M349R (Unique 54230)

**Instructor:** Gustavo Cepparo

Project 5

Juan Acosta

JA45384

Fall 2018

**Problem 1:**

1. From the plot we can see that the energy bills seem to oscillate as time progresses (variability), but also presents an overall trend where it decreases and then increases.
2. The data appears to present a seasonal variability where it oscillates approximately constantly throughout the plot.
3. Analysis:
   1. From the plot trend we can understand the quadratic portion of the model, and from the seasonality we can understand the dummies to be the seasonal cycle quarters. Therefore, Q1 will represent the first quarter of the seasonal cycle, Q2 the second quarter of the seasonal cycle, and Q3 the third quarter of the seasonal cycle. Finally, understanding that since they are dummies, when all three are zero we will get information for the fourth quarter of the seasonal cycle.
   2. From 6.35 we can see that all variables are significant from their p values, and despite Q2 being “non-significant” because it is a dummy variable it must remain in the model, yet we can state that approximately on the average the effect of seasonality during Q2 and Q4 is the same
   3. From the table 6.35:

ŷ = 276.64 – 7.4683\*t+0.30123\*t2 + 65.771\*Q1 – 37.870\*Q2 – 127.61\*Q3

From table 6.9 we can see that t=41 is in Q1 and t= 42 in Q2. Therefore:

ŷ41 = 276.64 – 7.4583\*41+0.30123\*412 + 65.771\*1 – 37.870\*0 – 127.61\*0 = 542.99

ŷ42 = 276.64 – 7.4583\*42+0.30123\*422 + 65.771\*0 – 37.870\*1 – 127.61\*0 = 456.89

* 1. From the data in 6.35 we can extract the point forecasts and prediction intervals:

|  |  |
| --- | --- |
| Point Forecast | 95% Prediction Interval |
| 542.9878 | (400.94 , 685.03) |
| 456.8910 | (312.33 , 601.45) |
| 385.2961 | (237.91 , 532.68) |
| 531.6563 | (381.15 , 682.16) |

* 1. We see that the 6.35 output has a Durbin Watson statistic D = 0.840.

We proceed with the Durbin Watson test at α = 0.05:

H0: Error terms are uncorrelated

Ha: Error terms are correlated

Following the procedure from the book we define k = 5, n = 40, and find the bounds for DW at α = 0.05 are (1.23 , 1.79). Therefore, given D = 0.840 we must reject the null hypothesis in favor of the alternative that there is autocorrelation.

1. From the ARIMA output we can see that the ϕ\_hat = 0.59408. And from the t statistic we can see that the approximated p values is less than .05 and we can assume this value is statistically significant and different from 0.
2. From the Arima output we can see that the p values for time (0.1011) and Q2 (0.0787) indicate that they are not statistically significant. Nonetheless, because of their correlation of time with time squared and the relationship of Q2 with the other dummies they are required in the model.
3. The point forecast as seen on the output table 6.36 are:

ŷ41 = 605.3285

ŷ42 = 505.6717

ŷ43 = 426.9411

ŷ44 = 569.9732

The 95% prediction intervals as seen on the output table 6.36 are:

ŷ41: (506.8378 , 703.8193)

ŷ42: (391.1114 , 620.2320)

ŷ43: (307.2232 , 546.6591)

ŷ44: (448.4872 , 691.4592)

1. From the model:

yt = β0 + β1t + β2t2 + β3Q1 + β4Q2 + β5Q3 + ϕ1ԑt-1 + a1

The prediction model is:

ŷt = β0 + β1t + β2t2 + β3Q1 + β4Q2 + β5Q3 + ϕ1ԑt-1

From the 6.36 table we can obtain all the estimated parameters:

ŷt = 283.95 – 9.22\*t + 0.35\*t2 + 70.11\*Q1 – 35.43\*Q2 – 126.53\*Q3 + 0.59\*ԑt-1

1. Using the model above, and values from the table 6.9 we can calculate the point estimations from part f:

ԑ40 = ŷ40 - (283.95 – 9.22\*40 + 0.35\*402 + 70.11\*0– 35.43\*0 – 126.53\*0) = 59.0501

ŷ41 = 283.95 – 9.22\*41 + 0.35\*412 + 70.11\*1– 35.43\*0 – 126.53\*0 + 0.59\*59.0501

