

Time Series – Descriptive Methods

Summary

The **Descriptive Methods** procedure creates various tables and plots for time series data. A *time series* consists of a set of sequential numeric data taken at equally spaced intervals, usually over a period of time or space. The procedure plots the data and displays the autocorrelations, partial autocorrelations, and sample periodogram. Tests are conducted to determine whether or not the observations could be samples from a random or “white noise” process. If a second time series is supplied, crosscorrelations between the two series are also calculated and displayed.

Sample StatFolio: *tsdescribe.sgp*

Sample Data:

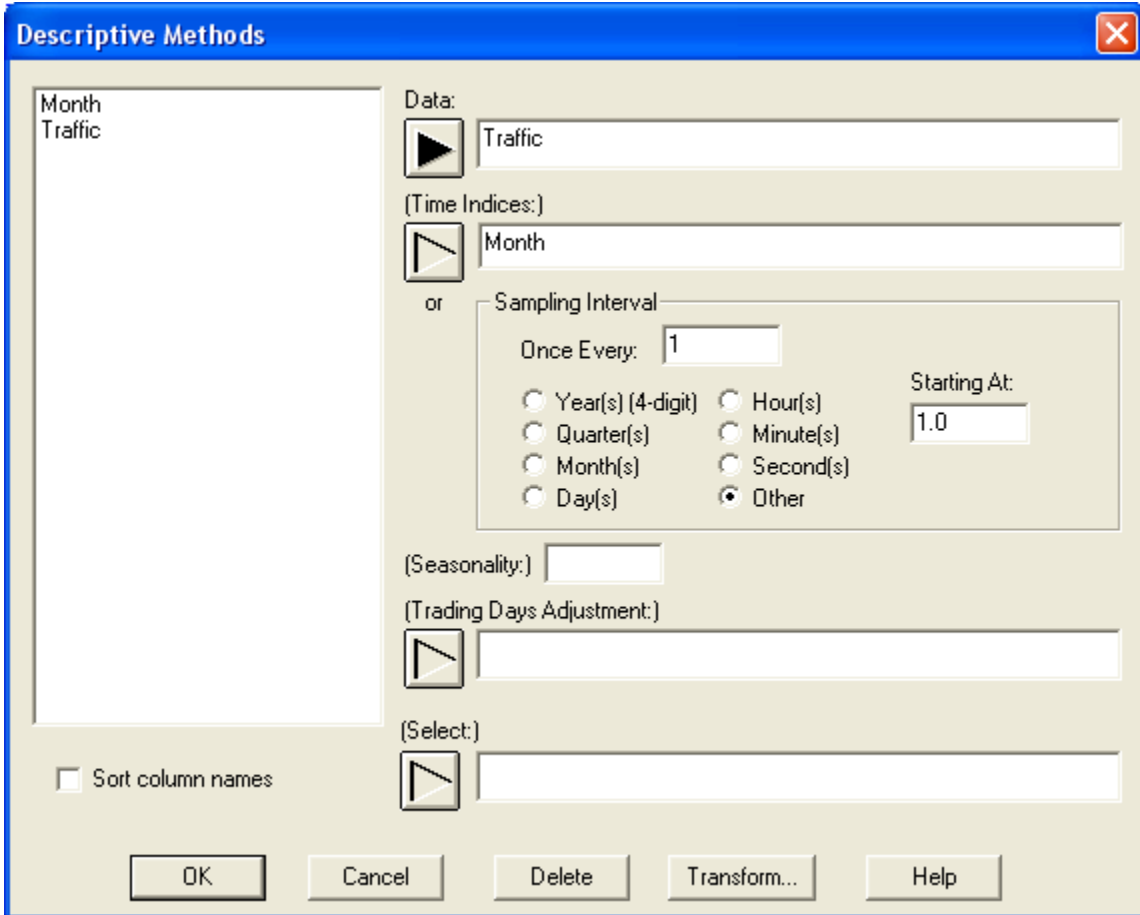
The file *golden gate.sgd* contains monthly traffic volumes on the Golden Gate Bridge in San Francisco for a period of $n = 168$ months from January, 1968 through December, 1981. The table below shows a partial list of the data from that file:

<i>Month</i>	<i>Traffic</i>
1/68	73.637
2/68	77.136
3/68	81.481
4/68	84.127
5/68	84.562
6/68	91.959
7/68	94.174
8/68	96.087
9/68	88.952
10/68	83.479
11/68	80.814
12/68	77.466
1/69	75.225
...	...

The data were obtained from a publication of the Golden Gate Bridge.

Data Input

The data input dialog box requests the name of the column containing the time series data:



- **Data:** numeric column containing n equally spaced numeric observations.
- **Time indices:** time, date or other index associated with each observation. Each value in this column must be unique and arranged in ascending order.
- **Sampling Interval:** If time indices are not provided, this defines the interval between successive observations. For example, the data from the Golden Gate Bridge were collected once every *month*, beginning in 01/68.
- **Seasonality:** the length of seasonality s , if any. The data is seasonal if there is a pattern that repeats at a fixed period. For example, monthly data such as traffic on the Golden Gate Bridge have a seasonality of $s = 12$. Hourly data that repeat every day have a seasonality of $s = 24$. If no entry is made, the data is assumed to be nonseasonal ($s = 1$).

- **Trading Days Adjustment:** a numeric variable with n observations used to normalize the original observations, such as the number of working days in a month. The observations in the *Data* column will be divided by these values before being plotted or analyzed.
- **Select:** subset selection.

