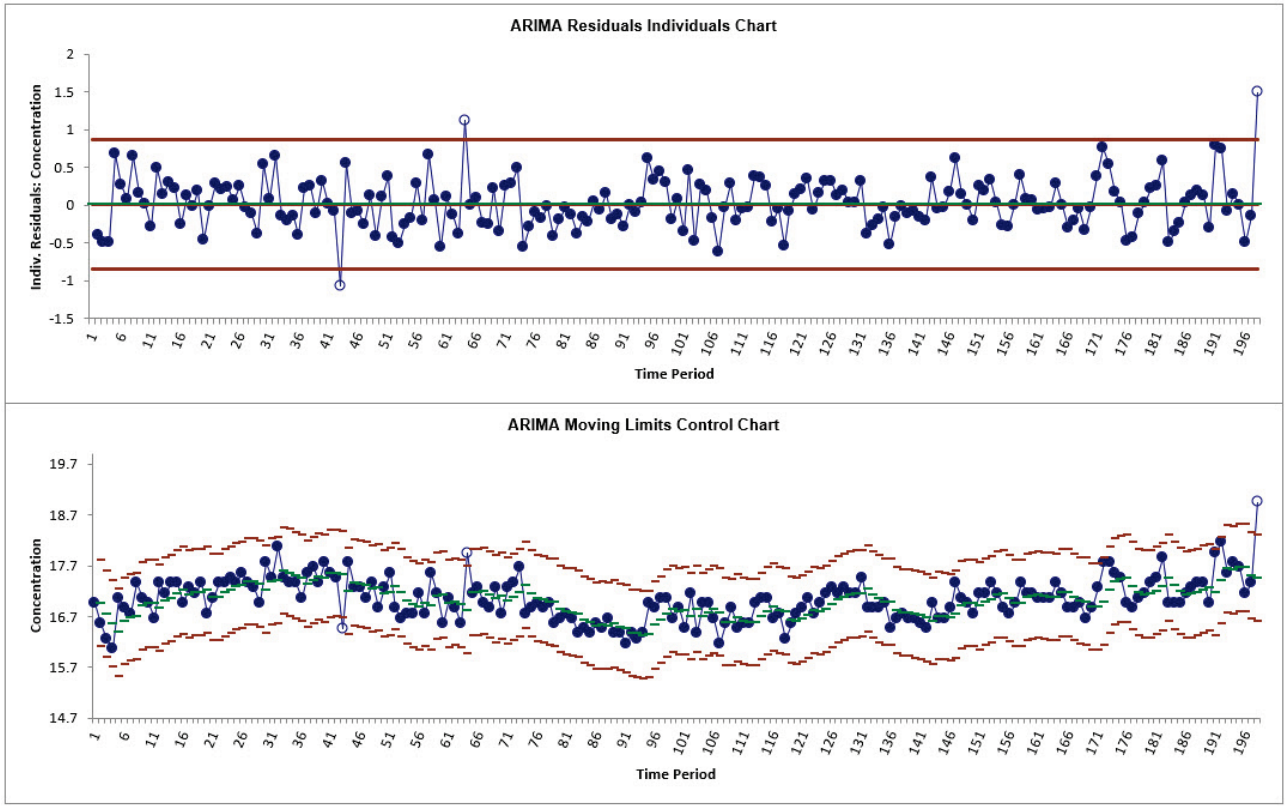
196	195	17.7
197	196	17.2
198	197	17.4
199	198	19

13. Hacer click en la pestaña **ARIMA Control Charts** (de existir más de una carta de control en la hoja de trabajo, por favor seleccione la carta a donde se agregue el dato).

14. Hacer click en **SigmaXL Chart Tools > Add Data to this Control Chart**



15. Se actualizan ahora con nuevos datos la Carta de Control de Individuales de los Residuos y las Cartas de Límites Móviles, y se muestra esto como un punto del dato fuera-de-control:



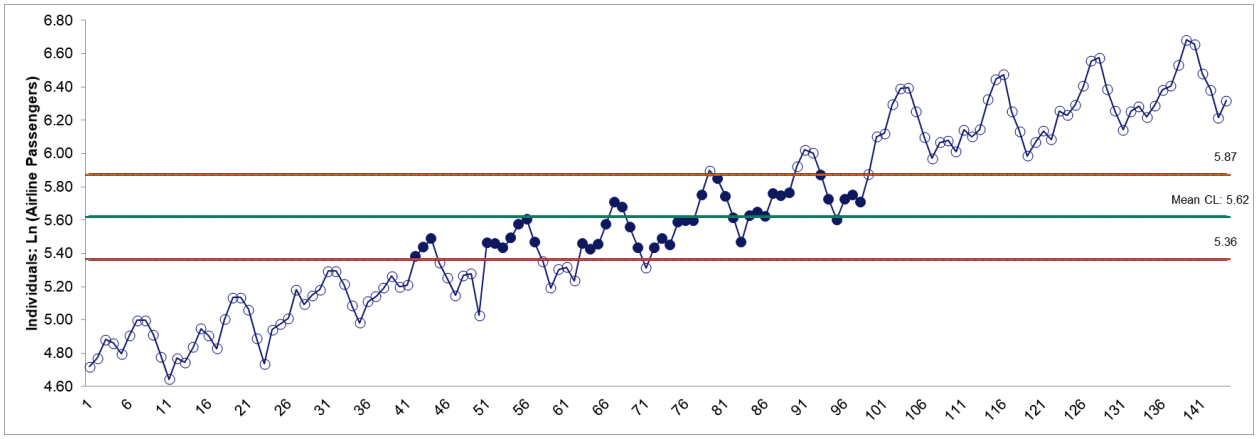
Se recomienda renombrar el libro de trabajo a **Chemical Process Concentration – Series A_AddData2.xlsx**, de tal manera que en el uso posterior los datos de la Concentración no incluyan el punto del dato añadido.

Pasajeros Mensuales de la Aerolínea - Modificado para Cartas de Control

16. Abrir **Pasajeros Mensuales de la Aerolínea - Modificado para Cartas de Control.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional desde enero 1949 hasta diciembre 1960. Se aplica una transformación Ln (evitando el uso de una transformación Box-Cox), se agrega un valor atípico en el dato 50 (-0.25) y se aplica un cambio de nivel (+0.25), con el inicio en 100. Se agregaron variables codificadas para ayudar a distinguir un valor atípico en comparación con un cambio de nivel y se analizarán posteriormente con el uso del Pronóstico ARIMA con Predictoras.

17. Anteriormente se apreció que este proceso tiene una autocorrelación significativa con una fuerte tendencia y estacionalidad. Para ver el impacto en una carta de control, se construirá una carta de individuales con los datos sin procesar. Hacer click en **SigmaXL > Control Charts > Individuals**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

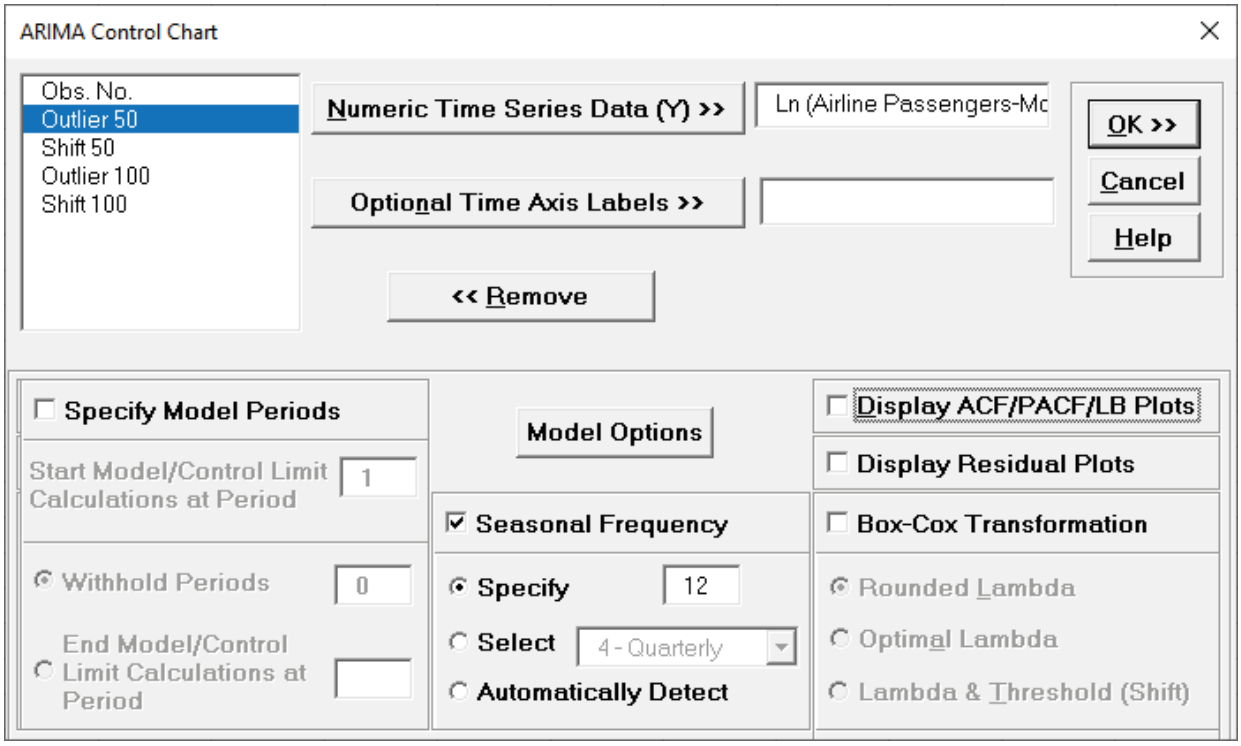
18. Seleccionar *Ln (Airline Passengers)*, Hacer click en **Numeric Data Variable (Y) >>**. Hacer click en **OK**. Se genera una Carta de Control de Individuales:



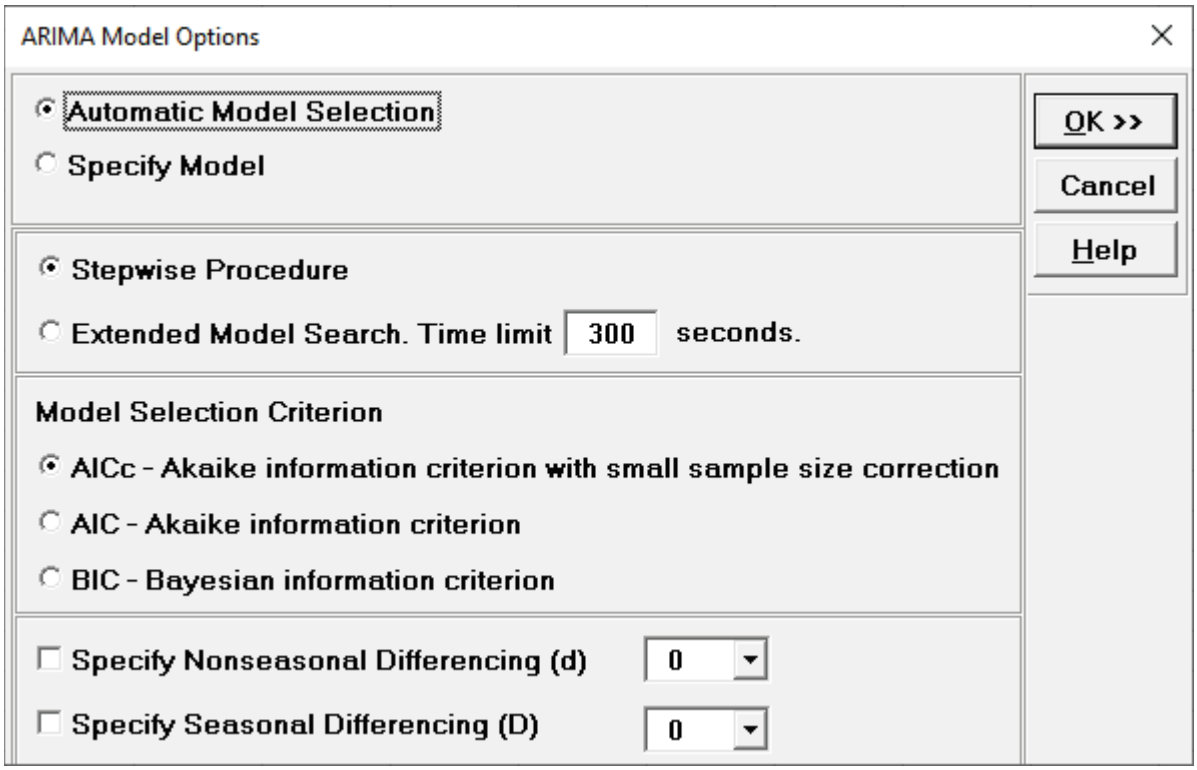
Con una fuerte tendencia, estacionalidad y autocorrelación positiva, esta carta de control carece de sentido.

19. Hacer click ahora en la pestaña de la **Hoja 1** y en **SigmaXL > Time Series Forecasting > ARIMA Control Chart > Control Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

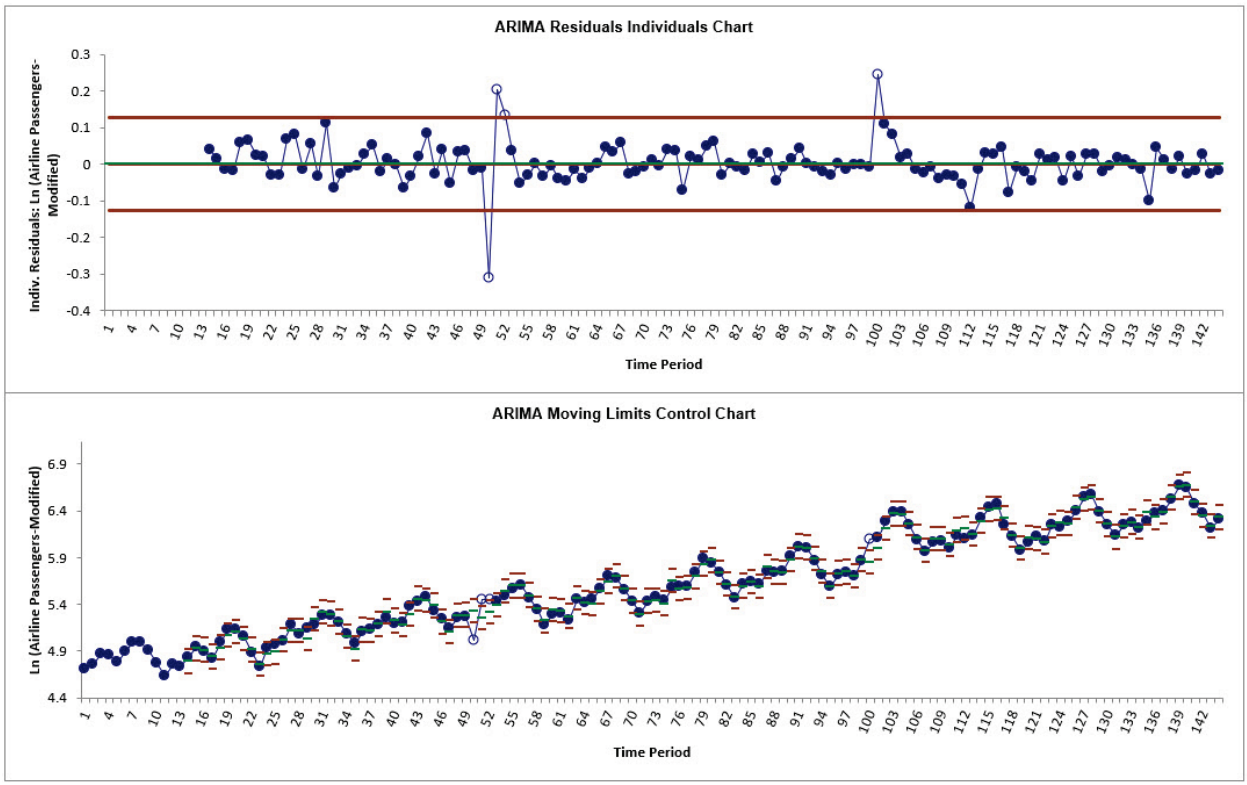
20. Seleccionar *Ln (Airline Passengers-Modified)*, hacer click en **Numeric Time Series Data (Y) >>**. Desmarcar **Display ACP/PACF/LB Plots** y **Display Residual Plots.** Marcar **Seasonal Frequency** con **Specify** = 12**.** Desmarcar **Specify Model Periods** y **Box-Cox Transformation**.



21. Hacer click en **Model Options**.

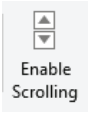


22. Se empleará por defecto **Automatic Model Selection** con **AICc** como el **Model Selection Criterion**. Hacer click en **OK** para retornar al diálogo de la Carta de Control ARIMA. Hacer click en **OK**. Se generan las cartas de control ARIMA:



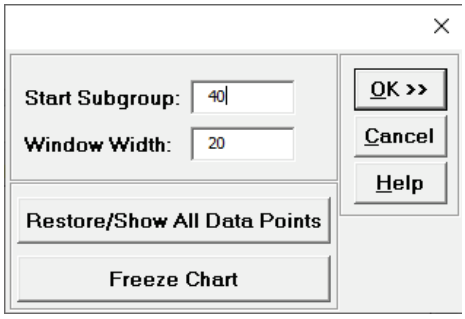
Se puede apreciar ahora en forma clara los puntos de datos fuera de control en 50, 51, 52 y 100 de la carta de individuales. Estos son casi los mismos que aquellos proporcionados en las Cartas de Control de Suavización Exponencial (el 52 estaba bajo-control). La Carta de Control de Individuales de los Residuos de ARIMA no muestra puntos de datos en los Períodos de Tiempo 1 al 13 debido a la diferencia no estacional y estacional.

23. Con la finalidad de visualizar los puntos de la carta de Límites Móviles se empleará el desplazamiento. Hacer click en **SigmaXL Chart Tools > Enable Scrolling**

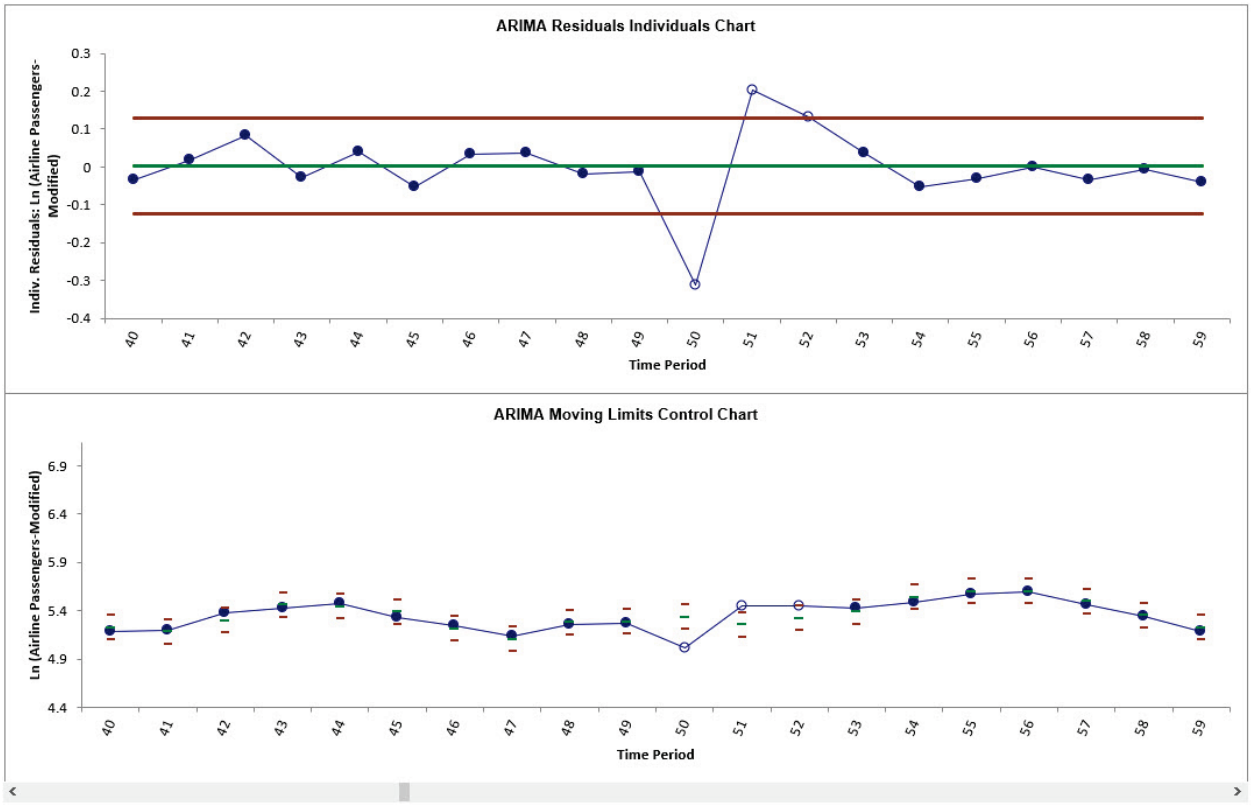


Puede que se le indique con un mensaje de advertencia que el formato personalizado de la carta se borrará. Puede evitar ver esta advertencia al marcar **Save this choice as default and do not show this form again**.

