QUESTION 1:

[A] Run the program unmodified. From the correlations, and keeping in mind the offset, what do you think is a typical delay (in days) between an order being called in and the resulting shipment?

The correlation between shipments and calls lagged by 1 is roughly: 0.98. However, recall the calls data set begins on Feb. 10, 1978; whereas, the shipments data set begins on following day. Hence, the firm has a shipping delay of 2 days.

<table>
<thead>
<tr>
<th>Earns Correlation Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob &gt;</td>
</tr>
<tr>
<td>Number of Observations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>shpmnts</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>calls</td>
<td>-0.07321</td>
</tr>
<tr>
<td></td>
<td>0.5004</td>
</tr>
<tr>
<td></td>
<td>87</td>
</tr>
</tbody>
</table>

| call1              | 0.98665 |
|                    | <.0001  |
|                    | 86     |

| call2              | -0.03692 |
|                    | 0.7373   |
|                    | 85     |

| call3              | -0.13244 |
|                    | 0.2298   |
|                    | 84     |

[B] Plot the graph of shipments versus the lagged calls (lagged as you decided above).
QUESTION 2:
Using the information in the class notes, create date variables for both data sets. Then merge by date as suggested (remove the * from the by date statement) Plot calls and shipments against date.
**QUESTION 3:**
Now, either before or after merging, use PROC EXPAND to fill in some values for shipments where they are missing. In this way, get a dataset with date, calls and shipments where missing shipments have been imputed. Repeat the correlation and the plot from part (1) with the newly completed shipments data. Comment on the effect of the imputations.

The correlation between shipments and calls lagged by 2 is roughly: 0.88; whereas, the correlation between shipments and calls lagged by 1 is no longer significant, as a result from merging the data sets by date. Moreover, the correlation structure has become slightly less significant by imputing the missing values in the shipment series by PROC EXPAND. A graph of shipments versus calls lagged by 2 is below.

|                      | Pearson Correlation Coefficients | Prob > |r| under H0: Rho=0 | Number of Observations |
|----------------------|----------------------------------|---------|----------------|------------------------|
|                      |                                  |         |                |                        |
| ship2                |                                  |         |                |                        |
| calls                | -0.09089                         | 0.3709  | 99             |                        |
| call1                | 0.00101                          | 0.9921  | 100            |                        |
| call2                | 0.88388                          | <.0001  | 99             |                        |
| call3                | 0.07364                          | 0.4711  | 98             |                        |

![Graph of shipments versus calls lagged by 2](image-url)
QUESTION 4:
Give the equation that is used to interpolate between the 5th and 7th value of shipments.

From the 5th equation below we can readily attain the imputed value for 16FEB1978.

\[ G(X) = 121 + 27.10X - 48.54X^2 + 13.24X^3 \]
\[ G(X=0) = 121 = 15FEB1978 \]
\[ G(X=1) = 112.81 = 16FEB1978 \]

<table>
<thead>
<tr>
<th>Obs</th>
<th>VARNAME</th>
<th>METHOD</th>
<th>OBSERVED</th>
<th>date</th>
<th>CONSTANT</th>
<th>LINEAR</th>
<th>QUAD</th>
<th>CUBIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ship</td>
<td>SPLINE</td>
<td>10FEB1978</td>
<td>172.146</td>
<td>-71.1462</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>ship</td>
<td>SPLINE</td>
<td>11FEB1978</td>
<td>101.000</td>
<td>-71.1462</td>
<td>74.9693</td>
<td>-19.8231</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ship</td>
<td>SPLINE</td>
<td>13FEB1978</td>
<td>100.000</td>
<td>-9.1462</td>
<td>-43.9693</td>
<td>36.1155</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ship</td>
<td>SPLINE</td>
<td>14FEB1978</td>
<td>83.000</td>
<td>11.2616</td>
<td>64.3771</td>
<td>-37.6387</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>ship</td>
<td>SPLINE</td>
<td>15FEB1978</td>
<td>121.000</td>
<td>27.0996</td>
<td>-48.5391</td>
<td>13.2446</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ship</td>
<td>SPLINE</td>
<td>17FEB1978</td>
<td>87.000</td>
<td>-8.1211</td>
<td>30.9287</td>
<td>-6.8076</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ship</td>
<td>SPLINE</td>
<td>18FEB1978</td>
<td>103.000</td>
<td>33.3135</td>
<td>10.5059</td>
<td>-22.8194</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ship</td>
<td>SPLINE</td>
<td>19FEB1978</td>
<td>124.000</td>
<td>-14.1329</td>
<td>-57.9523</td>
<td>29.0853</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ship</td>
<td>SPLINE</td>
<td>20FEB1978</td>
<td>81.000</td>
<td>-42.7818</td>
<td>29.3034</td>
<td>1.4784</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ship</td>
<td>SPLINE</td>
<td>21FEB1978</td>
<td>69.000</td>
<td>20.2602</td>
<td>33.7386</td>
<td>-15.9343</td>
<td></td>
</tr>
</tbody>
</table>

