**Getting Started**

To use PROC FORECAST, specify the input and output data sets and the number of periods to forecast in the PROC FORECAST statement, then list the variables to forecast in a VAR statement.

For example, suppose you have monthly data on the sales of some product, in a data set, named PAST, as shown in Figure 12.1, and you want to forecast sales for the next 10 months.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Obs** | **date** | **sales** | | --- | --- | --- | | 1 | JUL89 | 9.5161 | | 2 | AUG89 | 9.6994 | | 3 | SEP89 | 9.2644 | | 4 | OCT89 | 9.6837 | | 5 | NOV89 | 10.0784 | | 6 | DEC89 | 9.9005 | | 7 | JAN90 | 10.2375 | | 8 | FEB90 | 10.6940 | | 9 | MAR90 | 10.6290 | | 10 | APR90 | 11.0332 | | 11 | MAY90 | 11.0270 | | 12 | JUN90 | 11.4165 | | 13 | JUL90 | 11.2918 | | 14 | AUG90 | 11.3475 | | 15 | SEP90 | 11.2913 | | 16 | OCT90 | 11.3771 | | 17 | NOV90 | 11.5457 | | 18 | DEC90 | 11.6433 | | 19 | JAN91 | 11.9293 | | 20 | FEB91 | 11.9752 | | 21 | MAR91 | 11.9283 | | 22 | APR91 | 11.8985 | | 23 | MAY91 | 12.0419 | | 24 | JUN91 | 12.3537 | | 25 | JUL91 | 12.4546 | |

**Figure 12.1:** Example Data Set PAST

The following statements forecast 10 observations for the variable SALES using the default STEPAR method and write the results to the output data set PRED:

**proc forecast data=past lead=10 out=pred;**

**var sales;**

**run;**

The following statements use the PRINT procedure to print the data set PRED:

**proc print data=pred;**

**run;**

The PROC PRINT listing of the forecast data set PRED is shown in Figure 12.2.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Obs** | **\_TYPE\_** | **\_LEAD\_** | **sales** | | --- | --- | --- | --- | | 1 | FORECAST | 1 | 12.6205 | | 2 | FORECAST | 2 | 12.7665 | | 3 | FORECAST | 3 | 12.9020 | | 4 | FORECAST | 4 | 13.0322 | | 5 | FORECAST | 5 | 13.1595 | | 6 | FORECAST | 6 | 13.2854 | | 7 | FORECAST | 7 | 13.4105 | | 8 | FORECAST | 8 | 13.5351 | | 9 | FORECAST | 9 | 13.6596 | | 10 | FORECAST | 10 | 13.7840 | |

**Figure 12.2:** Forecast Data Set PRED

***Giving Dates to Forecast Values***

Normally, your input data set has an ID variable that gives dates to the observations, and you want the forecast observations to have dates also. Usually, the ID variable has SAS date values. (See Chapter 2, "Working with Time Series Data," for information on using SAS date values.) The ID statement specifies the identifying variable.

If the ID variable contains SAS date values, the INTERVAL= option should be used on the PROC FORECAST statement to specify the time interval between observations. (See Chapter 3, "Date Intervals, Formats, and Functions," for more information on time intervals.) The FORECAST procedure uses the INTERVAL= option to generate correct dates for forecast observations.

The data set PAST, shown in Figure 12.1, has monthly observations and contains an ID variable DATE with SAS date values identifying each observation. The following statements produce the same forecast as the preceding example and also include the ID variable DATE in the output data set. Monthly SAS date values are extrapolated for the forecast observations.

**proc forecast data=past interval=month lead=10 out=pred;**

**var sales;**

**id date;**

**run;**

***Computing Confidence Limits***

Depending on the output options specified, multiple observations are written to the OUT= data set for each time period. The different parts of the results are contained in the VAR statement variables in observations identified by the character variable \_TYPE\_ and by the ID variable.

For example, the following statements use the OUTLIMIT option to write forecasts and 95% confidence limits for the variable SALES to the output data set PRED. This data set is printed with the PRINT procedure.

