An Introduction to gRey methods by using R

Tan Xi
Department of Statistics, NUFEE
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

\[ \Delta_{0i}(k) = |Y_0(k) - Y_i(k)| \]

<table>
<thead>
<tr>
<th>Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\Delta_1)</td>
<td>0.000</td>
<td>0.112</td>
<td>0.072</td>
<td>0.225</td>
<td>0.660</td>
<td>0.931</td>
</tr>
<tr>
<td>(\Delta_2)</td>
<td>0.0000</td>
<td>0.0417</td>
<td>0.9050</td>
<td>1.0867</td>
<td>1.6083</td>
<td>2.5417</td>
</tr>
</tbody>
</table>
• **Step 3:** Compute the two-step minimum and maximum of the absolute subtraction sequences

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

\[
\Delta_{\text{max}} = \max_i \max_k \left| Y_0(k) - Y_i(k) \right| = 0
\]
Step 4: Compute coefficients of Grey incidence

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

of which the distinguishing coefficient \( \rho \) is 0.5.

<table>
<thead>
<tr>
<th>Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma(Y_0(k), Y_1(k)) )</td>
<td>1.0000</td>
<td>0.9190</td>
<td>0.9464</td>
<td>0.8496</td>
<td>0.6582</td>
<td>0.5772</td>
</tr>
<tr>
<td>( \gamma(Y_0(k), Y_2(k)) )</td>
<td>1.0000</td>
<td>0.9683</td>
<td>0.5841</td>
<td>0.5391</td>
<td>0.4414</td>
<td>0.3333</td>
</tr>
</tbody>
</table>
• **Step 5:** Compute the degree of Grey incidence

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

\[ \gamma(Y_0, Y_1) = 0.8251 \quad \gamma(Y_0, Y_2) = 0.6444 \]

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
  \[ \Delta_{0i}(k) = |Y_0(k) - Y_i(k)| \]
- Compute the two-step minimum and maximum of the absolute subtraction sequences
  \[ \Delta_{\text{min}} = \min_i \min_k |Y_0(k) - Y_i(k)| \quad \Delta_{\text{max}} = \max_i \max_k |Y_0(k) - Y_i(k)| \]
- Compute coefficients of Grey incidence
  \[ \gamma(Y_0(k), Y_i(k)) = \frac{\Delta_{\text{min}} + \rho \cdot \Delta_{\text{max}}}{\Delta_{0i}(k) + \rho \cdot \Delta_{\text{max}}} \]
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, $X^{(0)} = \{X^{(0)}(1), X^{(0)}(2), \ldots, X^{(0)}(n)\}$
- When this sequence is subjected to the Accumulating Generation Operation (AGO), the following sequence $X^{(1)}$ is obtained

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

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:

\[
X^{(0)}_p(k + H) = \left[ X^{(0)}(1) - \frac{b}{a} \right] e^{-a(k+H-1)} (1 - e^a)
\]
How to test the accuracy of the GM(1,1) Model?

• Residual Tests

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

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

$$\gamma(\hat{X}^{(0)}, X^{(0)}) > 0.6 \text{, when } \rho = 0.5 .$$
How to test the accuracy of the GM(1,1) Model?

- **C and P Criteria**

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

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

\[
C = \frac{S_2}{S_1} = 0.01887908
\]

\[
P = p\{\left|\Delta^{(0)}(i) - \bar{\Delta}^{(0)}\right| < 0.6745S_1\}\]
How to test the accuracy of the GM(1,1) Model?

• C and P Criteria

<table>
<thead>
<tr>
<th>( P )</th>
<th>( C )</th>
<th>判别结果</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 0.95</td>
<td>&lt; 0.35</td>
<td>好</td>
</tr>
<tr>
<td>&gt; 0.80</td>
<td>&lt; 0.50</td>
<td>合格</td>
</tr>
<tr>
<td>&gt; 0.70</td>
<td>&lt; 0.65</td>
<td>勉强合格</td>
</tr>
<tr>
<td>( \leq 0.70 )</td>
<td>( \geq 0.65 )</td>
<td>不合格</td>
</tr>
</tbody>
</table>
An easy example executed by R Program

Suppose the original sequence is:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X^{(0)}(i)$</td>
<td>26.7</td>
<td>31.5</td>
<td>32.8</td>
<td>34.1</td>
<td>35.8</td>
<td>37.5</td>
</tr>
</tbody>
</table>

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:

<table>
<thead>
<tr>
<th>$X^{(i)}(k)$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>26.7</td>
<td>58.2</td>
<td>91.0</td>
<td>125.1</td>
<td>160.9</td>
<td>198.4</td>
</tr>
</tbody>
</table>

• Step2: Construct the matrix $B$ and the vector $Y_n$

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

$$Y_n = \begin{pmatrix}
X^{(0)}(2) \\
X^{(0)}(3) \\
X^{(0)}(4) \\
X^{(0)}(5) \\
X^{(0)}(6)
\end{pmatrix} = \begin{pmatrix}
31.5 \\
32.8 \\
34.1 \\
35.8 \\
37.5
\end{pmatrix}$$
• Step 3: Compute $B^T B$, $(B^T B)^{-1}$ and $B^T Y_n$.

