

## Steps to Using SAS:

### A. GETTING STARTED:

1. Visit SAS University Edition Website: [http://www.sas.com/en\\_us/software/university-edition/download-software.html](http://www.sas.com/en_us/software/university-edition/download-software.html)
2. Determine which Operating System you are using: **Windows, OS X, or LINUX**. Choose the respective tab.



### Before You Begin

3. Install Oracle VirtualBox virtualization software on your machine



**Note: Once installed, be sure to choose “Linux Red Hat 64-bit” as your operating system. Do not choose your actual operating system, as choosing anything other than “Linux Red Hat 64-bit” may cause an execution error.**

4. Download the SAS University Edition vApp.

---



## Download the SAS® University Edition vApp.

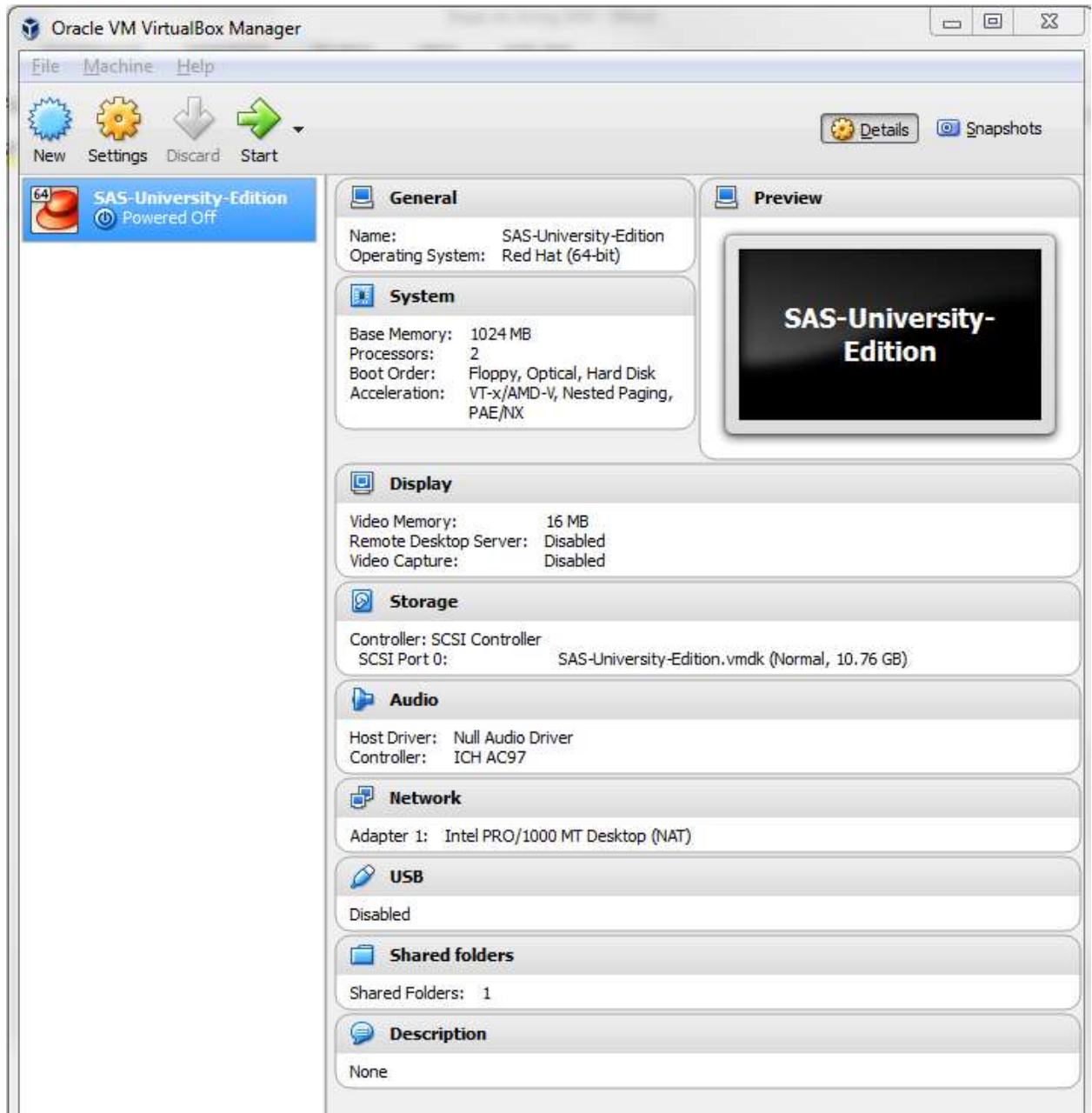
Once you click the download button below, you'll be prompted to:

- 1 Create a SAS profile if you don't already have one. If you already have a SAS profile, log in.
- 2 Accept the user licensing agreement.
- 3 Begin the download. If your browser asks whether you want to save or open the file, click **Save** to save the file in your Downloads directory.

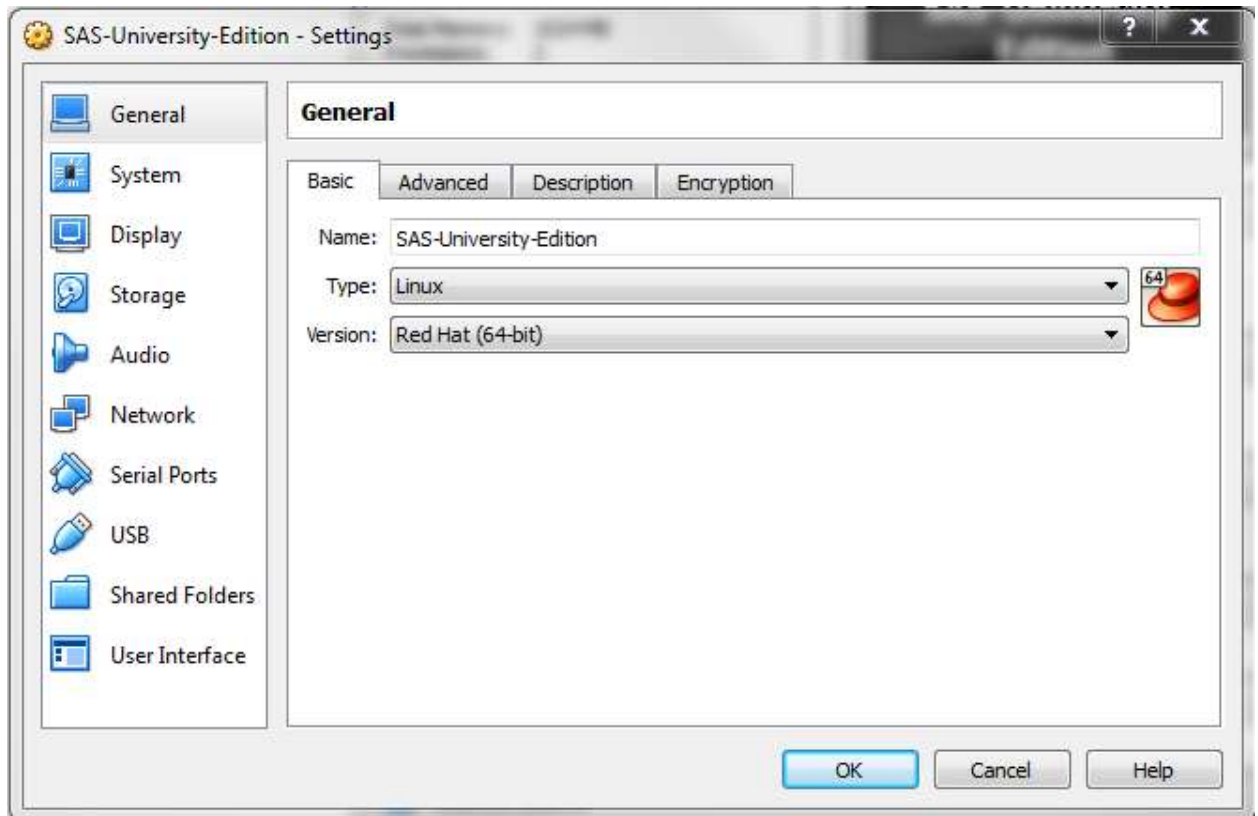
*Note: This application can only be opened through your virtual machine. Therefore, after successful installation of your Virtual Machine, e.g., Oracle's or VMWare's, open up the VM software to run SAS University.*

5. If any errors are encountered, be sure to check out SAS videos and pdfs. Also, copy the error code and paste into Google in search for a solution.