QUESTION 5:
Explain what these are doing:

(obs=10) : prints first 10 observations of a data set in proc print.
the label option in proc print: prints the variable label NOT the variable name in the output window.
call1-call3 (versus call1 call3 without the "."): call1-call3 = call1 call2 call3
call1 call3 = call1 call3.

QUESTION 6: (optional)
Try this in the merged dataset.

PROC ARIMA; I VAR=SHPMNTS CROSSCOR=(CALLS); RUN;
APPENDIX:

Turn in complete SAS program:

```sas
dm 'log; clear; output; clear('; QUIT;
options ps=80 ls=100 pageno=1;
goptions reset=global border ftext=swiss gunit=cm htext=0.4 htitle=0.5;
goptions gsfmode=replace gaccess=gsasfile;

/*;*********************************************************************;
AUTHOR: R. Adam Hoppes
COURSE: ST 730, FALL 2004
DATE:  8/22/2004
UPDATE: 
PURPOSE: Homework 01:
    review date methods, cubic splines, merging, and plots
*********************************************************************;*/
title1 'ST 730: Homework 1: R.A. Hoppes';

** QUESTION 1 **********;
** UNMODIFIED;
/*/Calls to a shipping firm. Calls taken daily, 7 days a week.
    First date is Feb. 10, 1978*/
Data calls; input calls @@;
cards;
  85 101  82 120  89 104 125  83 70
  81 117  81 130 119  95  93 115  77 111
 124  91 140  83  73 118  66  82 101  92
 107 110 114 116  81  93  98  84 108 101
 105  90  84 112 115  79 104  72 101 104
  87 117 131 109  97 143  86  86 112 117
  84  87  89  77 153 101 115  72 101  83
 114 114  80 112 130 127 120  79  42 102
  97  94 129 100  89  91  99 103  81
  79  76 106 111  93 102 116  82  99  86
; /*Calls to a shipping firm. Shipments made daily, 7 days a week.
    First date is Feb. 11, 1978 */
Data ship; Input shpmnts @@;
cards;
 101  85 100  83 121  .  87 103 124  81
  69  . 117  74 124 115  94  . 114  72
 111 117  92 139  .  71 113  66  86 105
  91 113 112  . 115  80 100  95  85 108
 102  .  90  84 116 121  78 102  71 101
 103  90 124 134 112  95 149  83  86 110
 118  .  92  89  73 148  . 111  .  75
  83 115 115  . 115  . 124 119  85  37
 101  .  94 126 103  .  88  92  99 102
  82  88  79 106 110  91 105 115  85  98
;
data all;
merge calls ship; * by date;
call1=lag(calls); call2=lag2(calls); call3=lag3(calls);
label call1 = "Calls lagged by 1";
label call2 = "Calls lagged by 2";
label call3 = "Calls lagged by 3";
label shpmnts="Shipments";
run;
```
proc print data=all(obs=10) label;run;
proc corr; var shpmnts; with calls call1-call3;run;
*END CODE;

*Append the date set all with the new variable t;
*the variable _n_ is a SAS defined variable;
data all;
   set all;
   t = _n_;
run; quit;