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

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

• Step 4: Solve the vector of parameters by using the least square estimate.

\[
\hat{\alpha} = \begin{pmatrix} -0.043804 \\ 29.541220 \end{pmatrix} \quad a = -0.043804 \quad \mu = 29.541220
\]
• Step 5: Construct the GM(1,1) prediction Model

\[
\frac{dX^{(1)}}{dt} + aX^{(1)} = \mu \\
\frac{dX^{(1)}}{dt} - 0.043804X^{(1)} = 29.541220
\]

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

\[
\hat{X}^{(1)}(k + 1) = \left( X^{(0)}(1) - \frac{\mu}{a} \right) e^{-ak} + \frac{\mu}{a} \\
X^{(0)}(1) = 26.7, \quad \frac{\mu}{a} = -674.3883
\]

\[
\hat{X}^{(1)}(k + 1) = 701.0883e^{0.043804k} - 674.3883
\]
Test the accuracy of the GM(1,1) Model

- Residual Test

<table>
<thead>
<tr>
<th>Computed Values1</th>
<th>Actual Values</th>
<th>Computed Values2</th>
<th>Actual Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{X}^{(1)}(1) = 26.70000$</td>
<td>$X^{(1)}(1) = 26.7$</td>
<td>$\hat{X}^{(0)}(1) = 26.70000$</td>
<td>$X^{(0)}(1) = 26.7$</td>
</tr>
<tr>
<td>$\hat{X}^{(1)}(2) = 58.09337$</td>
<td>$X^{(1)}(2) = 58.2$</td>
<td>$\hat{X}^{(0)}(2) = 31.39337$</td>
<td>$X^{(0)}(2) = 31.5$</td>
</tr>
<tr>
<td>$\hat{X}^{(1)}(3) = 90.89246$</td>
<td>$X^{(1)}(3) = 91.0$</td>
<td>$\hat{X}^{(0)}(3) = 32.79910$</td>
<td>$X^{(0)}(3) = 32.8$</td>
</tr>
<tr>
<td>$\hat{X}^{(1)}(4) = 125.16024$</td>
<td>$X^{(1)}(4) = 125.1$</td>
<td>$\hat{X}^{(0)}(4) = 34.26778$</td>
<td>$X^{(0)}(4) = 34.1$</td>
</tr>
</tbody>
</table>
• Residual Test

<table>
<thead>
<tr>
<th>绝对误差序列</th>
<th>相对误差序列</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta^{(0)}(i) =</td>
<td>X^{(0)}(i) - \hat{X}^{(0)}(i)</td>
</tr>
<tr>
<td>( \Delta^{(0)}(1) = 0.000000 )</td>
<td>( \Phi(1) = 0.000000% )</td>
</tr>
<tr>
<td>( \Delta^{(0)}(2) = 0.106635 )</td>
<td>( \Phi(2) = 0.339673% )</td>
</tr>
<tr>
<td>( \Delta^{(0)}(3) = 0.000901 )</td>
<td>( \Phi(3) = 0.002748% )</td>
</tr>
<tr>
<td>( \Delta^{(0)}(4) = 0.167778 )</td>
<td>( \Phi(4) = 0.489609% )</td>
</tr>
<tr>
<td>( \Delta^{(0)}(5) = 0.002222 )</td>
<td>( \Phi(5) = 0.006207% )</td>
</tr>
<tr>
<td>( \Delta^{(0)}(6) = 0.094624 )</td>
<td>( \Phi(6) = 0.252970% )</td>
</tr>
</tbody>
</table>
• The Test of the degree of Grey incidence

<table>
<thead>
<tr>
<th>Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma(\hat{X}^{(0)}(i), X^{(0)}(i))$</td>
<td>1.000000</td>
<td>0.440307</td>
<td>0.989374</td>
<td>0.333333</td>
<td>0.974196</td>
<td>0.469932</td>
</tr>
</tbody>
</table>

$$\gamma(\hat{X}^{(0)}, X^{(0)}) = \frac{1}{n} \sum_{i=1}^{n} \gamma(\hat{X}^{(0)}(i), X^{(0)}(i)) = 0.70119$$
• C and P Criteria

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

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

\[
C = \frac{S_2}{S_1} = 0.01887908 \quad S_0 = 2.546241
\]

\[
e_i = |\Delta^{(0)}(i) - \bar{\Delta}^{(0)}|
\]

\[
= \{0.06202667, 0.04460833, 0.06112567, 0.10575133, 0.05980467, 0.03259733\}
\]

\[
P = p\{\left|\Delta^{(0)}(i) - \bar{\Delta}^{(0)}\right| < 0.6745S_1\} = 1
\]
GM(1,1) Model can be used to predict.

<table>
<thead>
<tr>
<th></th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{X}^{(0)}$</td>
<td>39.08032</td>
<td>40.83026</td>
<td>42.65855</td>
<td>44.56872</td>
<td>46.56442</td>
</tr>
<tr>
<td>$\hat{X}^{(0)}(7)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{X}^{(0)}(8)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{X}^{(0)}(9)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{X}^{(0)}(10)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{X}^{(0)}(11)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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:

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