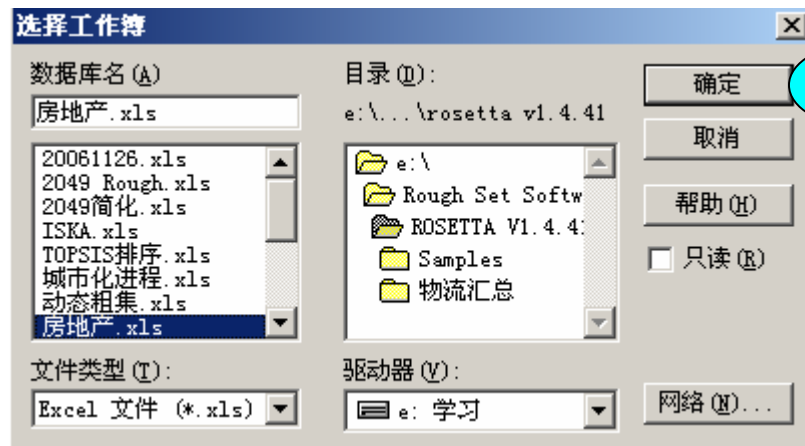
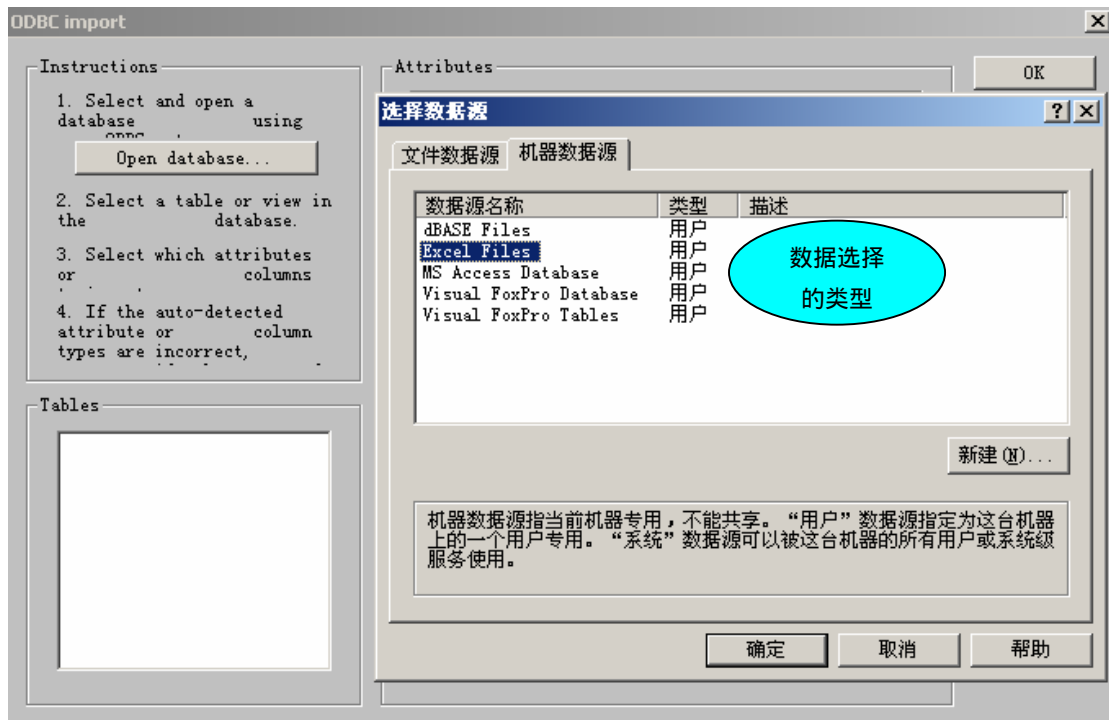
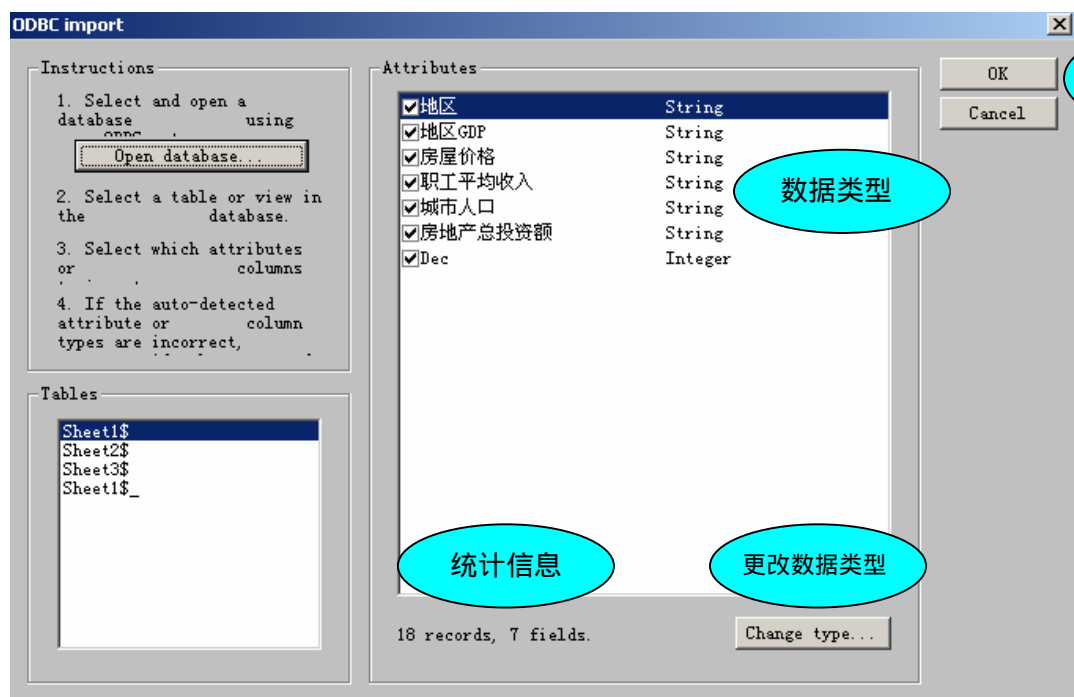