ŷ41 = 605.3285

ԑ41 = ŷ41 - (283.95 – 9.22\*41 + 0.35\*412 + 70.11\*1– 35.43\*0 – 126.53\*0) = 35.0796

ŷ42 = 283.95 – 9.22\*42 + 0.35\*422 + 70.11\*0– 35.43\*1 – 126.53\*0 + 0.59\*35.0796

ŷ42 = 505.6717

ԑ42 = ŷ42 - (283.95 – 9.22\*42 + 0.35\*422 + 70.11\*0– 35.43\*1 – 126.53\*0) = 20.8390

ŷ43 = 283.95 – 9.22\*43 + 0.35\*432 + 70.11\*0– 35.43\*0 – 126.53\*1 + 0.59\*20.8390

ŷ43 = 426.9411

ԑ43 = ŷ43 - (283.95 – 9.22\*43 + 0.35\*432 + 70.11\*0– 35.43\*0 – 126.53\*1) = -42.7177

ŷ44 = 283.95 – 9.22\*44 + 0.35\*442 + 70.11\*0– 35.43\*0 – 126.53\*0 - 0.59\*42.7177

ŷ44 = 569.9732

For the formula for prediction intervals we must define a few values:

s = 50.25132

ϕ1 = 0.59408

z = 1.96

Therefore, the prediction intervals are calculated:

ŷ41 = 605.3285 +- [1.96\*(50.25132)]

ŷ41: (506.8378 , 703.8193)

ŷ42 = 505.6717 +- [1.96\*(50.25132)\*sqrt(1+(0.59408)2)]

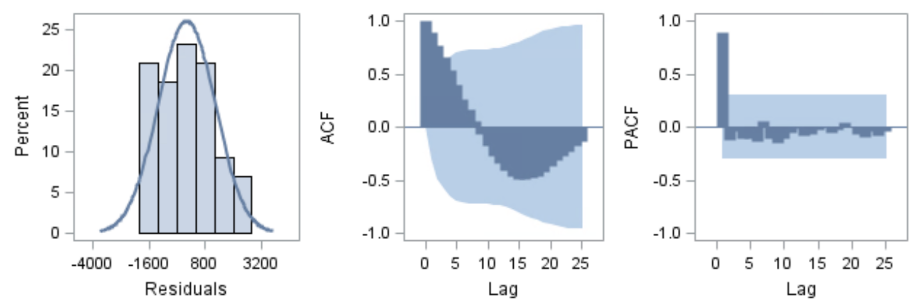
ŷ42: (391.1114 , 620.2320)

ŷ43 = 426.9411 +- [1.96\*(50.25132)\*sqrt(1+(0.59408)2+(0.59408)4)]

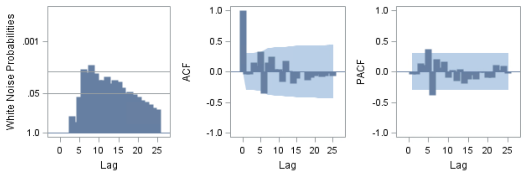
ŷ43: (307.2232 , 546.6591)

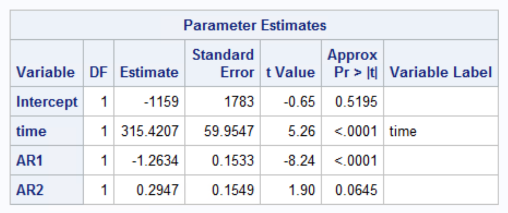
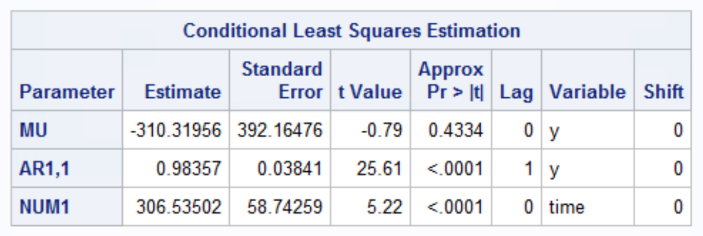
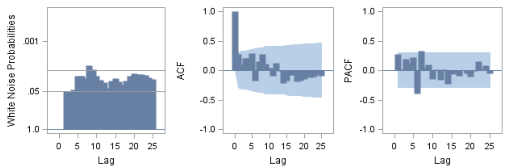
ŷ44 = 569.9732 +- [1.96\*(50.25132)\*sqrt(1+(0.59408)2+(0.59408)4+(0.59408)6)]