24. Hacer click en **OK.** Se aparece el diálogo del desplazamiento lo que permite especificar el **Start Subgroup** y **Window Width.** Registrar **Start Subgroup**=40 y **Window Width**=20 para visualizar los primeros dos puntos de datos fuera de control.



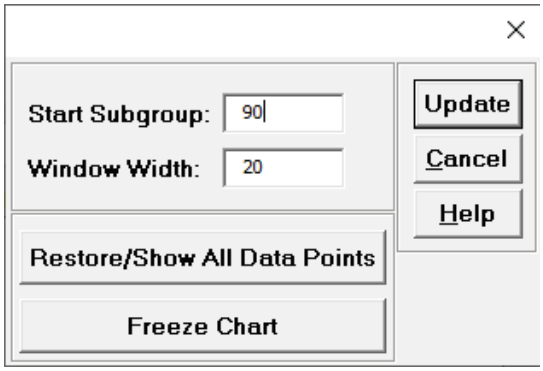
25. Hacer click en **OK.** Esto permite ampliar la vista sobre los puntos 50, 51, y 52 fuera de control.



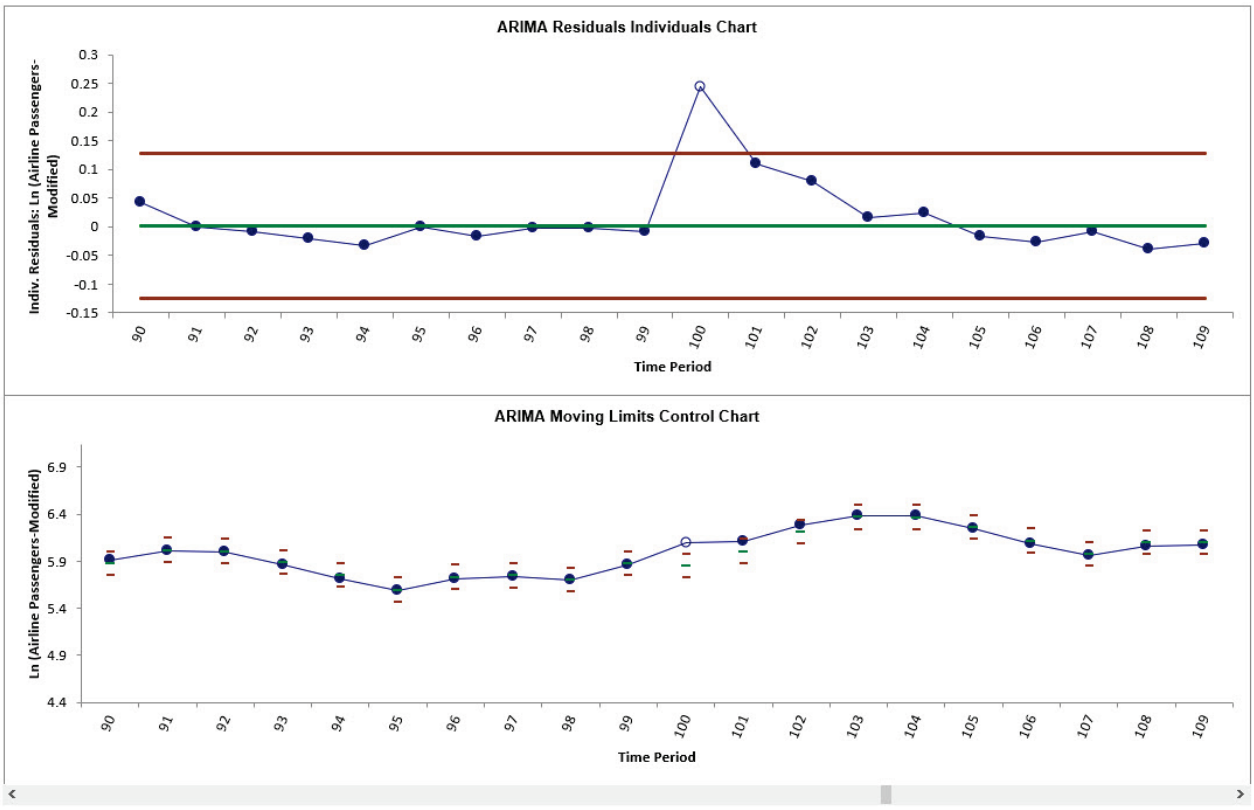
La observación **50** es menor que lo esperado por el modelo de pronóstico ARIMA. Las observaciones **51** y **52** son mayores que lo esperado.

Consejo: El desplazamiento mantiene el ajuste mínimo y máximo del eje Y original. Puede cambiar esto a automático haciendo click en el eje Y, haciendo click con el botón derecho en **Format Axis**, haciendo click en **Bounds Minimum Reset** y **Bounds Maximum Reset.** Esto cambia la configuración del eje a Auto para que cuando se desplace o actualice el eje Y se ajuste automáticamente también.

26. Registrar ahora **Start Subgroup** = 90 y **Window Width** = 20 para visualizar el tercer punto de los datos fuera de control.



27. Hacer click en **Update**.



La observación **100** es mayor que la esperada del modelo de pronóstico ARIM.

28. Hacer click en **Cancel** para salir del diálogo de desplazamiento**.**

29. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA:

ARIMA Model: Ln (Airline Passengers-Modified) - Model Automatically Selected
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

30. El Resumen del Modelo ARIMA es:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	1
SMA Order (Q)	1
Seasonal Frequency	12
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

El modelo ARIMA (0,1,1) (0,1,1) (sin constante) fue automáticamente seleccionado en base al criterio AICc. Esto es lo mismo que se obtuvo en el resumen del modelo ARIMA al utilizar una muestra de retención y concuerda con los resultados mostrados en los datos de Pasajeros Mensuales de la Aerolínea - Serie G a los que se aplicó diferencias en forma manual y corresponde al modelo empleado en Box y Jenkins, Capítulo 9., "Análisis de Series Temporales Estacionales".

No existen predictoras. La Frecuencia Estacional = 12; el Criterio de Selección del Modelo = "AICc" y la Transformación Box-Cox = "N/A".

31. Las Estimaciones de los Parámetros y las Estadísticas del Modelo ARIMA son ligeramente diferentes del análisis de Estimaciones de Parámetros porque se ha introducido un valor atípico y un cambio de nivel, también aquí se está utilizando todos los datos, es decir, no hay períodos de retención. Nótese que previamente se empleó una Transformación Box-Cox con Lambda=0 y aquí se emplea el Ln de los datos.

32. Las métricas de la Precisión del Pronóstico son:

Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	131	
RMSE	0.055615558	
MAE	0.036285476	
MAPE	0.648152199	
MASE	0.249152676	

Tener en cuenta que estos errores de previsión son muy diferentes del análisis de las métricas de Precisión del Pronóstico, donde los errores del pronóstico se calcularon sobre los datos sin procesar versus los valores finales pronosticados, pero aquí se ha utilizado el Ln de los datos de Pasajeros de la Aerolínea.

33. Se volverá ahora a analizar los datos para determinar si las señales fuera de control se deben a un valor atípico o a un cambio de nivel del proceso. Hacer click en la pestaña de la **Hoja 1**.

Obs. No.	Ln (Airline Passengers-Modified)	Outlier 50	Shift 50	Outlier 100	Shift 100
49	5.278114659	0	0	0	0
50	5.028114659	1	1	0	0
51	5.463831805	0	1	0	0
52	5.459585514	0	1	0	0
53	5.433722004	0	1	0	0
54	5.493061443	0	1	0	0
55	5.575949103	0	1	0	0
99	5.874930731	0	1	0	0
100	6.10220248	0	1	1	1
101	6.122117789	0	1	0	1
102	6.295005314	0	1	0	1
103	6.392037406	0	1	0	1
104	6.396329258	0	1	0	1
105	6.251414878	0	1	0	1

La primera señal fuera de control se produjo en la Obs. No = 50, así que se crearon dos variables codificadas. *Outlier50* está codificada de manera que la Obs. No. 50 = 1 y todos los demás valores son 0. *Shift50* está codificada para que la Obs. No. 1 a la 49 = 0, y la Obs. No. 50 a la 144 sean 1.

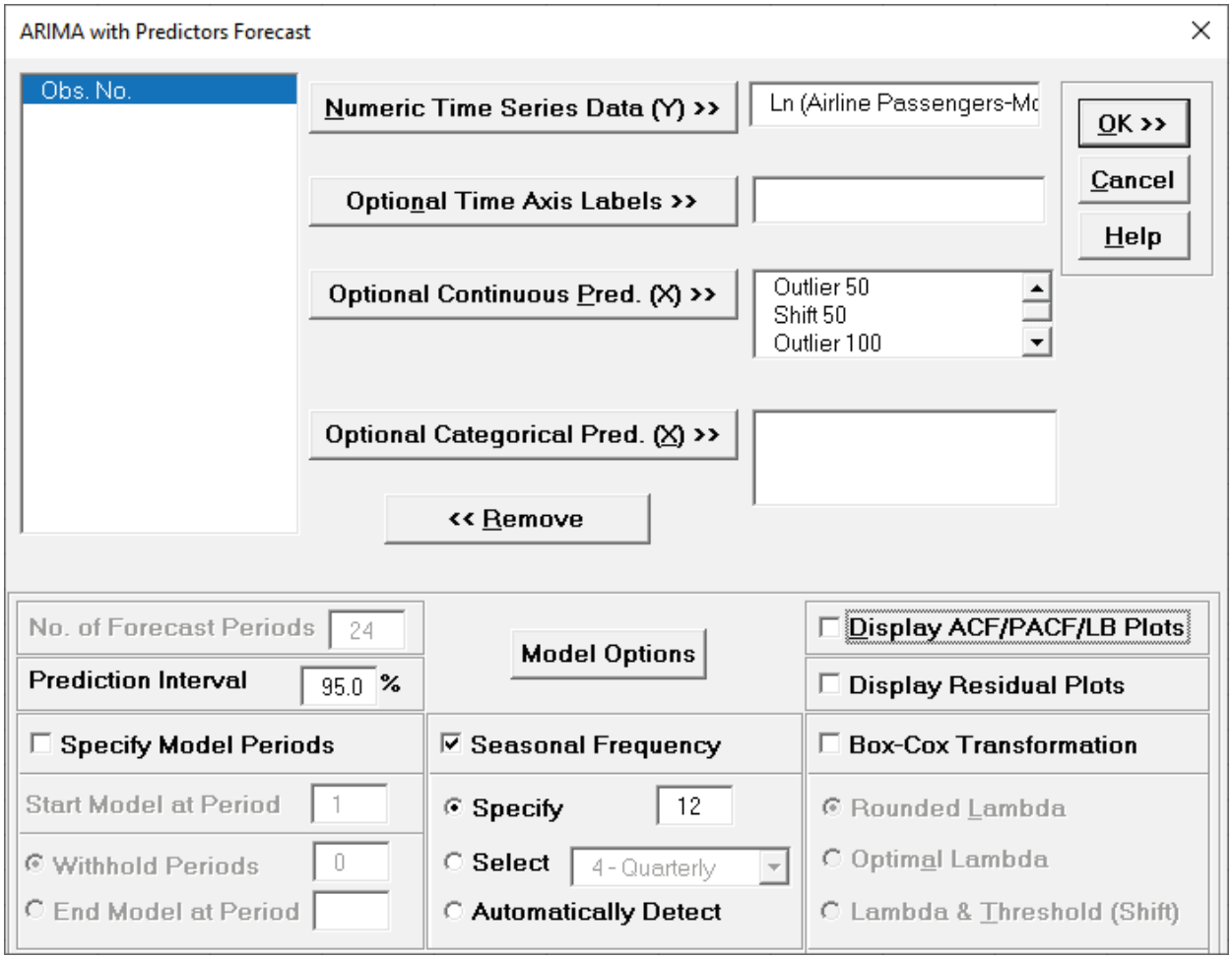
Para simplificar el análisis, se asumirá que las señales 51 y 52 fuera de control están relacionadas con la 50, por lo que se las ignorará. La siguiente señal fuera de control ocurre en la 100, así que de manera similar se crearon otras dos variables codificadas. *Outlier100* está

codificada de manera que la Obs. No. 100 = 1 y todos los demás valores son 0. *Shift100* está codificado de manera que la Obs. No. 1 a la 99 = 0, y la Obs. No. 100 a 144 sean 1.

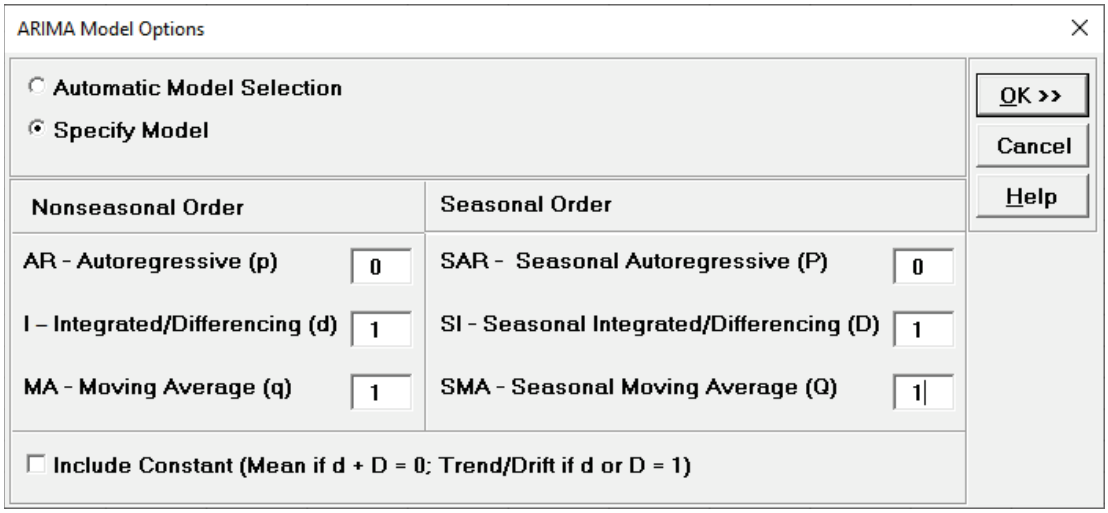
Dado que están codificadas como 0, 1 se las considerará como Continua.

34. Hacer click en **SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast with Predictors**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

35. Seleccionar *Ln (Airline Passengers – Modified)*, hacer click en **Numeric Time Series Data (Y) >>** Seleccionar *Outlier50 to Shift100*, hacer click en **Optional Continuous Pred. >>** Desmarcar **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Marcar **Seasonal Frequency** con **Specify** = 12. Desmarcar **Specify Model Periods** y **Box-Cox Transformation**. Se hace uso por defecto de **Prediction Interval** = 95.0 %.



36. Hacer click en **Model Options**. Con la adición de las predictoras codificadas, se pretende asegurarse que el modelo ARIMA es el mismo que se utiliza en la carta de control. Seleccionar **Specify Model**. Especificar **Nonseasonal Order I – Integrated/Differencing (d)** = 1 y **MA – Moving Average (q)** = 1. Especificar **Seasonal Order SI – Seasonal Integrated/Differencing (D)** = 1 y **SMA – Seasonal Moving Average (Q)** = 1. Desmarcar **Include Constant**.



37. Hacer click en **OK** para retornar al diálogo del Pronóstico ARIMA. Hacer click en **OK**. Se genera el informe del pronóstico ARIMA.

38. Desplazarse hacia abajo para visualizar las Estimaciones de los Parámetros:

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.395854258	0.086277876	4.588132	**0.0000**
SMA_1	0.556926317	0.073521579	7.575005	**0.0000**
Outlier 50	-0.296267312	0.032292885	9.174384	**0.0000**
Shift 50	0.044130277	0.035505653	1.242908	0.2162
Outlier 100	-0.000597974	0.031822469	0.018791	0.9850
Shift 100	0.249993132	0.035384072	7.065132	**0.0000**

Se puede apreciar que *Outlier50* y *Shift100* son significativas definiendo a la Obs. No. 50 como un valor atípico y a la 100 como un cambio de nivel. Esto es, por supuesto, lo que se esperaba ya que así es como se modificaron con Ln los datos de los Pasajeros de la Aerolínea.

Este método para la identificación de valores atípicos versus cambios de nivel sirve como complemento en el proceso de conocimiento y en la búsqueda de causas asignables que se utilizan en el clásico CEP.

Cartas de Control ARIMA con Predictoras

Se genera una carta de control de individuales utilizando los residuos del modelo de pronóstico ARIMA con Predictoras.

El modelo ARIMA con Predictoras permite las predictoras continuas o categóricas. Véase para mayor información **Pronóstico ARIMA con Predictoras**.

La carta de Límites Móviles utiliza la predicción de un paso como la línea central, por lo que los límites de control se moverán con la línea central. Si se hace uso de una transformación Box-Cox, se aplica una transformación inversa para calcular los límites de control.

Las populares funciones "Mostrar los últimos 30" y "Desplazamiento" de las herramientas de las Cartas en SigmaXL están disponibles para estas cartas de control. Actualmente, la función "Añadir Datos" no es compatible en la Carta de Control ARIMA con Predictoras.

Para mayores detalles y referencias, véase el Apéndice: **Cartas de Control para Datos Autocorrelacionados**.

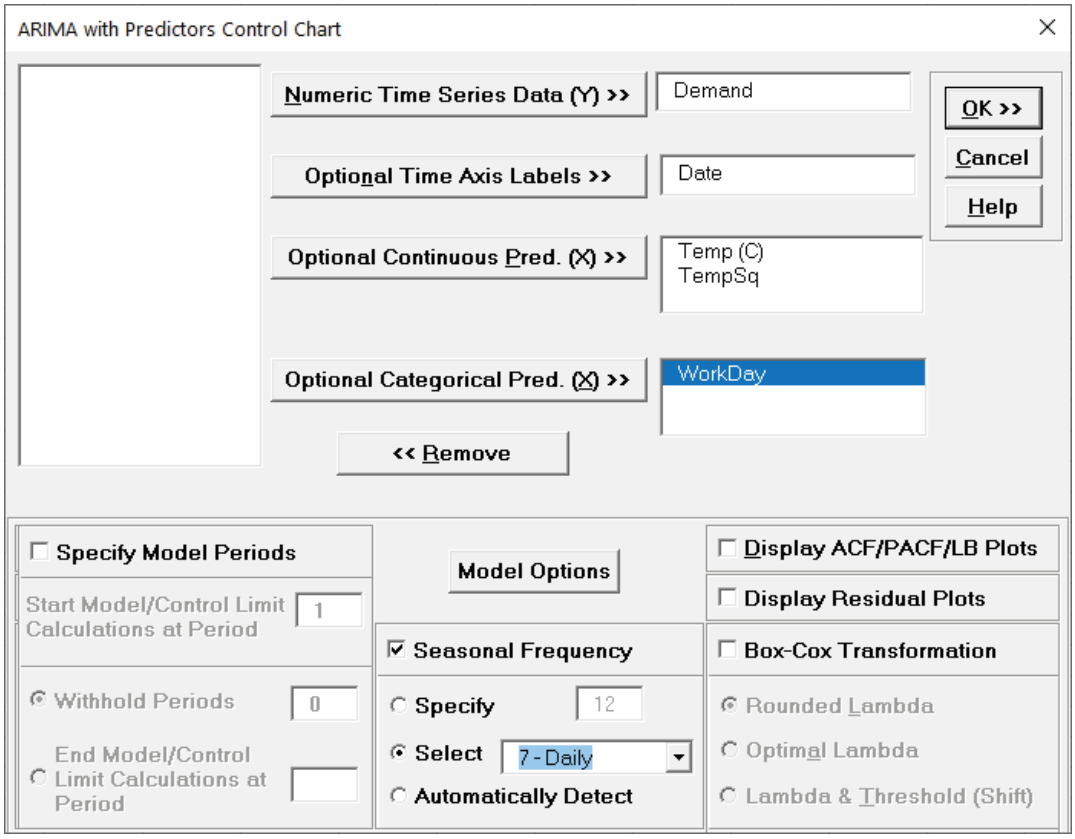
Tener en cuenta que una Carta de Control de Rango Móvil y la Prueba para Causas Especiales no están disponibles aquí, pero el usuario puede almacenar y Seleccionar Residuos, y luego crear con **SigmaXL > Control Charts > Individuals & Moving Range**.