Analysis Summary

The *Analysis Summary* displays the number of observations in the time series and the length of seasonality:

Descriptive Methods - Traffic	
Data variable: Traffic (Golden Gate Bridge Traffic Volume)	
Number of observations = 168	
Start index = 1/68	
Sampling interval = 1.0 month(s)	
Length of seasonality = 12	

Note: a limited amount of missing data is permitted, providing there are not too many missing values close together. Missing values are replaced by interpolated values according to the method outlined in the *Calculations* section.

Data Table

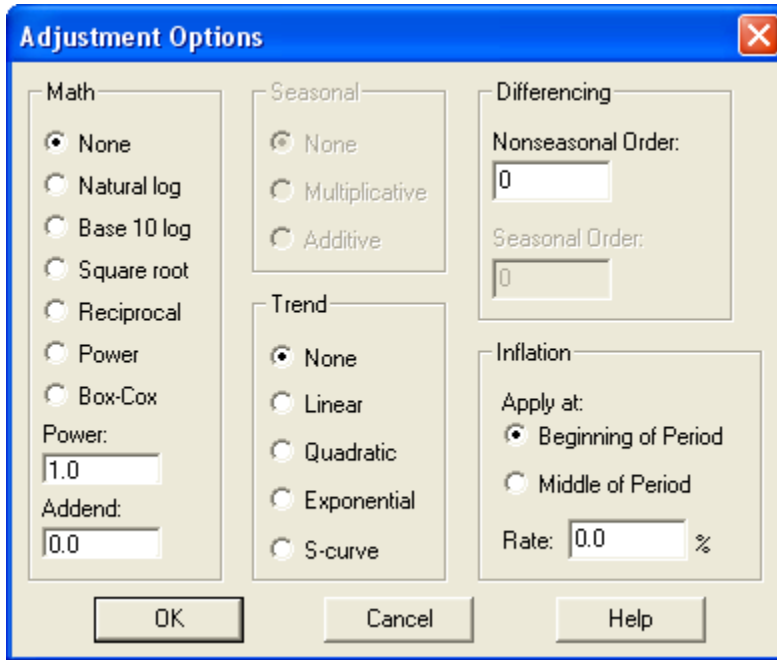
The *Data Table* displays the input data:

Data Table for Traffic		
<i>Period</i>	<i>Data</i>	<i>Adjusted</i>
1/68	73.637	1.8671
2/68	77.136	1.88726
3/68	81.481	1.91106
4/68	84.127	1.92494
5/68	84.562	1.92718
6/68	91.959	1.96359
7/68	94.174	1.97393
8/68	96.087	1.98266
9/68	88.952	1.94916
10/68	83.479	1.92158
11/68	80.814	1.90749
12/68	77.466	1.88911
...

- **Period:** the sample index t .
- **Data:** the observation y_t .
- **Adjusted:** the adjusted data y'_t , if an adjustment has been specified using *Analysis Options*.

Analysis Options

Analysis Options permits the data to be transformed before being plotted or analyzed:



- **Math:** transforms the data by performing the indicated mathematical operation. With the exception of the Box-Cox transformation, the selections are self-explanatory. The Box-Cox transformation is used when necessary to make the data more Gaussian. For a detailed discussion, see the documentation for the *Box-Cox Transformations* procedure.
- **Seasonal:** seasonally adjust the data using the indicated method. Seasonal adjustments are designed to remove any seasonal component from the data. The methods used are discussed in the documentation for the *Seasonal Decomposition* procedure.
- **Trend:** remove a trend by fitting and subtracting the indicated type of trend. The transformed data are the residuals from the trend line. The equations for each type of trend discussed in the documentation for the *Forecasting* procedure.
- **Differencing:** transform the data by taking nonseasonal differences of order d and/or seasonal differences of order D . Differencing is sometimes used to stabilize a nonstationary time series that does not have a constant mean. A nonseasonal difference of order d is created by subtracting consecutive observations d times. For example, a first difference ($d = 1$) is given by

$$y'_t = y_t - y_{t-1} \quad (1)$$

while a second difference ($d = 2$) is given by

$$y'_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \quad (2)$$

A seasonal difference of order 1 ($D = 1$) is given by

$$y'_t = y_t - y_{t-s} \quad (3)$$

A seasonal difference of order 2 ($D = 2$) is given by

$$y'_t = (y_t - y_{t-s}) - (y_{t-s} - y_{t-2s}) \quad (4)$$

- **Inflation:** adjusts the data for inflation using the specified inflation rate λ . If applied at the beginning of the period, the adjustment is

$$y'_t = \frac{y_t}{(1 + \lambda)^{(t-t_0+1)}} \quad (5)$$

where t_0 is the index of the first observation. If applied at the middle of the period, the adjustment is

$$y'_t = \frac{y_t}{(1 + \lambda)^{(t-t_0+0.5)}} \quad (6)$$

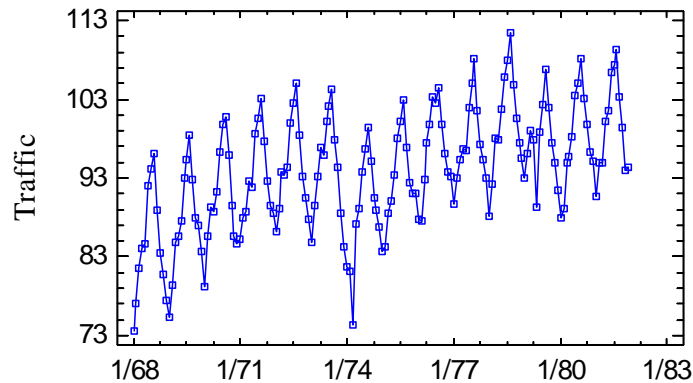
If more than one transformation is requested, they are applied in the following order:

1. trading day adjustment
2. inflation adjustment
3. math adjustment
4. seasonal adjustment
5. trend adjustment
6. differencing

Horizontal Time Sequence Plot

The *Horizontal Time Sequence Plot* displays the time series data in sequential order:

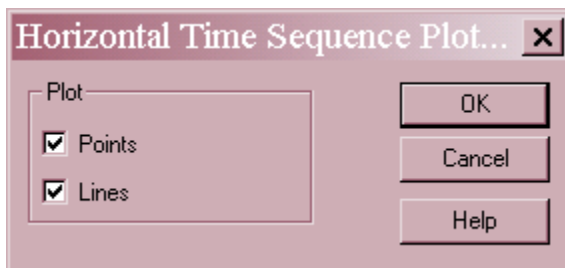
Time Series Plot for Traffic



The traffic data contains several very interesting features:

1. A general upward trend.
2. A regular yearly cycle around that trend, with peak traffic occurring in the summer months.
3. Dramatic changes to the trend line occurring late in 1973, when the Arab oil embargo made gasoline difficult to get. A similar though less dramatic change occurred during 1978.

Pane Options

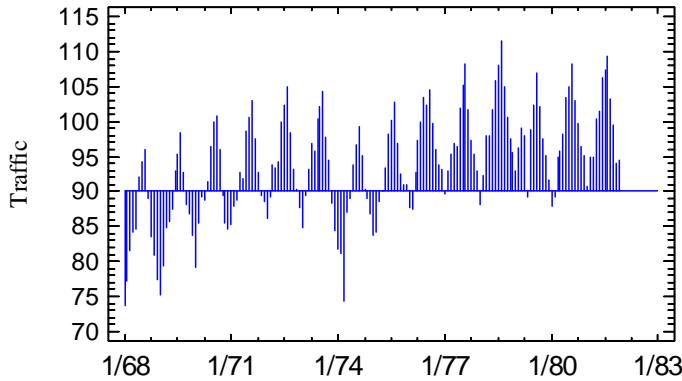


- **Points:** plot point symbols at the location of each observation.
- **Lines:** connect the observations with a line.

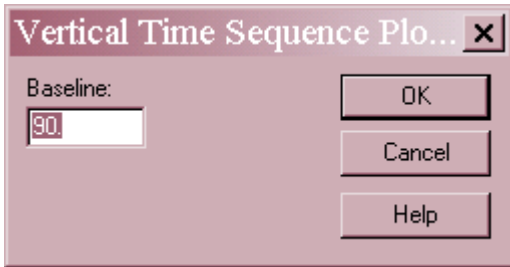
Vertical Time Sequence Plot

The *Vertical Time Sequence Plot* displays the time series by drawing vertical lines from a baseline to each observation:

Time Series Plot for Traffic



Pane Options



- **Baseline:** position from which vertical lines are drawn.

Autocorrelations

An important tool in modeling time series data is the autocorrelation function. The autocorrelation at lag k measures the strength of the correlation between observations k time periods apart. The sample lag k autocorrelation is calculated from

$$r_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2} \tag{7}$$

The *Autocorrelations* pane displays the sample autocorrelations together with large lag standard errors and probability limits:

Estimated Autocorrelations for Traffic				
			Lower 95.0%	Upper 95.0%
Lag	Autocorrelation	Std. Error	Prob. Limit	Prob. Limit
1	0.840362	0.0771517	-0.151215	0.151215
2	0.601514	0.119832	-0.234866	0.234866
3	0.329537	0.136627	-0.267785	0.267785
4	0.0932316	0.141279	-0.276902	0.276902
5	-0.108987	0.141645	-0.277619	0.277619
6	-0.169318	0.142143	-0.278596	0.278596
7	-0.11148	0.143339	-0.280939	0.280939
8	0.0723794	0.143854	-0.281949	0.281949
9	0.274421	0.14407	-0.282373	0.282373
10	0.490762	0.147149	-0.288407	0.288407
11	0.664932	0.156589	-0.306909	0.306909
12	0.735963	0.172579	-0.338249	0.338249
13	0.618444	0.190346	-0.373072	0.373072
14	0.411699	0.201953	-0.395821	0.395821
15	0.173765	0.206888	-0.405494	0.405494
16	-0.0384742	0.207755	-0.407193	0.407193
17	-0.219497	0.207797	-0.407276	0.407276
18	-0.280273	0.209173	-0.409972	0.409972
19	-0.229326	0.211397	-0.41433	0.41433
20	-0.0606596	0.212872	-0.417223	0.417223
21	0.126304	0.212975	-0.417424	0.417424
22	0.338969	0.21342	-0.418297	0.418297
23	0.506746	0.216601	-0.424532	0.424532
24	0.580971	0.223547	-0.438145	0.438145

The standard error for r_k is calculated on the assumption that the autocorrelations have “died out” by lag k and are equal to 0 at all lags greater or equal to k . It is calculated from:

$$se[r_k] = \sqrt{\frac{1}{n} \left\{ 1 + 2 \sum_{i=1}^{k-1} r_i^2 \right\}} \tag{8}$$

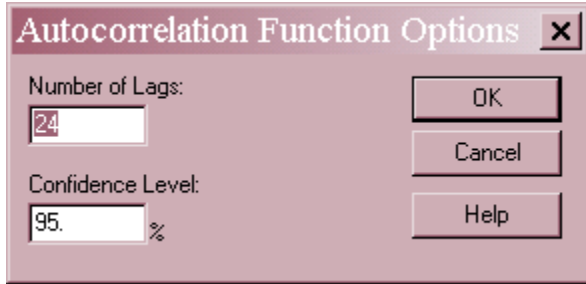
This standard error is then used to calculate $100(1-\alpha)\%$ probability limits around zero, using a critical value of the standard normal distribution:

$$0 \pm z_{\alpha/2} se[r_k] \tag{9}$$

If $\alpha = 0.05$, any sample autocorrelations that fall outside these limits are statistically significantly different from 0 at the 5% significance level. The StatAdvisor highlights any such autocorrelations in red.