6. Open up your Virtual Machine. *(This example is for Oracle VM VirtualBox)*

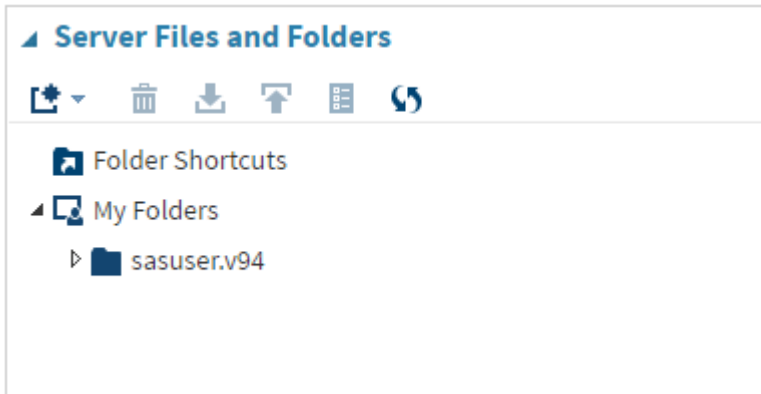


7. Go to Settings and select “Linux Red Hat 64-bit” under the “General” tab.

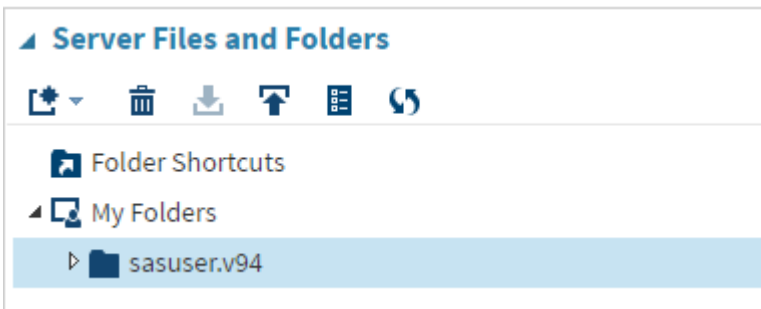


## B. USING SAS STUDIO

1. Choose “Server Files and Folders”

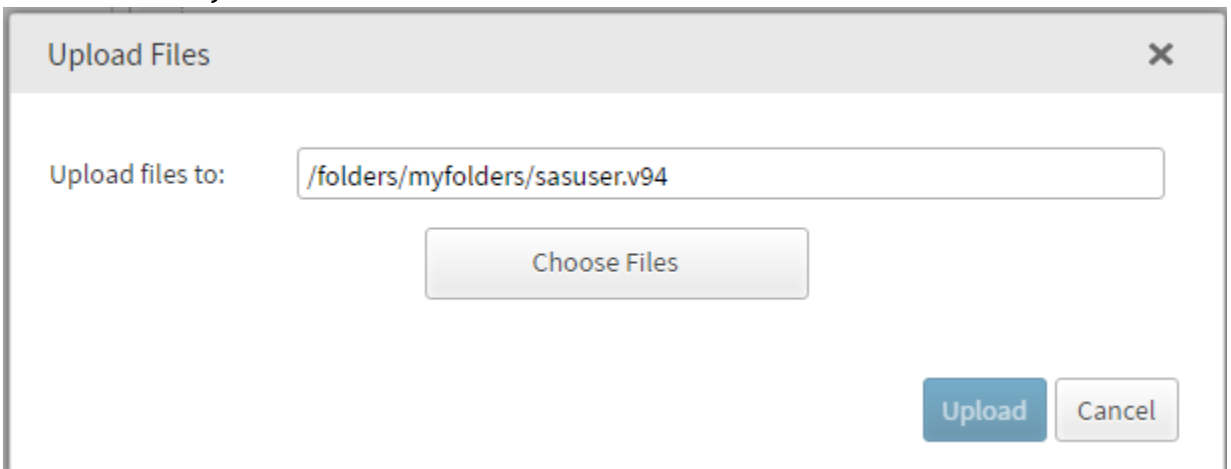


2. Select the sasuser.v94 folder

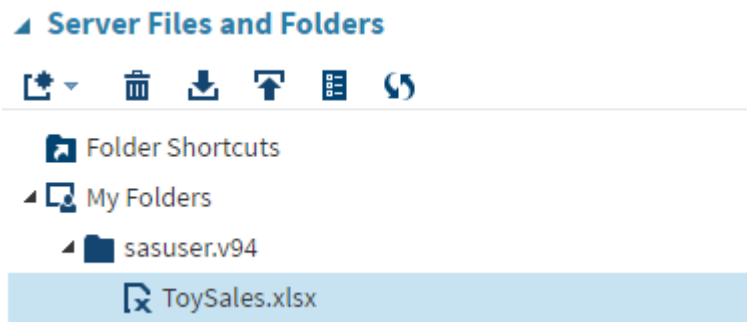


3. Click on the upload  button:

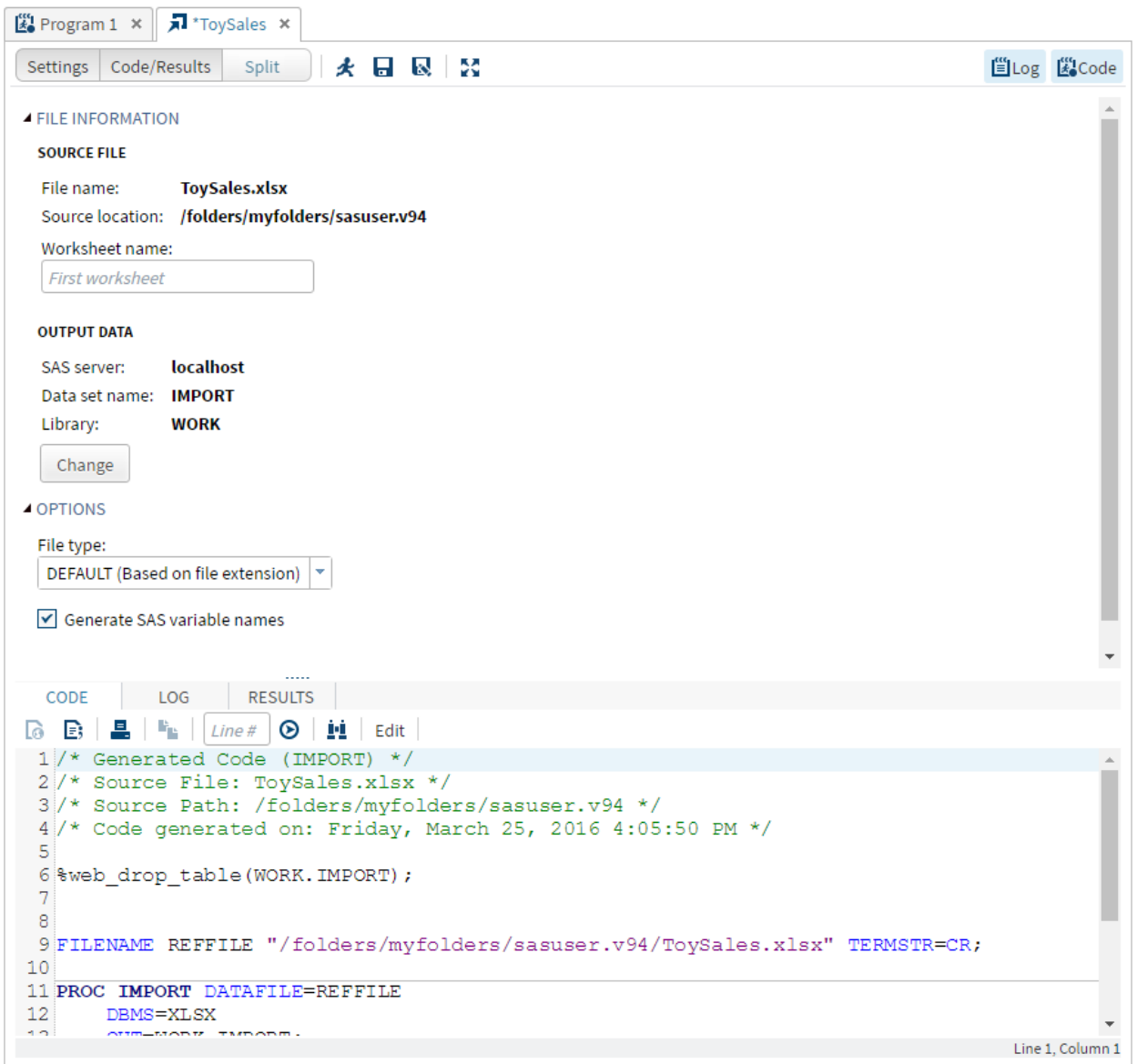
4. Choose the file you want:



5. Double click the ToySales.xlsx file



6. The ToySales Tab window opens on the right side.



7. List the name of the Worksheet you wish to import in the “Worksheet name:” field. *(In this case, the ToySales.xlsx file has only one worksheet, so inserting the worksheet name is not necessary. But had there been multiple sheets, then the desired worksheet for which to import would need to be specified.)*

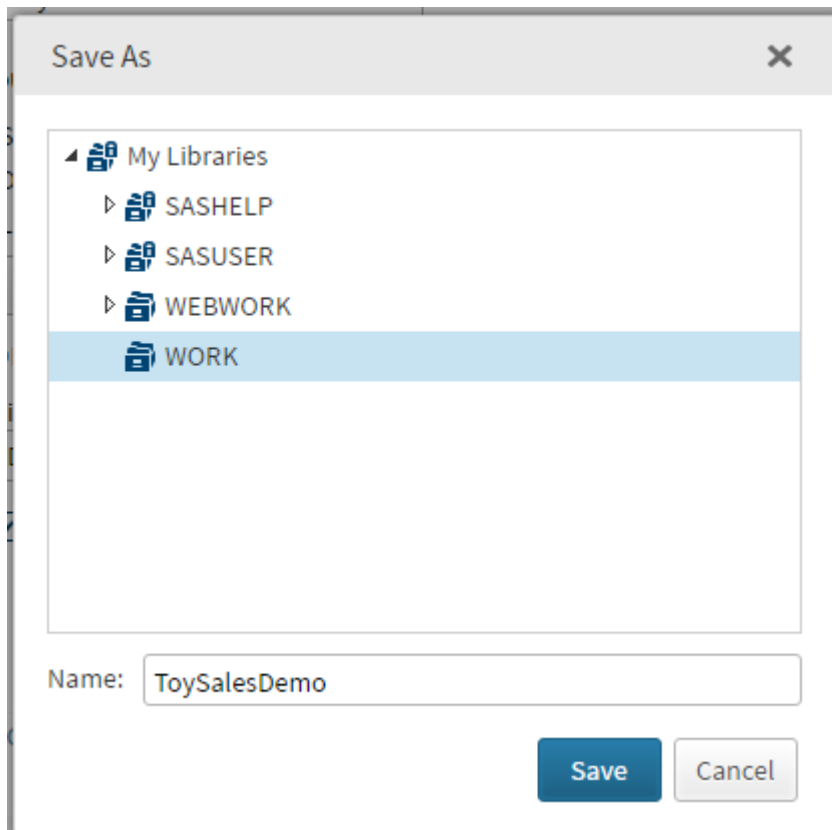
**SOURCE FILE**


File name: **ToySales.xlsx**  
Source location: **/folders/myfolders/sasuser.v94**  
Worksheet name:

8. Click on the “Change” button within the “OUTPUT DATA” area and “Save as” your file in your desired location with your desired name. *(In this demonstration, I saved the file as ToySalesDemo and saved it within the “Work” section.)*

**OUTPUT DATA**


SAS server: **localhost**  
Data set name: **ToySalesDemo**  
Library: **WORK**



9. Click on the “Run” button:  Your file data should output into the Results tab at the bottom of your screen, if you are in “Split” mode, which will indicate a successful import.