Demanda Diaria de Electricidad con Predictoras – ElecDaily

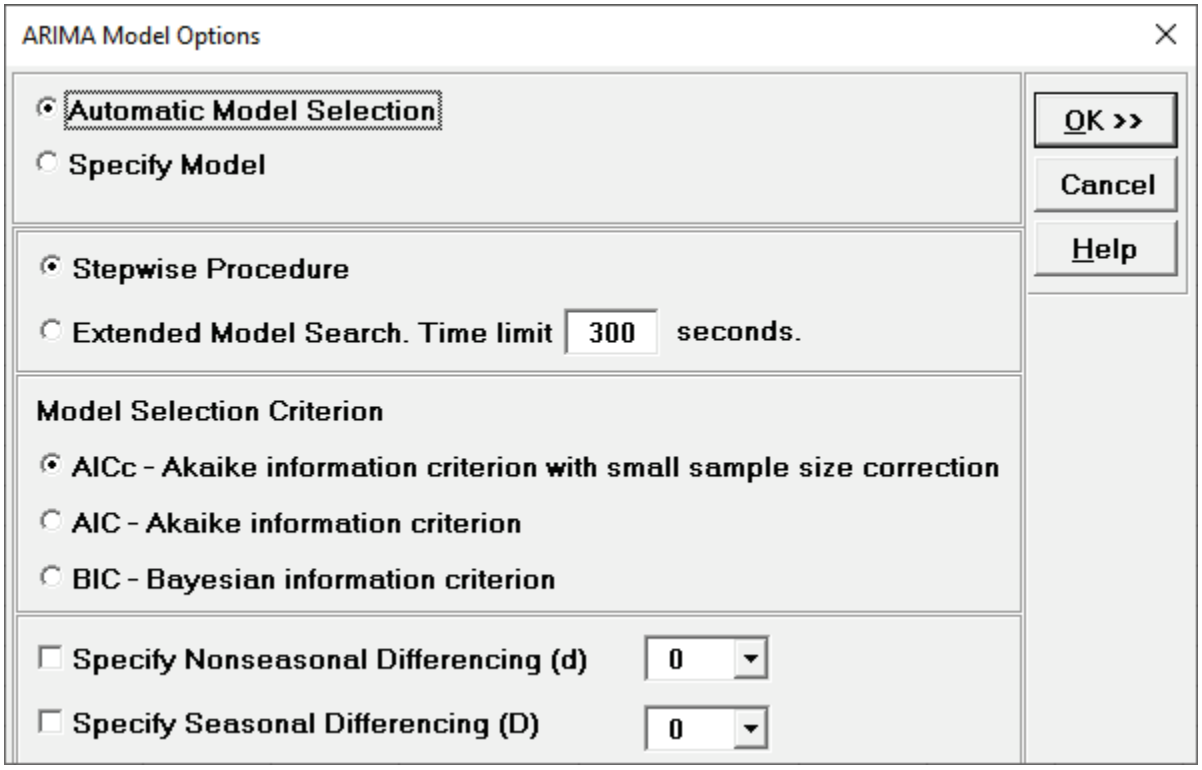
1. Abrir **Daily Electricity Demand with Predictors – ElecDaily.xlsx** (pestaña **Hoja 1**). Estos datos son la demanda diaria de electricidad (GW) para el estado de Victoria, Australia, cada día durante el año 2014 (véase Hyndman, fpp2, Sección 9.3, **Ejemplo: Pronóstico de la demanda de electricidad**, https://otexts.com/fpp2/forecasting.html). La Temp (C) es la máxima temperatura diaria en grados centígrados para la ciudad de Melbourne. TempSq es Temperatura al cuadrado. WorkDay asume el valor 1 para días laborables, de lo contrario es 0. Estos datos tienen una frecuencia estacional = 7 (observaciones por semana). Véase para estos datos Cartas de Corrida, Gráficos FAC/FACP, Gráfico de Densidad Espectral y Gráficos de Descomposición de Tendencia Estacional.

2. Hacer click en **SigmaXL > Time Series Forecasting > ARIMA Control Chart > Control Chart with Predictors**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

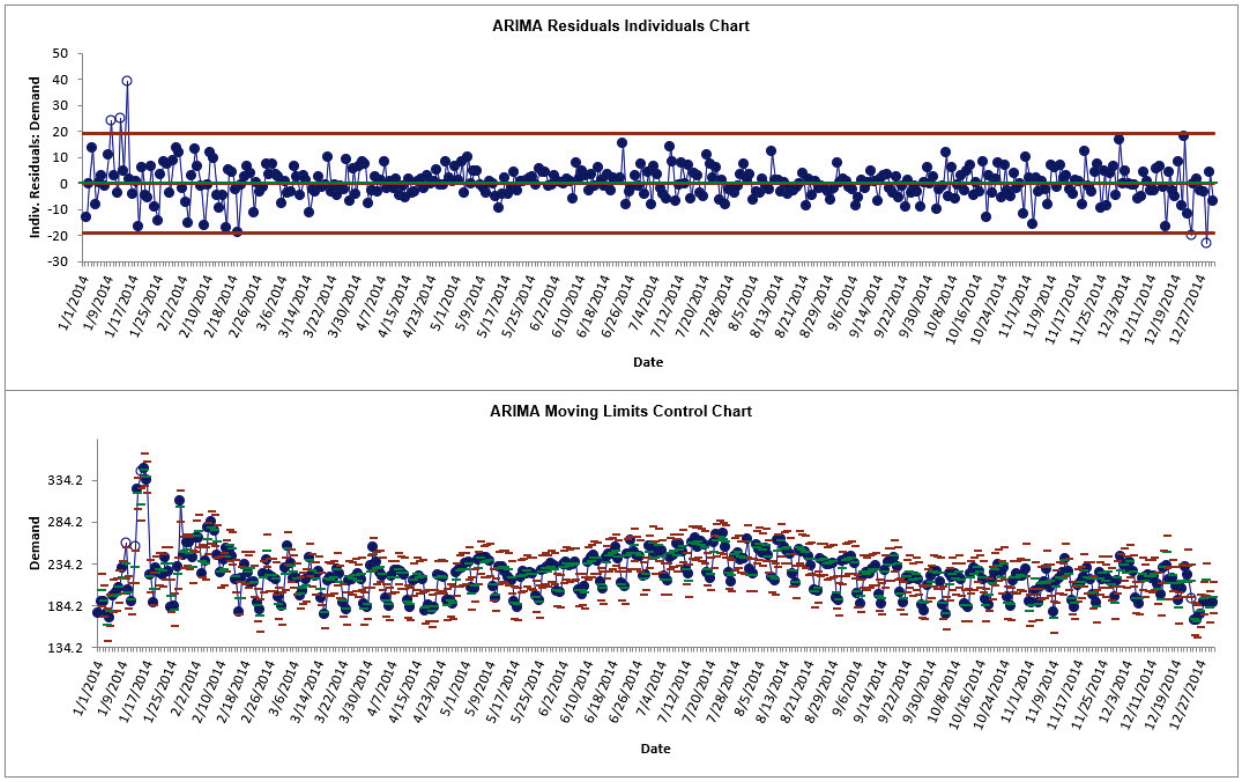
3. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**; Seleccionar *Date*, hacer click en **Optional X-Axis Labels >>**; Seleccionar *Temp (C) y TempSq*, hacer click en **Optional Continuous Pred. >>**; Seleccionar *WorkDay*, hacer click en **Optional Categorical Pred. >>**. Desmarcar **Display ACF/PACF/LB Plots** y **Display Residual Plots**. Marcar **Seasonal Frequency** with **Select** = 7 - Daily (or **Specify** = 7). Desmarcar **Specify Model Periods y Box-Cox Transformation**.



4. Hacer click en **Model Options**. Seleccionar **Automatic Model Selection**. Se hace uso por defecto: **Stepwise Procedure** y **Model Selection Criterion: AICc – Akaike information criterion with small sample size correction**, desmarcar **Specify Nonseasonal Differencing (d)** y **Specify Seasonal Differencing (D).**



5. Hacer click en **OK** para retornar al diálogo ARIMA with Predictors Control Chart. Hacer click en **OK**. Este es un modelo complejo, de tal manera que el tiempo de cómputo tomará aproximadamente uno a dos minutos. Se generan las cartas de control ARIMA con Predictoras:



Se puede apreciar señales fuera de control a mediados de enero con una demanda mayor que la predicha por el modelo y a finales de diciembre con una demanda menor que la predicha por el modelo.

6. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA:

ARIMA Model: Demand - Model Automatically Selected
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

7. El Resumen del Modelo ARIMA es:

ARIMA Model Summary	
AR Order (p)	2
I Order (d)	1
MA Order (q)	2
SAR Order (P)	2
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	7
Include Constant	0
No. of Predictors	3
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

Esto es un resumen de la información del modelo: ARIMA (2,1,2) (2,0,0) sin constante y 3 predictoras. Frecuencia Estacional = 7; Criterio de Selección del Modelo = “AICc” y Transformación Box-Cox = “N/A” debido a que la Box-Cox Transformation estaba Desmarcada.

Este es el mismo modelo utilizado en el Resumen del Modelo ARIMA en el Pronóstico ARIMA DEM, y concuerda con el modelo que aparece en Hyndman, fpp2, Sección 9.3, **Ejemplo: Pronóstico de la demanda de electricidad**, https://otexts.com/fpp2/forecasting.html.

ARIMA Carta de Control Descomposición Estacional Múltiple (DEM)

ARIMA no tiene un límite de frecuencia teórico, pero para la eficiencia computacional y para minimizar la pérdida potencial de observaciones por diferenciación, recomendamos usar ARIMA - Descomposición Estacional Múltiple (DME) para la frecuencia estacional mayor de 52 (o con frecuencias múltiples). El componente estacional se elimina primero a través de la descomposición, un modelo ARIMA no estacional se ajusta al resto (tendencia +), y luego se vuelve a añadir el componente estacional.

Como su nombre indica, la Descomposición Estacional Múltiple (DSM) también da cabida a la estacionalidad múltiple, por ejemplo, los datos de media hora con una frecuencia estacional de 48 observaciones por día y 336 observaciones por semana.

Para este método de pronóstico se crea un gráfico de control individual de los residuos. La carta de Límites Móviles utiliza la predicción de un paso como la línea central. Si se utiliza una transformación Box-Cox, entonces se aplica una transformación inversa para calcular los límites de control.

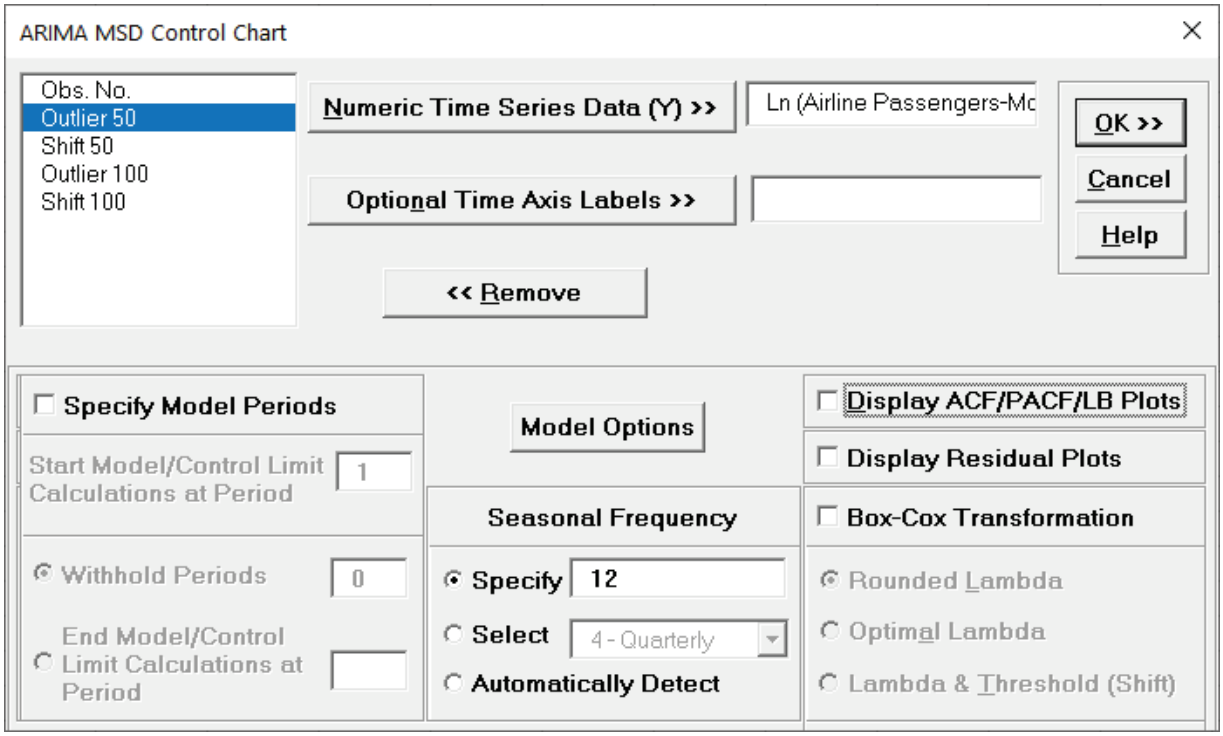
Las populares funciones "Añadir datos", "Mostrar los últimos 30" y "Desplazamiento" de las Herramientas de Cartas de SigmaXL están disponibles para estas cartas de control. Para "Add Data", los modelos de series de tiempo no se reajustan, sino que se utilizan para calcular los valores de los residuos de los nuevos datos.

Pasajeros Mensuales de la Aerolínea - Modificado para Cartas de Control

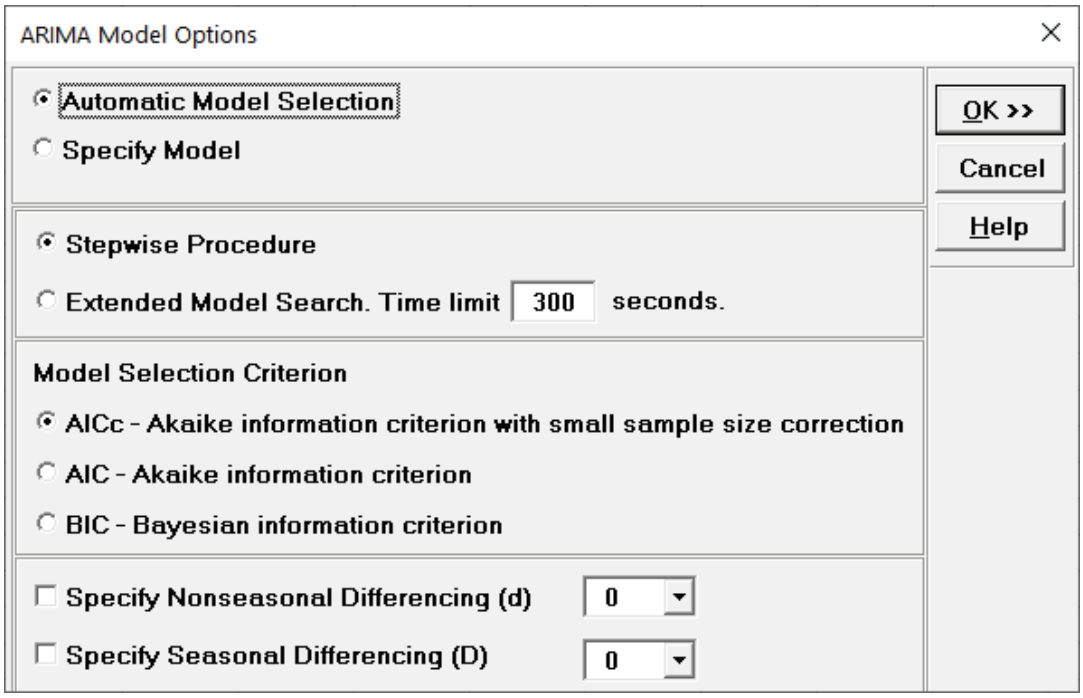
1. Abrir **Monthly Airline Passengers – Modified for Control Charts.xlsx** (pestaña **Hoja 1**). Estos datos son una Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional desde enero 1949 hasta diciembre 1960. Se aplica una transformación Ln (evitando el uso de una transformación Box-Cox), se agrega un valor atípico en el dato 50 (-0.25) y se aplica un cambio de nivel (+0.25), con el inicio en 100. La opción Descomposición Estacional Múltiple (DEM) no es necesaria para estos datos, pero como un ejemplo de introducción, se empleará esto para comparar el análisis previo.

2. Hacer click en **SigmaXL > Time Series Forecasting > ARIMA Control Chart > Multiple Seasonal Decomposition Control Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

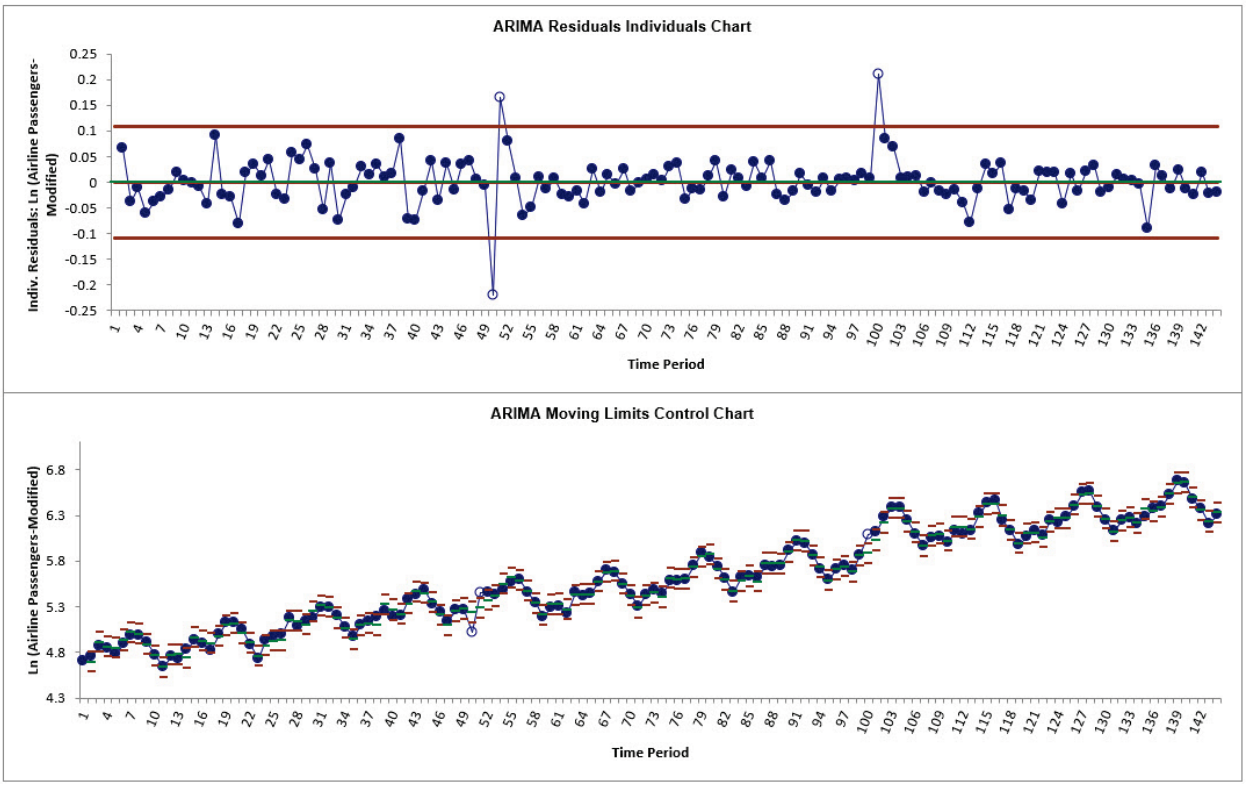
3. Seleccionar *Ln (Airline Passengers-Modified)*, hacer click en **Numeric Time Series Data (Y) >>**. Desmarcar **Display ACF/PACF/LB Plots** y **Display Residual Plots.** Marcar **Seasonal Frequency** with **Specify** = 12**.** Desmarcar **Specify Model Periods** y **Box-Cox Transformation**.



4. Hacer click en **Model Options**.



5. Se empleará por defecto **Automatic Model Selection** con **AICc** como el **Model Selection Criterion**. Hacer click en **OK** para retornar al diálogo ARIMA Control Chart. Hacer click en **OK**. Se generan las cartas de control ARIMA (DEM):



Se puede apreciar de forma clara los puntos de datos fuera de control en 50, 51 y 100 en carta de Individuales de los Residuos. Esto es similar a lo observado en las cartas de control ARIMA con las Cartas de Control regulares ARIMA.

6. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA DEM:

ARIMA Model (Multiple Seasonal Decomposition): Ln (Airline Passengers-Modified) - Model Automatically Selected
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

7. El Resumen del Modelo ARIMA es:

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	12 Decomposition
Include Constant	1
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

   Esto es un resumen de la información del modelo para los datos desestacionalizados: ARIMA (0,1,1) con una constante. Frecuencia Estacional = 12 utilizando la Descomposición y el Criterio de Selección del Modelo = "AICc". No existen términos estacionales en el modelo. La Transformación Box-Cox es "N/A".

