An IntRoduction to gRey methods by using R

Tan Xi
Department of Statistics, NUFE
Nov 6, 2009

Contents:

- A brief introduction to Grey Methods
- An analysis of Degree of Grey Incidence
- Grey GM(1, 1) Model

  (1)Construction

  (2)Test the accuracy of the model
- Some Envisions

A brief introduction to Grey Methods

- **Who?**

  Professor Deng! A Chinese!
- **When?**

  Long ago… About 1970s!
- **What?**

  White, Grey, Black?
- **Application?**

  Many realms! Economics, Physics, Social Science, and the list will go on!

An easy example——Step by Step

Suppose that the original sequence is

$$Y_0 = \{\ 8, 8.8, 16, 18, 24, 32\ \}$$

comparative sequences are：

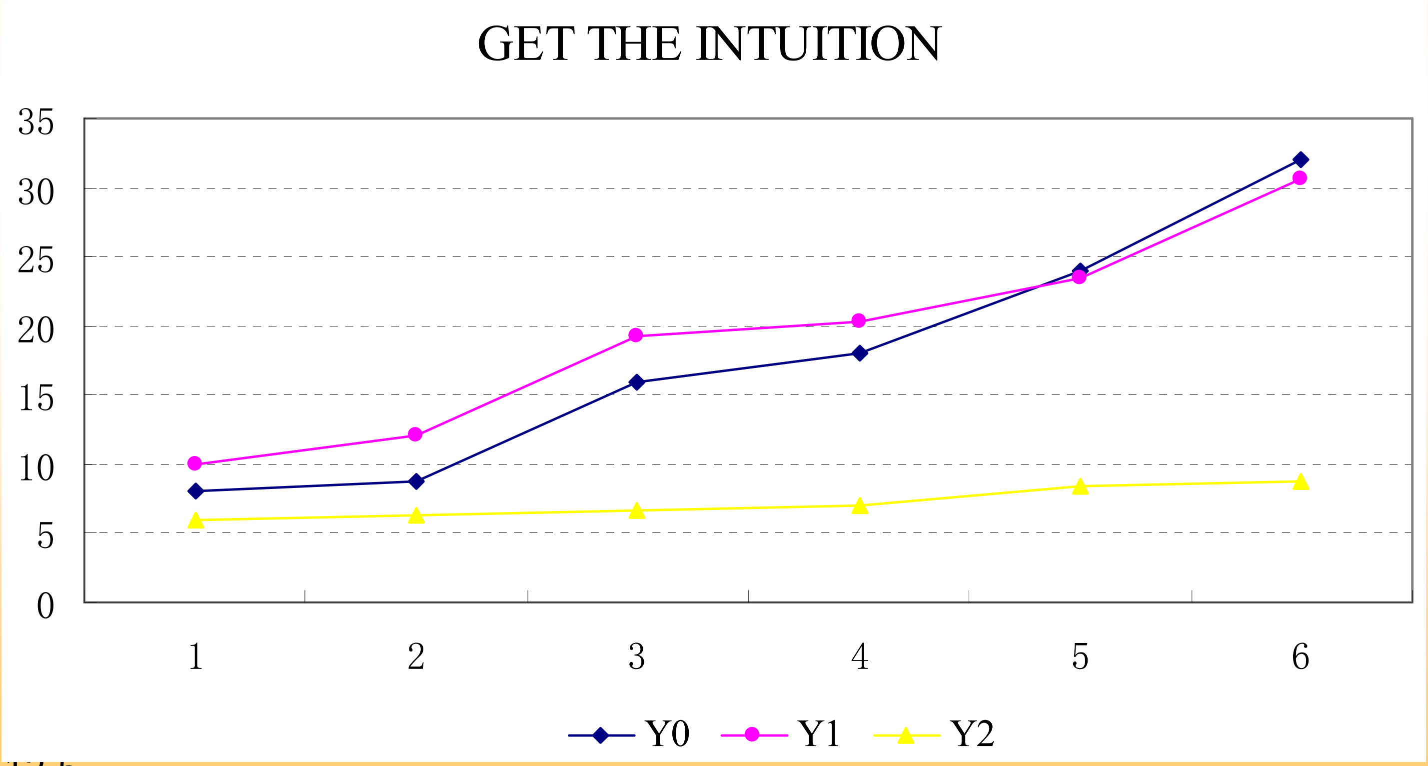
$$Y_1 = \{\ 10, 12.12, 19.28, 20.25, 23.4, 30.69\ \}$$