File type:

CODE LOG RESULTS OUTPUT DATA



The CONTENTS Procedure

Data Set Name	WORK_TOYSALESDMO	Observations	298
Member Type	DATA	Variables	3
Engine	V9	Indexes	0
Created	03/26/2018 12:00:18	Observation Length	24
Last Modified	03/26/2018 12:00:18	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	UTF-8 Unicode (UTF-8)		



10. Click on the Tasks dropdown menu to expand options:

**Tasks**

My Tasks


Tasks

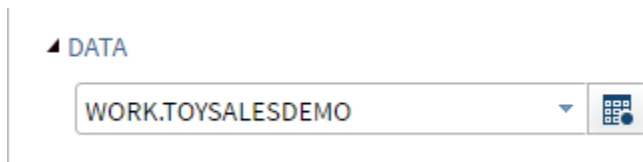
- Data
  - List Table Attributes
  - Characterize Data
  - Describe Missing Data
  - List Data
  - Transpose Data
  - Stack/Split Columns
  - Filter Data
  - Select Random Sample
  - Partition Data
  - Sort Data
  - Rank Data
  - Transform Data
  - Standardize Data
- Graph
- Combinatorics and Probability
- Statistics
- Forecasting
  - Time Series Data Preparation
  - Time Series Exploration
  - Modeling and Forecasting

11. Click on the Forecasting dropdown menu.

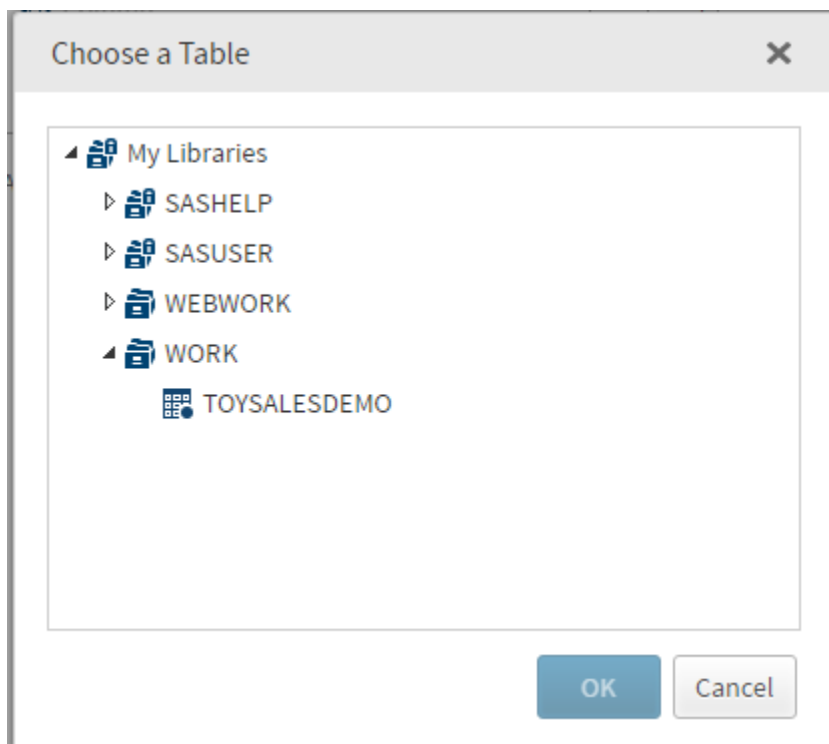
- ▲ Forecasting
  - ▶ Time Series Data Preparation
  - ▶ Time Series Exploration
  - ▶ Modeling and Forecasting

12. Click on Time Series Data Preparation

a. Click on the table  button under the DATA section and select your file:






b. Choose your table. (*I saved my file as TOYSALESDEMO under the "WORK" category*).




- c. Under ROLES, click on the **+** button and select your dependent variable, “units”

▲ ROLES



\* Time series variable  

 units

Treatment of missing values:  
 


- d. Click the arrow adjacent ADDITIONAL ROLES to expand menu. Click on the **+** button for Time ID and choose the variable, “date.” (*The software recognizes the intervals and season lengths in the data and pre-populates the fields*)



▲ ADDITIONAL ROLES



Time ID (1 item)  



 date


▲ Properties

Interval:  

Multiplier:   

Shift:   

Season length:   

- e. After variables have been declared and data prepped, click on the  button along the top, beneath the “Time Series Data Preparation” tab.



- f. You can view the data in the OUTPUT DATA section in either the Code/Results tab or Split tab, beneath the main tab, “Time Series Data Preparation.”

The screenshot displays the 'Time Series Data Preparation' software interface. The left pane is configured for the dataset 'WORK.TOYSALESDEMO'. Under 'ROLES', 'units' is selected as the time series variable. Under 'ADDITIONAL ROLES', 'date' is selected as the Time ID. Properties are set to Interval: Week, Multiplier: 1, Shift: 1, and Season length: 52. The 'Group analysis by' section shows 'Column' selected.

The right pane shows the 'OUTPUT DATA' section. The table is named 'WORK.TSPREP0002' and is viewed in 'Column names' format. It contains 298 rows and 2 columns: 'date' and 'units'. The data shows weekly sales units starting from Sun, 4 Jan 2004.

	date	units
1	Sun, 4 Jan 2004	1423.18793
2	Sun, 11 Jan 2004	1469.3087168
3	Sun, 18 Jan 2004	1434.3697249
4	Sun, 25 Jan 2004	1433.5880442
5	Sun, 1 Feb 2004	1419.4391848
6	Sun, 8 Feb 2004	1375.9159565
7	Sun, 15 Feb 2004	1415.2178473
8	Sun, 22 Feb 2004	1398.6898665
9	Sun, 29 Feb 2004	1438.1224341
10	Sun, 7 Mar 2004	1456.4292693
11	Sun, 14 Mar 2004	1428.9398509
12	Sun, 21 Mar 2004	1369.5139347
13	Sun, 28 Mar 2004	1393.1952398
14	Sun, 4 Apr 2004	1429.0585361
15	Sun, 11 Apr 2004	1420.8138113
16	Sun, 18 Apr 2004	1330.0670225
17	Sun, 25 Apr 2004	1380.4441576
18	Sun, 2 May 2004	1391.1878566
19	Sun, 9 May 2004	1358.4146986
20	Sun, 16 May 2004	1398.229181
21	Sun, 23 May 2004	1373.3742512
22	Sun, 30 May 2004	1398.2436813
23	Sun, 6 Jun 2004	1404.5042247
24	Sun, 13 Jun 2004	1370.2524322
25	Sun, 20 Jun 2004	1406.159838
26	Sun, 27 Jun 2004	1449.7561079
27	Sun, 4 Jul 2004	1396.1668312
28	Sun, 11 Jul 2004	1377.8414313

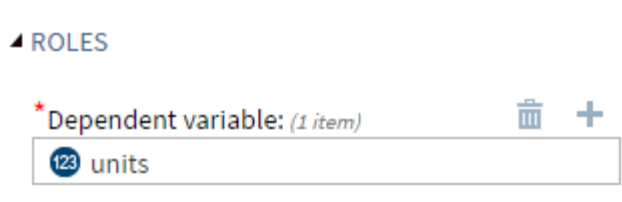
13. Click on Time Series Exploration task.

- ▲ Forecasting
  - Time Series Data Preparation
  - Time Series Exploration
  - Modeling and Forecasting

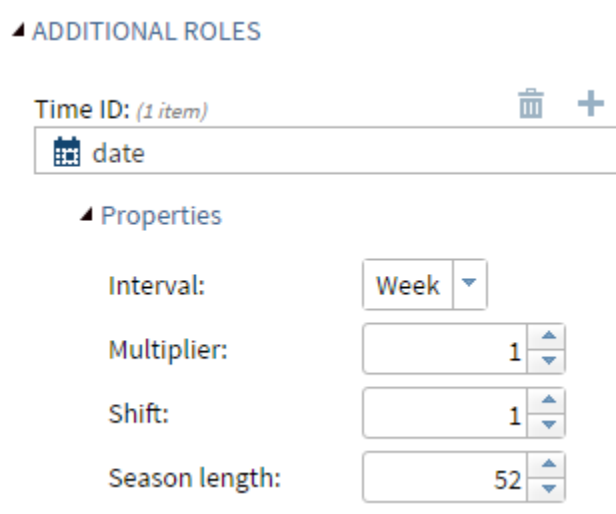
a. *The file you prepped in “Time Series Data Preparation” should already be preloaded in your DATA field. This will remain true as you go through all “Forecasting” tools.*



b. Choose “units” under the Dependent variable in the ROLES section.




c. Expand the ADDITIONAL ROLES menu and select “date” in the Time ID field.