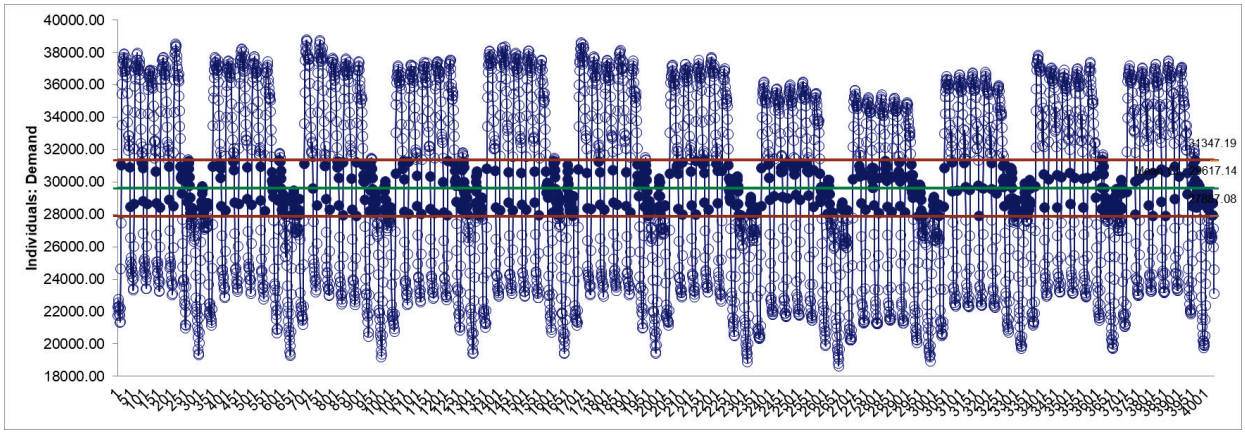
8. No se revisarán las Estimaciones de los Parámetros, las Estadísticas del Modelo y la Precisión del Pronóstico ya que se aproximan a los valores del ARIMA DEM observados anteriormente, aunque se puede observar ligeras diferencias que se deben a la introducción de un valor atípico y un cambio de nivel, así también como el empleo de todos los datos ahora, es decir, no hay períodos de retención. Previamente se empleó una Transformación Box-Cox con Lambda=0 y aquí se está utilizando el Ln de los datos.

Demanda de Media Hora de Electricidad Estacional Múltiple – Taylor

9. Abrir **Half-Hourly Multiple Seasonal Electricity Demand - Taylor.xlsx** (pestaña de la **Hoja 1**). Esta es la demanda de electricidad de media hora (MW) en Inglaterra y Wales desde el lunes, 5 de junio, 2000 hasta el Domingo, 27 de agosto, 2000 (taylor, R forecast). Estos datos tienen una múltiple estacionalidad con frecuencia = 48 (observaciones diarias) y 336 (observaciones semanales), con un total de 4032 observaciones. Véase para estos datos Carta de Corrida, Gráficos FAC/FACP, Gráfico de Densidad Espectral y Gráficos de Descomposición de Tendencia Estacional.

10. Se construirá primero una Carta de Control de Individuales clásica con los datos sin procesar. Hacer click en **SigmaXL > Control Charts > Individuals**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

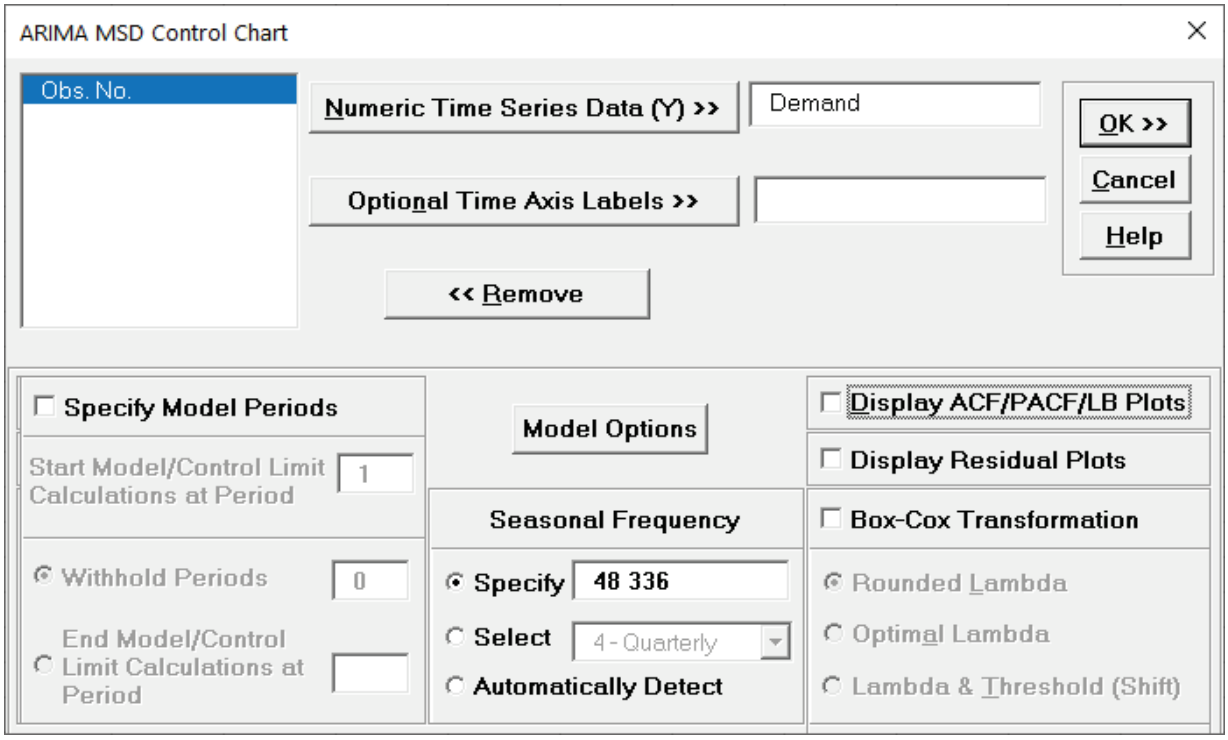
11. Seleccionar *Demand*, hacer click en **Numeric Data Variable (Y) >>**. Hacer click en **OK**. Se genera una Carta de Control de Individuales:



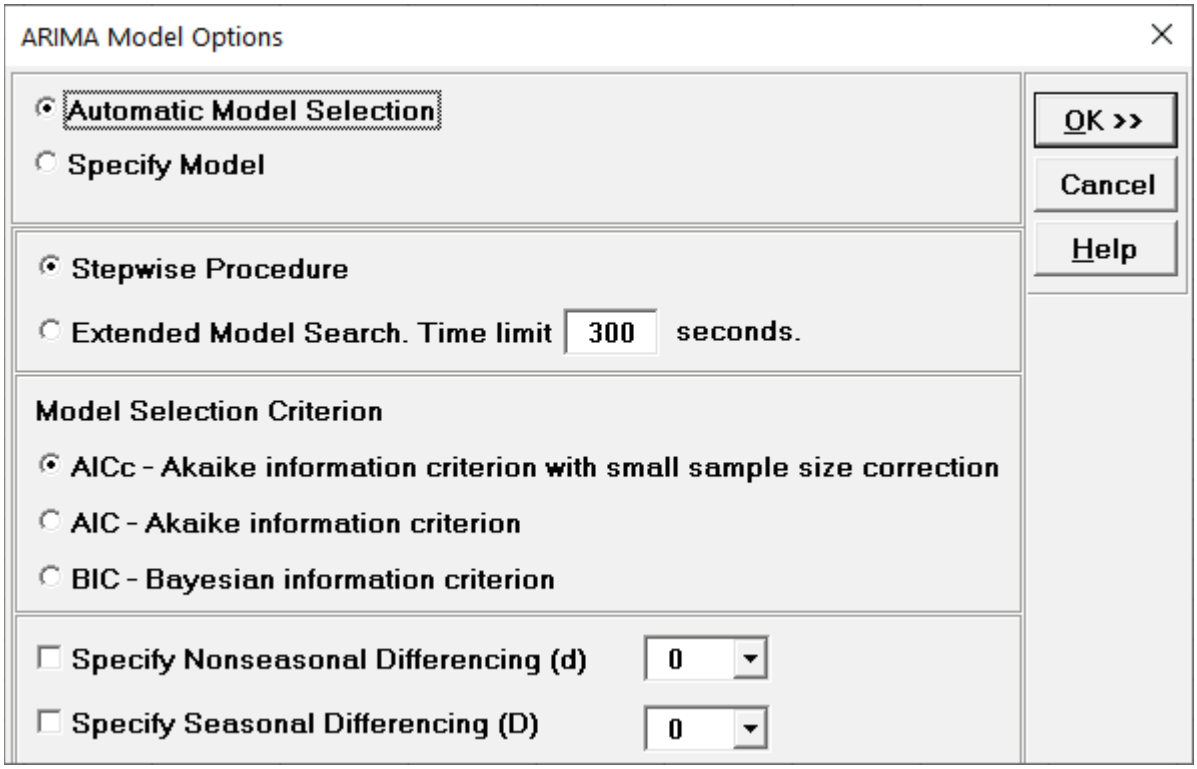
Con la estacionalidad de alta frecuencia, esta carta de control carece de sentido.

12. Hacer click en **SigmaXL > Time Series Forecasting > ARIMA Control Chart > Multiple Seasonal Decomposition Control Chart.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

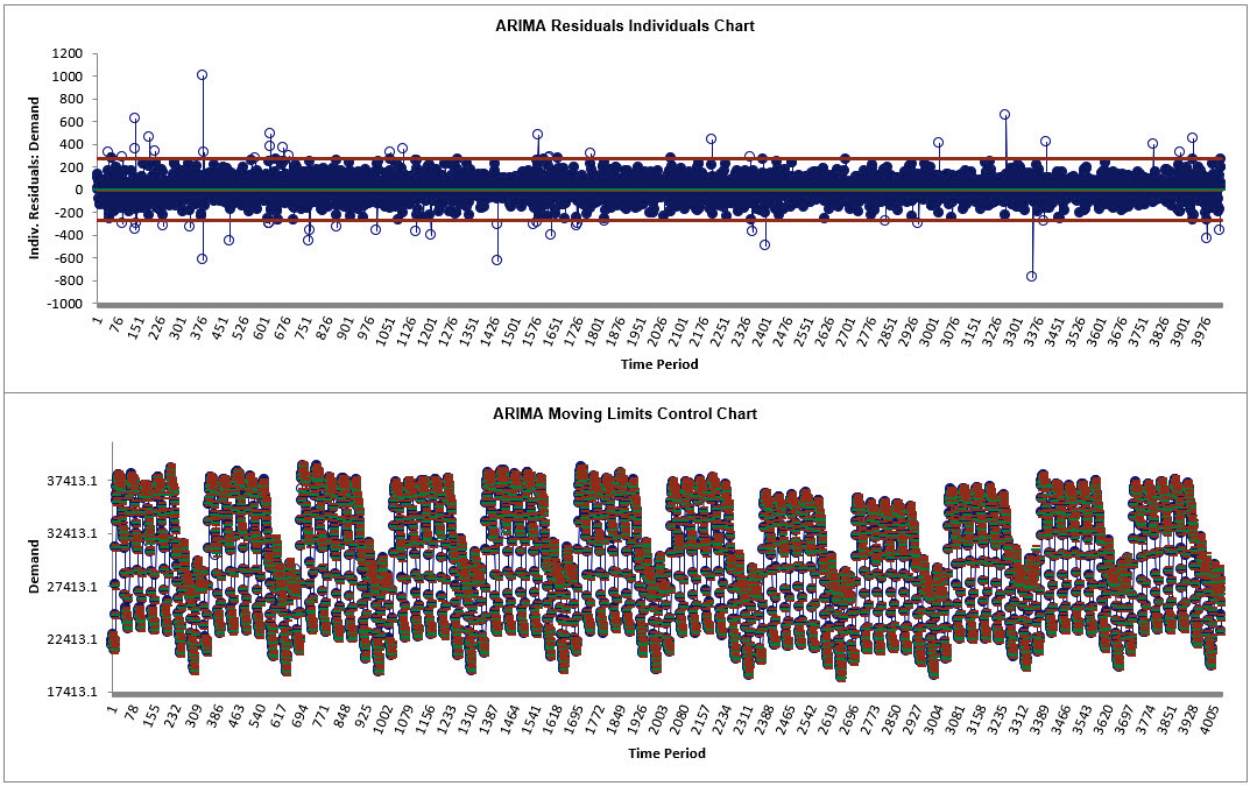
13. Seleccionar *Demand*, hacer click en **Numeric Time Series Data (Y) >>**. Desmarcar **Display ACF/PACF/LB Plots** and **Display Residual Plots**. Marcar **Seasonal Frequency** with **Specify** = 48 336**.** Desmarcar **Specify Model Periods** y **Box-Cox Transformation**.



14. Hacer click en **Model Options**.



15. Se empleará por defecto **Automatic Model Selection** con **AICc** como el **Model Selection Criterion**. Hacer click en **OK** para retornar al diálogo ARIMA MSD Control Chart. Hacer click en **OK**. Se generan las cartas de control de ARIMA DEM:



Se dispone ahora de una carta que puede utilizarse para identificar las causas asignables. El número de señales fuera de control se ha reducido drásticamente. Esto es similar a lo observado con las Cartas de Control de la Suavización Exponencial DEM.

16. Desplazarse hacia abajo para visualizar el encabezado del Modelo ARIMA DEM:

ARIMA Model (Multiple Seasonal Decomposition): Demand - Model Automatically Selected
Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

17. El Resumen del Modelo ARIMA es:

ARIMA Model Summary	
AR Order (p)	2
I Order (d)	1
MA Order (q)	0
SAR Order (P)	0
SI Order (D)	0
SMA Order (Q)	0
Seasonal Frequency	48, 336 Decomposition
Include Constant	0
No. of Predictors	0
Model Selection Criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

Este es un resumen de la información del modelo para los datos desestacionalizados: ARIMA (2,1,0) sin una constante. Seasonal Frequency = 48,336 al hacer uso del Decomposition and Model Selection Criterion = "AICc". No hay términos estacionales en el modelo. Este es el mismo modelo utilizado en el Resumen del Modelo ARIMA DEM pero que utilizó una muestra de retención de 96.

18. No se revisará las Estimaciones de los Parámetros, las Estadísticas del Modelo ni la Precisión del Pronóstico ya que se aproximan a los valores de las Estimaciones de los Parámetros del ARIMA DEM, aunque hay que tener en cuenta que ahora se están utilizando todos los datos, es decir, no hay períodos de retención.

Utilitarios – Datos en Diferencia

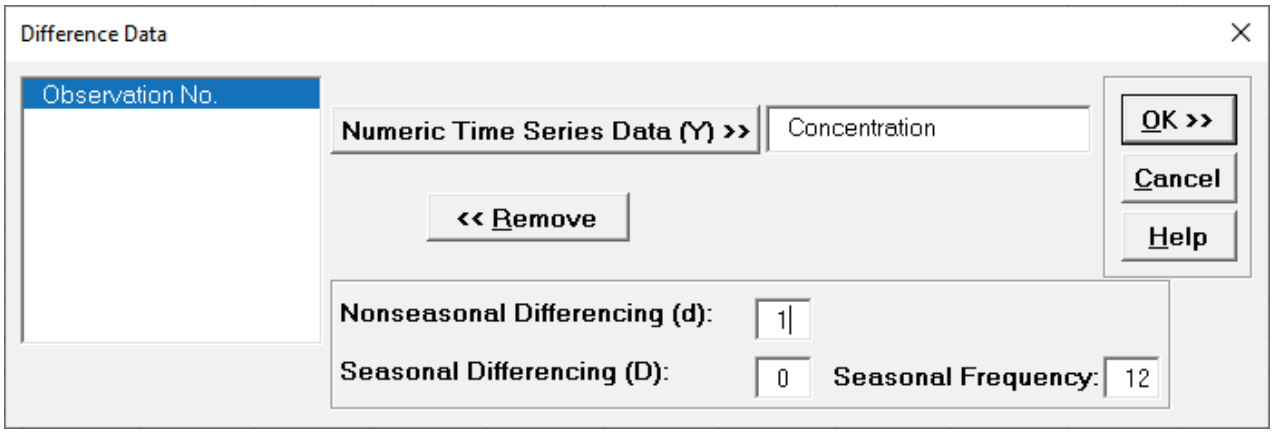
La Diferencia No Estacional y la Estacional se utilizan para convertir un proceso en estacionario. Esto es realizado automáticamente en ARIMA, pero este utilitario lo realiza manualmente. Se demostró aquello previamente en la sección **Carta de Corrida**.

Concentración del Proceso Químico - Serie A

1. Abrir **Chemical Process Concentration – Series A.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie A de Box y Jenkins, un conjunto de 197 valores de la concentración de un proceso químico tomados a intervalos de 2 horas.

2. Hacer click en la pestaña de la **Hoja 1**. Hacer click en **SigmaXL > Time Series Forecasting > Utilities > Difference Data.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Concentration*, Hacer click en **Numeric Time Series Data (Y) >>**. Ingrese 1 en **Nonseasonal Differencing (d)**.



4. Hacer click en **OK**. Se genera una nueva hoja con el orden d = 1 datos diferenciados (Y_2 - Y_1, Y_3-Y_2, …).

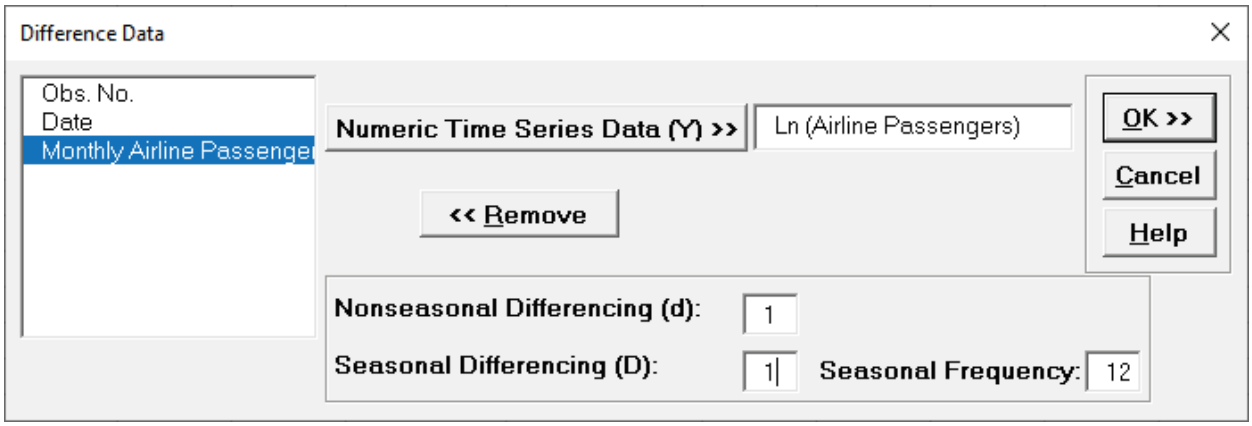
Concentration	Concentration: d = 1, D = 0	
17		
16.6	-0.4	
16.3	-0.3	
16.1	-0.2	
17.1	1	
16.9	-0.2	
16.8	-0.1	
17.4	0.6	
17.1	-0.3	
17	-0.1	
16.7	-0.3	
17.4	0.7	

Pasajeros Mensuales de la Aerolínea - Serie G

5. Abrir **Monthly Airline Passengers - Series G.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie G de Box y Jenkins, el total mensual de pasajeros de una aerolínea internacional de enero 1949 a diciembre 1960 y es uno del conjunto de datos más populares que se emplean en los cursos de introducción del pronóstico con series temporales.

6. Hacer click en la pestaña de la **Hoja 1**. Hacer click en **SigmaXL > Time Series Forecasting > Utilities > Difference Data.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

7. Seleccionar *Ln (Airline Passengers)*, hacer click en **Numeric Time Series Data (Y) >>**. Ingrese 1 en **Nonseasonal Differencing (d)**; ingrese 1 en **Seasonal Differencing (D)**; se especifica **12 como Seasonal Frequency.**



8. Hacer click en **OK**. Se genera una nueva hoja con el orden estacional D = 1 dato diferenciado (Y_{13} - Y_1, Y_{14}-Y_2, ...) y orden no estacional d=1 ($D_2 – D_1$, $D_3 – D_2$, ...).