For the traffic data, note that there are significant values for the first 3 lags and also in the vicinity of $s = 12$ and $2s = 24$. The significant values at low lags occur because observations close together in time are correlated. The significant lags around 12 and 24 are caused by the strong seasonal pattern.

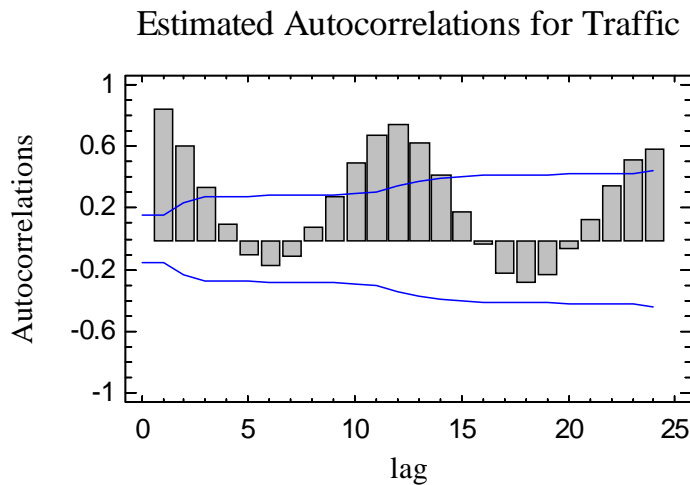
Pane Options



- **Number of lags:** maximum lag k at which to calculate the autocorrelation.
- **Confidence level:** value of $100(1-\alpha)\%$ used to calculate the probability limits.

Autocorrelation Function

The *Autocorrelation Function* plot displays the sample autocorrelations and probability limits:



Bars extending beyond the upper or lower limit correspond to statistically significant autocorrelations.

Partial Autocorrelations

Another important tool in modeling time series data is the partial autocorrelation function. The partial autocorrelations are used to help identify the proper order of autoregressive model to use to describe an observed time series. The sample lag k partial autocorrelation $\hat{\phi}_{kk}$ is calculated from the sample autocorrelations as described in the *Calculations* section. The *Partial Autocorrelations* pane displays the sample partial autocorrelations together with large lag standard errors and probability limits:

Estimated Partial Autocorrelations for Traffic				
Lag	Partial Autocorrelation	Std. Error	Lower 95.0% Prob. Limit	Upper 95.0% Prob. Limit
1	0.840362	0.0771517	-0.151215	0.151215
2	-0.356356	0.0771517	-0.151215	0.151215
3	-0.220746	0.0771517	-0.151215	0.151215
4	-0.0325883	0.0771517	-0.151215	0.151215
5	-0.132289	0.0771517	-0.151215	0.151215
6	0.294114	0.0771517	-0.151215	0.151215
7	0.174468	0.0771517	-0.151215	0.151215
8	0.309939	0.0771517	-0.151215	0.151215
9	0.0754963	0.0771517	-0.151215	0.151215
10	0.219082	0.0771517	-0.151215	0.151215
11	0.252541	0.0771517	-0.151215	0.151215
12	0.0188762	0.0771517	-0.151215	0.151215
13	-0.267034	0.0771517	-0.151215	0.151215
14	-0.0880152	0.0771517	-0.151215	0.151215
15	-0.00778138	0.0771517	-0.151215	0.151215
16	-0.023474	0.0771517	-0.151215	0.151215
17	-0.137019	0.0771517	-0.151215	0.151215
18	-0.0313481	0.0771517	-0.151215	0.151215
19	-0.0791926	0.0771517	-0.151215	0.151215
20	0.0714297	0.0771517	-0.151215	0.151215
21	-0.0340022	0.0771517	-0.151215	0.151215
22	0.12536	0.0771517	-0.151215	0.151215
23	0.0894574	0.0771517	-0.151215	0.151215
24	0.0420213	0.0771517	-0.151215	0.151215

The standard error for $\hat{\phi}_{kk}$ is calculated from:

$$se[\hat{\phi}_{kk}] = \sqrt{\frac{1}{n}} \tag{10}$$

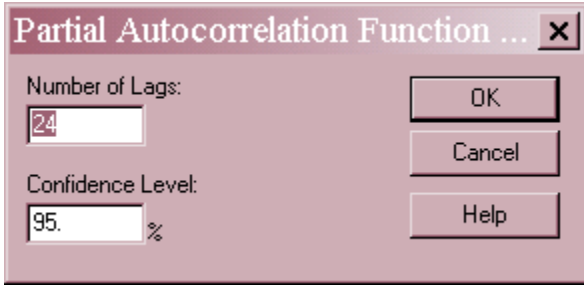
This standard error is then used to calculate $100(1-\alpha)\%$ probability limits around zero, using a critical value of the standard normal distribution:

$$0 \pm z_{\alpha/2} se[\hat{\phi}_{kk}] \tag{11}$$

If $\alpha = 0.05$, any sample partial autocorrelations that fall outside these limits are statistically significantly different from 0 at the 5% significance level. The StatAdvisor highlights any such partial autocorrelations in red.

For the traffic data, note that there are significant values scattered throughout the first 13 lags. This implies that it would take a rather complicated autoregressive model to describe the observed data, which is not surprising given its nonstationary (trending) nature.