- d. Then click on the Analyses tab to choose settings for your model. *(Please observe, different types of charts/graphs may be chosen here within the “Analyses” tab.)*
- a. *Time Series*
  - b. *Series Histogram*
  - c. *Seasonal Cycles*

*Please select the analysis needed for your intended objective. The default settings, as chosen by the software, are shown below and this demonstration uses such.)*

The screenshot shows the 'Time Series Exploration' application window. The 'ANALYSES' tab is active, displaying various analysis options. Under 'SERIES PLOTS', 'Time Series' is selected. Under 'AUTOCORRELATION ANALYSIS', 'Perform autocorrelation analysis' is selected, with 'Default plots' chosen for display and 'Use default value' for the number of lags. Other analysis options are currently unchecked.

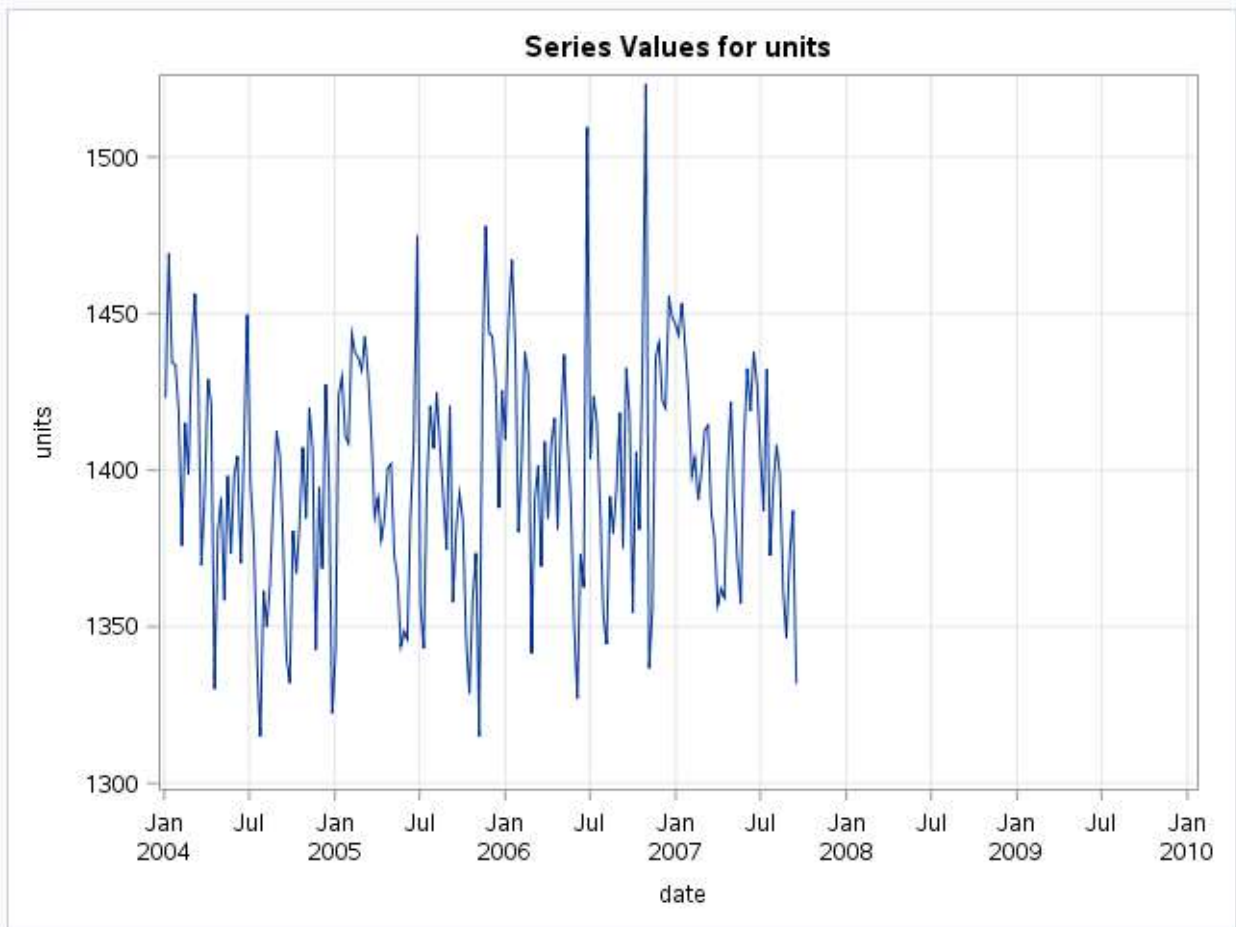
- e. After variables have been declared and settings made, click on the  run button along the top, beneath the “Time Series Exploration” tab.

This close-up shows the top toolbar of the application. The 'run' button, represented by a blue play icon, is the focus of the instruction. It is positioned between the 'Split' button and the 'Log' and 'Code' buttons.

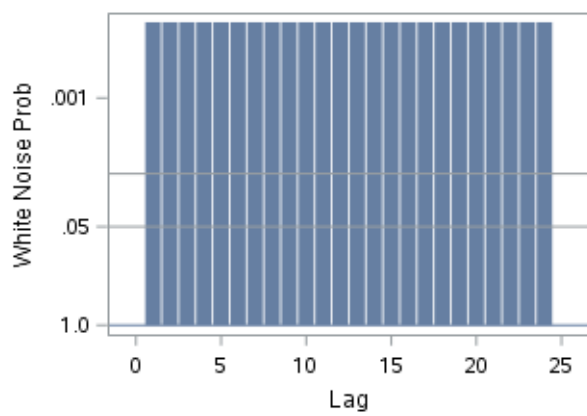
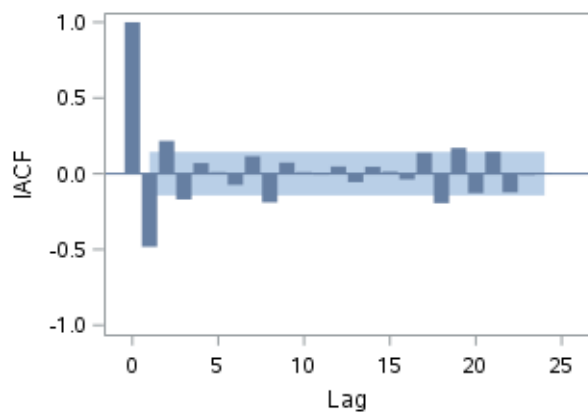
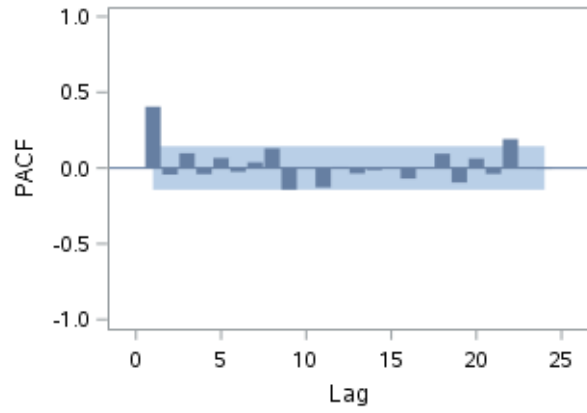
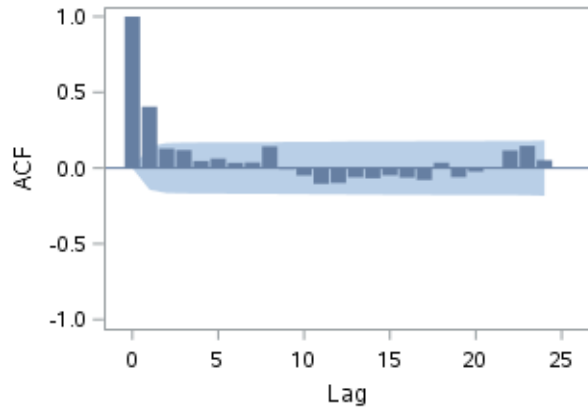
- f. To view the results and charts, click on the “Code/Results” tab, then click on “Results”

Input Data Set	
Name	WORK.TEMPSORTED
Label	
Time ID Variable	date
Time Interval	WEEK
Length of Seasonal Cycle	52

Variable Information	
Name	units
Label	units
First	Sun, 4 Jan 2004
Last	Sun, 16 Sep 2007
Number of Observations Read	298



### Correlations for units





14. Click on Modeling and Forecasting task

- ▾ Forecasting
  - Time Series Data Preparation
  - Time Series Exploration
  - Modeling and Forecasting

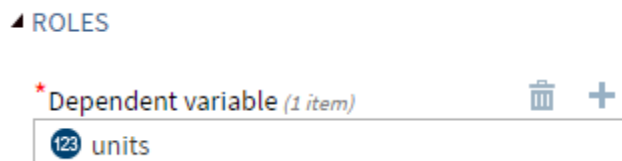
a. Your prepped file should be preloaded in your DATA field.



b. Select the “Settings” tab option to display all options available. (Be sure the DATA tab is selected so you may declare your variables.)



c. Under ROLES, choose “units” as your Dependent variable



d. Expand the ADDITIONAL ROLES menu and choose “date” in the Time ID field.

e. Click on the MODEL tab.



f. Select the dropdown menu for “Forecasting model type:” and select ARIMA

The screenshot shows a software interface with a top navigation bar containing 'Settings', 'Code/Results', and 'Split' tabs, along with icons for home, save, print, and zoom. Below this is a secondary navigation bar with 'DATA', 'MODEL', 'OPTIONS', 'OUTPUT', and 'INFORMATION' tabs. The 'MODEL' tab is active, showing a tree view with 'MODEL' expanded to 'Forecasting model type: ARIMA'. Underneath, 'Model Settings' is expanded to 'ARIMA', which includes three spinners for 'Autoregressive order (p):', 'Differencing order (d):', and 'Moving average order (q):', all currently set to 0. Below these is a section for 'Seasonal ARIMA' with three more spinners for 'Autoregressive order (P):', 'Differencing order (D):', and 'Moving average order (Q):', also set to 0. At the bottom, there is a checked checkbox labeled 'Include intercept in model'.