$$Y_2 = \{\ 6, 6.35, 6.57, 6.98, 8.35, 8.75\ \}$$

Which of the comparative sequences is much closer to the original series Y_0 ?

Before computing the exact values, you can get the intuition by looking at the graph.

(Ads: Where is Xie? ☺).



- **Step1:** Initialize all sequences

$$X_0 = \{1, 1.1, 2, 2.25, 3, 4\}$$
$$X_1 = \{1, 1.212, 1.928, 2.205, 2.34, 3.069\}$$
$$X_2 = \{1, 1.0583, 1.0950, 1.1633, 1.3917, 1.4583\}$$

- **Step2 :** Compute the absolute subtraction sequences $\Delta_{0i}(k) = |Y_0(k) - Y_i(k)|$

Table1 Absolute subtraction sequences

Number	1	2	3	4	5	6
Δ_1	0.000	0.112	0.072	0.225	0.660	0.931
Δ_2	0.0000	0.0417	0.9050	1.0867	1.6083	2.5417

- **Step3:** Compute the two-step minimum and maximum of the absolute subtraction sequences

$$\Delta_{\min} = \min_i \min_k \left| Y_0(k) - Y_i(k) \right| = 2.5417$$

$$\Delta_{\max} = \max_i \max_k \left| Y_0(k) - Y_i(k) \right| = 0$$

- **Step4:** Compute coefficients of Grey incidence

Formula: $$\gamma(Y_0(k), Y_i(k)) = \frac{\Delta_{\min} + \rho \cdot \Delta_{\max}}{\Delta_{oi}(k) + \rho \cdot \Delta_{\max}}$$

of which the distinguishing coefficient ρ is 0.5.

Number	1	2	3	4	5	6
$\gamma(Y_0(k), Y_1(k))$	1.0000	0.9190	0.9464	0.8496	0.6582	0.5772
$\gamma(Y_0(k), Y_2(k))$	1.0000	0.9683	0.5841	0.5391	0.4414	0.3333

- **Step5:** Compute the degree of Grey incidence

$$\gamma(Y_0, Y_i) = \frac{1}{n}\sum_{k=1}^{n}\gamma(Y_0(k), Y_i(k))$$

$$\gamma(Y_0, Y_1) = 0.8251 \qquad \gamma(Y_0, Y_2) = 0.6444$$

And so our intuition is right!

The computed results show that Y_0 is much more closer to Y_1 than to Y_2 , which is in coincidence with our intuition!

Recap:

- Initialize all sequences.
- Compute the absolute subtraction sequences

$$\Delta_{0i}(k) = \left| Y_0(k) - Y_i(k) \right|$$

- Compute the two-step minimum and maximum of the absolute subtraction sequences

$$\Delta_{\min} = \min_i \min_k \left| Y_0(k) - Y_i(k) \right| \qquad \Delta_{\max} = \max_i \max_k \left| Y_0(k) - Y_i(k) \right|$$

- Compute coefficients of Grey incidence

$$\gamma(Y_0(k), Y_i(k)) = \frac{\Delta_{\min} + \rho \cdot \Delta_{\max}}{\Delta_{oi}(k) + \rho \cdot \Delta_{\max}}$$

Recap:

- Bingo! !

Compute the degree of Grey incidence:

$$\gamma(Y_0, Y_i) = \frac{1}{n}\sum_{k=1}^{n}\gamma(Y_0(k), Y_i(k))$$

All steps in one function

Are you bored or puzzled with these steps??