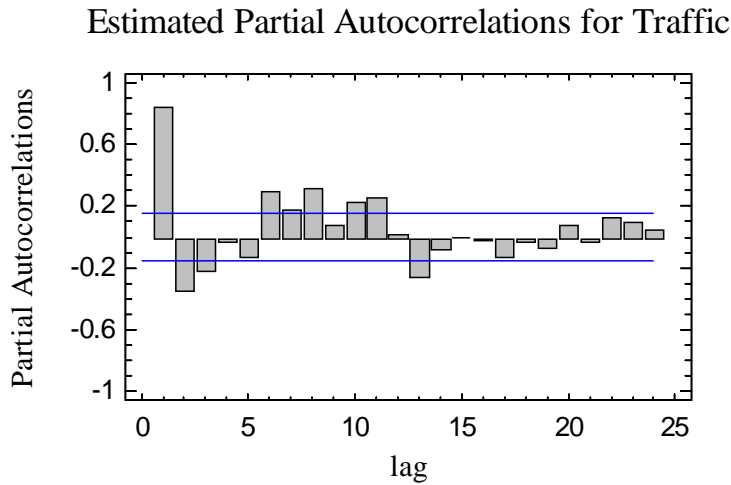
Pane Options



- **Number of lags:** maximum lag k at which to calculate the partial autocorrelation.
- **Confidence level:** value of $100(1-\alpha)\%$ used to calculate the probability limits.

Partial Autocorrelation Function

The *Partial Autocorrelation Function* plots the sample partial autocorrelations and probability limits:



Bars extending beyond the upper or lower limit correspond to statistically significant partial autocorrelations.

Periodogram

The autocorrelations and partial autocorrelations describe the behavior of the data in the *time domain*, i.e., by estimating statistics based on the amount of time between observations. It is also useful to examine the data in the *frequency domain*, by considering how much variability exists at different frequencies. It has been shown that any discrete time series can be represented as the sum of a set of sines and cosines at a set of frequencies called the Fourier frequencies. A typical component has the form

$$a_i \cos(2\pi f_i t) + b_i \sin(2\pi f_i t) \tag{12}$$

where f_i is the i -th Fourier frequency. The i -th Fourier frequency is

$$f_i = \frac{i}{n} \tag{13}$$

for $i = 0, 1, \dots, n/2$ if n is even and $i = 0, 1, \dots, (n-1)/2$ if n is odd.

The periodogram calculates the power in the data at each Fourier frequency by calculating:

$$I(f_i) = \frac{n}{2} (a_i^2 + b_i^2) \tag{14}$$

which is scaled so that the sum of the periodogram ordinates across all of the Fourier frequencies except for $i = 0$ yields the sum of squared deviations of the time series about its mean, i.e., $\sum_{i=1}^n (y_i - \bar{y})^2$. In effect, the periodogram generates an analysis of variance by frequency.

The *Periodogram* pane displays the following table:

i	Frequency	Period	Ordinate	Cumulative Sum	Integrated Periodogram
0	0.0		1.57558E-23	1.57558E-23	1.66608E-27
1	0.00595238	168.0	1387.62	1387.62	0.146731
2	0.0119048	84.0	866.251	2253.87	0.238332
3	0.0178571	56.0	465.451	2719.32	0.28755
4	0.0238095	42.0	90.789	2810.11	0.297151
5	0.0297619	33.6	447.388	3257.5	0.344459
6	0.0357143	28.0	68.8937	3326.39	0.351744
7	0.0416667	24.0	60.3328	3386.72	0.358124
8	0.047619	21.0	28.0432	3414.77	0.361089
9	0.0535714	18.6667	36.3759	3451.14	0.364936
10	0.0595238	16.8	61.0357	3512.18	0.37139
11	0.0654762	15.2727	40.4935	3552.67	0.375672
12	0.0714286	14.0	24.073	3576.74	0.378217
13	0.077381	12.9231	1.28899	3578.03	0.378354
14	0.0833333	12.0	4968.08	8546.11	0.903696
15	0.089285157	1511.2	30.471915	8576.58	0.906918
...

The table includes:

- **Frequency:** the i -th Fourier frequency $f_i = i/n$.

- **Period:** the period associated with the Fourier frequency, given by $1/f_i$. This is the number of observations in a complete cycle at that frequency.
- **Ordinate:** the periodogram ordinate $I(f_i)$.
- **Cumulative Sum:** the sum of the periodogram ordinates at all frequencies up to and including the i -th.
- **Integrated Periodogram:** the cumulative sum divided by the sum of the periodogram ordinates at all of the Fourier frequencies. This column represents the proportion of the power in the time series at or below the i -th frequency.

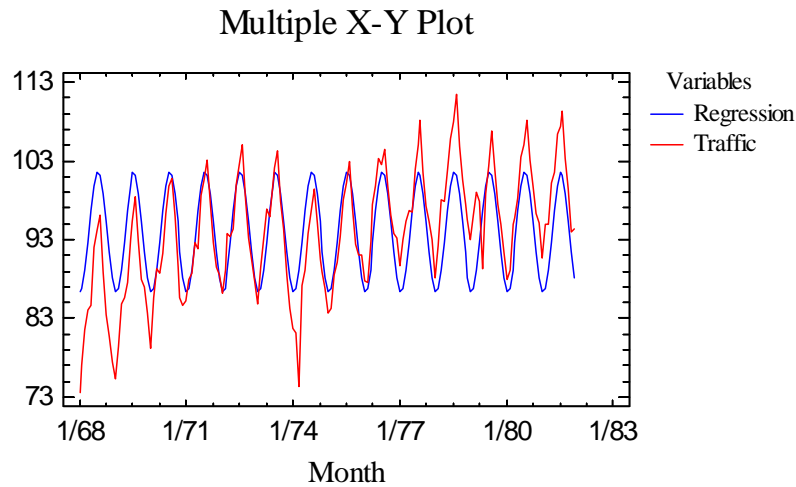
For example, the 14-th Fourier frequency corresponds to an oscillation with a period of 12 months. There is a very large ordinate at that frequency, since the data tend to rise and fall on a yearly basis. If one were to fit a regression model at that frequency, it would take the form

$$Y_t = c + a \cos\left(\frac{2\pi t}{12}\right) + b \sin\left(\frac{2\pi t}{12}\right) + e_t \tag{15}$$

where c is a constant and e_t is the error term. Fitting this model using the *Multiple Regression* procedure yields:

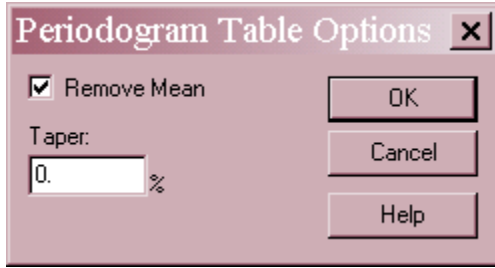
$$\hat{Y}_t = 93.9783 - 4.94209 \cos\left(\frac{2\pi t}{12}\right) + 5.89233 \sin\left(\frac{2\pi t}{12}\right) \tag{16}$$

A scatterplot of this model is shown below:



Notice how much of the variability has been explained by that single component.

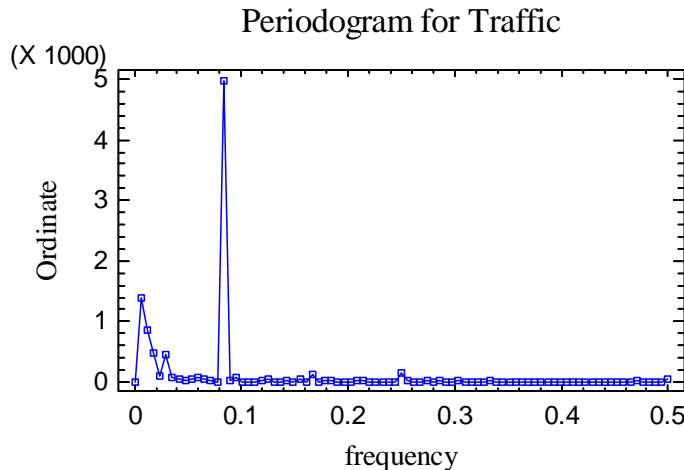
Pane Options



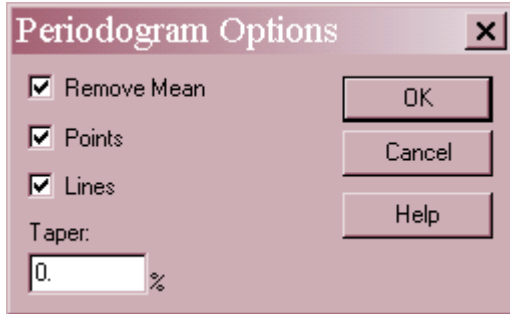
- **Remove mean:** check to subtract the mean from the time series before calculating the periodogram. If the mean is not removed, the ordinate at $i = 0$ is likely to be very large.
- **Taper:** percent of the data at each end of the time series to which a data taper will be applied before the periodogram is calculated. Following Bloomfield (2000), STATGRAPHICS uses a cosine taper that downweights observations close to $i = 1$ and $i = n$. This is useful for correcting bias if the periodogram ordinates are to be smoothed in order to create an estimate of the underlying spectral density function.

Periodogram Plot

The *Periodogram Plot* displays the periodogram ordinates:



Note the huge spike at a frequency of 1/12 months. Two small bumps can be observed at the first and second harmonics (2/12 and 3/12), since the seasonal oscillation is not purely sinusoidal. There is also some power at very small frequencies, caused by the trends and sudden changes in the traffic time series.

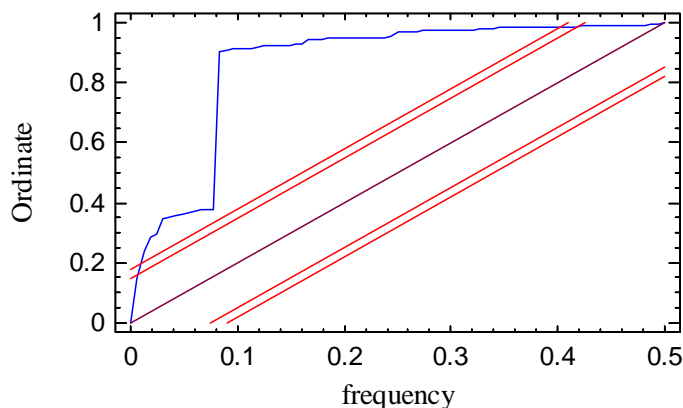
Pane Options

- **Remove mean:** check to subtract the mean from the time series before calculating the periodogram.
- **Points:** if checked, point symbols will be displayed.
- **Lines:** if checked, the ordinates will be connected by a line.
- **Taper:** percent of the data at each end of the time series to which a data taper will be applied before the periodogram is calculated.

Integrated Periodogram

The *Integrated Periodogram* displays the cumulative sums of the periodogram ordinates, divided by the sum of the ordinates over all of the Fourier frequencies:

Periodogram for Traffic



A diagonal line is included on the plot, together with 95% and 99% Kolmogorov-Smirnov bounds. If the time series is purely random, the integrated periodogram should fall within those bounds 95% and 99% of the time. For the traffic data, it is safe to conclude that the data does not form a random time series.