**proc forecast data=past interval=month lead=10**

**out=pred outlimit;**

**var sales;**

**id date;**

**run;**

**proc print data=pred;**

**run;**

The output data set PRED is shown in Figure 12.3.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Obs** | **date** | **\_TYPE\_** | **\_LEAD\_** | **sales** | | --- | --- | --- | --- | --- | | 1 | AUG91 | FORECAST | 1 | 12.6205 | | 2 | AUG91 | L95 | 1 | 12.1848 | | 3 | AUG91 | U95 | 1 | 13.0562 | | 4 | SEP91 | FORECAST | 2 | 12.7665 | | 5 | SEP91 | L95 | 2 | 12.2808 | | 6 | SEP91 | U95 | 2 | 13.2522 | | 7 | OCT91 | FORECAST | 3 | 12.9020 | | 8 | OCT91 | L95 | 3 | 12.4001 | | 9 | OCT91 | U95 | 3 | 13.4039 | | 10 | NOV91 | FORECAST | 4 | 13.0322 | | 11 | NOV91 | L95 | 4 | 12.5223 | | 12 | NOV91 | U95 | 4 | 13.5421 | | 13 | DEC91 | FORECAST | 5 | 13.1595 | | 14 | DEC91 | L95 | 5 | 12.6435 | | 15 | DEC91 | U95 | 5 | 13.6755 | | 16 | JAN92 | FORECAST | 6 | 13.2854 | | 17 | JAN92 | L95 | 6 | 12.7637 | | 18 | JAN92 | U95 | 6 | 13.8070 | | 19 | FEB92 | FORECAST | 7 | 13.4105 | | 20 | FEB92 | L95 | 7 | 12.8830 | | 21 | FEB92 | U95 | 7 | 13.9379 | | 22 | MAR92 | FORECAST | 8 | 13.5351 | | 23 | MAR92 | L95 | 8 | 13.0017 | | 24 | MAR92 | U95 | 8 | 14.0686 | | 25 | APR92 | FORECAST | 9 | 13.6596 | | 26 | APR92 | L95 | 9 | 13.1200 | | 27 | APR92 | U95 | 9 | 14.1993 | | 28 | MAY92 | FORECAST | 10 | 13.7840 | | 29 | MAY92 | L95 | 10 | 13.2380 | | 30 | MAY92 | U95 | 10 | 14.3301 | |

**Figure 12.3:** Output Data Set

***Form of the OUT= Data Set***

The OUT= data set PRED, shown in Figure 12.3, contains three observations for each of the 10 forecast periods. Each of these three observations has the same value of the ID variable DATE, the SAS date value for the month and year of the forecast.

The three observations for each forecast period have different values of the variable \_TYPE\_. For the \_TYPE\_=FORECAST observation, the value of the variable SALES is the forecast value for the period indicated by the DATE value. For the \_TYPE\_=L95 observation, the value of the variable SALES is the lower limit of the 95% confidence interval for the forecast. For the \_TYPE\_=U95 observation, the value of the variable SALES is the upper limit of the 95% confidence interval.

You can control the types of observations written to the OUT= data set with the PROC FORECAST statement options OUTLIMIT, OUTRESID, OUTACTUAL, OUT1STEP, OUTSTD, OUTFULL, and OUTALL. For example, the OUTFULL option outputs the confidence limit values, the one-step-ahead predictions, and the actual data, in addition to the forecast values. See the sections "Syntax" and "OUT= Data Set" later in this chapter for more information.

***Plotting Forecasts***

The forecasts, confidence limits, and actual values can be plotted on the same graph with the GPLOT procedure. Use the appropriate output control options on the PROC FORECAST statement to include in the OUT= data set the series you want to plot. Use the \_TYPE\_ variable in the GPLOT procedure PLOT statement to separate the observations for the different plots.