Ln (Airline Passengers)	Ln (Airline Passengers): d = 1, D = 1
4.718498871	
4.770684624	
4.882801923	
4.859812404	
4.795790546	
4.905274778	
4.997212274	
4.997212274	
4.912654886	
4.779123493	
4.644390899	
4.770684624	
4.744932128	
4.836281907	0.039164025
4.94875989	0.000360685
4.905274778	-0.020495594
4.828313737	-0.012939182

Utilitarios – Datos de Rezagos

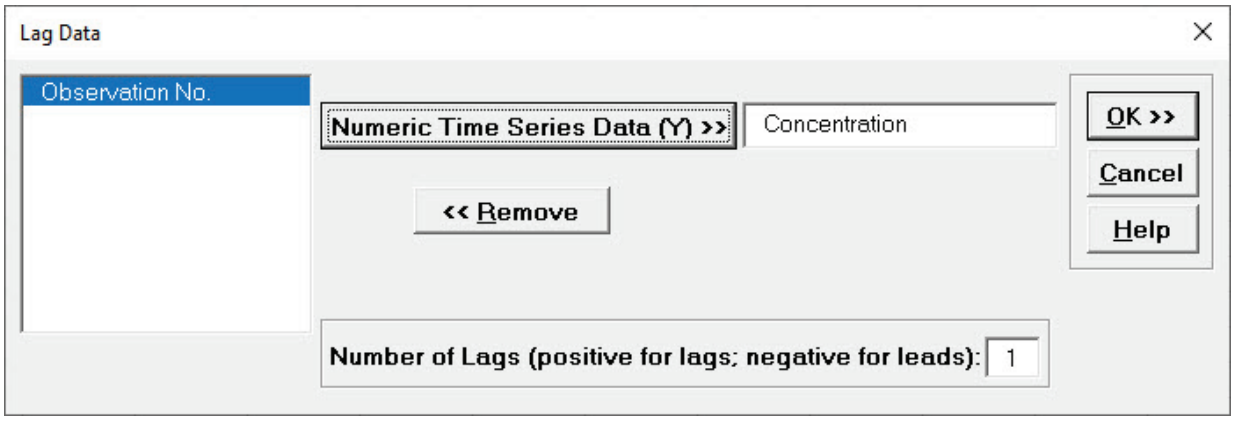
El utilitario de los Datos de Rezago es un provechoso utilitario para crear datos de columna de Rezagos (o Dirección). Esto se demostró anteriormente en la sección de Gráficos de Autocorrelación (FAC/FACP).

Concentración del Proceso Químico - Serie A

1. Abrir **Chemical Process Concentration – Series A.xlsx** (pestaña de la **Hoja 1**). Estos datos son una Serie A de Box y Jenkins, un conjunto de 197 valores de la concentración de un proceso químico tomados a intervalos de 2 horas.

2. Hacer click en **SigmaXL > Time Series Forecasting > Utilities > Lag Data**. Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Concentration*, hacer click en **Numeric Time Series Data (Y) >>**. Utilice por defecto **Number of Lags** = 1.



4. Hacer click en **OK**. Se crea una nueva hoja con los datos del Rezago 1:

Concentration	Concentration: Lag = 1
17	
16.6	17
16.3	16.6
16.1	16.3
17.1	16.1
16.9	17.1

5. Hacer click en el menú **Recall SigmaXL Dialog** o presione **F3** para retornar al último diálogo. Registrar **Lag** = 2. Hacer click en **OK**. Repetir **Lag** = 3. Las columnas de los rezagos se agregan como sigue:

Concentration	Concentration: Lag = 1	Concentration: Lag = 2	Concentration: Lag = 3
17			
16.6	17		
16.3	16.6	17	
16.1	16.3	16.6	17
17.1	16.1	16.3	16.6
16.9	17.1	16.1	16.3
16.8	16.9	17.1	16.1

Utilitarios – Interpolar los Valores Ausentes

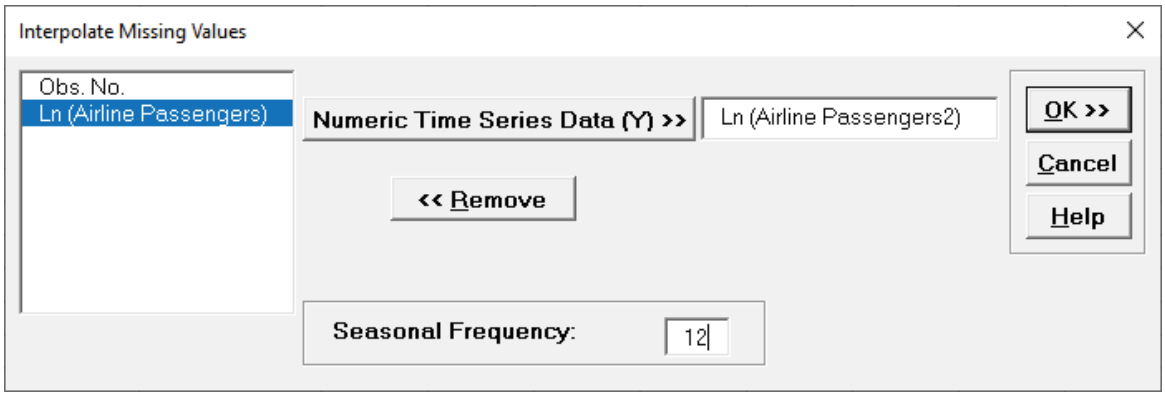
El utilitario para interpolar los valores Ausentes se mostró anteriormente. Véase para mayores detalles el Apéndice: **Interpolación Lineal de Valores Ausentes Ajustados Estacionalmente**. Es utilizado automáticamente en la Suavización Exponencial cuando están presentes valores ausentes. Aunque existe cierta robustez contra la presencia de valores ausentes, si el número de valores ausentes es muy grande, entonces la estimación del modelo y la precisión del pronóstico se verán degradadas.

Pasajeros Mensuales de la Aerolínea – Valores Ausentes

1. Abrir **Monthly Airline Passengers – Missing Values.xlsx** (pestaña de la **Hoja 1**). Ln (Airline Passengers2) tiene valores ausentes en las observaciones 50 y 100.

2. Hacer click en la pestaña de la **Hoja 1**. Hacer click en **SigmaXL > Time Series Forecasting > Utilities > Interpolate Missing Values.** Cerciorarse en seleccionar la tabla completa de los datos. De lo contrario, marcar **Use Entire Data Table**. Hacer click en **Next**.

3. Seleccionar *Ln (Airline Passengers2)*, hacer click en **Numeric Time Series Data (Y) >>**. Especificar **Seasonal Frequency** como 12.



4. Hacer click en **OK**. Se genera una nueva hoja con los valores interpolados linealmente con un ajuste estacional los cuales se resaltan en amarillo.

Ln (Airline Passengers2)	Y: Interpolated (12)
4.718498871	4.718498871
4.770684624	4.770684624
4.882801923	4.882801923
4.859812404	4.859812404
	5.254047791
	5.846836235

Y50 es estimado como 5.254 (versus el valor original de 5.278). Y100 se estima en 5.847 (versus el valor original de 5.852).

SigmaXL Apéndice

SigmaXL® Version 9.0 Feature List Summary

Compatible with Excel 2019 for Windows and Mac

Menu Layout: Classical or DMAIC

Recall Last Dialog

Data Manipulation:

- Subset by Category, Number, Date or Random
- Stack Subgroups Across Rows
- Stack and Unstack Columns
- Standardize Data
- Random Number Generator
  - Normal
  - Uniform (Continuous & Integer)
  - Lognormal
  - Weibull
  - Triangular
- Box-Cox Transformation

Templates & Calculators:

- DMAIC & DFSS Templates
  - Team/Project Charter
  - SIPOC Diagram
  - Flowchart Toolbar
  - Data Measurement Plan
  - Cause & Effect (Fishbone) Diagram and Quick Template
  - Cause & Effect (XY) Matrix with Pareto
  - Failure Mode & Effects Analysis (FMEA) with RPN Sort
  - Quality Function Deployment (QFD)
  - Pugh Concept Selection Matrix
  - Control Plan
- Lean Templates
  - Takt Time Calculator
  - Value Analysis/Process Load Balance
  - Value Stream Mapping
- Graphical Templates
  - Pareto Chart, Histogram, Run Chart
- Statistical Templates
  - Sample Size – Discrete and Continuous
  - Minimum Sample Size for Robust t-Tests and ANOVA
  - 1 Sample Z-test & Confidence Interval for Mean
  - 1 Sample t-Test & Confidence Interval for Mean
  - 2 Sample t-Test & Confidence Interval (Compare 2 Means)
  - 1 Sample Equivalence Test for Mean
  - 2 Sample Equivalence Test (Compare 2 Means)
  - 1 Sample Chi-Square Test and CI for Standard Deviation
  - 2 Sample F-Test and CI (Compare 2 Standard Deviations)
  - 1 Proportion Test and Confidence Interval
  - 2 Proportions Test and Confidence Interval (with Fisher Exact P-Value)
  - 2 Proportions Equivalence Test
  - 1 Poisson Rate Test and Confidence Interval
  - 2 Poisson Rates Test and Confidence Interval
  - 2 Poisson Rates Equivalence Test
  - One-Way Chi-Square Goodness-of-Fit Test (with Exact and Monte Carlo P-Value)
- Probability Distribution Calculators
  - Normal, Lognormal, Exponential, Weibull
  - Binomial, Poisson, Hypergeometric
- Measurement Systems Analysis (MSA) Templates
  - Type 1 Gage Study
  - Gage Bias and Linearity Study
  - Gage R&R Study – with Multi-Vari Analysis
  - Attribute Gage R&R (Attribute Agreement Analysis)
- Process Sigma Level – Discrete and Continuous
- Process Capability & Confidence Intervals
- Tolerance Interval Calculator (Normal Exact)
- DOE Templates
  - 2 to 5 Factors
  - 2-Level Full and Fractional-Factorial designs
  - Main Effects & Interaction Plots
- Taguchi DOE Templates
  - L4, L8, L9, L12, L16, L18, L27
  - Signal-to-Noise Ratios: Nominal is Best, Nominal is Best (Variance Only), Nominal is Best (Mean Square Deviation with Target), Larger is Better, Smaller is Better
  - Pareto of Deltas (Effects) and ANOVA SS (Sum-of-Squares) % Contribution (for Main Effects and Two-Way Interactions)
  - Main Effects & Interaction Plots
- Control Chart Templates
  - Individuals, C-Chart
  - G and T Rare Events**
  - Exponentially Weighted Moving Average (EWMA)**

** New features in SigmaXL® Version 9.0

SigmaXL® Version 9.0 Feature List Summary

Statistical Tools:

- P-values turn red when results are significant: (p-value < alpha)
- Descriptive Statistics including Anderson-Darling Normality test, Skewness and Kurtosis with p-values
- Descriptive Statistics options:
  - Percentile Report and Percentile Ranges
  - Percentile Confidence and Tolerance Intervals
  - Additional Descriptive Statistics and Normality Tests (Shapiro-Wilk and Doornik-Hansen Normality)
  - Outlier (Boxplot and Grubbs) and Randomness (Nonparametric Runs) Tests
- 1 Sample t-test and confidence intervals
- Paired t-test, 2 Sample t-test
- 2 Sample comparison tests:
  - Reports AD Normality, F-test and Levene's for variance, t-test assuming equal and unequal variance, Mann-Whitney test for medians.
  - Recommended tests are highlighted based on sample size, normality, and variance
- One-Way ANOVA and Means Matrix
  - Multiple Comparison of Means Probability Methods (Post-Hoc): Fisher, Tukey, Dunnett with Control
- Automatic Assumptions Check for One Sample, Two-Sample, Paired T-tests and One-Way ANOVA
  - Text report with color highlight: Green (OK), Yellow (Warning) and Red (Serious Violation)
  - Test each sample for Normality. If not, check minimum sample size for robustness of test
  - Check each sample for Outliers: Potential (Tukey's Boxplot 1.5*IQR); Likely (2.2*IQR); Extreme (3*IQR)
  - Randomness (Nonparametric Runs Test)
  - Equal Variance (for 2 or more samples)
- Two-Way ANOVA (Balanced and Unbalanced)
- Equal Variance Tests (Bartlett, Levene and Welch's ANOVA)
  - Multiple Comparison of Variances Probability Methods (Post-Hoc): F-Test (with Bonferroni Correction), Levene, Tukey ADM (Absolute Deviations from Median)
  - Welch Multiple Comparison of Means Probability Methods (Post-Hoc): Welch Pairwise, Games-Howell
- Correlation Matrix (Pearson and Spearman's Rank Correlation)
  - Automatic normality check for correlation utilizing the powerful Doornik-Hansen bivariate normality test
  - Yellow highlight to recommend significant Pearson or Spearman correlation - Pearson is highlighted if the data are bivariate normal, otherwise Spearman is highlighted
- Multiple Linear Regression:
  - Accepts continuous and/or categorical (discrete) predictors
  - Interactive Predicted Response Calculator with 95% Confidence Interval and 95% Prediction Interval
  - Residual Plots: histogram, normal probability plot, residuals vs. time, residuals vs. predicted and residuals vs. X factors
  - Residual types include Regular, Standardized, Studentized (Deleted t) and Cook's Distance (Influence), Leverage and DFITS
  - Highlight of significant outliers in residuals
  - Durbin-Watson Test for Autocorrelation in Residuals with p-value
  - ANOVA report for categorical predictors
  - Pure Error and Lack-of-Fit report
  - Collinearity Variance Inflation Factor (VIF) and Tolerance report
  - Fit Intercept is optional
- Binary and Ordinal Logistic Regression
  - Powerful and user-friendly logistic regression.
  - Report includes a calculator to predict the response event probability for a given set of input X values.
  - Categorical (discrete) predictors can be included in the model in addition to continuous predictors.
  - Model summary and goodness of fit tests include Likelihood Ratio Chi-Square, Pseudo R-Square, Pearson Residuals Chi-Square, Deviance Residuals Chi-Square, Observed and Predicted Outcomes – Percent Correctly Predicted.
  - Stored data includes Event Probabilities, Predicted Outcome, Observed-Predicted, Pearson Residuals, Standardized Pearson Residuals, and Deviance Residuals.
- Chi-Square Test (Stacked Column data and Two-Way Table data)
  - With Fisher Exact (utilizing permutations and fast network algorithms) and Monte Carlo P-Values
  - Options: Advanced Tests and Measures of Association for Nominal & Ordinal Categories
- Nonparametric Tests:
  - 1 Sample Sign and 1 Sample Wilcoxon
  - 2 Sample Mann-Whitney
  - Kruskal-Wallis and Mood's Median Test (with graph of Group Medians and 95% Median Confidence Intervals)
  - Runs Test
  - With Exact and Monte Carlo P-Value
- Power and Sample Size Calculators for:
  - 1 and 2 Sample t-Test
  - One-Way ANOVA
  - 1 Proportion Test, 2 Proportions Test
  - The Power and Sample Size Calculators allow you to solve for Power (1 – Beta), Sample Size, or Difference (specify two, solve for the third).
- Power and Sample Size Chart. Quickly create a graph showing the relationship between Power, Sample Size and Difference.

** New features in SigmaXL® Version 9.0

SigmaXL® Version 9.0 Feature List Summary

Design of Experiments:

- Generate 2-Level Factorial and Plackett-Burman Screening Designs
  - User-friendly dialog box
  - 2 to 19 Factors; 4,8,12,16,20 Runs
  - Unique "view power analysis as you design"
  - Randomization, Replication, Blocking and Center Points
- Basic DOE Templates
  - 2 to 5 Factors, 2-Level Full and Fractional-Factorial designs
  - Automatic update to Pareto of Coefficients
  - Easy to use, ideal for training
- Main Effects & Interaction Plots
- Analyze 2-Level Factorial and Plackett-Burman Screening Designs
  - Used in conjunction with Recall Last Dialog, it is very easy to iteratively remove terms from the model
  - Interactive Predicted Response Calculator with 95% Confidence Interval and 95% Prediction Interval.
  - ANOVA report for Blocks, Pure Error, Lack-of-Fit and Curvature
  - Collinearity Variance Inflation Factor (VIF) and Tolerance report
  - Residual plots: histogram, normal probability plot, residuals vs. time, residuals vs. predicted and residuals vs. X factors
  - Highlight of significant outliers in residuals
  - Durbin-Watson Test for Autocorrelation
- Contour & 3D Surface Plots
- Response Surface Designs
  - 2 to 5 Factors
  - Central Composite and Box-Behnken Designs
  - Easy to use design selection sorted by number of runs

Control Charts:

- Control Chart Selection Tool
- Individuals, Individuals & Moving Range
- X-Bar & R, X-Bar & S
- I-MR-R, I-MR-S (Between/Within)
- P, NP, C, U
- P' and U' (Laney) to handle overdispersion
- Control charts include a report on tests for special causes. Special causes are also labeled on the control chart data point. Set defaults to apply any or all of Tests 1-8.
- Tests for Special Causes: Support for Varying Subgroup Sizes (Moving Limits), Historical Groups and MR/Range/StDev Charts (Tests 1-4)**
- Process Capability report (Pp, Ppk, Cp, Cpk) is available for I, I-MR, X-Bar & R, X-bar & S charts.
- Add data to existing charts for operator ease of use!
- Scroll through charts with user defined window size
- Advanced Control Limit options: Subgroup Start and End; Historical Groups (e.g. split control limits to demonstrate before and after improvement)
- Exclude data points for control limit calculation
- Add comment to data point for assignable cause
- ± 1, 2 Sigma Zone Lines
- Control charts for Nonnormal data (Individuals)
  - Box-Cox and Johnson Transformations
  - 16 Nonnormal distributions supported (see Process Capability)
  - Individuals chart of original data with percentile based control limits
  - Individuals/Moving Range chart for normalized data with optional tests for special causes
- Control chart templates:
  - Individuals, C-Chart
  - G and T Rare Events**
  - Exponentially Weighted Moving Average (EWMA)**

Reliability/Weibull Analysis:

- Weibull Analysis
  - Complete and Right Censored data
  - Least Squares and Maximum Likelihood
  - Output includes percentiles with confidence intervals, survival probabilities, and Weibull probability plot.

** New features in SigmaXL® Version 9.0

SigmaXL® Version 9.0 Feature List Summary

Time Series Forecasting and Control Charts for Autocorrelated Data:**

- Run Chart
- Autocorrelation Function (ACF)/Partial Autocorrelation (PACF) Plots
- Cross Correlation (CCF) Plots
- Seasonal Trend Decomposition Plots
- Spectral Density Plot
- Exponential Smoothing:
  - Forecast with Prediction Intervals
  - Exponential Smoothing models use Rob Hyndman's taxonomy:
  - Additive/Multiplicative Error
  - Additive/Additive Damped Trend
  - Additive/Multiplicative Seasonal
  - This includes all of the classical exponential smoothing models such as Simple/Single/EWMA, Double and Holt-Winters
  - Multiple Seasonal Decomposition (MSD) option: useful for high frequency and/or multiple frequency data, such as Monthly with frequency = 12, Daily with frequency = 7 and Hourly with frequency = 24
- Exponential Smoothing Control Chart for autocorrelated data:
  - Control Chart based on Exponential Smoothing residuals
  - Add Data, Show Last 30 Data Points, Enable Scroll options
  - MSD option
- Autoregressive Integrated Moving Average (ARIMA):
  - Forecast with Prediction Intervals
  - ARIMA Forecast with Predictors (Continuous and/or Categorical)
  - MSD option
- ARIMA Control Chart for autocorrelated data:
  - ARIMA Control Chart with Predictors
  - Add Data, Show Last 30 Data Points, Enable Scroll options
  - MSD option
- Utilities:
  - Difference Data
  - Lag Data
  - Interpolate Missing Values (seasonally adjusted linear interpolation)
- Time Series Forecasting Model Features:
  - ARIMA and Exponential Smoothing models are fully automatic or user specified
  - Utilizes modern State Space and Kalman Filter models for accurate parameter estimation
  - ARIMA estimates missing values with Kalman Filter; Exponential Smoothing uses seasonally adjusted linear interpolation
  - Automatic Box-Cox Transformation
  - Automatic seasonal frequency detection
- Model diagnostics:
  - ACF/PACF Plots
  - Ljung-Box p-values
  - Log-Likelihood, AIC, AICc, BIC, Residual StDev
  - Residual plots (histogram, normal probability, residual versus fits, residuals versus order)
  - Residuals Nonparametric Randomness Runs Test
- Forecast Accuracy:
  - In-Sample (Estimation) one-step-ahead forecast errors (RMSE, MAE, MASE, MAPE)
  - Out-of-Sample (Withhold) one-step-ahead forecast errors
  - Out-of-Sample (Withhold) multi-step-ahead forecast errors
  - Evaluated using the benchmark standard M4 forecast competition data, a total of 100,000 data sets with Yearly, Quarterly, Monthly, Weekly, Daily and Hourly data. Using a hybrid average of automatic Exponential Smoothing and ARIMA, SigmaXL (unofficially) ranked 10th out of 60 in the Overall Weighted Average forecast accuracy score, ahead of Forecast Pro, Crystal Ball and Autobox.