Tests For Randomness

The *Tests for Randomness* pane displays the results of additional tests run to determine whether or not the time series is purely random:

Tests for Randomness of Traffic	
(1) Runs above and below median	
Median = 94.323	
Number of runs above and below median = 32	
Expected number of runs = 85.0	
Large sample test statistic $z = 8.12529$	
P-value = 4.44089E-16	
(2) Runs up and down	
Number of runs up and down = 47	
Expected number of runs = 111.667	
Large sample test statistic $z = 11.8052$	
P-value = 0.0	
(3) Box-Pierce Test	
Test based on first 24 autocorrelations	
Large sample test statistic = 677.926	
P-value = 0.0	

Three tests are performed:

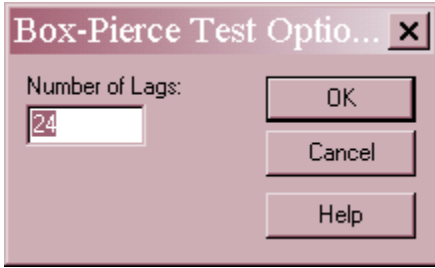
1. *Runs above and below median*: counts the number of times the series goes above or below its median. This number is compared to the expected value for a random time series. A trending series, such as the traffic data, is likely to show significantly less runs than expected. Small P-values (less than 0.05 if operating at the 5% significance level) indicate that the time series is not purely random.
2. *Runs up and down*: counts the number of times the series goes up or down. This number is compared to the expected value for a random time series. A series with a strong oscillation, such as the traffic data, is likely to show significantly less runs than expected. Small P-values indicate that the time series is not purely random.
3. *Box-Pierce Test*: constructs a test statistic based on the first k sample autocorrelations by calculating:

$$Q = n \sum_{i=1}^k r_i^2 \quad (17)$$

This statistic is compared to a chi-squared distribution with k degrees of freedom. As with the other two tests, small P-values indicate that the time series is not purely random.

Again, there is no question that the traffic series contains significant non-random structure.

Pane Options



- **Number of Lags:** number of lags k to include in the Box-Pierce test.

Crosscorrelations

The *Crosscorrelations* pane displays crosscorrelations between the main time series and a second series, specified using *Pane Options*. The crosscorrelation between one time series Y at time t and a second time series X at time $t-k$ is denoted by $c_{xy}(k)$. A typical use of crosscorrelations is in identifying “leadings indicators” or an input-output relationship. For example, Box, Jenkins and Reinsel (1994) present data from the input and output of a gas furnace at 9 second intervals, contained in the file *furnace.sfb*. The data consist of:

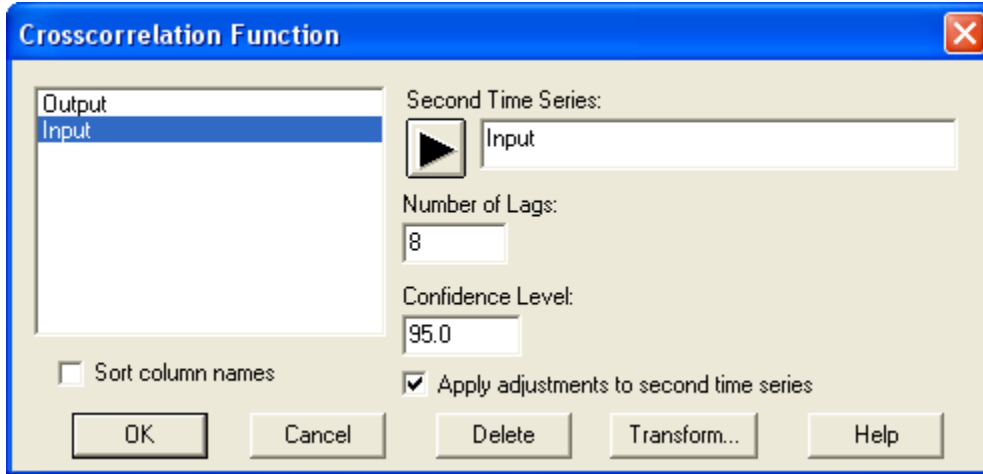
1. Output series Y : % Co2 in outlet gas
2. Input series X : input gas rate in cubic feet per minute

The crosscorrelation table is shown below:

Lag	Crosscorrelation
-8	-0.179456
-7	-0.206068
-6	-0.226716
-5	-0.242871
-4	-0.260351
-3	-0.286432
-2	-0.328542
-1	-0.393467
0	-0.484451
1	-0.598405
2	-0.725033
3	-0.84282
4	-0.924592
5	-0.95032
6	-0.914593
7	-0.82932
8	-0.71652

Some large negative correlations are noticeable, peaking at $k = 5$. This suggests that *increases* in the input gas rate cause *decreases* in the % of Co2 in the outlet gas rate, peaking about 45 seconds later.

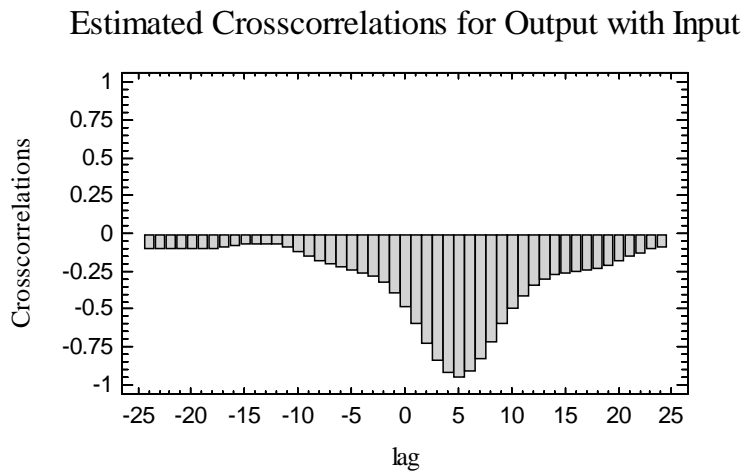
Pane Options



- **Second Time Series:** the observations for the X time series.
- **Number of Lags:** maximum lag k (both positive and negative) at which to calculate the crosscorrelations
- **Confidence Level:** value of $100(1-\alpha)\%$ used to calculate the probability limits.
- **Apply adjustments to second time series:** whether the adjustments applied to the first time series should also be applied to the second time series.