In this example, the OUTFULL option is used, and the resulting output data set contains the actual and predicted values, as well as the upper and lower 95

**proc forecast data=past interval=month lead=10**

**out=pred outfull;**

**id date;**

**var sales;**

**run;**

**proc gplot data=pred;**

**plot sales \* date = \_type\_ /**

**haxis= '1jan90'd to '1jan93'd by qtr**

**HREF='15jul91'd;**

**symbol1 i=none v=star; /\* for \_type\_=ACTUAL \*/**

**symbol2 i=spline v=circle; /\* for \_type\_=FORECAST \*/**

**symbol3 i=spline l=3; /\* for \_type\_=L95 \*/**

**symbol4 i=spline l=3; /\* for \_type\_=U95 \*/**

**where date >= '1jan90'd;**

**run;**

The \_TYPE\_ variable is used in the GPLOT procedure's PLOT statement to make separate plots over time for each type of value. A reference line marks the start of the forecast period. (Refer to *SAS/GRAPH Software: Reference, Volume 2, Version 7, First Edition* for more information on using PROC GPLOT.) The WHERE statement restricts the range of the actual data shown in the plot. In this example, the variable SALES has monthly data from July 1989 through July 1991, but only the data for 1990 and 1991 are shown in the plot.

The plot is shown in Figure 12.4.

|  |
| --- |
| forgs04.gif (4665 bytes) |

**Figure 12.4:** Plot of Forecast with Confidence Limits

***Plotting Residuals***

You can plot the residuals from the forecasting model using PROC GPLOT and a WHERE statement.

1. Use the OUTRESID option or the OUTALL option in the PROC FORECAST statement to include the residuals in the output data set.
2. Use a WHERE statement to specify the observation type of 'RESIDUAL' in the PROC GPLOT code.

The following example adds the OUTRESID option to the preceding example and plots the residuals:

**proc forecast data=past interval=month lead=10**

**out=pred outfull outresid;**

**id date;**

**var sales;**

**run;**

**proc gplot data=pred;**

**where \_type\_='RESIDUAL';**

**plot sales \* date /**

**haxis= '1jan89'd to '1oct91'd by qtr;**

**symbol1 i=circle;**

**run;**

The plot of residuals is shown in Figure 12.5.

|  |
| --- |
| forgs05.gif (4568 bytes) |

**Figure 12.5:** Plot of Residuals

***Model Parameters and Goodness-of-Fit Statistics***

You can write the parameters of the forecasting models used, as well as statistics measuring how well the forecasting models fit the data, to an output SAS data set using the OUTEST= option. The options OUTFITSTATS, OUTESTTHEIL, and OUTESTALL control what goodness-of-fit statistics are added to the OUTEST= data set.

For example, the following statements add the OUTEST= and OUTFITSTATS options to the previous example to create the output statistics data set EST for the results of the default stepwise autoregressive forecasting method:

**proc forecast data=past interval=month lead=10**

**out=pred outfull outresid**

**outest=est outfitstats;**

**id date;**

**var sales;**

**run;**

**proc print data=est;**

**run;**

The PRINT procedure prints the OUTEST= data set, as shown in Figure 12.6.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Obs** | **\_TYPE\_** | **date** | **sales** | | --- | --- | --- | --- | | 1 | N | JUL91 | 25 | | 2 | NRESID | JUL91 | 25 | | 3 | DF | JUL91 | 22 | | 4 | SIGMA | JUL91 | 0.2001613 | | 5 | CONSTANT | JUL91 | 9.4348822 | | 6 | LINEAR | JUL91 | 0.1242648 | | 7 | AR1 | JUL91 | 0.5206294 | | 8 | AR2 | JUL91 | . | | 9 | AR3 | JUL91 | . | | 10 | AR4 | JUL91 | . | | 11 | AR5 | JUL91 | . | | 12 | AR6 | JUL91 | . | | 13 | AR7 | JUL91 | . | | 14 | AR8 | JUL91 | . | | 15 | SST | JUL91 | 21.28342 | | 16 | SSE | JUL91 | 0.8793714 | | 17 | MSE | JUL91 | 0.0399714 | | 18 | RMSE | JUL91 | 0.1999286 | | 19 | MAPE | JUL91 | 1.2280089 | | 20 | MPE | JUL91 | -0.050139 | | 21 | MAE | JUL91 | 0.1312115 | | 22 | ME | JUL91 | -0.001811 | | 23 | MAXE | JUL91 | 0.3732328 | | 24 | MINE | JUL91 | -0.551605 | | 25 | MAXPE | JUL91 | 3.2692294 | | 26 | MINPE | JUL91 | -5.954022 | | 27 | RSQUARE | JUL91 | 0.9586828 | | 28 | ADJRSQ | JUL91 | 0.9549267 | | 29 | RW\_RSQ | JUL91 | 0.2657801 | | 30 | ARSQ | JUL91 | 0.9474145 | | 31 | APC | JUL91 | 0.044768 | | 32 | AIC | JUL91 | -77.68559 | | 33 | SBC | JUL91 | -74.02897 | | 34 | CORR | JUL91 | 0.9791313 | |