ŷ44: (448.4872 , 691.4592)

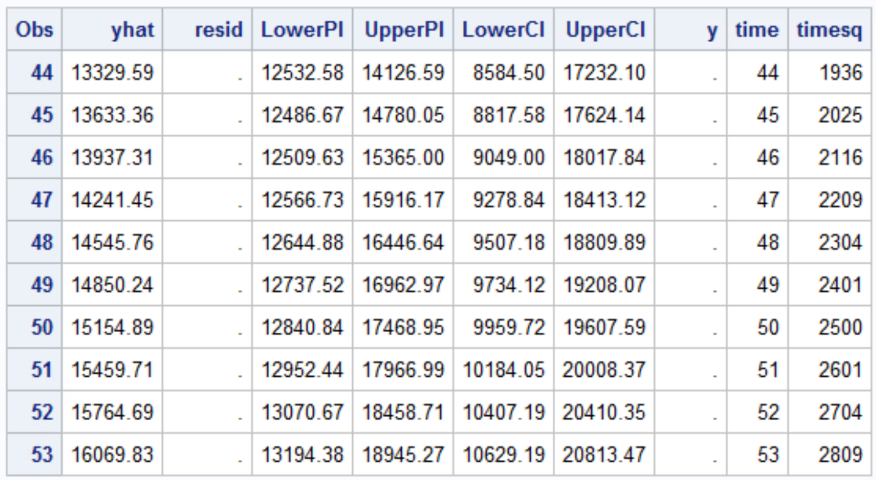
**Problem 2:**

From the initial ACF plot we can see that it is not dying down quickly.

Therefore, we will fit an AR(1) model and an AR(2) model and decide wh ich one is better:

 AR(1) AR(2)

From the graphs and tables we can see that the AR(2) is too much, hence we will accept the AR(1) as the better fit.

Forecasting:

**Problem 3:**

1. From the SAS output 6.37 we can see that the p value of M5 seems to be much higher than any other, which could imply that it is not important for the model.
2. Again from the SAS output 6.37 we can see that the predicted values are:

ŷ169\* = 5.3489

ŷ170\*= 5.2641

1. Using the SAS output from figure 6.37 we can write the model:

ŷt\*= 4.80732 + 0.00352\*t – 0.05247\*M1 – 0.14079\*M2 – 0.10710\*M3 + 0.04988\*M4 + 0.02542\*M5 + 0.19017\*M6 + 0.38245\*M7 + 0.41337\*M8 + 0.07142\*M9 + 0.05064\*M10 – 0.14194\*M11

Therefore, since the Mi’s are dummy variables, we can calculate ŷ169\* and ŷ170\*:

ŷ169\* = 4.80732 + 0.00352\*169 – 0.05247\*1 – 0.14079\*0 – 0.10710\*0 + 0.04988\*0 + 0.02542\*0 + 0.19017\*0 + 0.38245\*0 + 0.41337\*0 + 0.07142\*0 + 0.05064\*0 – 0.14194\*0

ŷ169\* = 5.3489

ŷ170\*= 4.80732 + 0.00352\*170 – 0.05247\*0 – 0.14079\*1 – 0.10710\*0 + 0.04988\*0 + 0.02542\*0 + 0.19017\*0 + 0.38245\*0 + 0.41337\*0 + 0.07142\*0 + 0.05064\*0 – 0.14194\*0

ŷ170\*= 5.2641

1. To predict ŷ169 = (ŷ169\*)4 we proceed to raise both the value and the PI to the fourth power:

ŷ169 = (5.3489)4 = 818.5739

PI[ŷ169]: [5.29134 , 5.40654] = [783.8799 , 854.4071]