Crosscorrelation Plot

The *Crosscorrelation Plot* displays the estimated crosscorrelations:



Note the large negative correlation at positive lags.

Save Results

The following results can be saved to the datasheet:

1. *Data* – the original observations, together with any interpolated replacements for missing values.
2. *Adjusted data* – time series data after any adjustments have been made.
3. *Period labels* – identification of time t for each observation.
4. *Autocorrelations* – sample autocorrelations.
5. *Partial autocorrelations* – sample partial autocorrelations.
6. *Crosscorrelations* – sample crosscorrelations.
7. *Periodogram ordinates* – calculated periodogram ordinates.
8. *Fourier frequencies* – Fourier frequencies corresponding to the periodogram ordinates.

Calculations

Missing Data

A limited number of missing values are permitted, as long as there are not too many missing values close together. Before the data is analyzed, missing values are replaced by interpolated values, determined using the following rule:

1. If y_t , the observation at time t , is missing, find the two observations in the same season that precede time t (y_{t-s} and y_{t-2s}) and the two observations in the same season that come after time t (y_{t+s} and y_{t+2s}).
2. If none of the four observations are missing, then the replacement value for y_t is:

$$y_t = \frac{-3y_{t-2s} + 12y_{t-s} + 12y_{t+s} - 3y_{t+2s}}{18} \quad (18)$$

3. If y_{t+2s} is missing but the other three are not, then the replacement value for y_t is:

$$y_t = \frac{-y_{t-2s} + 3y_{t-s} + y_{t+s}}{3} \quad (19)$$

4. If y_{t+s} is missing but the other three are not, then the replacement value for y_t is:

$$y_t = \frac{-3y_{t-2s} + 8y_{t-s} + y_{t+s}}{6} \quad (20)$$

5. If y_{t-s} is missing but the other three are not, then the replacement value for y_t is:

$$y_t = \frac{y_{t-2s} + 8y_{t+s} - 3y_{t+2s}}{6} \quad (21)$$

6. If y_{t-2s} is missing but the other three are not, then the replacement value for y_t is:

$$y_t = \frac{y_{t-s} + 3y_{t+s} - y_{t+2s}}{3} \quad (22)$$

7. If y_{t+s} and y_{t+2s} are missing but the other two are not, then the replacement value for y_t is:

$$y_t = -y_{t-2s} + 2y_{t-s} \quad (23)$$

8. If y_{t-s} and y_{t+2s} are missing but the other two are not, then the replacement value for y_t is:

$$y_t = \frac{y_{t-2s} + 2y_{t+s}}{3} \quad (24)$$

9. If y_{t-s} and y_{t+s} are missing but the other two are not, then the replacement value for y_t is:

$$y_t = \frac{y_{t-2s} + y_{t+2s}}{2} \quad (25)$$

10. If y_{t-2s} and y_{t+2s} are missing but the other two are not, then the replacement value for y_t is:

$$y_t = \frac{y_{t-s} + y_{t+s}}{2} \quad (26)$$

11. If y_{t-2s} and y_{t+s} are missing but the other two are not, then the replacement value for y_t is:

$$y_t = \frac{2y_{t-s} + y_{t+2s}}{3} \quad (27)$$

12. If y_{t-2s} and y_{t-s} are missing but the other two are not, then the replacement value for y_t is:

$$y_t = 2y_{t+s} - y_{t+2s} \quad (28)$$

If more than 2 of the four observations are missing, an error message will be displayed and the analysis will not be performed.

The interpolated values are designed to perfectly reproduce a quadratic trend (if only one observation is missing) or a linear trend (if two observations are missing), provided no noise is present.

Partial Autocorrelations

$$\hat{\phi}_{kk} = \begin{cases} r_1 & k = 1 \\ \frac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_j} & \text{for } k > 1 \end{cases} \quad (29)$$

where

$$\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk} \hat{\phi}_{k-1,k-j} \quad \text{for } j = 1, 2, \dots, k-1 \quad (30)$$

Runs Tests

Refer to the documentation for the *Runs Charts* procedure.

Crosscorrelations

$$r_{xy}(k) = \frac{c_{xy}(k)}{s_x s_y} \quad (31)$$

where

$$c_{xy}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (x_t - \bar{x})(y_{t+k} - \bar{y}) \quad \text{for } k = 0, 1, 2, \dots \quad (32)$$

$$c_{xy}(k) = \frac{1}{n} \sum_{t=1}^{n+k} (y_t - \bar{y})(x_{t-k} - \bar{x}) \quad \text{for } k = 0, -1, -2, \dots \quad (33)$$

$$\bar{x} = \frac{\sum_{t=1}^n x_t}{n} \quad (34)$$

$$\bar{y} = \frac{\sum_{t=1}^n y_t}{n} \quad (35)$$

$$s_x = \sqrt{c_{xx}(0)} \quad (36)$$

$$s_y = \sqrt{c_{yy}(0)} \quad (37)$$