** New features in SigmaXL® Version 9.0

Distribuciones

Distribución Beta

PDF

$$\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

CDF

$$\int_{-\infty}^{x}\frac{1}{B(\alpha,\beta)}t^{\alpha-1}(1-t)^{\beta-1}dt$$

donde *B* es la función Beta:

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Rango $0 \leq x \leq 1$.

Parámetro de Forma 1 $\alpha>0$.

Parámetro de Forma 2 $\beta>0$.

Γ(β) es la función Gamma y se describe abajo en la Distribución Gamma.

Distribución Beta con Umbral Inferior/Superior

PDF

$$\frac{1}{B(\alpha,\beta)}\left(\frac{x-\theta_1}{\theta_2-\theta_1}\right)^{\alpha-1}\left(\frac{\theta_2-x}{\theta_2-\theta_1}\right)^{\beta-1}$$

CDF

$$\int_{-\infty}^{\frac{x-\theta_1}{\theta_2-\theta_1}}\frac{1}{B(\alpha,\beta)}t^{\alpha-1}(1-t)^{\beta-1}dt$$

donde *B* es la función Beta.

Rango $0 \leq x \leq 1$.

Parámetro de Forma 1 α>0.

Parámetro de Forma 2 β>0.

Umbral Inferior θ_1

Umbral Superior θ_2

Notas: La estimación de la Distribución Beta de 4 parámetros es realizada en dos partes. En la primera parte, los parámetros iniciales se derivan con el uso del método de los momentos. Los umbrales de los parámetros entonces se fijan utilizando estos valores y los Parámetros de forma se estiman por medio de máxima verosimilitud.

Referencia para la Distribución Beta con la Estimación de Parámetro Umbral Inferior/Superior:

Wang, J.Z. (2005). "A note on Estimation in the Four Parameter Beta Distribution", *Comm in Stats Simulation and computation*, Vol. 34 pp. 495- 501.

Distribución Box-Cox con Umbral

PDF

$$\frac{(x-\theta)^{\lambda-1}}{\sigma\sqrt{2\pi}}e^{-(((x-\theta)^{\lambda}-1)/\lambda-\mu)^2/2\sigma^2}$$

CDF

$$\int_{-\infty}^{[(x-\theta)^{\lambda}-1]/\lambda}\frac{1}{\sigma\sqrt{2\pi}}e^{-(t-\mu)^2/2\sigma^2}dt$$

Rango $0 < x - \theta < \infty$.

Parámetro de localización, μ, la media.

Parámetro de escala, $\sigma > 0$, la desviación estándar.

Parámetro de forma λ.

Parámetro del umbral $\theta < \min(x)$.

Notas: La probabilidad concentrada se utiliza en la estimación de MV. Esto implica que los parámetros de localización y escala no se estiman con libertad, sino que se derivan como la media y la desviación estándar de la variable transformada por el método Box-Cox. Los parámetros estimados λ and θ luego se utilizan en la transformación Box-Cox (Potencia). Véase **Transformaciones** más abajo.

Distribución Exponencial

PDF

$$\frac{e^{-x/\alpha}}{\alpha}$$

CDF

$$1 - e^{-x/\alpha}$$

Rango $0 \leq x < \infty$.

Parámetro de escala, $\alpha > 0$, la media.

Distribución Exponencial con Umbral

PDF

$$\frac{e^{-(x-\theta)/\alpha}}{\alpha}$$

CDF

$$1 - e^{-(x-\theta)/\alpha}$$

Rango $0 \leq x < \infty$.

Parámetro de escala, $\alpha > 0$, la media.

Parámetro del umbral $\theta < \min(x)$.

Distribución Gamma

PDF

$$\frac{(x/\alpha)^{\beta-1}e^{-x/\alpha}}{\alpha\Gamma(\beta)}$$

CDF

$$\int_0^{x/\alpha} \frac{e^{-t}t^{\beta-1}}{\Gamma(\beta)}dt$$

donde Γ(β) es la función Gamma:

$$\Gamma(\beta) = \int_0^{\infty} e^{-t}t^{\beta-1}dt$$

Rango $0 \leq x < \infty$.

Parámetro de escala $\alpha > 0$.

Parámetro de forma $\beta > 0$.

Distribución Gamma con Umbral

PDF

$$\frac{[(x-\theta)/\alpha])^{\beta-1}e^{-(x-\theta)/\alpha}}{\alpha\Gamma(\beta)}$$

CDF

$$\int_0^{(x-\theta)/\alpha} \frac{e^{-t}t^{\beta-1}}{\Gamma(\beta)}dt$$

Rango $0 \leq x - \theta < \infty$.

Parámetro de escala $\alpha > 0$.

Parámetro de forma $\beta > 0$.

Parámetro del umbral $\theta < \min(x)$.

Distribución Media Normal

PDF

$$\frac{\sqrt{2/\pi}}{\sigma^2}e^{-x^2/2\sigma^2}$$

CDF

$$\int_{-\infty}^{x} \frac{\sqrt{2/\pi}}{\sigma}e^{-t^2/2\sigma^2}dt - 1$$

Rango $0 \leq x < \infty$.

Parámetro de escala, $\sigma > 0$, la desviación estándar

Referencia para la Aplicación de la Distribución Media-Normal:

Chou, C., & H. Liu, (1998). "Properties of the half-normal Distribution and its application to quality control", *Journal of Industrial Technology* Vol. 14(3) pp. 4-7

Distribución

PDF

$$\frac{1}{\sigma} e^{-(x-\mu)/\sigma} e^{-e^{-(x-\mu)/\sigma}}$$

CDF

$$e^{-e^{-(x-\mu)/\sigma}}$$

Rango $-\infty < x < \infty$.

Parámetro de localización, μ, la moda.

Parámetro de escala $\sigma > 0$.

Distribución Logística

PDF

$$\frac{e^{(x-\mu)/\sigma}}{\sigma(1+e^{(x-\mu)/\sigma})^2}$$

CDF

$$\frac{1}{1+e^{-(x-\mu)/\sigma}}$$

Rango $-\infty < x < \infty$.

Parámetro de localización, μ, la media.

Parámetro de escala $\sigma > 0$.

Distribución Log logística

PDF

$$\frac{e^{(\ln(x)-\mu)/\sigma}}{x\sigma(1+e^{(\ln(x)-\mu)/\sigma})^2}$$

CDF

$$\frac{1}{1+e^{-(\ln(x)-\mu)/\sigma}}$$

Rango $0 < x < \infty$.

Parámetro de localización, μ, la media.

Parámetro de escala $\sigma > 0$.

Distribución Log logística con Umbral

PDF

$$\frac{e^{(\ln(x-\theta)-\mu)/\sigma}}{(x-\theta)\sigma(1+e^{(\ln(x-\theta)-\mu)/\sigma})^2}$$

CDF

$$\frac{1}{1+e^{-(\ln(x-\theta)-\mu)/\sigma}}$$

Rango $0 < x < \infty$.

Parámetro de localización, μ, la media.

Parámetro de escala $\sigma > 0$.

Parámetro del umbral $\theta < \min(x)$.

Distribución Log normal

PDF

$$\frac{1}{x\sqrt{2\pi\sigma^2}} e^{-.5(\ln(x)-\mu)^2/\sigma^2}$$

CDF

$$\int_{-\infty}^{x} \frac{1}{t\sigma\sqrt{2\pi}} e^{-.5(\ln(t)-\mu)^2/\sigma^2} dt$$

Rango $0 < x < \infty$.

Parámetro de escala, μ, la media de Ln (x).

Parámetro de forma, $\sigma > 0$, la desviación estándar de Ln (x) .

Distribución Log normal con Umbral

PDF

$$\frac{1}{(x-\theta)\sqrt{2\pi\sigma^2}} e^{-.5(\ln(x-\theta))-\mu)^2/\sigma^2}$$

CDF

$$\int_{-\infty}^{x-\theta} \frac{1}{t\sigma\sqrt{2\pi}} e^{-.5(\ln(t)-\mu)^2/\sigma^2} dt$$

Rango $0 < x < \infty$.

Parámetro de escala, μ, la media de Ln (x).

Parámetro de forma, $\sigma > 0$, la desviación estándar de Ln (x) .

Parámetro del umbral $\theta < \min(x)$.

Referencia sobre la Distribución Log Normal con Umbral – Estimación de los Parámetros:

Giesbrecht, F. and A.H. Kempthorne (1966). "Maximum Likelihood Estimation in the Three-parameter Lognormal Distribution", *Journal of the Royal Statistical Society*, B 38, pp. 257-264.

Distribución Normal

PDF

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-.5(x-\mu)^2/\sigma^2}$$

CDF

$$\int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-.5(t-\mu)^2/\sigma^2} dt$$

Rango $-\infty < x < \infty$.

Parámetro de localización, μ, la media.

Parámetro de escala, $\sigma > 0$, la desviación estándar.

Nota: En consistencia con otros reportes sobre SigmaXL tal como Estadística Descriptiva, la desviación estándar se estima como la desviación estándar de la muestra con n-1 (en lugar de n).

Distribución del Valor Extremo Más Pequeño

PDF

$$\frac{1}{\sigma}e^{(x-\mu)/\sigma}e^{-e^{(x-\mu)/\sigma}}$$

CDF

$$1-e^{-e^{(x-\mu)/\sigma}}$$

Rango $-\infty < x < \infty$.

Parámetro de localización, μ, la media.

Parámetro de escala $\sigma > 0$.

Distribución Weibull

PDF

$$\frac{\beta x^{\beta-1}}{\alpha^{\beta}}e^{-(x/\alpha)^{\beta}}$$

CDF

$$1-e^{-(x/\alpha)^{\beta}}$$

Rango $0 \le x < \infty$.

Parámetro de escala, $\alpha > 0$, la vida característica.

Parámetro de forma $\beta > 0$.

Distribución Weibull con Umbral

PDF

$$\frac{\beta x^{\beta-1}}{\alpha^{\beta}} e^{-[(x-\theta)/\alpha]^{\beta}}$$

CDF

$$1 - e^{-[x-\theta)/\alpha]^{\beta}}$$

Rango $0 \leq x < \infty$.

Parámetro de escala, $\alpha > 0$, la vida característica.

Parámetro de forma $\beta > 0$.

Parámetro del umbral $\theta < \min(x)$.

Referencia sobre la Distribución Weibull con Umbral – Estimación de los Parámetros:

Lockhart, R.A. and M.A. Stephens (1994)."Estimation and Tests of Fit for the Three-parameter Weibull Distribution", *Journal of the Royal Statistical Society*, Vol.56 (3), pp. 491-500.

Transformaciones

Transformación Box-Cox (Potencia)

$\lambda \neq 0$

$$\begin{aligned} z &= x^{\lambda} \\ x &= z^{1/\lambda} \end{aligned}$$

$\lambda = 0$

$$\begin{aligned} z &= \ln(x) \\ x &= e^{z} \end{aligned}$$

Rango $0 < x < \infty$.

Parámetro de forma λ.

Nota: El Parámetro Óptimo de forma, λ, se deriva con el uso de una búsqueda en la cual la función de criterio es la desviación estándar de la variable estandarizada transformada.

Transformación Box-Cox (Potencia) con Umbral

$\lambda \neq 0$

$$\begin{aligned} z &= (x-\theta)^{\lambda} \\ x &= z^{1/\lambda} + \theta \end{aligned}$$

$\lambda = 0$

$$\begin{aligned} z &= \ln(x-\theta) \\ x &= e^{z} + \theta \end{aligned}$$

Rango $0 < x - \theta < \infty$.

Parámetro de forma λ.

Parámetro del umbral $\theta < \min(x)$.

Nota: Los parámetros λ y θ se estiman utilizando EMV como se describe arriba en la **Distribución Box-Cox con Umbral.**

Transformación Johnson

La Transformación Johnson es una de las tres familias de la Distribución: SB (limitada), SL (log normal), y SU (no limitada) y los parámetros asociados para transformar los datos a ser distribuidos normalmente. La metodología sigue a Chou et al (1998) y utiliza el valor-p de Anderson Darling como criterio de normalidad.

Transformación Johnson - S_B

$$z = \gamma + \eta \ln\left(\frac{x-\epsilon}{\lambda+\epsilon-x}\right)$$

$$x = \epsilon + \frac{1}{1+e^{-(z-\gamma)/\eta}}$$

Z is $N(0, 1)$

Rango $\varepsilon < x < \varepsilon + \lambda$.

Parámetro de localización $\eta > 0$.

Parámetro de escala $\lambda > 0$.

Parámetro de forma γ ilimitado.

Parámetro de forma ε ilimitado.

Transformación Johnson - S_L

$$\begin{aligned} z &= \gamma + \eta \ln(x - \epsilon) \\ x &= \epsilon + e^{(z-\gamma)/\eta} \end{aligned}$$

Z is N (0, 1)

Rango $x > \varepsilon$

Parámetro de localización $\eta > 0$.

Parámetro de forma γ ilimitado.

Parámetro de forma ε ilimitado.

Transformación Johnson - S_U

$$\begin{aligned} z &= \gamma + \eta \sinh^{-1}[(x - \epsilon)/\lambda] \\ x &= \epsilon + \lambda \sinh[(z - \gamma)/\eta] \end{aligned}$$

Z is N (0, 1)

Rango $-\infty < x < \infty$.

Parámetro de localización $\eta > 0$.

Parámetro de escala $\lambda > 0$.

Parámetro de forma γ ilimitado.

Parámetro de forma ε ilimitado.

Referencias para la Transformación Johnson:

Choy. A.M. Polansky, and R.L. Mason (1998). "Transforming Nonnormal Data to Normality in Statistical Process Control," *Journal of Quality Technology*, Vol. 30(2), pp. 133-141.

David, H.A. (1981). *Order Statistics*, John Wiley & Sons, New York

Tadikamalla, P., R. (1980). "Notes and Comments: On Simulating Nonnormal Distributions", *Psychometrika,* Vol. 45(2), pp. 273-279.

Pronósticos de Series de Tiempo y Cartas de Control para Datos Autocorrelacionados

Autocorrelación (FAC), Autocorrelación Parcial (FACP) y Correlación Cruzada (FCC)

Así como la correlación mide el alcance de una relación lineal entre dos variables, la autocorrelación mide la relación lineal entre los valores rezagados de los datos [9].

El coeficiente de autocorrelación (r_k) en el rezago k para un conjunto de datos y está dado por:

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \overline{y})(y_{t-k} - \overline{y})}{\sum_{t=1}^{T} (y_t - \overline{y})^2}$$

de donde T es la longitud de la serie de tiempo (esto es, el número de observaciones). A esto se lo conoce como la Función de Autocorrelación (FAC) [9].

El usuario puede proporcionar el máximum rezago (k_{max}); el valor por defecto es el máximo de $10\log(T)$ $3 * frecuencia\ estacional$, pero que no se exceda de $T/3$.

La autocorrelación parcial es similar a la autocorrelación excepto que se ajusta a la correlación inherente en los rezagos, ejemplos: y_t and y_{t-2} podría estar correlacionada simplemente porque ambas están conectadas a y_{t-1}, en lugar de nueva información que esté contenida en y_{t-2} [9].

La autocorrelación parcial es la autocorrelación entre y_t y y_{t+k} después de remover cualquier dependencia lineal sobre $y_{t+1}, y_{t+2}, \ldots, y_{t+k-1}$; de esta manera mide la correlación entre y_t y y_{t+k} que no es considerada por un proceso Autorregresivo AR(k-1). Cada autorrelación parcial se puede estimar como el último coeficiente en un modelo autorregresivo. Específicamente, α_k, el kth coeficiente de autocorrelación parcial, es igual a la estimación de ϕ_k en un modelo AR(k) [9].

El coeficiente de autocorrelación parcial (ϕ_k) en el rezago $k, k > 1$ está dado por:

$$\hat{y}_{t+k} = \beta_1 y_{t+k-1} + \beta_2 y_{t+k-2} + \cdots + \beta_{k-1} y_{t+1}$$
$$\hat{y}_t = \beta_1 y_{t+1} + \beta_2 y_{t+2} + \cdots + \beta_{k-1} y_{t+k-1}$$
$$\phi_k = \Phi(y_{t+k} - \hat{y}_{t+k}, y_t - \hat{y}_t)$$

de donde Φ es la función de correlación.

Si hay valores perdidos, los coeficientes de autocorrelación y de autocorrelación parcial se evalúan saltándose las correlaciones que incluyen valores ausentes y ajustando el recuento de muestras.

Bajo la hipótesis nula, los coeficientes de autocorrelación y autocorrelación parcial de la muestra se distribuyen de forma aproximadamente normal, con una media de cero, y una varianza de $1/n$; así, los intervalos de confianza pueden utilizarse para probar la hipótesis nula de no correlación. El gráfico de barras de la secuencia de autocorrelaciones con intervalos de confianza se denomina gráfico de la Función de Autocorrelación (FAC) (a veces también llamado correlograma). El gráfico de barras de la secuencia de autocorrelaciones parciales se denomina gráfico de la Función de autocorrelación parcial (FACP).

La Correlación Cruzada es el método estándar para estimar el grado de correlación de dos series. El gráfico de barras de la secuencia de correlaciones cruzadas se llama gráfico de la Función de correlación cruzada (FCC). El coeficiente de correlación cruzada $(r_{x,y})$ para los conjuntos de datos x e y viene dado por:

$$r_{x,y} = \frac{1}{n\sigma_x\sigma_y}\sum_{t=1}^{n}(x_t - \bar{x})(y_t - \bar{y})$$

de donde n es el número de observaciones, $\bar{x}$ y $\bar{y}$ son las medias, σ_x y σ_y son los errores estándares de x y y, respectivamente. El coeficiente de correlación cruzada se puede estimar entre los valores rezagados de x y y; por ejemplo, entre y_{t+k} y x_t. Los rangos del coeficiente de correlación se encuentran entre +1 (correlación positiva perfecta), y -1 (correlación negativa perfecta).

Si x está autocorrelacionada, la CCF se afecta por la estructura de la serie de tiempo y cualquier tendencia en común que las series de x y y pueden tener a lo largo del tiempo. El preblanqueo soluciona este problema al remover la autocorrelación y las tendencias. Los Datos del Preblanqueo realiza lo siguiente:

1. Se aplica el ARIMA Automático a la entrada x; se graban los residuos.
2. Se filtra los datos de y con el uso del mismo modelo y valores de los parámetros; se graban los residuos.
3. Se crea la CCF utilizando x y los residuos de y.

Para mayores detalles, véase “El Preblanqueo como una Ayuda para la Interpretación de la CCF” https://newonlinecourses.science.psu.edu/stat510/lesson/9/9.1.

La Prueba de Ljung-Box

Además de observar el gráfico de FAC, se puede también realizar una prueba formal para la autocorrelación al considerar el conjunto completo de los valores de r_k como grupo, en lugar de tratarlos a cada uno en forma separada. Este es la Prueba de Ljung-Box Q:

$$Q = T(T+2)\sum_{k=1}^{h}(T-k)^{-1}r_k^2,$$

de donde h es el máximo rezago considerado y T es el número de observaciones.

Si las autocorrelaciones vinieran de una serie de ruido blanco, entonces Q^ tendría una distribución χ^2 con $(h-k)$ Los grados de libertad, donde k es el número de parámetros en el modelo [9]. Esto también se conoce como una prueba de Portmanteau.

Transformación Box-Cox

El procedimiento Box-Cox transforma el vector x como sigue:

$$y = \frac{(x-\tau)^{\lambda}-1}{\lambda}$$

de donde $\tau < x_{min}$ es un parámetro del umbral, tal que $x-\tau$ sea siempre positivo. El objetivo de este procedimiento es transformar datos no normales en datos que se aproximen a la normal con el uso de Programación Secuencial Cuadrática (PSC) para maximizar una función de verosimilitud. La verosimilitud incluye la tendencia y las variables ficticias de estacionalidad.

Si λ es cero, entonces $y = \mathrm{Ln}\,(x)$. λ se restringe a ser >= 0 y <= 1. Si se selecciona Lambda Redondeada, las alternativas para λ son $\mathrm{Ln}\,(x)$ o $\sqrt{x}$. Se ha comprobado que este enfoque conservador genera los mejores resultados con datos rivales. El usuario puede también especificar cualquier valor de λ y el umbral.

Si se especifica un umbral inválido, se especifica un valor del umbral como $x_{min}-\delta$, de donde $\delta = \min(0.001, 0.001(x_{max}-x_{min}))$.

Si no se especifica un parámetro del umbral, y x tiene valores negativos, entonces τ se estima junto a λ; se seleccionan estos parámetros para maximizar la verosimilitud restringida con el uso de la PSC. Se emplea la restricción para asegurarse que $x-\tau$ sea siempre positiva. No se incluyen ni la tendencia ni las variables ficticias estacionales en la verosimilitud para este caso.