1. To predict ŷ170 = (ŷ170\*)4 we proceed to raise both the value and the PI to the fourth power:

ŷ170 = (5.2641)4 = 767.8856

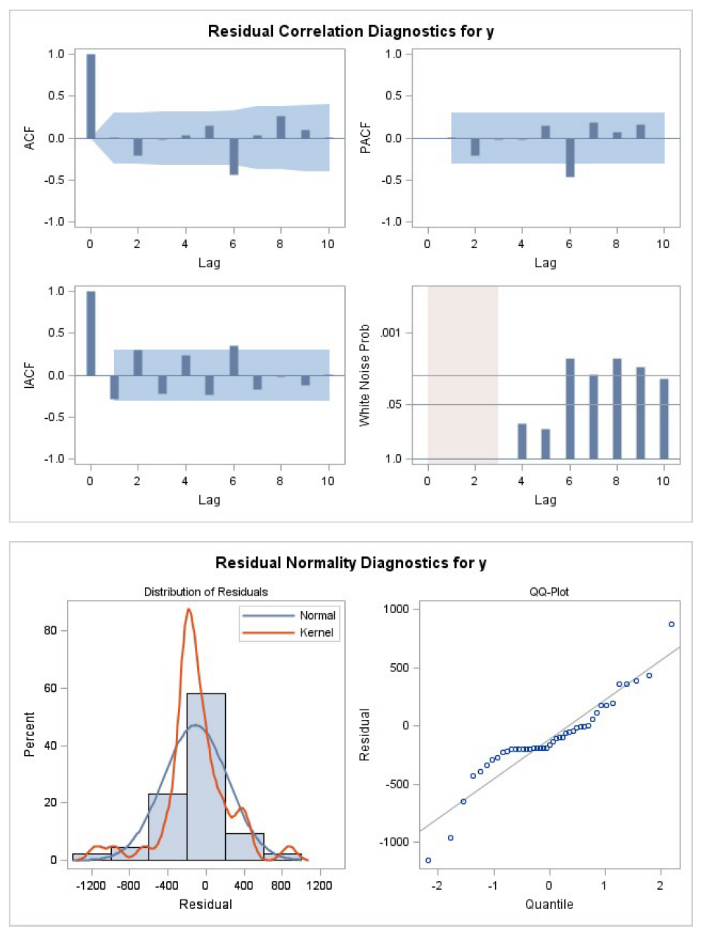
PI[ŷ170]: [5.20654 , 5.32174] = [734.8243 , 802.0502]

1. From the SAS output 6.37 we can get the Durbin-Watson d-statistic = 1.262 and we proceed with the Durbin Watson test at α = 0.05:

H0: Error terms are uncorrelated

Ha: Error terms are correlated

Following the procedure from the book we define k = 12, n = 168, and find the bounds for DW at α = 0.05 are (1.61 , 1.89). Therefore, given the d-statistic = 1.262 we must reject the null hypothesis in favor of the alternative that there is autocorrelation.

**Problem 4:**

1. p=(**1**)
2. p=(**1**,**2**)
3. p=(**1**) q=(**1**)
4. q=(**1**)
5. q=(**1**,**2**)
6. p=(**1**,**2**) q=(**1**)

From running several models (code below), the first one performed well with reasonable white noise, 2 was not good, 3,4,5 had very high white noise, and 6 had overall good ACF and PACF graphs as well as reasonable white noise and was preferred over the first model.

**Problem 5 (9.3):**

1. From figure 9.15 we can see that the values die down slowly, therefore, the time series values are nonstationary.
2. Using the formula and data points from 9.16(a) we have:

Where:

k = 3

n = 90

b = 1

r1 = 0.96846

r2 = 0.93652

r3 = 0.90414

Therefore:

1. First defining the necessary equations and data points (note zt values obtained from table 9.8):

(From figure 9.16)

b = 2

n = 90

For k = 11:

For k = 22:

For k = 33:

1. We can see on both the SAC and SPAC that the spikes cut off quickly. Therefore, the first differences are stationary.
2. On figure 9.16 we can see that on both the SAC and SPAC get “cut off” after lag 1

