CREATE

CREATE produces new series as a function of existing series.

Notation

The following notation is used throughout this chapter unless otherwise stated:

Existing Series	$X_1, \ldots, X_n$
New Series	$Y_1, \ldots, Y_n$

Cumulative Sum (CSUM(*X*))

$$Y_j = \sum_{i=1}^{j} X_i \qquad j = 1, \ldots, n$$

Differences of Order *m* (DIFF(*X*,*m*))

Define

$$Z_j(k) = Z_j(k-1) - Z_{j-1}(k-1) \qquad k = 1, \ldots, m \qquad j = k+1, \ldots, n$$

with

$$Z_j(0) = X_j \qquad j = 1, \ldots, n$$

then

$$Y_j = \begin{cases} Z_j(m) & j = m+1, \ldots, n \\ \text{SYSMIS} & \text{otherwise} \end{cases}$$

Lag of Order *m* (LAG(*X*,*m*))

$$Y_j = \begin{cases} X_{j-m} & j = m+1,\ldots,n \\ \text{SYSMIS} & j = 1,\ldots,m \end{cases}$$

Lead of Order *m* (LEAD(*X*,*m*))

$$Y_j = \begin{cases} X_{j+m} & j = 1,\ldots,n-m \\ \text{SYSMIS} & j = n-m+1,\ldots,n \end{cases}$$

Moving Average of Length *m* (MA(*X*,*m*))

If m is odd, define

$$q = \frac{m-1}{2}$$

then

$$Y_j = \begin{cases} \sum_{k=-q}^{q} X_{j+k} \big/ m & j = q+1,\ldots,n-q \\ \text{SYSMIS} & \text{otherwise} \end{cases}$$

If m is even, define $q = m/2$ and

$$Z_j = \sum_{k=-q+1}^{q} X_{j+k} \big/ m \qquad j = q,\ldots,n-q$$

then

$$Y_j = \begin{cases} \left(Z_{j-1} + Z_j\right) \big/ 2 & j = q+1,\ldots,n-q \\ \text{SYSMIS} & \text{otherwise} \end{cases}$$

Running Median of Length *m* (*X*,*m*)

If m is odd,

$$q = \frac{m-1}{2}$$

$$Y_j = \begin{cases} \text{median}\left(X_{j-q}, X_{j-q+1}, \ldots, X_j, X_{j+1}, \ldots, X_{j+q}\right) & j = q+1, \ldots, n-q \\ \text{SYSMIS} & \text{otherwise} \end{cases}$$

If m is even, define

$$Z_j = \text{median}\left(X_{j-q+1}, \ldots, X_j, X_{j+1}, \ldots, X_{j+q}\right) \qquad j = q, \ldots, n-q$$

then

$$Y_j = \begin{cases} \left(Z_{j-1} + Z_j\right)/2 & j = q+1, \ldots, n-q \\ \text{SYSMIS} & \text{otherwise} \end{cases}$$

where

$$\text{median}\left(a_1, \ldots, a_k\right) = \begin{cases} a_{(l)} & \text{if } k \text{ is odd} \\ \left(a_{(l)} + a_{(l+1)}\right)/2 & \text{if } k \text{ is even} \end{cases}$$

$$l = \begin{cases} (k+1)/2 & \text{if } k \text{ is odd} \\ k/2 & \text{if } k \text{ is even} \end{cases}$$

and $a_{(1)} < a_{(2)} < \ldots < a_{(k)}$ is the ordered sample of $a_1, \ldots, a_k$.

Seasonal Differencing of Order *m* and Period *p* (SDIFF(*X*,*m*,*p*))

Define

$$Z_j(k) = Z_j(k-1) - Z_{j-p}(k-1) \qquad k = 1, \ldots, m \qquad j = pk+1, \ldots, n$$

where

$$Z_j(0) = X_j \qquad j = 1, \ldots, n$$

then

$$Y_j = Z_j(m) \qquad j = mp + 1, \ldots, n$$

The T4253H Smoothing Function (T4253H(*X*))

The original series is smoothed by a compound data smoother based on Velleman (1980). The smoother starts with:

- A running median of 4:

  Let *Z* be the smoothed series, then

$$Z_{j+1/2} = \text{median}\left(X_{j-1}, X_j, X_{j+1}, X_{j+2}\right) \qquad j = 2, \ldots, n-2$$

and

$$Z_{0.5}^{(1)} = X_1 \qquad Z_{1.5}^{(1)} = \text{median}\left(X_1, X_2\right) = \tfrac{1}{2}\left(X_1 + X_2\right)$$

$$Z_{n-1/2}^{(1)} = \text{median}\left(X_{n-1}, X_n\right) = \tfrac{1}{2}\left(X_{n-1} + X_n\right) \qquad Z_{n+1/2}^{(1)} = X_n$$

- A running median of *Z*:

$$Z_1^{(1)} = Z_{0.5} \qquad Z_n^{(1)} = Z_{n+1/2}$$

and

$$Z_j^{(1)} = \tfrac{1}{2}\left(Z_{j-1/2} + Z_{j+1/2}\right) \qquad j = 2, \ldots, n-1\,.$$

- A running median of 5 on $Z_1^{(1)},\ldots,Z_n^{(1)}$ from the previous step:

  Let $Z^{(2)}$ be the resulting series, then

$$Z_1^{(2)} = Z_1^{(1)} \quad Z_n^{(2)} = Z_n^{(1)}$$

$$Z_2^{(2)} = \text{median}\left(Z_1^{(1)}, Z_2^{(1)}, Z_3^{(1)}\right)$$

$$Z_{n-1}^{(2)} = \text{median}\left(Z_{n-2}^{(1)}, Z_{n-1}^{(1)}, Z_n^{(1)}\right)$$

and

$$Z_j^{(2)} = \text{median}\left(Z_{j-2}^{(1)}, Z_{j-1}^{(1)}, Z_j^{(1)}, Z_{j+1}^{(1)}, Z_{j+2}^{(1)}\right) \quad j = 3,\ldots,n-2$$

- A running median of 3 on $Z_1^{(2)},\ldots,Z_n^{(2)}$ from the previous step:

  Let $Z^{(3)}$ be the resulting series, then

$$Z_j^{(3)} = \text{median}\left(Z_{j-1}^{(2)}, Z_j^{(2)}, Z_{j+1}^{(2)}\right) \quad j = 2,3,\ldots,n-1$$

$$Z_1^{(3)} = \text{median}\left(3Z_2^{(3)} - 2Z_3^{(3)}, Z_1^{(2)}, Z_2^{(3)}\right)$$

$$Z_n^{(3)} = \text{median}\left(3Z_{n-1}^{(3)} - 2Z_{n-2}^{(3)}, Z_n^{(2)}, Z_{n-1}^{(3)}\right)$$

- Hanning (Running Weighted Averages):

$$Z_j^{(4)} = \tfrac{1}{4} Z_{j-1}^{(3)} + \tfrac{1}{2} Z_j^{(3)} + \tfrac{1}{4} Z_{j+1}^{(3)} \quad j = 2,\ldots,n-1$$

$$Z_1^{(4)} = Z_1^{(3)}, \quad Z_n^{(4)} = Z_n^{(3)}$$

- Residual:

$$D_i = X_i - Z_i^{(4)} \qquad i = 1,\ldots,n$$

- Repeat the previous steps on the residuals $D_1,\ldots,D_n$:
- Let $D_1^{(4)},\ldots,D_n^{(4)}$ be the final result.
- Final smooth:

$$Y_i = Z_i^{(4)} + D_i^{(4)} \qquad i = 1,\ldots,n$$

Prior Moving Averages of Length *m* (PMA(*X*,*m*))

$$Y_i = \begin{cases} \sum_{j=i-m}^{i-1} X_j / m & i = m+1,\ldots,n \\ \text{SYSMIS} & i = 1,\ldots,m \end{cases}$$

Fast Fourier Transform (FFT(*X*))

The discrete Fourier transform of a sequence $X = \{X_1,\ldots,X_n\}$ is defined as

$$Y_k = \frac{1}{n}\sum_{t=1}^{n} X_t \exp\{-i2\pi f_k (t-1)\}$$

$$= \frac{1}{n}\sum_{t=1}^{n} X_t\left[\cos(2\pi f_k (t-1)) - i\sin(2\pi f_k (t-1))\right]$$

$$= \frac{1}{n}\sum_{t=1}^{n} X_t \cos(2\pi f_k (t-1)) + i\left[-\frac{1}{n}\sum_{t=1}^{n} X_t \sin(2\pi f_k (t-1))\right]$$

$$= a_k + ib_k$$

Thus a, b are two sequences generated by FFT and they are called real and imaginary, respectively.

$$a_k = \frac{1}{n}\sum_{t=1}^{n} X_t \cos\left(2\pi f_k (t-1)\right) \quad k = 1,\ldots,r$$

$$b_k = -\frac{1}{n}\sum_{t=1}^{n} X_t \sin\left(2\pi f_k (t-1)\right) \quad k = 1,\ldots,r$$

where

$$r = \begin{cases} (n-1)/2 & \text{if } n \text{ is odd} \\ n/2 & \text{if } n \text{ is even} \end{cases}$$

and

$$a_0 = \overline{X} \qquad b_0 = -\frac{1}{n}\sum_{t=1}^{n} X_t \cos\left(\pi(t-1)\right)$$

Inverse Fast Fourier Transform of Two Series (IFFT(*a*,*b*))

The inverse Fourier Transform of two series $\{a, b\}$ is defined as

$$X_t = a_0 - b_0 \cos\left(\pi(t-1)\right) + 2\left[\sum_{k=1}^{q} a_k \cos\left(2\pi f_k (t-1)\right) - \sum_{k=1}^{q} b_k \sin\left(2\pi f_k (t-1)\right)\right]$$

References

Brigham, E. O. 1974. *The fast fourier transform*. Englewood Cliffs, N.J.: Prentice-Hall.

Monro, D. M. 1975. Algorithm AS 83: Complex discrete fast Fourier transform. *Applied Statistics*, 24: 153–160.

Monro, D. M., and Branch, J. L. 1977. Algorithm AS 117: The chirp discrete Fourier transform of general length. *Applied Statistics*, 26: 351–361.

Velleman, P. F. 1980 Definition and comparison of robust nonlinear data smoothing algorithms. *Journal of the American Statistical Associations*, 75: 609–615.

Velleman, P. F., and Hoaglin, D. C. 1981. *Applications, basics, and computing of exploratory data analysis*. Boston: Duxbury Press.