En ambos, ARIMA y la Suavización Exponencial, el usuario tiene la opción de invocar la transformación Box-Cox. Los valores del Pronóstico se especifican utilizando la inversa de la transformación.

Frecuencia Estacional

La Frecuencia Estacional es el número de observaciones por "ciclo" de unidad de tiempo. Algunos ejemplos incluyen:

- Datos trimestrales, frecuencia estacional = 4 observaciones por año
- Datos diarios, frecuencia estacional = 7 observaciones por semana
- Datos mensuales, frecuencia estacional = 12 observaciones por año
- Datos horarios, frecuencia estacional = 24 observaciones por día
- Datos de media hora, frecuencia estacional = 48 observaciones por día
- Datos semanales, frecuencia estacional = 52 observaciones por año
- Datos horarios, frecuencia estacional = 168 observaciones por semana
- Datos de media hora, frecuencia estacional = 336 observaciones por semana
- Datos diarios, frecuencia estacional = 365 observaciones por año (excluye años bisiesto).

La descomposición estacional debe utilizarse para frecuencias específicas superiores a 24.

SigmaXL sigue la convención utilizada en los objetos de series de tiempo R (ts) y el paquete de pronóstico R [12], pero tener en cuenta que es lo contrario de la definición de frecuencia en la física, o en el análisis de ingeniería de Fourier, donde el "período" es la longitud del ciclo, y la "frecuencia" es el inverso del período (véase https://robjhyndman.com/hyndsight/seasonal-periods/).

Descomposición de Tendencia Estacional

Se puede descomponer una serie temporal en tres componentes:

$$y_t = T_t + S_t + I_t$$

donde T_t es el componente de tendencia y cíclico, S_t es el componente estacional, e I_t es el residuo, el componente irregular o de ruido. TEL es un método versátil y robusto para descomponer las series temporales en estos tres componentes. TEL es un acrónimo de "Descomposición de Tendencia Estacional utilizando Loess". Loess es un suavizador de regresión ponderado localmente. El método TEL fue desarrollado por Cleveland, et al. [3]. Nótese que es estrictamente aditivo, pero uno puede simplemente especificar una Transformación Box-Cox con Lambda = 0 (Ln) para obtener un equivalente a la descomposición multiplicativa. A diferencia del uso de Box-Cox en el pronóstico, los resultados permanecen transformados, es decir, la transformación inversa no se aplica a los datos descompuestos - el modelo aditivo se mantiene.

Si no se especifica la frecuencia estacional, la tendencia se calcula utilizando el "Super Smoother" de Friedman [4], que utiliza la regresión lineal local con anchos de banda adaptables. La idea básica es estimar primero un número de anchos de banda fijos suavizados por medio de la regresión lineal local. Los residuos de la validación cruzada dejando uno afuera de cada una de esas estimaciones iniciales se suavizan entonces utilizando un ancho de banda constante. Se selecciona un conjunto

de anchos de banda suavizados y los resultados suavizados se interpolan linealmente y se suavizan con un ancho de banda fijo.

La Descomposición Estacional Múltiple (DEM) utiliza el TEL de forma iterativa y fue desarrollada por Hyndman, et al. [12].

Densidad Espectral y Detección Automática de la Frecuencia Estacional

La Densidad Espectral determina la frecuencia estacional de una serie temporal con el uso de análisis espectral mediante la Transformada Rápida de Fourier, utilizando una ventana uniforme específica (véase https://en.wikipedia.org/wiki/Window_function).

Se aplica automáticamente la prueba de ruido blanco de Fisher Kappa (véase Fuller [5], Capítulo 7) y, si es significativa (valor-p < 0.05), retorne la frecuencia redondeada con el mayor valor de la potencia espectral. Si no es significativa, se retorna la frecuencia unitaria (no estacional).

Una prueba de potencia estacional basada en la descomposición se aplica a la frecuencia redondeada (https://otexts.com/fpp2/seasonal-strength.html) y si el estadístico es >= 0.5, se emplea el valor de la frecuencia de arriba, de lo contrario se lo fija en 1. Esto sirve para distinguir entre series temporales cíclicas y estacionales.

En el caso de la opción Gráfica de Densidad Espectral o DEM, si la frecuencia estacional es > 1, el proceso anterior se repite dos veces más con la frecuencia dominante eliminada por la descomposición estacional, el análisis espectral aplicado al residuo y la prueba de potencia estacional aplicada a la nueva frecuencia dominante.

Interpolación Lineal de Valores Ausentes Ajustados Estacionalmente

La interpolación lineal se lleva a cabo estimando primero los valores ausentes mediante un modelo de regresión con tendencia y codificación de variables ficticias para las estaciones. Las series temporales se descomponen en tres componentes -tendencia, estacional y el resto- utilizando la descomposición STL para los datos estacionales. Se estima un punto de datos mediante interpolación lineal de los componentes no estacionales de los valores anteriores y posteriores, y luego se añade el componente estacional para la observación en cuestión [12, 17]. Para los datos no estacionales se utiliza la interpolación lineal.

Una vez que se ha especificado un modelo, los valores ausentes se estiman con el uso de un pronóstico de un paso hacia adelante.

Criterio de Información para la Comparación del Modelo

Los parámetros se estiman maximizando la función de Log-Verosimilitud (que es similar a la minimización de la suma de residuos al cuadrado). Dado un conjunto de modelos candidatos para los datos, el modelo preferido es aquel con el valor mínimo del Criterio de Información. El Criterio de Información premia la bondad del ajuste (evaluada por la función de verosimilitud), pero también incluye una penalización la cual es una función creciente del número de parámetros estimados. La penalización desalienta el exceso del ajuste, porque el aumento del número de parámetros en el modelo casi siempre mejora la bondad del ajuste

(véase https://en.wikipedia.org/wiki/Akaike_information_criterion).

El Criterio de Información de Akaike (AIC) es:

$$\text{AIC} = -2\text{Ln}\,(L) + 2(k+1)$$

donde L es la verosimilitud del modelo y k es el número total de parámetros (y los estados iniciales para la suavización exponencial) que se han estimado.

El AIC corregido para el sesgo de muestra pequeña (AICc) se define como:

$$\text{AIC}_\text{c} = \text{AIC} + \frac{2(k+1)(k+2)}{T-k-2}$$

El Criterio de Información Bayesiano (BIC) is:

$$\text{BIC} = \text{AIC} + (k+1)[\text{Ln}\,(T) - 2]$$

Se recomienda el AICc como Criterio de Información por defecto, basado en el performance de los errores de pronóstico con los datos de la competencia M3 [12].

Notas sobre el AIC de Rob Hyndman (véase https://robjhyndman.com/hyndsight/aic/):

- Se puede comparar modelos muy diferentes con el AIC, sin embargo, asegurarse que las probabilidades se calculan con los mismos datos. Por ejemplo, no se puede comparar un modelo ARIMA con diferenciación con un modelo ARIMA sin diferenciación, porque se pierde una o más observaciones por la diferenciación.

- Por una razón similar, no se puede comparar el AIC de un modelo de Suavización Exponencial con el AIC de un modelo ARIMA. Los cálculos de la función de verosimilitud son diferentes y los dos modelos tratan los valores iniciales de manera diferente. Por ejemplo, después de diferenciar, un modelo ARIMA se calcula con menos observaciones, mientras que un modelo de Suavización Exponencial se calcula siempre con el conjunto completo de

datos. Inclusive cuando los modelos son equivalentes (por ejemplo, un ARIMA (0,1,1) y un ETS (A, N, N), los valores del AIC serán diferentes

Precisión del Pronóstico

Como se ha señalado anteriormente, no se puede utilizar el criterio de información AIC/AICc para comparar los modelos ARIMA y ETS o los modelos ARIMA con diferentes valores d, D, pero sí se pueden comparar las métricas de precisión de los pronósticos. Un error de pronóstico e_t es la diferencia entre un valor observado y su pronóstico. No es lo mismo que un residuo del modelo porque los residuos pueden estar en diferentes unidades debido a la transformación de Box-Cox o a un modelo multiplicativo.

Los tipos de errores del pronóstico incluyen:

- Pronóstico Dentro-de-la-Muestra Un-Paso-Hacia-Adelante. Menos útil porque el modelo puede ser sobre ajustado.
- Fuera-de-la-Muestra (Retención) Un-Paso-Hacia-Adelante. Las estimaciones de los parámetros del modelo no usan datos con retención, pero los pronósticos se actualizan con cada nueva observación con retención.
- Fuera-de-la-Muestra (Retención) Multi-Pasos-Hacia-Adelante. Importante si se evalúa la precisión de los pronósticos sobre un horizonte. Se utiliza con pronósticos rivales.

SigmaXL hace uso de las siguientes métricas de precisión del pronóstico:

Raíz cuadrada del error cuadrático medio: las siglas en inglés, RMSE $= \sqrt{\text{media}(e_t^2)}$

Error medio absoluto: MAE $= \text{media}(|e_t|)$

Error porcentual medio absoluto: MAPE $= \text{media}\left(\left|\frac{100e_t}{y_t}\right|\right)$

Error en escala medio absoluto: MASE $= \text{media}(|e_t|)/escala$

donde la *escala* es el MAE del pronóstico naïve de la muestra o del pronóstico estacional naïve (establezca todos los pronósticos en el valor de la última observación/período). Un error de escala es menor que uno si surge de un pronóstico mejor que el promedio de los pronósticos naïve o estacionales naïve. Por el contrario, es mayor que uno si el pronóstico es peor que el pronóstico naïve promedio [9].

Suavización exponencial - ETS

Suavización exponencial simple:

Los pronósticos de la suavización exponencial simple se calculan utilizando promedios ponderados, donde los pesos decrecen exponencialmente mientras las observaciones provienen del pasado de tal manera que los pesos más pequeños se asocian con las observaciones más distantes:

$$\hat{y}_{t+1} = \alpha\ y_t + \alpha(1-\alpha)\ y_{t-1} + \alpha(1-\alpha)^2\ y_{t-2} + \cdots$$

donde $0 \leq \alpha \leq 1$ es el parámetro de suavización.

Una formulación equivalente de la suavización exponencial simple es:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t$$

con el valor inicial del pronóstico (nivel inicial) $\hat{y}_1$ a estimarse y se denota como el nivel l_0.

Hyndman proporciona una representación alternativa conocida como la forma componente [9]. Para la suavización exponencial simple, el único componente incluido es el nivel ℓ_t. Otros métodos también pueden incluir una tendencia b_t y/o un componente estacional s_t.

Las representaciones en forma de componente de los métodos de suavización exponencial comprenden una ecuación de pronóstico y una ecuación de suavización para cada uno de los componentes incluidos en el método. La forma componente de la suavización exponencial simple viene dada por:

$$\begin{array}{lll} \text{Ecuación del pronóstico} & \hat{y}_{t+h|t} & = \ell_t \\ \text{Ecuación de suavización} & \ell_t & = \alpha y_t + (1-\alpha)\ell_{t-1}, \end{array}$$

donde ℓ_t es el nivel (o el valor suavizado) de la serie en el tiempo t. $\hat{y}_{t+h|t}$ denota un pronóstico de h multi-pasos, tomando en consideración todas las observaciones hasta el tiempo t. Al fijar $h = 1$, un pronóstico de un paso hacia adelante proporciona el ajuste de los valores dentro de la muestra.

La forma de la "corrección del error" es:

$$\begin{aligned} \ell_t &= \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \\ &= \ell_{t-1} + \alpha e_t \end{aligned}$$

donde $e_t = y_t - \ell_{t-1} = y_t - \hat{y}_{t|t-1}$ es el residuo en el tiempo t. Las ecuaciones del modelo de error aditivo se pueden expresar como:

$$\begin{aligned} y_t &= \ell_{t-1} + \varepsilon_t \\ \ell_t &= \ell_{t-1} + \alpha\varepsilon_t. \end{aligned}$$

Los errores Multiplicativos son errores relativos de un-paso-hacia-adelante:

$$\varepsilon_t = \frac{y_t - \hat{y}_{t|t-1}}{\hat{y}_{t|t-1}}$$

Se pueden expresar las ecuaciones del error del modelo multiplicativo como sigue:

$$\begin{aligned} y_t &= \ell_{t-1}(1+\varepsilon_t) \\ \ell_t &= \ell_{t-1}(1+\alpha\varepsilon_t). \end{aligned}$$

Modelos de Error, Tendencia, Estacional (ETS):

Los modelos de error, tendencia, estacional (ETS) se expanden en una simple suavización exponencial para acomodar los componentes de tendencia y estacionales, así como los errores aditivos o multiplicativos. La Suavización Exponencial Simple es un modelo de error. El modelo de error, de tendencia, es el lineal de Holt, también conocido como suavización exponencial doble. El Modelo de Error, de Tendencia Estacional, es el Holt-Winters, también conocido como suavización exponencial triple.

Rob Hyndman ha desarrollado una taxonomía completa que describe todas las combinaciones de los modelos de suavización exponencial de forma coherente (véase fpp2):

- Error:
  - Aditivo o Multiplicativo. El error multiplicativo es un error relativo: $e_t = \frac{y_t - \hat{y}_t}{\hat{y}_t}$
  - Los pronósticos puntuales desarrollados por los modelos son idénticos si utilizan los mismos valores de los parámetros de la suavización. Sin embargo, el multiplicativo generará diferentes intervalos de pronósticos para acomodar el cambio en la varianza.
  - Una alternativa al multiplicativo es el uso de la transformación Ln (transformación Box-Cox con Lambda = 0).
  - Los modelos de error incluyen el parámetro de suavización α y el valor del nivel inicial.

- Tendencia:
  - Ninguna, Aditiva, Aditiva Amortiguada
  - No se recomienda la tendencia multiplicativa ya que tiende a producir malos pronósticos, por lo que no se incluye en SigmaXL
  - Los modelos de tendencia agregan un parámetro de suavización β y un valor de tendencia inicial.

- Los modelos de tendencia amortiguados agregan un parámetro de suavización φ que "amortigua" la tendencia a una línea plana en algún momento del futuro.

- Estacional:
  - Ninguno, Aditivo, Multiplicativo
    - Se prefiere el método aditivo cuando las variaciones estacionales son aproximadamente constantes a lo largo de la serie, en cambio es preferible el método multiplicativo cuando las variaciones estacionales cambian de manera proporcional con respecto al nivel de la serie.

  - Los modelos estacionales agregan un parámetro de suavización *γ* y valores estacionales iniciales.
    - # de valores iniciales estimados = frecuencia estacional – 1
    - limitado a la suma a 0 para la frecuencia aditiva o estacional *m* para la multiplicativa

Resumen de los Modelos ETS en SigmaXL:

Abreviado (Error, Tendencia, Estacional)	Método
(A, N, N)	Suavización Exponencial Simple con Errores Aditivos – Promedio Móvil Ponderado Exponencial (EWMA)
(M, N, N)	Suavización Exponencial Simple con Errores Multiplicativos
(A, A, N)	Método de Tendencia Aditiva con Errores Aditivos (Holt Lineal)
(M, A, N)	Método de Tendencia Aditiva con Errores Multiplicativos (Holt Lineal)
(A, A, A)	Método de Tendencia Aditiva, Estacional Aditivo con Errores Aditivos (Holt-Winters)
(M, A, A)	Método de Tendencia Aditiva, Estacional Aditivo con Errores Multiplicativos (Holt-Winters)
(A, N, A)	Método Estacional Aditivo con Errores Aditivos
(M, N, A)	Método Estacional Aditivo con Errores Multiplicativos
(A, Ad, N)	Método de Tendencia Atenuada con Errores Aditivos
(M, Ad, N)	Método de Tendencia Atenuada con Errores Multiplicativos
(A, Ad, A)	Método de Tendencia Atenuada Aditiva Estacional Aditiva con Errores Aditivos
(M, Ad, A)	Método de Tendencia Atenuada Aditiva Estacional Aditiva con Errores Multiplicativos
(M, A, M)	Método de Tendencia Aditiva Estacional Multiplicativo con Errores Multiplicativos (Holt-Winters)
(M, N, M)	Método Estacional Multiplicativo con Errores Multiplicativos
(M, Ad, M)	Método de Tendencia Atenuada Aditiva Estacional Multiplicativo con Errores Multiplicativos (Holt-Winters)

Ecuaciones del Modelo ETS:

Las siguientes son las ecuaciones para cada uno de los modelos (conocidos también como “modelo de innovaciones de espacio de estado”) [9]:

ADDITIVE ERROR MODELS

Trend	Seasonal		
	N	**A**	**M**
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
$\mathbf{A_d}$	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal		
	N	**A**	**M**
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
$\mathbf{A_d}$	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

Selección Automática del Modelo (Identificación):

El proceso de identificación se utiliza para determinar qué modelo de ETS se debe utilizar. El mejor modelo se determina mediante la estimación de todas las combinaciones posibles de error, tendencia, estacionalidad y atenuación. Se recomienda el AICc como criterio de información por defecto, basado en el rendimiento del error pronosticado con los datos de la competencia M3.

Algunas de las combinaciones de modelos conducen a la inestabilidad numérica y no se consideran en el proceso de selección: (A, N, M) (A, A, M) (A, Ad, M).

Si se utiliza una transformación Box-Cox, no se consideran los modelos multiplicativos.

Estimación:
Dada la existencia de un modelo específico de ETS, los valores de los pesos (α para el error, β para la tendencia, γ para la estación, y φ para la atenuación) se estiman utilizando la máxima probabilidad exacta restringida en el espacio de estado que hace uso de la recursión, empleando el algoritmo simplex de Nelder Mead. En efecto, el objetivo es minimizar la suma residual de los cuadrados:

$$SSE = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2$$

La optimización devuelve los pesos (α para el error, β para la tendencia, γ para la estacionalidad y φ para la atenuación), así como los valores del estado inicial. Hay una serie de restricciones en los pesos (además de los pesos que se encuentran cada uno entre 0 y 1) que se requieren para la estabilidad [8, 10]. Si se encuentran valores ausentes, se utiliza la interpolación lineal ajustada estacionalmente. Se permite una transformación Box-Cox antes de la estimación, aunque dicha transformación sólo debe utilizarse para un modelo ETS aditivo.

Las covariables no se pueden estimar como parte de una estimación integrada del ETS (véase https://robjhyndman.com/hyndsight/ets-regressors/ para una discusión sobre este tema).

Los valores iniciales se toman de las estimaciones de OLS de una regresión de y contra una constante y las covariables escaladas.

Pronóstico:
Una vez que se especifica un modelo, los valores futuros de la serie pueden ser evaluados fácilmente. Dados los datos $y_1 \ldots y_3$, entonces la media condicional de $y_{(n+1)}$ se deriva usando el pronóstico puntual. El proceso se repite dinámicamente. Se utiliza un valor inicial de 0,5 para α, 0,05 para β, 0,025 para γ y 0,978 para φ. Para las covariables, los valores iniciales se toman de las estimaciones de OLS de una regresión de y contra una constante y las covariables escaladas.

Véase https://otexts.com/fpp2/ets-forecasting.html para una breve discusión sobre los intervalos de predicción del modelo ETS. Los detalles completos de la fórmula son muy complejos y se encuentran en Hyndman et al. [11].

Promedio Móvil Autorregresivo Integrado - ARIMA

El modelo Promedio Móvil Autorregresivo Integrado (ARIMA) fue desarrollado por Box y Jenkins [2]. La determinación automática del orden del mejor modelo con SigmaXL utiliza el método de Rob Hyndman [9,10,12].

Estacionariedad y Diferencias:

El ARIMA asume que la serie de tiempo es estacionaria, es decir, tiene la propiedad de que la media, la varianza y la estructura de autocorrelación no cambian con el tiempo. Si la media de una serie temporal no es estacionaria (por ejemplo, la tendencia), esto puede corregirse mediante la diferenciación, calculando las diferencias entre las observaciones consecutivas para los datos no estacionales y entre los períodos consecutivos para los datos estacionales (por ejemplo, enero de 2019 - enero de 2018, etc.). En el caso de los datos no estacionales, esto implicará típicamente 1 o 2 órdenes de diferenciación. Este orden es el término integrado d. Para los datos estacionales, esto implicará típicamente 1 orden de diferenciación. Este orden es el término integrado estacional D. Un término constante c es opcional:

- Si d+D=0, un término constante en el modelo es la media.
- Si d+D=1, un término constante en el modelo es una tendencia (deriva).
- Si d+D>1, un término constante sería una tendencia cuadrática o superior, por lo que la constante no debe incluirse.
- Se recomienda que d+D no sea > 3.
- Si la varianza no es estacionaria, utilice una transformación Box-Cox.

Modelo autorregresivo (AR):

En un modelo autorregresivo, se pronostica la variable de interés utilizando una combinación lineal de valores de la variable que ocurrieron en el pasado. El término autorregresivo indica que es una regresión de la variable contra sí misma:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

donde ε_t es ruido blanco. Esto es casi como una regresión múltiple, pero con valores de rezago de y_t como predictoras. Esto es a lo que se refiere a un modelo AR(p), un modelo autorregresivo de orden p (fpp2).