Alternatives:

- **The first:** R functions!

  I've involved all preceding steps in one function:

  灰色关联分析函数.R

  I'll show you how to use it!

- **The second:**

  Click-Mouse Statistical Packages ……

- **It's your choice! It's all up to you! For R-Users??**

GM（1, 1）Model

- GM(1, 1) type of Grey model is the most widely used in the literature, pronounced as “Grey Model First Order One Variable”.
- This model is a time series forecasting model. The differential equations of the GM(1, 1) model have time-varying coefficients.

How to construct the GM(1,1) Model?

- Consider a time sequence $X^{(0)}$, which has n observations, $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \cdots, X^{(0)}(n)\}$
- When this sequence is subjected to the Accumulating Generation Operation (AGO), the following sequence $X^{(1)}$ is obtained

$$X^{(1)} = \{X^{(1)}(1), X^{(1)}(2), \cdots, X^{(1)}(n)\}$$

where

$$X^{(1)}(k) = \sum_{i=1}^{k} X^{(0)}(i)$$

How to construct the GM(1,1) Model?

- The grey difference equation of GM(1,1) is defined as follows:

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = \mu$$

The least square estimate sequence of the grey difference equation of GM(1,1) is defined as follows:

$$\hat{\alpha} = \begin{pmatrix} a \\ \mu \end{pmatrix} = (B^T B)^{-1} B^T Y_n$$

How to construct the GM(1,1) Model?

- Solve the grey difference equation of GM(1,1), the predicted GM(1,1) Model can be obtained:

$$\hat{X}^{(1)}(k+1) = \left(X^{(0)}(1) - \frac{\mu}{a} \right) e^{-ak} + \frac{\mu}{a}$$

- To obtain the predicted value of the primitive data at time $(k+H)$, the IAGO is used to establish the following grey model:

$$X_p^{(0)}(k+H) = [X^{(0)}(1) - \frac{b}{a}] e^{-a(k+H-1)} (1 - e^{a})$$

How to test the accuracy of the GM(1,1) Model?

- Residual Tests

$$\Delta^{(0)}(i) = \left| X^{(0)}(i) - \hat{X}^{(0)}(i) \right| \quad i = 1, 2, \cdots, n$$

$$\Phi(i) = \frac{\Delta^{(0)}(i)}{X^{(0)}(i)} \times 100\% \quad i = 1, 2, \cdots, n$$

How to test the accuracy of the GM(1,1) Model?

- The Test of the degree of Grey incidence

$$\gamma(\hat{X}^{(0)}, X^{(0)}) = \frac{1}{n}\sum_{i=1}^{n}\gamma(\hat{X}^{(0)}(i), X^{(0)}(i))$$

According to experience, the GM(1,1) Model is qualified if

$\gamma(\hat{X}^{(0)}, X^{(0)}) > 0.6$, when $\rho = 0.5$.

How to test the accuracy of the GM(1,1) Model?

- C and P Criteria

$$S_1 = \sqrt{\frac{\sum [X^{(0)}(i) - \bar{X}^{(0)}]^2}{n-1}}$$

$$S_2 = \sqrt{\frac{\sum [\Delta^{(0)}(i) - \bar{\Delta}^{(0)}]^2}{n-1}}$$

$$\text{C} = \frac{\text{S}_2}{\text{S}_1} = 0.01887908$$

$$P = p\{\left|\Delta^{(0)}(i) - \bar{\Delta}^{(0)}\right| < 0.6745 S_1\}$$

How to test the accuracy of the GM(1,1) Model?

- C and P Criteria

P	C	判别结果
>0.95	<0.35	好
>0.80	<0.50	合格
>0.70	<0.65	勉强合格
≤ 0.70	≥ 0.65	不合格

An easy example executed by R Program

Suppose the original sequence is：

	1	2	3	4	5	6
$X^{(0)}(i)$	26.7	31.5	32.8	34.1	35.8	37.5
	$X^{(0)}(1)$	$X^{(0)}(2)$	$X^{(0)}(3)$	$X^{(0)}(4)$	$X^{(0)}(5)$	$X^{(0)}(6)$

Construct the GM(1,1) Model and predict the values of 7~11th Periods.