Rosetta 软件基本用法中文指导书

西南交通大学经管学院 刘盾博士

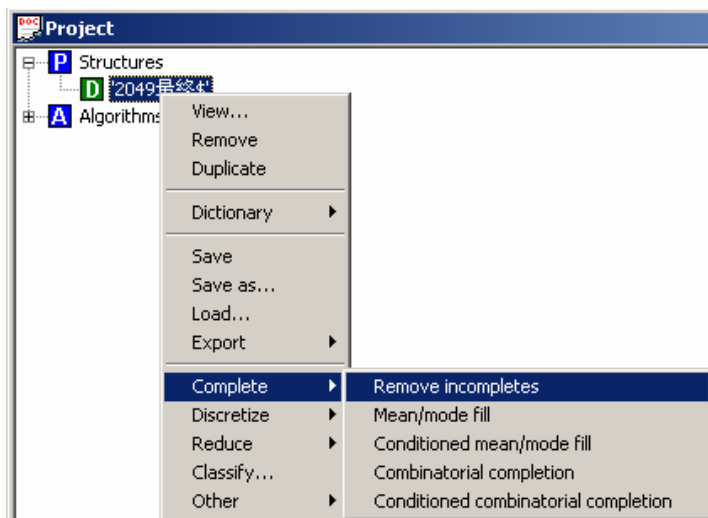
1. 数据的读入



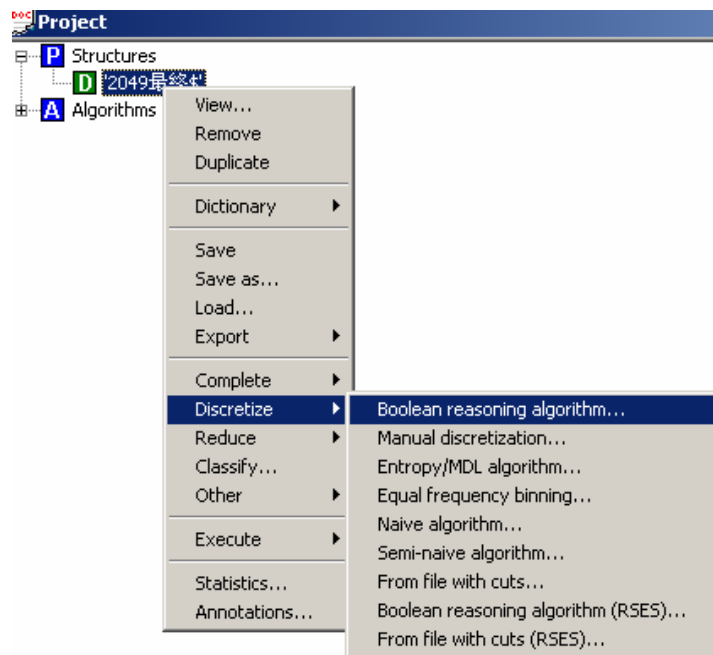


	地区GDP	房屋价格	职工平均收入	城市人口	房地产总投资额	Dec
1	A	A	A	A	A	1
2	D	C	E	C	D	4
3	D	B	B	E	E	2
4	D	D	E	D	C	4
5	B	D	B	C	C	2
6	B	B	D	B	B	3
7	E	D	F	E	E	5
8	E	E	E	D	E	4
9	D	D	E	D	D	4
10	C	E	E	C	D	4
11	C	D	E	B	C	4
12	D	D	E	D	D	4
13	C	C	D	C	C	3