Modelo de Promedio Móvil (MA):

En lugar de utilizar en una regresión los valores del pasado de una variable de pronóstico, un modelo de promedio móvil utiliza errores de pronósticos del pasado en un modelo similar a una regresión:

$$y_t = c + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q}$$

Esto es a lo que se refiere a un modelo MA(q), un modelo de promedio móvil de orden q (fpp2).

Modelo Promedio Móvil Autorregresivo Integrado (ARIMA):

Si se combina las diferencias con un modelo de auto regresión con uno de promedio móvil, se obtiene un modelo ARIMA no estacional:

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} + \varepsilon_t$$

de donde y'_t es la serie en diferencias. Este es un modelo ARIMA (p, d, q) , de donde p es el número de términos autorregresivos, d es el grado de diferenciación y q es el número de términos del promedio móvil.

$$\underset{\uparrow}{(1 - \phi_1 B - \cdots - \phi_p B^p)} \quad \underset{\uparrow}{(1 - B)^d y_t} \quad = \quad \underset{\uparrow}{c + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t}$$

AR(p) — d differences — MA(q)

ARIMA Estacional:

Para el caso estacional, el modelo consiste en términos similares a los componentes no estacionales del modelo, pero que implican retrocesos del período estacional. El modelo estacional es ARIMA (P, D, Q), y el combinado es ARIMA $(p, d, q)\ (P, D, Q)$.

Por ejemplo, un modelo ARIMA(1,1,1)(1,1,1) para datos trimestrales (frecuencia estacional = 4), se puede escribir como [9]:

$$(1 - \phi_1 B)\ (1 - \Phi_1 B^4)(1 - B)(1 - B^4)y_t = (1 + \theta_1 B)\ (1 + \Theta_1 B^4)\varepsilon_t.$$

Selección Automática del Modelo (Identificación):

El proceso de identificación es necesario para determinar el grado de diferenciación (d, D) requerido para generar una serie estacionaria, y para luego determinar el orden del modelo (p, q, P, Q). Esto puede hacerse manualmente utilizando el correlograma (ACF) y el autocorrelograma parcial (PACF) tanto antes como después de la diferenciación. Este proceso está automatizado en SigmaXL.

En primer lugar, el grado de diferenciación estacional (D) se determina con la prueba de Potencia Estacional (SEAS) [12, 19] que utiliza la razón entre el componente estacional y el componente sin tendencia y lo compara con el valor crítico del SEAS. Si la relación excede el valor crítico, esto implica el rechazo de la hipótesis nula sin raíz unitaria, lo que implica $D = 1$, y los datos se diferencian estacionalmente.

El grado de diferenciación para el componente no estacional (d) se determina entonces utilizando la prueba de la raíz unitaria de KPSS [12, 13]. Los datos se prueban para una raíz unitaria, y si se encuentra tal, los datos se diferencian, y el proceso continúa; se detiene cuando no se encuentra una raíz unitaria significativa o d=2. SigmaXL utiliza el parámetro de rezago de truncamiento de Schwert = floor(12*(n/100) ^.25) para n≤70 y el Kernel Espectral Cuadrático para n>70 [21, 22]. Si d=0 y n≤ 500, se aplica una prueba adicional que compara la desviación estándar del (0,0,0) con el modelo de la constante con la desviación estándar de (0,1,0) con la constante. Si d=1 produce un valor de desviación estándar inferior, entonces se utiliza como orden para la desviación no estacional (véase https://people.duke.edu/~rnau/411arim2.htm). Aunque existe un riesgo potencial de sobre diferenciación, esta prueba adicional da como resultado una mejora drástica en la precisión del pronóstico utilizando datos de competencia M3 y M4. Si se piensa que el proceso es estacionario y SigmaXL devuelve $d = 1$, el usuario debe especificar manualmente un modelo $d = 0$y comparar los dos modelos.

Existen otras pruebas para la estacionariedad estacional y no estacional (raíz-unitaria), pero Hyndman recomienda SEAS y KPSS en base a la evaluación de la precisión del pronóstico con los datos de la competencia M3 (véase https://robjhyndman.com/hyndsight/forecast83/ y https://robjhyndman.com/hyndsight/unit-root-tests/). Actualmente, SigmaXL no incluye pruebas alternativas de raíz unitaria/estacionariedad. Los valores estadísticos de las pruebas SEAS y KPSS se utilizan internamente pero no se reportan.

Una vez comprobado el grado de diferenciación estacional y no estacional, la selección del modelo implica la elección de los componentes AR y MA - p, q, P, Q - así como si una constante debe o no formar parte del modelo. Un método para elegir el mejor modelo consiste en estimar exhaustivamente todas las combinaciones posibles, sujetos a los límites superiores de los términos AR/MA. Un enfoque más eficiente es estimar una serie de modelos ARIMA utilizando un orden escalonado sugerido por Hyndman y Khandakar [10]. El modelo preferido es el que tiene la métrica de ajuste más baja, siendo el predeterminado el AICc. Sólo se consideran los modelos que son estacionarios e invertibles.

Estimación:

Dada la existencia de un modelo específico de ARIMA (p, d, q, P, D, Q), se estiman los valores de los términos AR y MA. Se permiten las covariables (que se escalan automáticamente), así como una constante si no hay diferencias, y una tendencia si d y D es igual a { 0 1 } o { 1 0 }. Hay dos modelos generales disponibles - la suma condicional de cuadrados (CSS) y el espacio de estado Kalman de máxima probabilidad. La probabilidad para el modelo CSS es:

$$logL = \sum_i \left[\log(2\pi)/2 - 0.5\log(\sigma^2) - e_i^2/2\sigma^2\right]$$

El CSS siempre se utiliza para las estimaciones iniciales y se usa si $n > 500$ o la frecuencia >12. Los residuos utilizados se derivan de las series temporales diferenciadas - para el CSS, se utiliza el filtrado AR/MA para generar los residuos, mientras que para Kalman, que es un estimador recursivo, los residuos se derivan de un proceso AR/MA de máxima probabilidad. En cada caso, el proceso de máxima verosimilitud utiliza el algoritmo restringido simplex de Nelder Mead. Las restricciones imponen un valor máximo de unidad a los términos AR/MA mediante el uso de una transformación tanh, así como mediante una prueba de raíz unitaria para asegurar que los términos AR son estacionarios y los términos MA son invertibles.

Para el CSS, si se encuentran valores ausentes, se utiliza el mayor Rango contiguo. Para Kalman, los valores ausentes se evalúan saltándose la fase de actualización en el modelo basado en el estado.

Los límites de confianza en los coeficientes se derivan de la diagonal del inverso del Hessiano del producto exterior del gradiente (OPG).

Las covariables (predictoras) se tratan en el modelo tradicional ARIMAX añadiéndolas al lado derecho antes de diferenciarlas. Para estimar un modelo ARIMA con covariables, los términos de las covariables se restan de los datos antes de evaluar los residuos en cada iteración de la evaluación de la verosimilitud. Los parámetros de los términos de las covariables se estiman simultáneamente con todos los demás parámetros. Para la predicción y el pronóstico, se genera el proceso inverso, y los términos de las covariables se vuelven a sumar a los valores pronosticados. Así pues, el efecto de las variables exógenas es siempre aditivo, independientemente del modelo ARIMA.

Se utiliza un valor inicial de 0,1 para cada parámetro, con las siguientes excepciones: Para el término constante, se utiliza la media si no hay diferencia, y la media de los datos diferenciados, ajustada por el número de términos AR, en caso contrario. Para las covariables, los valores iniciales se toman de las estimaciones de los OLS de un modelo de ARX con m términos de AR, covariables escaladas y tendencia, donde m es la frecuencia estacional.

Las Estimaciones de los Parámetros del ARIMA incluyen pruebas de significación; los valores-P < 0.05 son significativos y están resaltados en rojo. Esto es útil para el refinamiento del modelo con múltiples predictoras. Tener en cuenta que para la selección del orden de los modelos AR/MA, se debe utilizar el mínimo AICc en lugar de las pruebas de significación [25].

Ajustado Dentro-de-la-Muestra (Un-Paso-Hacia-Adelante):

Los valores pronosticados para un proceso ARIMA se calculan con el uso del proceso de Kalman sobre los datos diferenciados y sin tendencia, y_d. Para el punto de datos iniciales, el proceso de Kalman utiliza un valor predictivo y_p de cero; los valores predictivos subsiguientes se derivan del proceso de Kalman. Los residuos calculados se derivan como la diferencia entre y_d e y_p, y los datos

predichos, que son predicciones de un paso hacia adelante, se derivan de los datos reales más el residuo calculado.

Los pronósticos para un proceso ARIMA se calculan usando el proceso de Kalman o el proceso CSS sobre los datos en diferencias y sin tendencia, y_d. Para los puntos de datos iniciales, el proceso de Kalman utiliza un valor predictivo y_p de cero; los valores predictivos subsiguientes se derivan del proceso de Kalman; estos valores iniciales se excluyen bajo el CSS. Los residuos calculados se derivan como la diferencia entre y_d y y_p, y los datos pronosticados, después de eliminar la tendencia. Estas son predicciones de un-paso-hacia-adelante, y se derivan de los datos reales más el residual calculado. Si se realiza una diferencia, los primeros valores pronosticados de d + D se establecen como faltantes.

Pronósticos:

Los coeficientes sin procesar se transforman en los vectores φ^* y θ^*, que pueden ser utilizados en las series temporales originales. Se realiza un pronóstico desde el final de la serie para el número especificados de los períodos de pronósticos. Por defecto, los pronósticos se evalúan dinámicamente, de tal manera que los valores pronosticados en cualquier punto proporcionan la base para la predicción del siguiente período. Una banda de confianza se genera para cada uno de los valores del pronóstico, que se basa en la desviación estándar del error del pronóstico para las observaciones $n + m$:

$$\sigma_{n+m} = \sqrt{\sigma_w^2 \sum_{j=0}^{m-1} \Psi_j^2}$$

donde σ_w^2 es la varianza de los residuos derivada a partir de las n observaciones, y Ψ_j es el peso de la observación $t - j$ cuando el residuo para la observación t se escribe como un modelo MA de orden infinito.

Evaluación de la Precisión de Pronósticos a través de Concursos de Pronósticos

Los concursos "M" organizados por Spyros Makridakis han tenido una enorme influencia en el área de los pronósticos. Han centraron la atención en los modelos que han generado buenos pronósticos, más que en las propiedades matemáticas de esos modelos.

El concurso "M3" ha incluido 3003 series temporales con varias frecuencias estacionales y áreas de aplicación. Los datos de M3 se han seguido utilizándose desde el 2000 para probar nuevos métodos de pronósticos de series temporales. La reciente competencia M4 ha incluido 100.000 series temporales (Véase https://robjhyndman.com/hyndsight/m4comp/ y https://www.mcompetitions.unic.ac.cy/).

Con el uso de un promedio híbrido de suavización exponencial automática y ARIMA, SigmaXL (extraoficialmente) se ha ubicado en el décimo lugar de 60 en el puntaje Promedio Ponderado

Global M4, por encima de tres conocidos paquetes comerciales de software de pronósticos.

Cartas de Control para Datos Autocorrelacionados

El control estadístico de procesos autocorrelacionados utiliza típicamente el modelo de pronóstico de un paso hacia adelante EWMA (Promedio Móvil Ponderado Exponencialmente) [1, 2, 7, 14, 15]. La fórmula utilizada para el EWM es la misma que la de la suavización exponencial simple, pero el parámetro de suavización λ es por lo general usado en lugar de α y X_t en lugar de y_t:

$$\text{EWMA}_{t+1} = \lambda X_t + (1-\lambda)\text{EWMA}_t$$

con el valor de pronóstico inicial EWMA_1 estimado como la media de los datos o el valor objetivo. En el caso de una carta de control EWMA con datos independientes, el parámetro de suavización λ está determinado por las características de la longitud media de recorrido deseada y es típicamente 0,2. En el caso de la CEP para datos autocorrelacionados, el parámetro de suavización y el nivel inicial se determinan reduciendo al mínimo los errores del pronóstico de la suma de cuadrados (residuos):

$$\text{SSE} = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T} e_t^2.$$

Utilizando métodos de minimización no lineal como el de Newton-Raphson o el Simplex de Nelder-Mead.

El modelo de series de tiempo pronostica el movimiento en la media y se traza una carta de control de los residuos para detectar las causas asignables. Si no se tiene en cuenta la autocorrelación, se producirán límites demasiado estrechos que resultarán en un exceso de falsas alarmas, o límites demasiado amplios que resultarán en fallos.

Woodall y Faltin proporcionan algunas pautas útiles para tratar la autocorrelación [20]:

- Si es posible, primero se debe intentar eliminar la fuente de la autocorrelación.
- Si no se puede eliminar directamente la fuente de la autocorrelación, entonces puede ser posible modelar la autocorrelación y utilizar un esquema de control de retroalimentación para reducir la variabilidad sobre un valor objetivo especificado.
- Si la fuente de la autocorrelación no puede eliminarse directamente, y el control de retroalimentación no es una opción viable, entonces es importante vigilar el proceso con cartas de control que no den señales repetidas debido a la presencia de la autocorrelación.

SigmaXL proporciona una opción para crear una carta de control de Individuales y de Límites Móviles (con un Pronóstico de Un-Paso-Hacia-Adelante) para los residuos de cualquiera de los modelos de pronóstico, lo que incluye:

- Suavización exponencial
- Descomposición Estacional Múltiple con Suavización exponencial
- ARIMA
- ARIMA con Predictoras
- Descomposición estacional múltiple con ARIMA

La carta de Límites Móviles de Residuos utiliza la predicción de un paso como la línea central. Si se hace uso de una transformación Box-Cox, se aplica una transformación inversa para calcular los límites de control. Si los residuos son de suavización exponencial multiplicativa, los límites de control son aproximados y las señales fuera de control pueden no coincidir exactamente con la Carta de Individuales. Si eso ocurre, se debe utilizar la Carta de Individuos para determinar qué puntos están fuera de control.

No se recomienda el uso de pruebas de Causas Especiales con estas cartas de control. Mientras que los residuos sean independientes, cualquier cambio sostenido en la media del proceso se ajusta rápidamente en el modelo de series de tiempo, por lo que las reglas de ejecución adicionales no funcionarían como se han diseñado para detectar pequeños cambios.

La distinción de una causa asignable como un valor atípico o como un desplazamiento sostenido puede hacerse añadiendo predictoras codificadas en un modelo ARIMA y se comparan los valores-p del coeficiente de regresión. Un posible valor atípico se designaría con un "1" y todas las demás observaciones con un "0". Para probar un desplazamiento, todos los valores hasta la observación de interés serían un "0", y los restantes se designan como un "1" [23].

La Carta de Rangos Móviles de Residuos no se incluye dado que tiene demasiadas falsas alarmas [24, pp. 114-115].

Si se desean pruebas para Causas Especiales o una Carta de Rangos Móviles, simplemente cree una Carta de Individuales y Rangos Móviles en los Residuos del Pronóstico. Tener en cuenta que la función "Agregar Datos" con actualización automática de los residuos no es compatible en este caso.

Referencias del Pronóstico de Series Temporales y Cartas de Control para Datos Autocorrelacionados

[1] Alwan, L.C., and Roberts, H.V. (1988), "Time Series Modeling for Statistical Process Control", *Journal of Business and Economic Statistics*, 6, 87-95.

[2] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2016). *Time Series Analysis, Forecasting and Control*, 5th edition, Wiley.

[3] Cleveland, R.B., Cleveland, W.S., McRae, J.E., and Terpenning, I. (1990) "STL: A Seasonal-Trend Decomposition Procedure Based on Loess", *Journal of Official Statistics*, Vol. 6(1), pp. 3-73.

[4] Friedman, J. H. (1984). "A variable span smoother. Laboratory for Computational Statistics", *Technical Report (5)*, Department of Statistics, Stanford University.

[5] Fuller, W.A. (1996). *Introduction to Statistical Time Series*, 2nd ed. John Wiley and Sons, New York.

[6] Gardner, G, Harvey, A.C. and Phillips, G.D.A. (1980). "Algorithm AS154. An algorithm for exact maximum likelihood estimation of autoregressive - moving average models by means of Kalman filtering". *Applied Statistics*, Vol. 29, pp. 311 - 322.

[7] Hunter, J.S. (1986), "The Exponentially Weighted Moving Average," *Journal of Quality Technology*, 18, 203-210.

[8] Hyndman, R.J., M. Akram, and B.C. Archibald, (2006). "The admissible parameter space for suavización exponencial models." *Annals of the Institute of Statistical Mathematics* Vol. 60. pp. 407-426.

[9] Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*, second edition, OTexts: Melbourne, Australia. OTexts.com/fpp2.

[10] Hyndman, R.J. and Y. Khandakar, (2008). "Automatic Time Series Forecasting: The forecast Package for R." *Journal of Statistical Software*, Vol. 27(3), pp. 1-22.

[11] Hyndman, R.J., A.B. Koehler, J. Keith Ord and R.D. Snyder (2008). *Forecasting with Suavización exponencial. The State Space Approach* Springer-Verlag, Berlin Heidelberg.

[12] Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2019). _forecast: Forecasting functions for time series and linear models_. R package version 8.9.

[13] Kwiatowski, D., Phillips, P.C.B, Schmid, P. and Shin, T. (1992). "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic series have a unit root", *Journal of Econometrics*, Vol. 54, pp. 159-178.

[14] Montgomery, D. C., and Mastrengelo, C.M. (1991), "Some Statistical Process Control Methods for Autocorrelated Data", *Journal of Quality Technology*, 23, 179-204.

[15] Montgomery, D. C. (2013). *Introduction to Statistical Quality Control*, 7th edition, Wiley.

[16] Montgomery, D. C., Jennings, C.L., and Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting*, second edition, Wiley.

[17] Moritz, S., Sarda, A., et al. (2015). "Comparison of different Methods for Univariate Time Series Imputation in R", https://arxiv.org/ftp/arxiv/papers/1510/1510.03924.pdf.

[18] *NIST/SEMATECH e-Handbook of Statistical Methods*,

https://www.itl.nist.gov/div898/handbook.

[19] Wang, X., Smith, K. and Hyndman, R. (2006) “Characteristic-Based Clustering for Time Series Data", *Data Mining and Knowledge Discovery*, Vol. 13, pp. 335-364.

[20] Woodall, W.H. and Faltin, F.W. “Autocorrelated Data and SPC”, *ASQC Statistics Division Newsletter*, 13(4).

[21] Hobijn, B., P. H. Franses, and M. Ooms. 1998. “Generalizations of the KPSS-test for stationarity”, *Econometric Institute Report 9802/A*, Econometric Institute, Erasmus University Rotterdam.

[22] Schwert, “G.W. Tests for Unit Roots: A Monte Carlo Investigation”, *Journal of Business and Economic Statistics*, 7, 1989, 147-160.

[23] The ARIMA Procedure – Detecting Outliers, SAS/ETS(R) 9.3 User's Guide, http://support.sas.com/documentation/cdl/en/etsug/63939/HTML/default/viewer.htm#etsug_arima_sect045.htm.

[24] Fuentes, Jesús Cuéllar (2008), “Statistical Monitoring of a Process with Autocorrelated Output and Observable Autocorrelated Measurement Error” (Doctoral Dissertation), Baylor University. https://baylor-ir.tdl.org/handle/2104/5185.

[25] Kostenko, A.V. and Hyndman, R.J. (2008), “Forecasting without significance tests”, https://robjhyndman.com/papers/sst2.pdf

Acerca de SigmaXL, Inc.

Se estableció en 1998, SigmaXL Inc. es un proveedor líder de Complementos de Excel de uso amigable de herramientas gráficas y estadísticas para Lean Six Sigma y simulación Monte Carlo.

Nuestro producto insignia, SigmaXL®, se diseñó desde cero como una herramienta costo-efectiva, potente y de fácil uso que permite a los usuarios medir, analizar, mejorar, y controlar sus servicios, procesos transaccionales y de manufactura. Como un complemento del conocido Microsoft Excel, SigmaXL® es ideal para el entrenamiento y aplicación de Lean Six Sigma, o para el uso en cursos de estadísticas en la universidad.

DiscoverSim™ permite cuantificar sus riesgos a través de la simulación Monte Carlo y minimizar su riesgo con optimización global. Las decisiones en los Negocios se toman con frecuencia bajo los supuestos del valor de una estimación puntual o un promedio, lo que genera resultados inesperados. DiscoverSim™ permite modelar la incertidumbre de sus entradas de tal manera que pueda conocer qué esperar con respecto a sus salidas.

Información de Contacto:

Soporte Técnico: 1-866-475-2124 (Llamada Sin Costo en América del Norte) o

1-519-579-5877

Ventas: 1-888-SigmaXL (888-744-6295)

E-mail: support@SigmaXL.com Web: www.SigmaXL.com