**For this example, we will use a 3-1-1 Arima model, so set Autoregressive order to 3, Differencing order to 1, and Moving average order to 1. All else will remain as defaults.**

**Additionally, please note the plethora of model options to choose from:**

1. **Random Walk**
2. **Moving Average**
3. **Exponential Smoothing**
4. **ARIMA**
5. **ARIMAX**
6. **Unobserved Components**

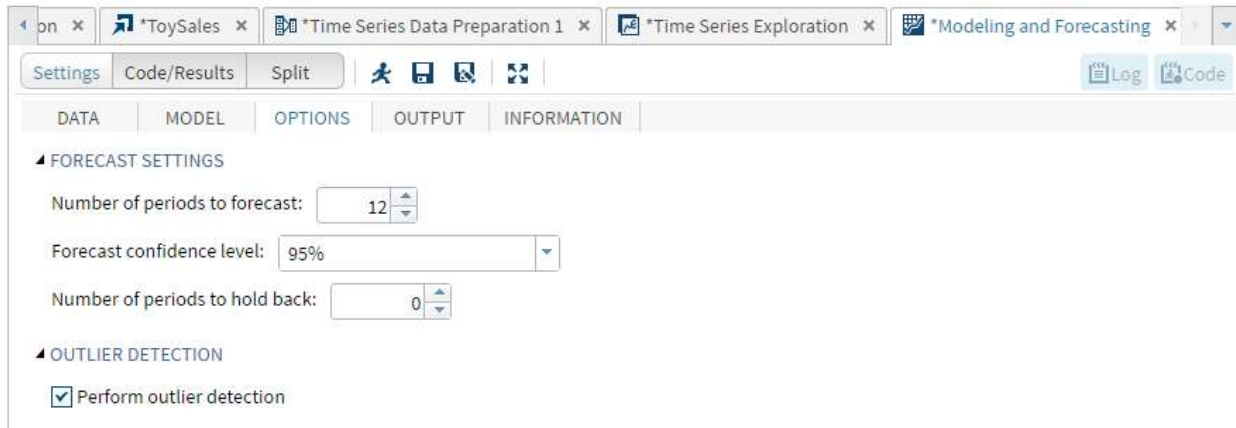
Select the model that best suits your intended objective:

The screenshot shows a software interface with three tabs: \*ToySales, \*Time Series Data Preparation, and \*Time Series Exploration. Below the tabs are buttons for Settings, Code/Results, and Split, along with navigation icons. The main panel has tabs for DATA, MODEL, OPTIONS, OUTPUT, and INFORMATION. The MODEL tab is active, showing a tree view of settings:

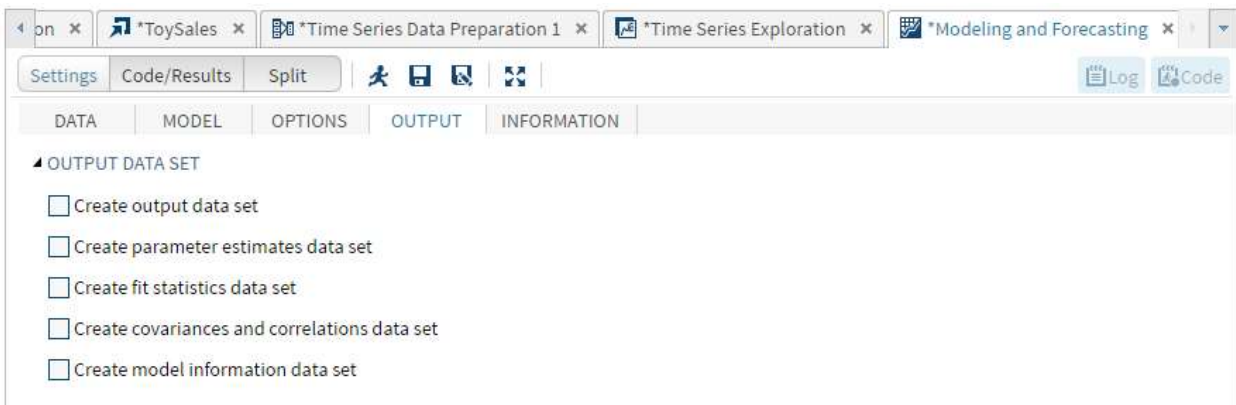
- MODEL
  - \*Forecasting model type: ARIMA (dropdown menu is open, showing options: Random walk, Moving average, Exponential smoothing, ARIMA (highlighted), ARIMAX, Unobserved components)
  - Model Settings
    - ARIMA
      - Autoregressive order
      - Differencing order (d)
      - Moving average order
    - Seasonal ARIMA
      - Autoregressive order (P): 0
      - Differencing order (D): 0
      - Moving average order (Q): 0
  - Plots
    - Select plots to display: Default plots


Additional settings include a checked checkbox for "Include intercept in model".

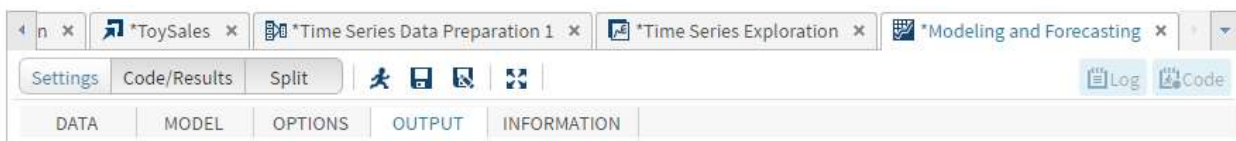
- g. Next, choose the **OPTIONS** tab and adjust settings (*This demonstration keeps everything at default*).



- h. Choose the **OUTPUT** tab and select the desired **DATA SETS** to be output.



- i. After declaring variables, selecting a model, determining options, and figuring out desired outputs, click on the  run button below the “Modeling and Forecasting” tab.



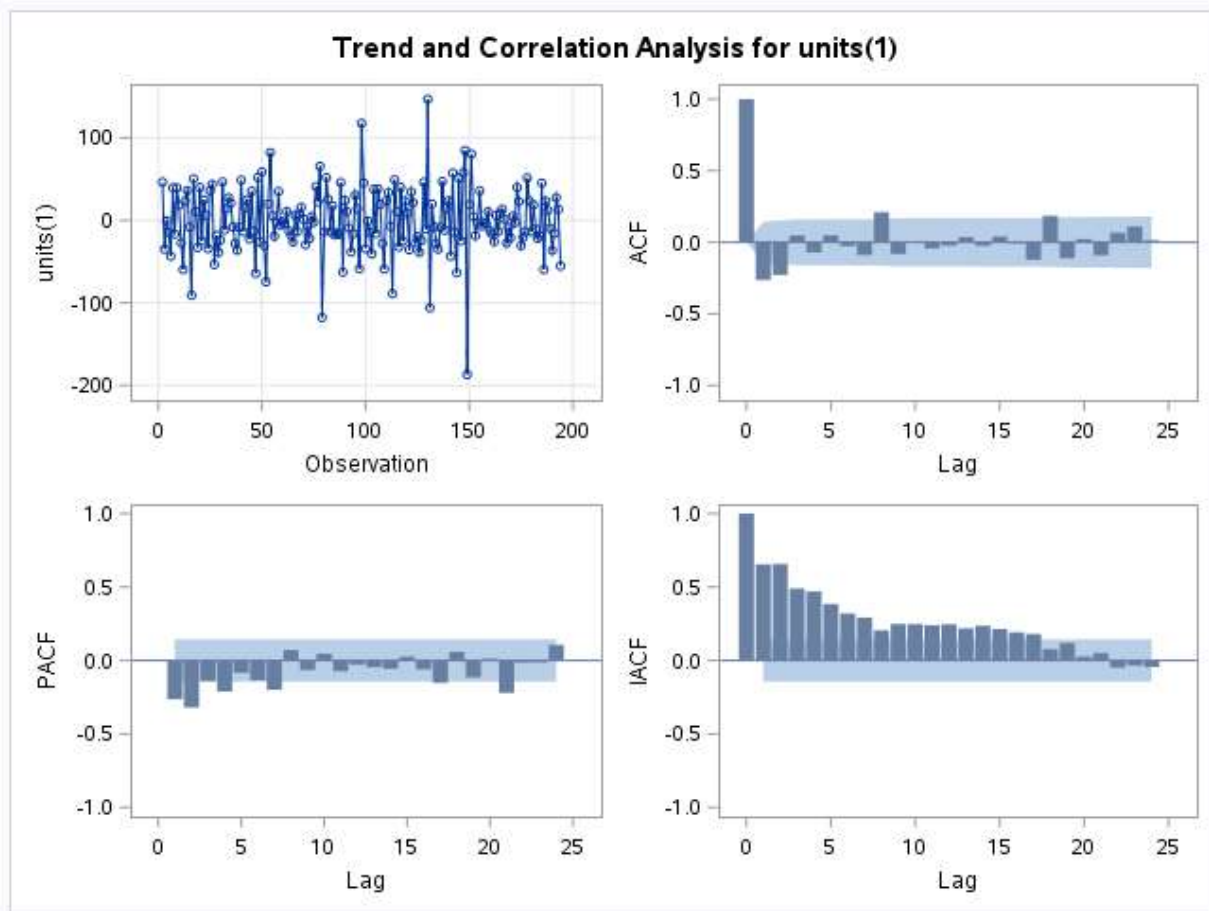
j. Click on “Code/Results,” then click on “RESULTS” to see charts and graphs