14	D	E	E	E	D	4
15	D	E	F	B	B	5
16	E	B	E	F	F	4
17	E	F	F	E	F	5
18	D	E	E	D	E	3

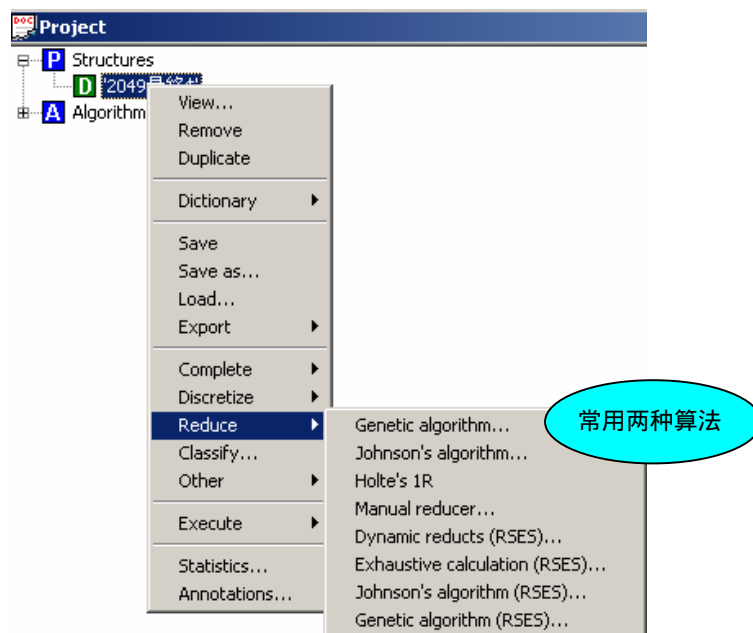
2. 数据的补齐



3. 数据的离散化



4. 数据的约简



利用遗传算法得到的约简：

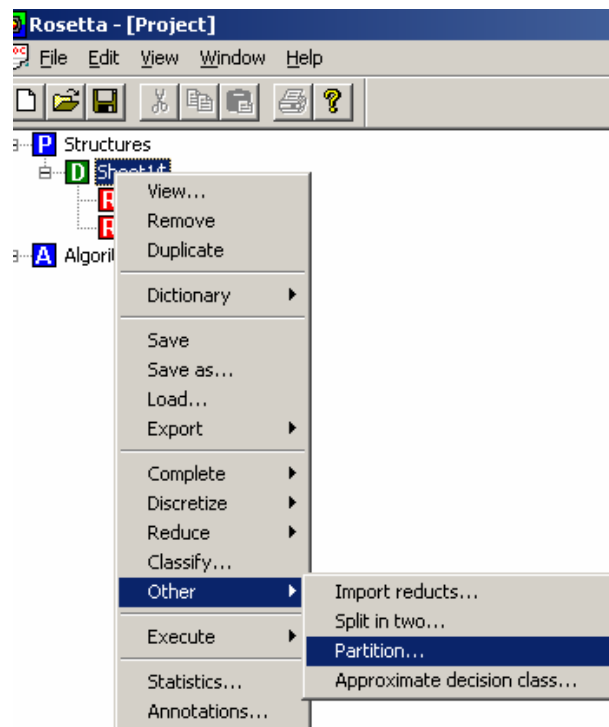
Reduct	Support	Length
{地区GDP, 城市人口, 房地产总投资额}	100	3
{地区GDP, 房屋价格, 房地产总投资额}	100	3
{地区GDP, 房屋价格, 城市人口}	100	3
{地区GDP, 职工平均收入, 房地产总投资额}	100	3