R program :

GM(1，1)模型建立、检验和预测.R

- Step1: Construct the AGO sequence:

	1	2	3	4	5	6
$X^{(1)}(k)$	26.7	58.2	91.0	125.1	160.9	198.4

- Step2: Construct the matrix B and the vector Y_n

$$B = \begin{pmatrix} -\frac{1}{2}[X^{(1)}(1)+X^{(1)}(2)] & 1 \\ -\frac{1}{2}[X^{(1)}(2)+X^{(1)}(3)] & 1 \\ -\frac{1}{2}[X^{(1)}(3)+X^{(1)}(4)] & 1 \\ -\frac{1}{2}[X^{(1)}(4)+X^{(1)}(5)] & 1 \\ -\frac{1}{2}[X^{(1)}(5)+X^{(1)}(6)] & 1 \end{pmatrix} = \begin{pmatrix} -42.45 & 1 \\ -74.60 & 1 \\ -108.05 & 1 \\ -143.00 & 1 \\ -179.65 & 1 \end{pmatrix}$$

$$Y_n = \begin{pmatrix} X^{(0)}(2) \\ X^{(0)}(3) \\ X^{(0)}(4) \\ X^{(0)}(5) \\ X^{(0)}(6) \end{pmatrix} = \begin{pmatrix} 31.5 \\ 32.8 \\ 34.1 \\ 35.8 \\ 37.5 \end{pmatrix}$$

- Step3: Compute $B^T B$, $(B^T B)^{-1}$ and $B^T Y_n$.

$$B^T B = \begin{pmatrix} 71765.09 & -547.75 \\ -547.75 & 5.00 \end{pmatrix} \qquad (B^T B)^{-1} = \begin{pmatrix} 0.000085 & 0.009316 \\ 0.009316 & 1.220591 \end{pmatrix}$$

$$B^T Y_n = \begin{pmatrix} -0.043804 \\ 29.541220 \end{pmatrix}$$

- Step4: Solve the vector of parameters by using the least square estimate.

$$\hat{\alpha} = \begin{pmatrix} -0.043804 \\ 29.541220 \end{pmatrix} \qquad a = -0.043804 \qquad \mu = 29.541220$$

- Step5: Construct the GM(1,1) prediction Model

$$\frac{dX^{(1)}}{dt} + aX^{(1)} = \mu$$

$$\frac{dX^{(1)}}{dt} - 0.043804X^{(1)} = 29.541220$$

$$\begin{cases} \hat{X}^{(1)}(k+1) = \left(X^{(0)}(1) - \frac{\mu}{a} \right) e^{-ak} + \frac{\mu}{a} \\ X^{(0)}(1) = 26.7, \quad \frac{\mu}{a} = \text{-}674.3883 \end{cases}$$

So the GM(1,1) prediction model is:

$$\hat{X}^{(1)}(k+1) = 701.0883e^{0.043804k} \text{-}674.3883$$

Test the accuracy of the GM(1,1) Model

- Residual Test

Computed Values1	Actual Values	Computed Values2	Actual Values
$\hat{X}^{(1)}(1)=26.70000$	$X^{(1)}(1)=26.7$	$\hat{X}^{(0)}(1)=26.70000$	$X^{(0)}(1)=26.7$
$\hat{X}^{(1)}(2)=58.09337$	$X^{(1)}(2)=58.2$	$\hat{X}^{(0)}(2)=31.39337$	$X^{(0)}(2)=31.5$
$\hat{X}^{(1)}(3)=90.89246$	$X^{(1)}(3)=91.0$	$\hat{X}^{(0)}(3)=32.79910$	$X^{(0)}(3)=32.8$
$\hat{X}^{(1)}(4)=125.16024$	$X^{(1)}(4)=125.1$	$\hat{X}^{(0)}(4)=34.26778$	$X^{(0)}(4)=34.1$