Settings Code/Results Split Log Code

CODE LOG RESULTS OUTPUT DATA

Name of Variable = units	
Period(s) of Differencing	1
Mean of Working Series	-0.47368
Standard Deviation	39.31342
Number of Observations	193
Observation(s) eliminated by differencing	1

Autocorrelation Check for White Noise									
To Lag	Chi-Square	DF	Pr > Chi Sq	Autocorrelations					
6	25.97	6	0.0002	-0.264	-0.228	0.047	-0.071	0.050	-0.027
12	38.47	12	0.0001	-0.090	0.210	-0.082	0.004	-0.042	-0.024
18	49.76	18	<.0001	0.036	-0.025	0.040	-0.009	-0.122	0.185
24	56.13	24	0.0001	-0.112	0.023	-0.091	0.066	0.110	0.014



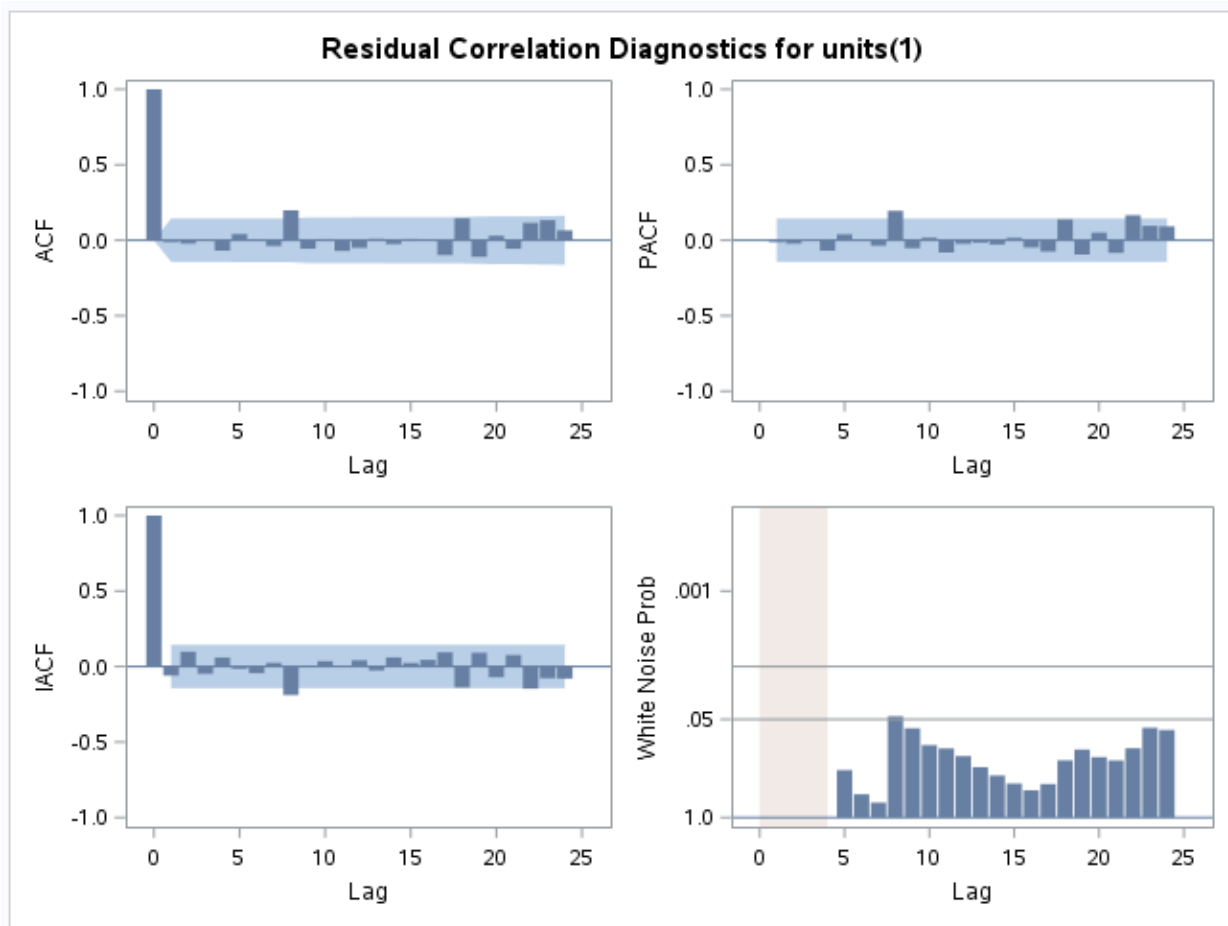
ARIMA Estimation Optimization Summary	
Estimation Method	Maximum Likelihood
Parameters Estimated	5
Termination Criteria	Maximum Relative Change in Estimates
Iteration Stopping Value	0.001
Criteria Value	10.2588
Maximum Absolute Value of Gradient	4471.754
R-Square Change from Last Iteration	0.105576
Objective Function	Log Gaussian Likelihood
Objective Function Value	-950.6
Marquardt's Lambda Coefficient	0.00001
Numerical Derivative Perturbation Delta	0.001
Iterations	8
Warning Message	Estimates may not have converged.

Maximum Likelihood Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr >  t	Lag
MU	0.03052	0.09910	0.31	0.7581	0
MA1,1	0.99994	7.24843	0.14	0.8903	1
AR1,1	0.44908	0.07505	5.98	<.0001	1
AR1,2	-0.07855	0.07941	-0.99	0.3228	2
AR1,3	0.12221	0.07470	1.64	0.1018	3

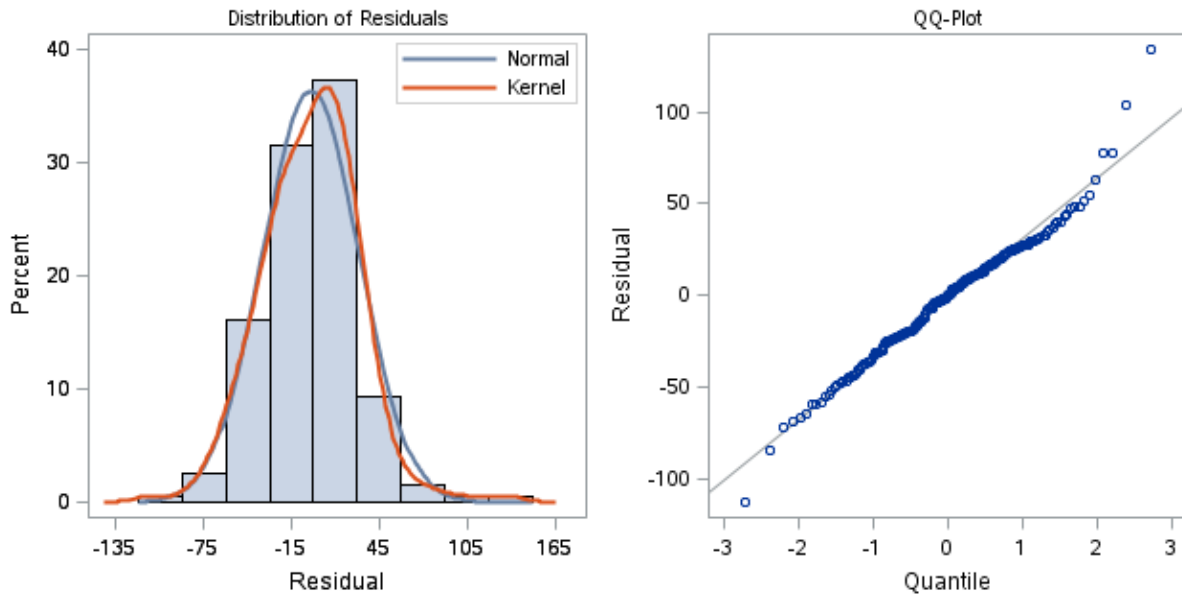
Constant Estimate	0.01548
Variance Estimate	1116.04
Std Error Estimate	33.40718
AIC	1911.199
SBC	1927.513
Number of Residuals	193

Correlations of Parameter Estimates					
Parameter	MU	MA1,1	AR1,1	AR1,2	AR1,3
MU	1.000	0.560	0.132	0.044	0.072
MA1,1	0.560	1.000	0.252	0.107	0.200
AR1,1	0.132	0.252	1.000	-0.346	0.104
AR1,2	0.044	0.107	-0.346	1.000	-0.349
AR1,3	0.072	0.200	0.104	-0.349	1.000