**Problem 5 (9.4):**

1. Considering both output from 9.15 we can see that because the original values are non-stationary we opt for the transformation:

zt = yt – yt-1

Moreover, from table 9.16 we have that said model dies down quickly after lag 1. Therefore we would like to use the nonseasonal autoregressive model of order 1:

zt = δ + ϕ1\*zt-1 + at­

1. From question (a) we have:

yt – yt-1 = δ + ϕ1\*zt-1 + at­

Therefore:

yt = δ + (ϕ1\*zt-1)\*yt-1 - ϕ1\*yt-2 + at­

1. The constant δ represents the “slope” of this model. Indicating that yt will change by δ each time period.
2. Calculating:
   1. Using values given and from table 9.7 the residual :

y1 = 235

y2 = 239

y3 = 244.09

249.656

* 1. Using observation 90 we predict t = 91, 92, 93:

Then given the model :

Calculating :

Calculating :

Calculating :

* 1. Calculating the prediction intervals from the predicted values :

PI[]: [1039.709 ± 5.4831]

PI[]: [1034.2259 , 1045.1921]

PI[]: [1049.399 ± 10.5683]

PI[]: [1038.8307 , 1059.9673]

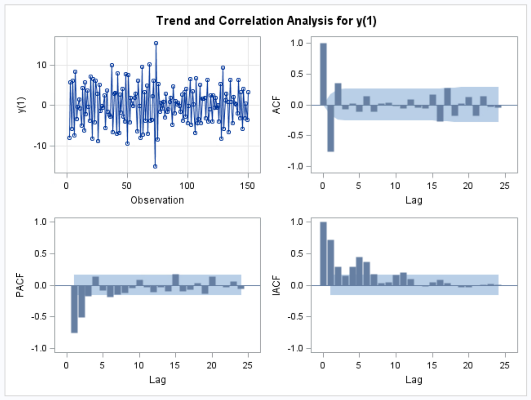
PI[]: [1058.74 ± 15.4975]

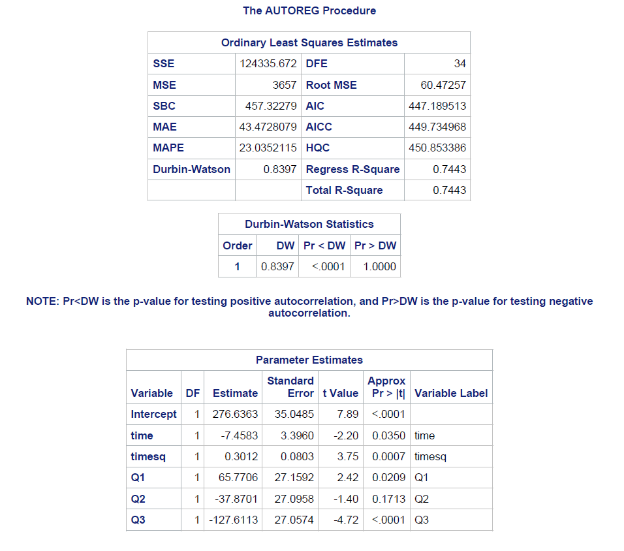
PI[]: [1043.2425 , 1074.2375]

**Problem 6:**

1. Looking at the SAC and SPAC we can see that the SAC has spikes at lag 1 and 2 and cuts off afterwards therefore a first differences could make sense. While the SPAC has spikes on 1, 2, 3 and 5, and an MA of 2 could make sense. Hence we have:

zt = yt – yt-1

zt = δ + at­ - θ1\*at-1 – θ2\*at-2

**Code:**

**Problem 1:**

ods html;

ods graphics on;

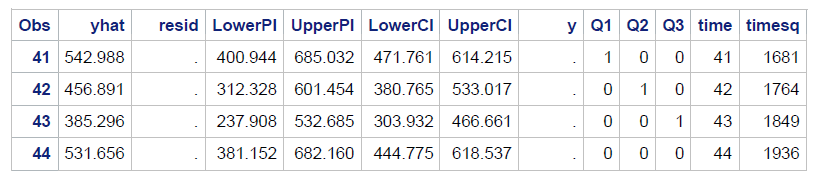
**proc** **print** data=energy;

**run**;

**proc** **autoreg** data=energy;

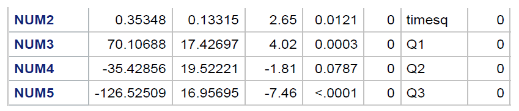
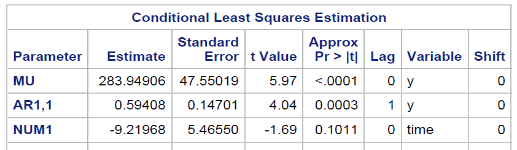
model y=time timesq Q1 Q2 Q3/dwprob;

output out=results p=yhat residual=resid lclm=LowerCI uclm=UpperCI lcl=LowerPI ucl=UpperPI;

