NDK_PCR_ANOVA

Last Modified on 03/14/2016 11:35 am CDT

- C/C++
- .Net

```
int __stdcall NDK_PCR_ANOVA ( double **  X,
                              size_t     nXSize,
                              size_t     nXVars,
                              LPBYTE     mask,
                              size_t     nMaskLen,
                              double *   Y,
                              size_t     nYSize,
                              double     intercept,
                              WORD       nRetType,
                              double *   retVal
                            )
```

Returns an array of cells for the i-th principal component (or residuals).

Returns

status code of the operation

Return values

NDK_SUCCESS	Operation successful
NDK_FAILED	Operation unsuccessful. See **Macros** for full list.

Parameters

[in]	**X**	is the independent variables data matrix, such that each column represents one variable
[in]	**nXSize**	is the number of observations (i.e. rows) in X
[in]	**nXVars**	is the number of variables (i.e. columns) in X
[in]	**mask**	is the boolean array to select a subset of the input variables in X. If missing (i.e. NULL), all variables in X are included.
[in]	**nMaskLen**	is the number of elements in mask
[in]	**Y**	is the response or the dependent variable data array (one dimensional array)
[in]	**nYSize**	is the number of elements in Y
[in]	**intercept**	is the constant or the intercept value to fix (e.g. zero). If missing (NaN), an intercept will not be fixed and is computed normally

[in]	**nRetType**	is a switch to select the return output: 1. SSR (sum of squares of the regression) 2. SSE (sum of squares of the residuals) 3. SST (sum of squares of the dependent variable) 4. MSR (mean squares of the regression) 5. MSE (mean squares error or residuals) 6. F-stat (test score) 7. Significance F (P-value of the test)
[out]	**retVal**	is the calculated statistics ANOVA output.

Remarks

1. The underlying model is described **here**.
2. $$\mathbf{y} = \alpha + \beta_1 \times \mathbf{PC}_1 + \dots + \beta_p \times \mathbf{PC}_p$$
3. The regression ANOVA table examines the following hypothesis: $$\mathbf{H}_o: \beta_1 = \beta_2 = \dots = \beta_p = 0$$ $$\mathbf{H}_1: \exists \beta_i \neq 0, i \in \left[1,0 \right]$$
4. In other words, the regression ANOVA examines the probability that the regression does NOT explain the variation in $\mathbf{y}$, i.e. that any fit is due purely to chance.
5. The MLR_ANOVA calculates the different values in the ANOVA tables as follows:
   $$\mathbf{SST}=\sum_{i=1}^N \left(Y_i - \bar Y \right)^2$$
   $$\mathbf{SSR}=\sum_{i=1}^N \left(\hat Y_i - \bar Y \right)^2$$
   $$\mathbf{SSR}=\sum_{i=1}^N \left(Y_i - \hat Y_i \right)^2$$ Where:
   - $\mathbf{PC}$ is the principal component.
   - N is the number of non-missing observations in the sample data.
   - $\bar Y$ is the empirical sample average for the dependent variable.
   - $\hat Y_i$ is the regression model estimate value for the i-th observation.
   - $\mathbf{SST}$ is the total sum of squares for the dependent variable.
   - $\mathbf{SSR}$ is the total sum of squares for the regression (i.e. $\hat y$) estimate.
   - $\mathbf{SSE}$ is the total sum of error (aka residuals ϵ) terms for the regression (i.e. $\epsilon = y - \hat y$) estimate.
   - $\mathbf{SST} = \mathbf{SSR} + \mathbf{SSE}$

   AND $$\mathbf{MSR} = \frac{\mathbf{SSR} }{p} \] \[\mathbf{MSE} = \frac{ \mathbf{SSE} }{N-p-1}$$ $$\mathbf{F\text{-}Stat} = \frac{\mathbf{MSR} }{\mathbf{MSE} }$$ Where:
   - p is the number of explanatory (aka predictor) variables in the regression.
   - $\mathbf{MSR}$ is the mean squares of the regression.
   - $\mathbf{MSE}$ is the mean squares of the residuals.
   - $\textrm{F-Stat}$ is the test score of the hypothesis.
   - $\textrm{F-Stat} \sim \mathbf{F}\left(p,N-p-1 \right)$
6. The sample data may include missing values.
7. Each column in the input matrix corresponds to a separate variable.
8. Each row in the input matrix corresponds to an observation.

9. Observations (i.e. row) with missing values in X or Y are removed.
10. The number of rows of the response variable (Y) must be equal to the number of rows of the explanatory variables (X).
11. The MLR_ANOVA function is available starting with version 1.60 APACHE.

Requirements

Header	SFSDK.H
Library	SFSDK.LIB
DLL	SFSDK.DLL

References

Hamilton, J .D.; Time Series Analysis , Princeton University Press (1994), ISBN 0-691-04289-6

Tsay, Ruey S.; Analysis of Financial Time Series John Wiley & SONS. (2005), ISBN 0-471-690740

See Also

[template("related")]