**Figure 12.6:** The OUTEST= Data Set for STEPAR Method

In the OUTEST= data set, the DATE variable contains the ID value of the last observation in the data set used to fit the forecasting model. The variable SALES contains the statistic indicated by the value of the \_TYPE\_ variable. The \_TYPE\_=N, NRESID, and DF observations contain, respectively, the number of observations read from the data set, the number of nonmissing residuals used to compute the goodness-of-fit statistics, and the number of nonmissing observations minus the number of parameters used in the forecasting model.

The observation having \_TYPE\_=SIGMA contains the estimate of the standard deviation of the one-step prediction error computed from the residuals. The \_TYPE\_=CONSTANT and \_TYPE\_=LINEAR contain the coefficients of the time trend regression. The \_TYPE\_=AR1, AR2, ..., AR8 observations contain the estimated autoregressive parameters. A missing autoregressive parameter indicates that the autoregressive term at that lag was not included in the model by the stepwise model selection method. (See the section "STEPAR Method" later in this chapter for more information.)

The other observations in the OUTEST= data set contain various goodness-of-fit statistics that measure how well the forecasting model used fits the given data. See "OUTEST= Data Set" later in this chapter for details.

***Controlling the Forecasting Method***

The METHOD= option controls which forecasting method is used. The TREND= option controls the degree of the time trend model used. For example, the following statements produce forecasts of SALES as in the preceding example but use the double exponential smoothing method instead of the default STEPAR method:

**proc forecast data=past interval=month lead=10**

**method=expo trend=2**

**out=pred outfull outresid**

**outest=est outfitstats;**

**var sales;**

**id date;**

**run;**

**proc print data=est;**

**run;**

The PRINT procedure prints the OUTEST= data set for the EXPO method, as shown in Figure 12.7.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | **Obs** | **\_TYPE\_** | **date** | **sales** | | --- | --- | --- | --- | | 1 | N | JUL91 | 25 | | 2 | NRESID | JUL91 | 25 | | 3 | DF | JUL91 | 23 | | 4 | WEIGHT | JUL91 | 0.1055728 | | 5 | S1 | JUL91 | 11.427657 | | 6 | S2 | JUL91 | 10.316473 | | 7 | SIGMA | JUL91 | 0.2545069 | | 8 | CONSTANT | JUL91 | 12.538841 | | 9 | LINEAR | JUL91 | 0.1311574 | | 10 | SST | JUL91 | 21.28342 | | 11 | SSE | JUL91 | 1.4897965 | | 12 | MSE | JUL91 | 0.0647738 | | 13 | RMSE | JUL91 | 0.2545069 | | 14 | MAPE | JUL91 | 1.9121204 | | 15 | MPE | JUL91 | -0.816886 | | 16 | MAE | JUL91 | 0.2101358 | | 17 | ME | JUL91 | -0.094941 | | 18 | MAXE | JUL91 | 0.3127332 | | 19 | MINE | JUL91 | -0.460207 | | 20 | MAXPE | JUL91 | 2.9243781 | | 21 | MINPE | JUL91 | -4.967478 | | 22 | RSQUARE | JUL91 | 0.930002 | | 23 | ADJRSQ | JUL91 | 0.9269586 | | 24 | RW\_RSQ | JUL91 | -0.243886 | | 25 | ARSQ | JUL91 | 0.9178285 | | 26 | APC | JUL91 | 0.0699557 | | 27 | AIC | JUL91 | -66.50591 | | 28 | SBC | JUL91 | -64.06816 | | 29 | CORR | JUL91 | 0.9772418 | |

**Figure 12.7:** The OUTEST= Data Set for METHOD=EXPO