**run**;

**proc** **print** data=results;

**run**;



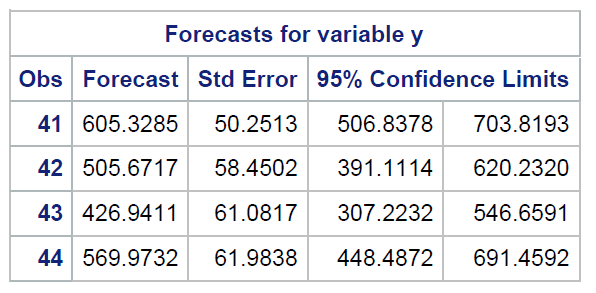
**proc** **arima** data=energy;

identify var=y crosscorr=(y time timesq Q1 Q2 Q3) noprint;

estimate p=(**1**) input=(time timesq Q1 Q2 Q3) printall plot;

forecast lead=**4** out=fcast;

**run**;



**proc** **print** data=fcast;

**run**;

ods graphics off;

ods html;

**Problem 2:**

ods html;

ods graphics on;

**proc** **print** data=japan;

**run**;

**proc** **autoreg** data=japan plots=(acf pacf);

model y=time/nlag=**1** dwprob method=uls;

output out=results p=yhat residual=resid lclm=LowerCI uclm=UpperCI lcl=LowerPI ucl=UpperPI;

**run**;

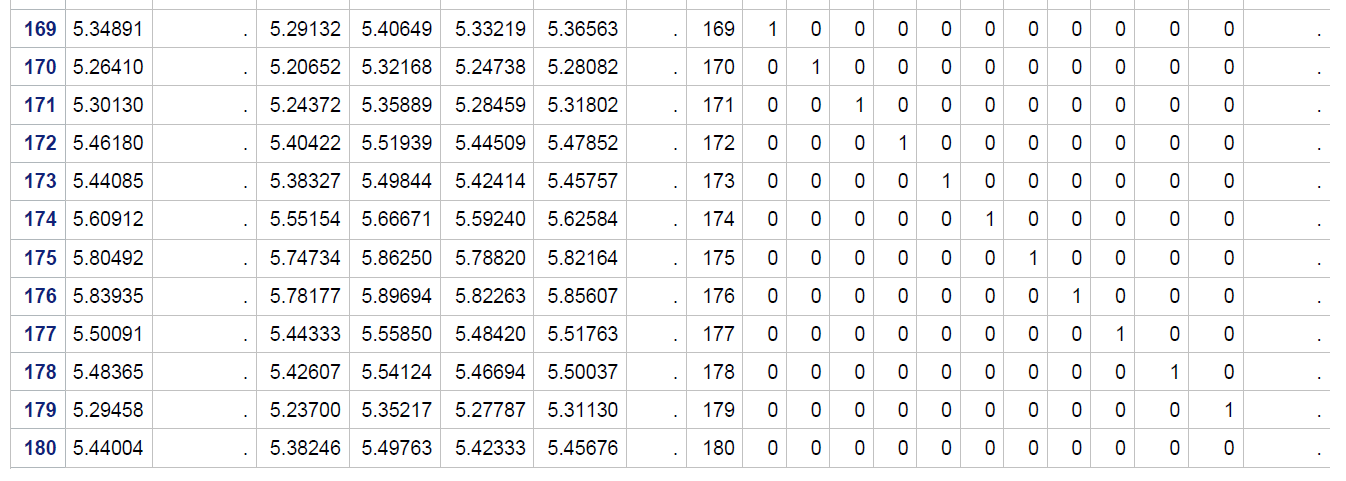
**proc** **print** data=results;

**run**;

ods graphics off;

ods html;

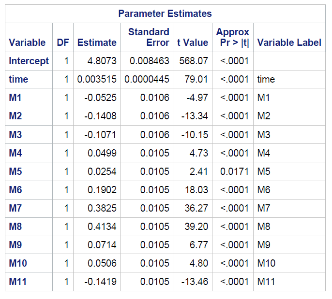
**Problem 3:**

ods html;

ods graphics on;