核：所有约简的交

利用 Johnson 算法得到的约简：

Reduct	Support	Length
{地区GDP, 房屋价格, 城市人口}	100	3

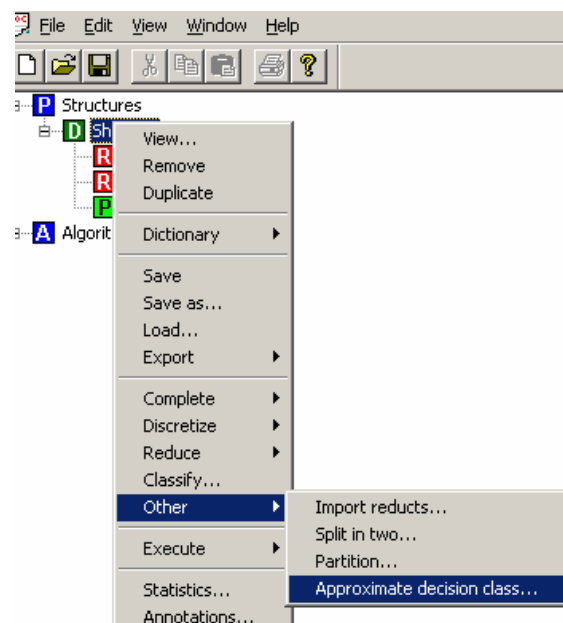
5. 上下近似集的计算



等价类的获取：

Eq. class	Cardinality
{1}	1
{2}	1
{3}	1
{9, 12}	2
{4}	1
{14}	1
{18}	1
{15}	1
{6}	1
{5}	1
{16}	1
{7}	1
{8}	1
{17}	1
{13}	1
{11}	1
{10}	1

上下近似集的获取：



Approximator

Approximation target

Decision: 1 Precision: 0

1
2
3
4
5

决策属性值的选取

OK
Cancel

Approximator

Approximation target

Decision: 1 Precision: 0

Attributes

☐ 地区GDP
☐ 房屋价格
☐ 职工平均收入
☐ 城市人口
☐ 房地产总投资额
☐ Dec

条件属性的选取

OK
Cancel
Invert

Approximator

Approximation target

Decision: 0 Precision: 0.2

Attributes...

容错率 μ 的值
 $0 \leq \mu \leq 0.5$

OK
Cancel

变精度粗集模型

Approximator

Approximation target

Decision: 4 Precision: 0

Attributes

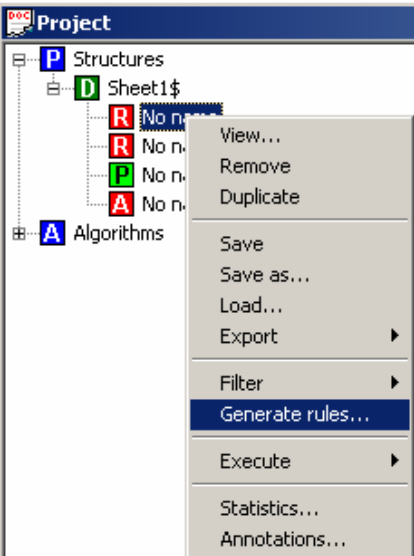
☒ 地区GDP
☒ 房屋价格
☒ 职工平均收入
☒ 城市人口
☒ 房地产总投资额
☐ Dec