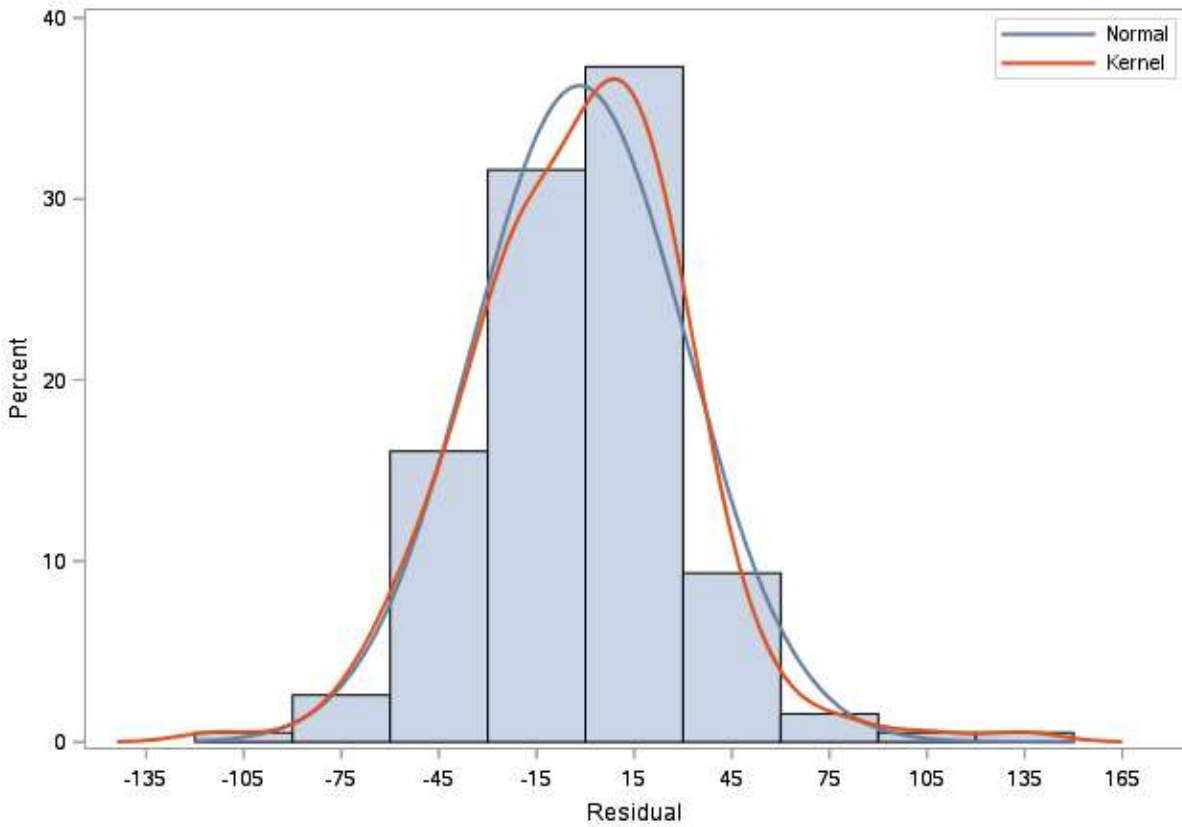
Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1.35	2	0.5099	-0.009	-0.020	0.003	-0.066	0.042	0.009
12	11.85	8	0.1583	-0.036	0.200	-0.054	0.010	-0.070	-0.049
18	18.59	14	0.1812	0.011	-0.024	0.010	-0.001	-0.097	0.146
24	29.84	20	0.0724	-0.109	0.030	-0.055	0.116	0.133	0.066
30	36.94	26	0.0756	-0.066	0.028	-0.103	0.049	-0.106	-0.041
36	39.92	32	0.1586	-0.017	-0.030	0.016	0.008	-0.095	0.044



### Residual Normality Diagnostics for units(1)

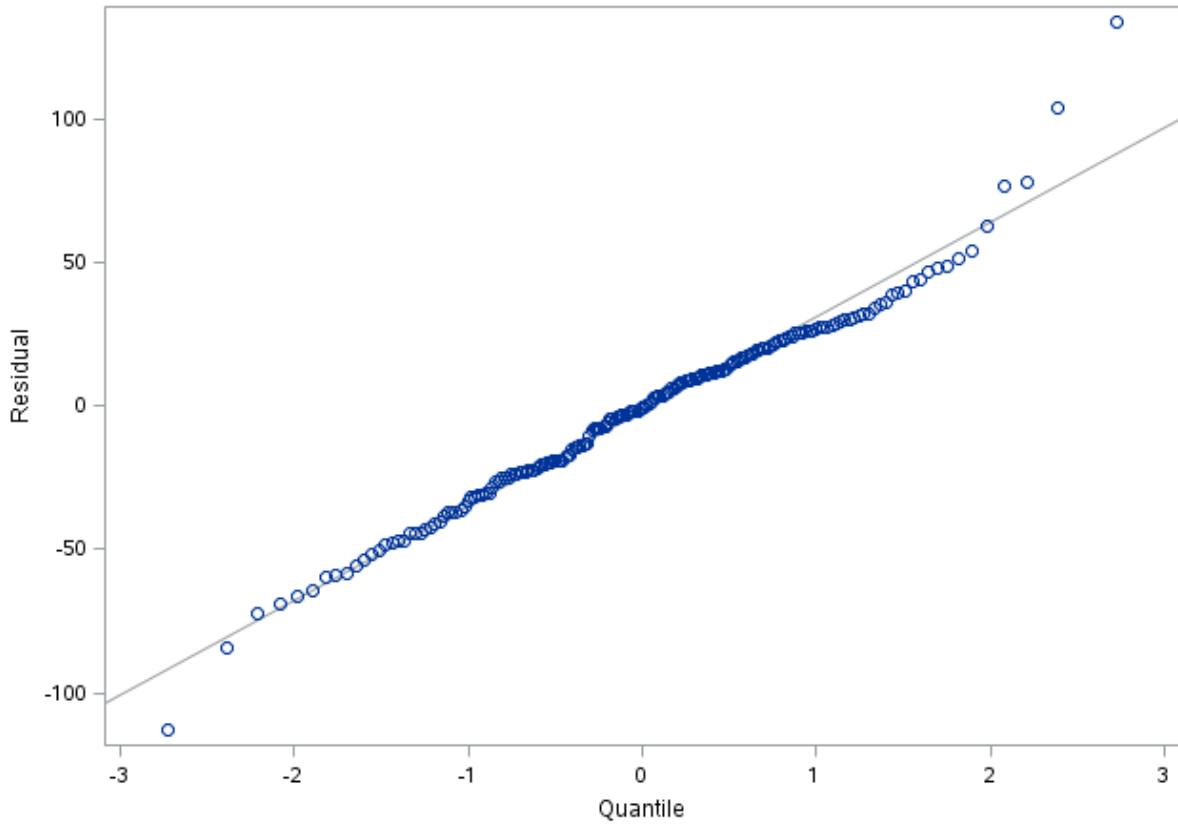


### Distribution of Residuals for units(1)





Residual Q-Q Plot for units(1)

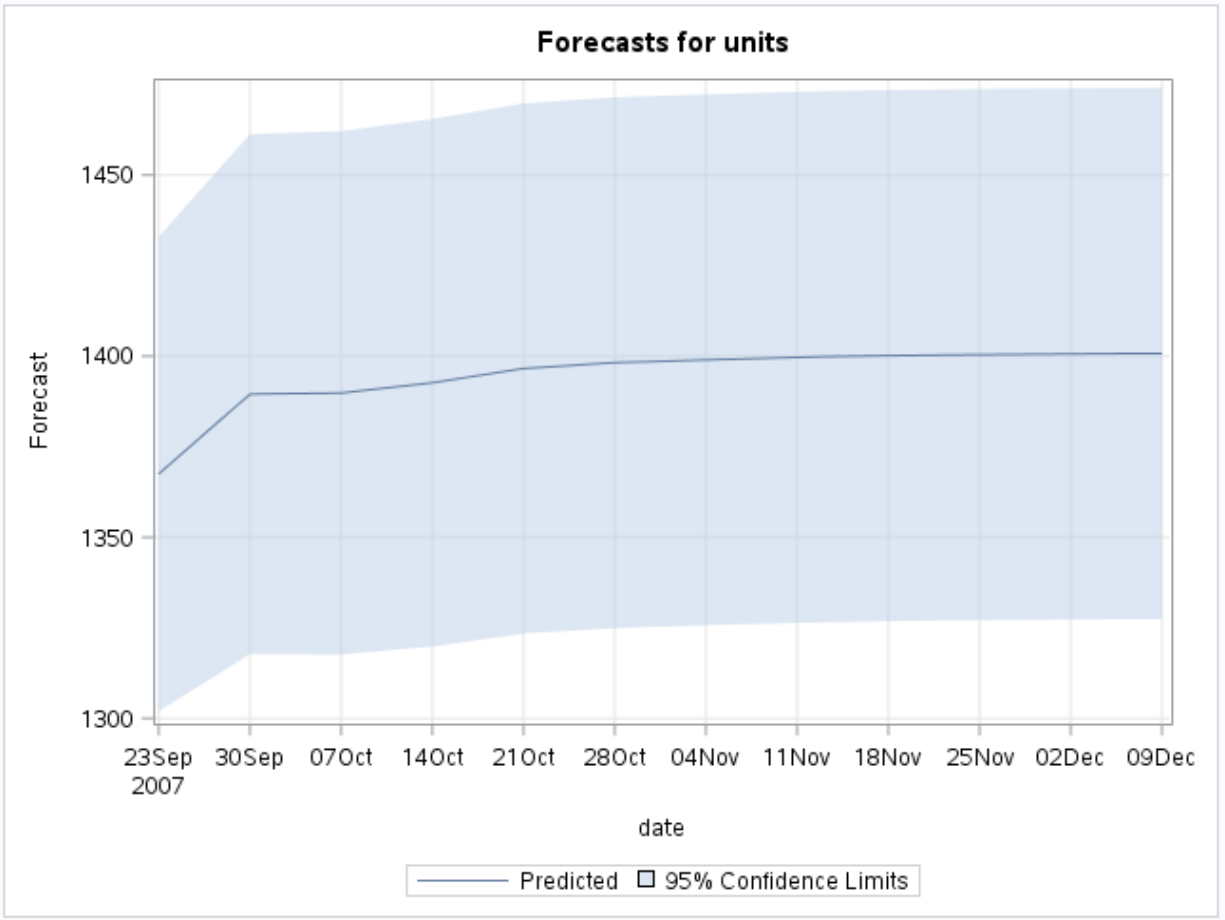


Model for variable units	
Estimated Mean	0.030517
Period(s) of Differencing	1

Autoregressive Factors	
Factor 1:	$1 - 0.44908 B^{**}(1) + 0.07855 B^{**}(2) - 0.12221 B^{**}(3)$