- Residual Test

绝对误差序列	相对误差序列
$\Delta^{(0)}(i) = \left\lvert X^{(0)}(i) - \hat{X}^{(0)}(i) \right\rvert$	$\Phi(i) = \frac{\Delta^{(0)}(i)}{X^{(0)}(i)} \times 100\%$
$\Delta^{(0)}(1) = 0.000000$	$\Phi(1) = 0.000000\%$
$\Delta^{(0)}(2) = 0.106635$	$\Phi(2) = 0.339673\%$
$\Delta^{(0)}(3) = 0.000901$	$\Phi(3) = 0.002748\%$
$\Delta^{(0)}(4) = 0.167778$	$\Phi(4) = 0.489609\%$
$\Delta^{(0)}(5) = 0.002222$	$\Phi(5) = 0.006207\%$
$\Delta^{(0)}(6) = 0.094624$	$\Phi(6) = 0.252970\%$

- The Test of the degree of Grey incidence

Number	1	2	3	4	5	6
$\gamma(\hat{X}^{(0)}(i), X^{(0)}(i))$	1.000000	0.440307	0.989374	0.333333	0.974196	0.469932

$$\gamma(\hat{X}^{(0)}, X^{(0)}) = \frac{1}{n}\sum_{i=1}^{n}\gamma(\hat{X}^{(0)}(i), X^{(0)}(i)) = 0.70119$$

- C and P Criteria

$$S_1 = \sqrt{\frac{\sum [X^{(0)}(i) - \bar{X}^{(0)}]^2}{n-1}} = 3.775006$$

$$S_2 = \sqrt{\frac{\sum [\Delta^{(0)}(i) - \bar{\Delta}^{(0)}]^2}{n-1}} = 0.07126863$$

$$\mathrm{C} = \frac{\mathrm{S}_2}{\mathrm{S}_1} = 0.01887908 \qquad \mathrm{S}_0 = 2.546241$$

$$e_i = \left| \Delta^{(0)}(i) - \bar{\Delta}^{(0)} \right|$$

$$= \{0.06202667,\ 0.04460833,\ 0.06112567,\ 0.10575133,\ 0.05980467,\ 0.03259733\}$$

$$P = p\{\left| \Delta^{(0)}(i) - \bar{\Delta}^{(0)} \right| < 0.6745 S_1\} = 1$$

GM(1,1) Model can be used to predict.

	7	8	9	10	11
$\hat{X}^{(0)}$	39.08032	40.83026	42.65855	44.56872	46.56442
	$\hat{X}^{(0)}(7)$	$\hat{X}^{(0)}(8)$	$\hat{X}^{(0)}(9)$	$\hat{X}^{(0)}(10)$	$\hat{X}^{(0)}(11)$

Some Envisions

- R package？

  Any existing R package?

  Or can we write the first one?

- Collaboration

  More R programs and R Functions On grey methods?

Reference:

[1]邓聚龙，灰理论基础[M].上海：华中科技大学出版社，2003.

[2]刘思峰，谢思明，灰色系统理论及其应用[M].北京：科学出版社，2008.

[3]王庚，现代数学建模方法[M].北京：科学出版社，2008.

[4]徐国祥，统计预测与决策[M].上海：上海财经大学出版社，2005.

[5] Erdal Kayacan, Baris Ulutas, Okyay Kaynak, Grey system theory-based models in time series prediction[J]. Expert Systems with Applications, 2010:1784–1789.

Acknowledgements

- I am grateful to all members of COS, without your excellent work, no R conferences could be held currently in China.
- I am also grateful to Professor Wang Gen for his help and suggestions on this topic.
- And thank everyone here for your patient listening and welcome any suggestions.

Thank you!

Contact Information:

Tel: 13770636679

Email: xitannj@gmail.com