** QUESTION 2 **********;
**plot shipments*call1**;
proc gplot data=all;
title2 'Shipments vs Lagged Calls (1) with reference data 2/13/1978: t=1';
plot shpmnts*call1 / frame haxis=axis1 vaxis=axis2 legend=legend1 grid;
symbol1 v=dot height=0.2 cv=green;
axis1 label = ('calls lagged by 1')
   length = 15;
axis2 label = (a=90 'Shipments')
   length = 15;
run;quit;
title2' ';

** QUESTION 3 **********;
data set1;
do i = 0 to 99;
   input calls @@;
   date = intnx('day', '10feb1978'd, i);
   format date date9.;
drop i;
output;
end;
cards;
   85 101  82 120  89  85 104 125  83 70
   81 117  81 130 119  95  93 115  77 111
  124  91 140  81 117 116  81  93  98  84 108 101
  107 110 114 116  81  93  98  84 108 101
  105  90  84 112 115  79 104  72 101 104
  87 117 131 109  97 143  86  86 112 117
  84  87  89  77 153 101 113 100  78  83
  114 114  80 112 130 127 120  79  42 102
  97  94 129 110  89  91  99 103  81
  79  76 106 111  93 102 116  82  99  86
;
run;

data set2;
do i = 0 to 99;
   input ship @@;
   date = intnx('day', '11feb1978'd, i);
   format date date9.;
drop i;
output;
end;
cards;
   101  85 100  83 121 .  87 103 124  81
   69 . 117  74 124 115  94 . 114 72
  111 117  92 139 .  71 113  66  86 105
  91 113 112 . 115  80 100  95  85 108
  102 .  90  84 116 121  78 102  71 101
;
run;

data set3;
  merge set1 set2; by date;
    t = _n_;  
    call1=lag(calls); call2=lag2(calls); call3=lag3(calls);
    label call1 = "Calls lagged by 1";
    label call2 = "Calls lagged by 2";
    label call3 = "Calls lagged by 3";
    label ship = "Shipments";
  run;

proc print data=set3(obs=10); run;

** PLOT 1: points**;
proc gplot data=set3;
  title2 'Calls and Shipments versus date';
  plot (calls ship)*date / overlay frame haxis=axis1 vaxis=axis2 legend=legend1 grid;
  symbol1 v=dot i=none height=0.2 cv=green;
  symbol2 v=star i=none height=0.2 cv=blue;
  axis1 label = ('Date' ) length = 15;
  axis2 label = (a=90 'Calls/Shipments' ) length = 15;
  legend1 label = none position = (middle bottom outside) across = 2 down = 1 mode = reserve frame offset = (0.5, 0.5) value = (j=l h=0.3 'Calls' 'Shipments');
run;quit;

** PLOT 2: i=join**;
proc gplot data=set3;
  title2 'Calls and Shipments versus date';
  plot (calls ship)*date / overlay frame haxis=axis1 vaxis=axis2 legend=legend1 grid;
  symbol1 v=none i=join height=0.2 cv=green;
  symbol2 v=none i=join height=0.2 cv=blue;
  axis1 label = ('Date' ) length = 15;
  axis2 label = (a=90 'Calls/Shipments' ) length = 15;
  legend1 label = none position = (middle bottom outside) across = 2 down = 1 mode = reserve frame offset = (0.5, 0.5) shape = line(1) value = (j=l h=0.3 'Calls' 'Shipments');
run;quit;

title2 ' ';
** QUESTION 3 **********;
proc expand data = set3 from=day to=day out=out1 outest=spline;
  label ship2 = interpolated shipments;
  id date;
run;

proc print data=out1(obs=10) label; run;
proc corr data=out1; var ship2; with calls call1-call3; run;

**plot shipments*call2**;
proc gplot data=out1;
  title2 'Interpolated Shipments vs Lagged Calls (2)';
  plot ship*call2 / frame haxis=axis1 vaxis=axis2 legend=legend1 grid;
  symbol1 v=dot i=none height=0.2 cv=green;
  axis1 label = ('calls lagged by 1') length = 15;
  axis2 label = (a=90 'Shipments') length = 15;
run;quit;
** QUESTION 4 **********;
proc print data=spline(obs=10); run;

** QUESTION 5 **********;
** QUESTION 6 **********;
proc arima data=set3;
  identify var=ship crosscor=(calls);
run;quit;