Moving Average Factors	
Factor 1:	$1 - 0.99994 B^{**}(1)$

Forecasts for variable units				
Obs	Forecast	Std Error	95% Confidence Limits	
195	1367.4965	33.4072	1302.0197	1432.9734
196	1389.5245	36.6220	1317.7466	1461.3023
197	1389.8657	36.8526	1317.6359	1462.0955
198	1392.6706	37.1580	1319.8422	1465.4990
199	1396.6109	37.3366	1323.4325	1469.7894
200	1398.2173	37.3785	1324.9567	1471.4779
201	1398.9875	37.3944	1325.6957	1472.2792
202	1399.7042	37.4030	1326.3957	1473.0127
203	1400.1773	37.4062	1326.8625	1473.4921
204	1400.4431	37.4074	1327.1261	1473.7602
205	1400.6284	37.4078	1327.3103	1473.9464
206	1400.7640	37.4081	1327.4456	1474.0824



Outlier Detection Summary	
Maximum number searched	4
Number found	4
Significance used	0.05

Outlier Details				
Obs	Type	Estimate	Chi-Square	Approx Prob>ChiSq
130	Additive	128.98820	19.17	<.0001
148	Additive	122.68660	18.29	<.0001
78	Additive	87.63701	9.93	0.0016
97	Additive	-79.99659	8.25	0.0041

- k. Click on the OUTPUT DATA tab to see your forecasted quantities, std deviations, confidence intervals, and residuals (differences).

The screenshot shows a software interface with a data table. The table has 7 columns: 'date', 'units', and 'FORECAST'. The 'FORECAST' column contains numerical values, and the 'date' column contains dates from 01/04/2004 to 07/04/2004. The 'units' column contains large numerical values. The interface also includes a 'Columns' list on the left with checkboxes for 'date', 'units', 'FORECAST', 'STD', 'L95', 'U95', and 'RESIDUAL'. A 'Property Value' table is visible at the bottom left.

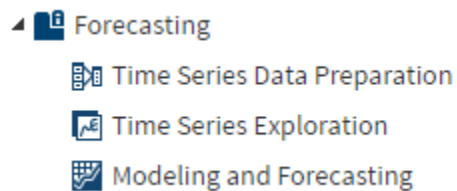
date	units	FORECAST
01/04/2004	1423.18793	
01/11/2004	1469.3087168	1423.2184471
01/18/2004	1434.3697249	1457.1291968
01/25/2004	1433.5880442	1431.7796648
02/01/2004	1419.4391848	1440.0989368
02/08/2004	1375.9159565	1426.8324531
02/15/2004	1415.2178473	1402.3955732
02/22/2004	1398.6898665	1423.0863653
02/29/2004	1438.1224341	1404.9712477
03/07/2004	1456.4292693	1431.6550182
03/14/2004	1428.9398509	1436.741289
03/21/2004	1369.5139347	1427.2190892
03/28/2004	1393.1952398	1400.9907365
04/04/2004	1429.0585361	1412.4494899
04/11/2004	1420.8138113	1420.4487675
04/18/2004	1330.0670225	1416.8595535
04/25/2004	1380.4441576	1376.4857596
05/02/2004	1391.1878566	1405.4469107
05/09/2004	1358.4146986	1394.5478882
05/16/2004	1398.229181	1383.4851317
05/23/2004	1373.3742512	1405.9201934
05/30/2004	1398.2436813	1386.262101
06/06/2004	1404.5042247	1404.7510125
06/13/2004	1370.2524322	1402.5774391
06/20/2004	1406.159838	1388.5432494
06/27/2004	1449.7561079	1408.7778554
07/04/2004	1396.1668312	1422.7982544

## Appendix

1. To find the Mean Squared Error, Root Mean Squared Error, Mean Absolute Percentage Error, or Maximum Percent Error, please select the following:

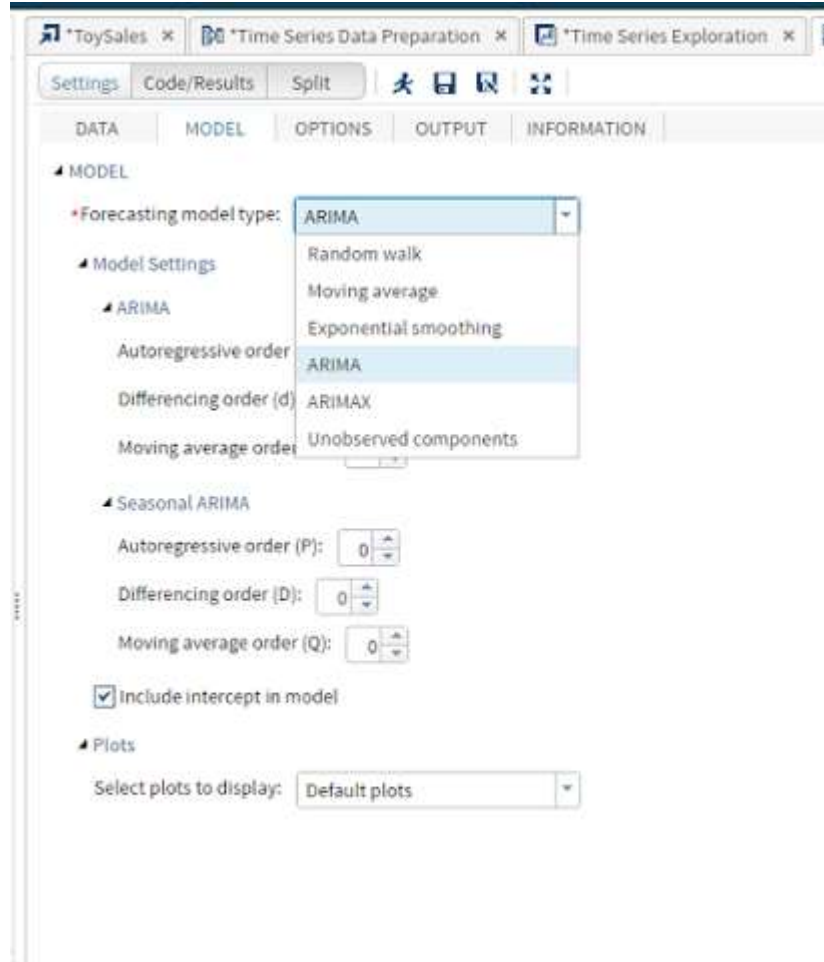
Fit Statistics Based on Residuals	
Mean Squared Error	1257.21999
Root Mean Squared Error	35.45730
Mean Absolute Percentage Error	1.94482
Maximum Percent Error	9.26984
R-Square	0.04581
Adjusted R-Square	0.04081
Random Walk R-Square	0.18655
Amemiya's Adjusted R-Square	0.02583
Number of non-missing residuals used for computing the fit statistics = 193	

- a. Click on Modeling and Forecasting task



- b. Your data should already be prepped and variables declared. If not, please refer to steps aforementioned above.

c. Under the “MODEL” tab, choose amongst the “Forecasting model type:”



d. Select the “Unobserved components” forecasting model.

The screenshot shows the 'Modeling and Forecasting 1' application window. The 'MODEL' tab is active, and the 'Forecasting model type' is set to 'Unobserved components'. Under 'Model Settings', the 'Include an irregular component' and 'Include a level component' checkboxes are checked. The 'Plots' section shows 'Default plots' selected.

e. Adjust settings to your preferences, then click the  run button.

f. Click on “Code/Results” and then “Results” to see the model.

The screenshot shows a software interface with a menu bar (Settings, Code/Results, Split) and a toolbar. The main content area displays several summary tables for the variable 'units'.

**Input Data Set**

Name	WORK_TEMPORATED
Time ID Variable	date

**Estimation Span Summary**

Variable	Type	First Obs	Last Obs	NObs	NMiss	Min	Max	Mean	Standard Deviation
units	Dependent	Sun, 4 Jan 2004	Sun, 16 Sep 2007	194	0	1314.93940	1523.43020	1397.78356	26.34474

**Forecast Span Summary**

Variable	Type	First Obs	Last Obs	NObs	NMiss	Min	Max	Mean	Standard Deviation
units	Dependent	Sun, 4 Jan 2004	Sun, 16 Sep 2007	194	0	1314.93940	1523.43020	1397.78356	26.34474

**Preliminary Estimates of the Free Parameters**

Component	Parameter	Estimate
Irregular	Error Variance	1051.30485
Level	Error Variance	394.25932

**Likelihood Based FH Statistics**

Statistic	Value
Full Log Likelihood	-982.8
Diffuse Part of Log Likelihood	-3E-13
Non-Missing Observations Used	194
Estimated Parameters	2
Initialized Diffuse State Elements	1
Normalized Residual Sum of Squares	193
AIC (smaller is better)	1928.1
BIC (smaller is better)	1935.7
AICC (smaller is better)	1929.3
HQIC (smaller is better)	1931.8