No name

	Universe	Upper	Lower	Boundary	Outside
1	{1}	{2}	{2}		{1}
2	{2}	{9, 12}	{9, 12}		{3}
3	{3}	{4}	{4}		{18}
4	{9, 12}	{14}	{14}		{15}
5	{4}	{16}	{16}		{6}
6	{14}	{8}	{8}		{5}
7	{18}	{11}	{11}		{7}
8	{15}	{10}	{10}		{17}
9	{6}				{13}
10	{5}				
11	{16}				
12	{7}				
13	{8}				
14	{17}				
15	{13}				
16	{11}				
17	{10}				

上近似集
下近似集
边界域
负域

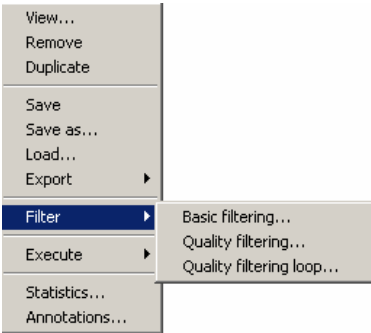
6. 规则生成



Option	Remove if
<i>RHS Support</i>	$l \leq support(\alpha \cdot (d = v)) \leq u$
<i>RHS Accuracy</i>	$l \leq accuracy(\alpha \rightarrow (d = v)) \leq u$
<i>RHS Coverage</i>	$l \leq coverage(\alpha \rightarrow (d = v)) \leq u$
<i>RHS Stability</i>	$l \leq stability(\alpha \rightarrow (d = v)) \leq u$
<i>Decision</i>	$v = v'$
<i>LHS Length</i>	$l \leq length(\alpha) \leq u$
<i>Condition</i>	$(a = v')$ occurs in α

LHS : Left hand side , 左边
RHS : Right hand side , 右边

Rule	RHS Support	RHS Accuracy	RHS Coverage	RHS Stability	LHS Length	RHS Length
地区GDP(A) AND 房屋价格(A) AND 城市人口(A) => Dec(1)	1	1.0	1.0	1.0	3	1
地区GDP(D) AND 房屋价格(C) AND 城市人口(C) => Dec(4)	1	1.0	0.111111	1.0	3	1
地区GDP(D) AND 房屋价格(B) AND 城市人口(E) => Dec(2)	1	1.0	0.5	1.0	3	1
地区GDP(D) AND 房屋价格(D) AND 城市人口(D) => Dec(4)	3	1.0	0.333333	1.0	3	1
地区GDP(B) AND 房屋价格(D) AND 城市人口(C) => Dec(2)	1	1.0	0.5	1.0	3	1
地区GDP(B) AND 房屋价格(B) AND 城市人口(B) => Dec(3)	1	1.0	0.333333	1.0	3	1
地区GDP(E) AND 房屋价格(D) AND 城市人口(E) => Dec(5)	1	1.0	0.333333	1.0	3	1
地区GDP(E) AND 房屋价格(E) AND 城市人口(D) => Dec(4)	1	1.0	0.111111	1.0	3	1
地区GDP(C) AND 房屋价格(E) AND 城市人口(C) => Dec(4)	1	1.0	0.111111	1.0	3	1
地区GDP(C) AND 房屋价格(D) AND 城市人口(B) => Dec(4)	1	1.0	0.111111	1.0	3	1
地区GDP(C) AND 房屋价格(C) AND 城市人口(C) => Dec(3)	1	1.0	0.333333	1.0	3	1
地区GDP(D) AND 房屋价格(E) AND 城市人口(E) => Dec(4)	1	1.0	0.111111	1.0	3	1
地区GDP(D) AND 房屋价格(E) AND 城市人口(B) => Dec(5)	1	1.0	0.333333	1.0	3	1
地区GDP(E) AND 房屋价格(B) AND 城市人口(F) => Dec(4)	1	1.0	0.111111	1.0	3	1
地区GDP(E) AND 房屋价格(F) AND 城市人口(E) => Dec(5)	1	1.0	0.333333	1.0	3	1
地区GDP(D) AND 房屋价格(E) AND 城市人口(D) => Dec(3)	1	1.0	0.333333	1.0	3	1
地区GDP(A) AND 职工平均收入(A) AND 房地产总投资额(A) => Dec(1)	1	1.0	1.0	1.0	3	1
地区GDP(D) AND 职工平均收入(E) AND 房地产总投资额(D) => Dec(4)	4	1.0	0.444444	1.0	3	1
地区GDP(D) AND 职工平均收入(E) AND 房地产总投资额(E) => Dec(2)	1	1.0	0.5	1.0	3	1
地区GDP(D) AND 职工平均收入(E) AND 房地产总投资额(C) => Dec(4)	1	1.0	0.111111	1.0	3	1
地区GDP(B) AND 职工平均收入(B) AND 房地产总投资额(C) => Dec(2)	1	1.0	0.5	1.0	3	1
地区GDP(B) AND 职工平均收入(D) AND 房地产总投资额(B) => Dec(3)	1	1.0	0.333333	1.0	3	1
地区GDP(E) AND 职工平均收入(F) AND 房地产总投资额(E) => Dec(5)	1	1.0	0.333333	1.0	3	1
地区GDP(E) AND 职工平均收入(E) AND 房地产总投资额(E) => Dec(4)	1	1.0	0.111111	1.0	3	1
地区GDP(C) AND 职工平均收入(E) AND 房地产总投资额(D) => Dec(4)	1	1.0	0.111111	1.0	3	1
地区GDP(C) AND 职工平均收入(E) AND 房地产总投资额(C) => Dec(4)	1	1.0	0.111111	1.0	3	1