**proc** **autoreg** data=qhotel;

model quarty=time M1 M2 M3 M4 M5 M6 M7 M8 M9 M10 M11/dwprob;

output out=results p=yhat residual=resid lclm=LowerCI uclm=UpperCI lcl=LowerPI ucl=UpperPI;

**run**;

**proc** **print** data=results;

**run**;

ods graphics off;

ods html;

**Problem 4:**

ods html;

ods graphics on;

**proc** **arima** data=japan;

identify var=y crosscorr=(time) noprint;

estimate p=(**1**) input=(time) printall plot;

forecast lead=**10** out=fcast;

**run**;

**proc** **arima** data=japan;

identify var=y crosscorr=(time) noprint;

estimate p=(**1**,**2**) input=(time) printall plot;

forecast lead=**10** out=fcast;

**run**;

**proc** **arima** data=japan;

identify var=y crosscorr=(time) noprint;

estimate p=(**1**) q=(**1**) input=(time) printall plot;

forecast lead=**10** out=fcast;

**run**;

**proc** **arima** data=japan;

identify var=y crosscorr=(time) noprint;

estimate q=(**1**) input=(time) printall plot;

forecast lead=**10** out=fcast;

**run**;

**proc** **arima** data=japan;

identify var=y crosscorr=(time) noprint;

estimate q=(**1**,**2**) input=(time) printall plot;

forecast lead=**10** out=fcast;

**run**;

**proc** **arima** data=japan;

identify var=y crosscorr=(time) noprint;

estimate p=(**1**,**2**) q=(**1**) input=(time) printall plot;

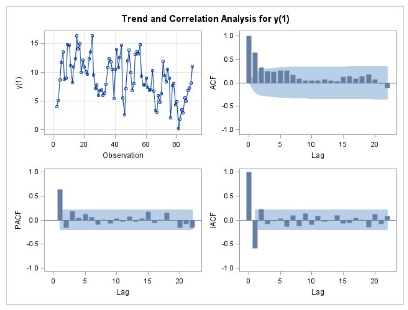
forecast lead=**10** out=fcast;

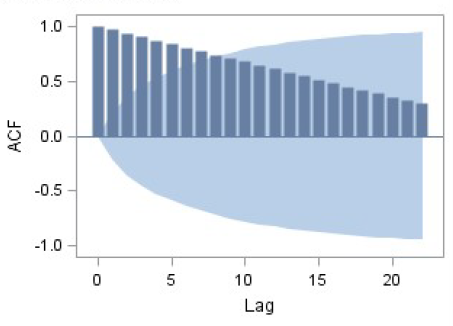
**run**;

ods graphics off;

ods html;

**Problem 5:**

ods html;

ods graphics on;

**proc** **arima** data=tooth;

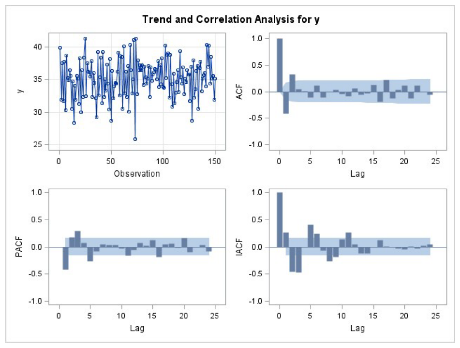
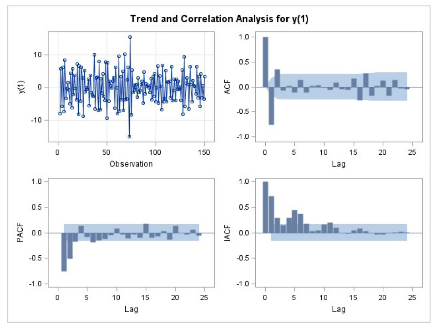
identify var=y;

identify var=y(**1**);

**run**;

ods graphics off;

ods html;

**Problem 6:**

ods html;

ods graphics on;

**proc** **arima** data=visc;

identify var=y;

identify var=y(**1**);

**run**;

ods graphics off;

ods html;