规则的挑选：

Rule filter

Criteria

☐ Remove rules with RMS support

☒ Remove rules with RMS accuracy

☒ Remove rules with RMS coverage

☐ Remove rules with RMS stability

☐ Remove rules with decision

☐ Remove rules with LMS length

☐ Remove rules with condition

Connective

Interpret the composed set of

☐ Conjunction (AND)

☒ Disjunction (OR)

☐ Invert logic

OK

Cancel

选取精度>0.75
和覆盖度>0.05
的规则

Rule quality filter

Quality measure

☐ Michalski Accuracy

☐ Coleman Normalization

☒ Torgo

☐ G2

☐ Brazdil

☐ J

☐ Pearson

☐ Kononenko

☐ Cohen

Filtering criterion

☒ Remove if quality is less than

☐ Remove if quality is greater than

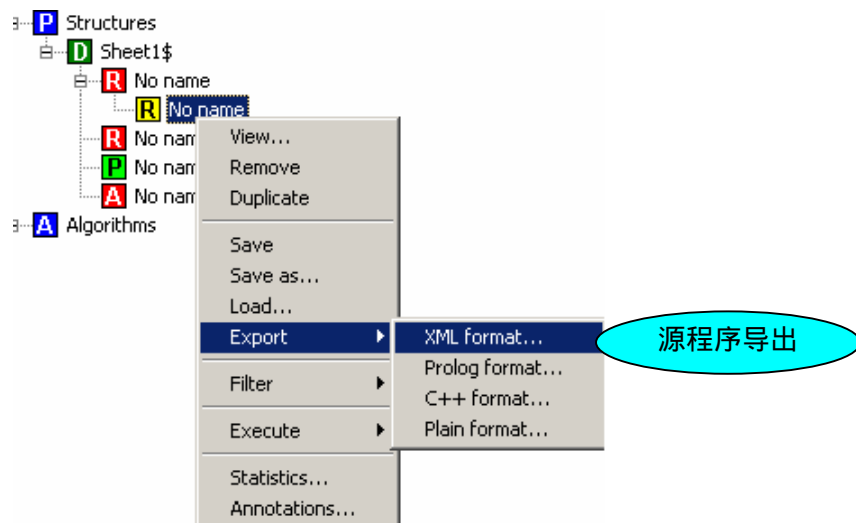
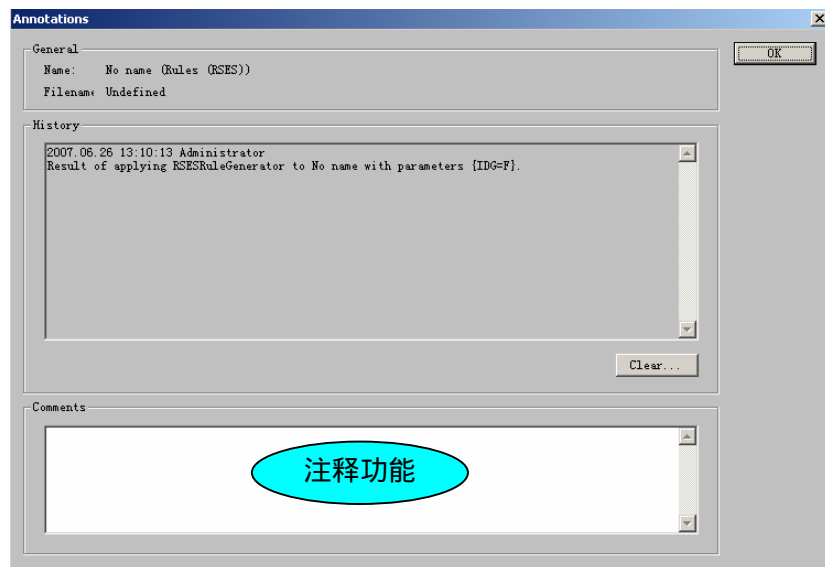
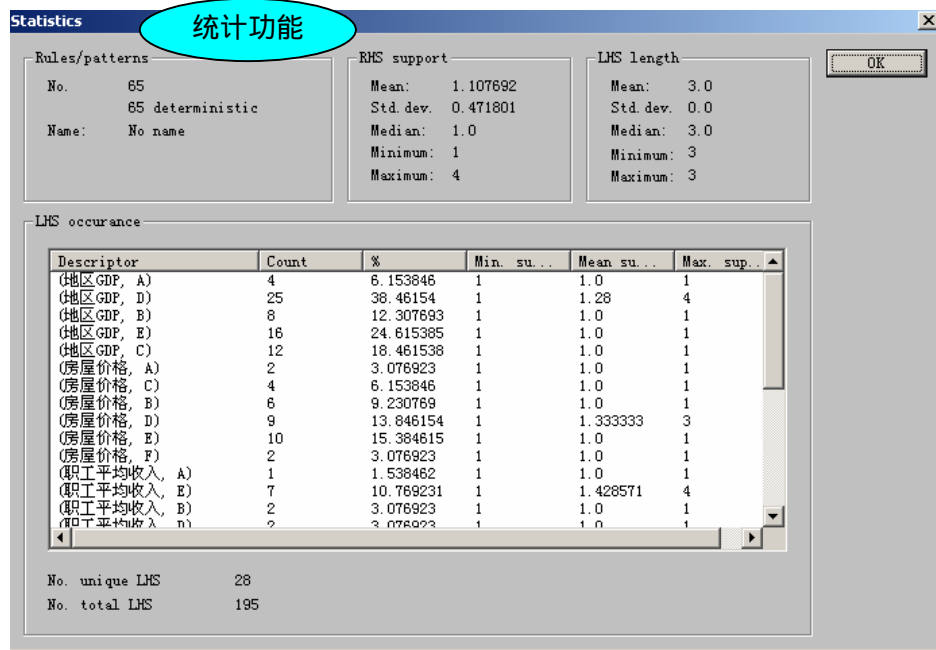
☒ Remove if quality is under

☐ Invert log

OK

Cancel

7. 其他功能



导出 C++程序：

```

                                                                    Undefined
//-----
// The header (.h) file starts here.
//=====

// Classifier generated by ROSETTA.
// Exported 2007.06.26 13:22:30 by Administrator.
//
// No name
// 65 rules.

#ifndef __CLASSIFIER_EXPORTED_FROM_ROSETTA_H__
#define __CLASSIFIER_EXPORTED_FROM_ROSETTA_H__

//-----
// Class.....: ROSETTAObject
// Author.....: Aleksander 蔡rn
// Date.....: 2007.06.26 13:22:30
// Description...: An object input to a ROSETTAClassifier.
//
//
//          For a legend to the encoding of slots for the
//          symbolic attributes, see the Lookup methods.
//=====

class ROSETTAObject {
public:

    // Variable slots.....
    int      地区GDP;
    int      房屋价格;
    int      职工平均收入;
    int      城市人口;
    int      房地产总投资额;

    // Dictionary lookup methods.....
    static int Lookup地区GDP(const char *text);
    static int Lookup房屋价格(const char *text);
    static int Lookup职工平均收入(const char *text);
    static int Lookup城市人口(const char *text);
    static int Lookup房地产总投资额(const char *text);

};
```

文件保存：

