Tato prezentace je součástí wiki-prezentace Metoda GUHA, LISp-Miner a typové úlohy
Je dostupná z této adresy
Verse 20. 8. 2019

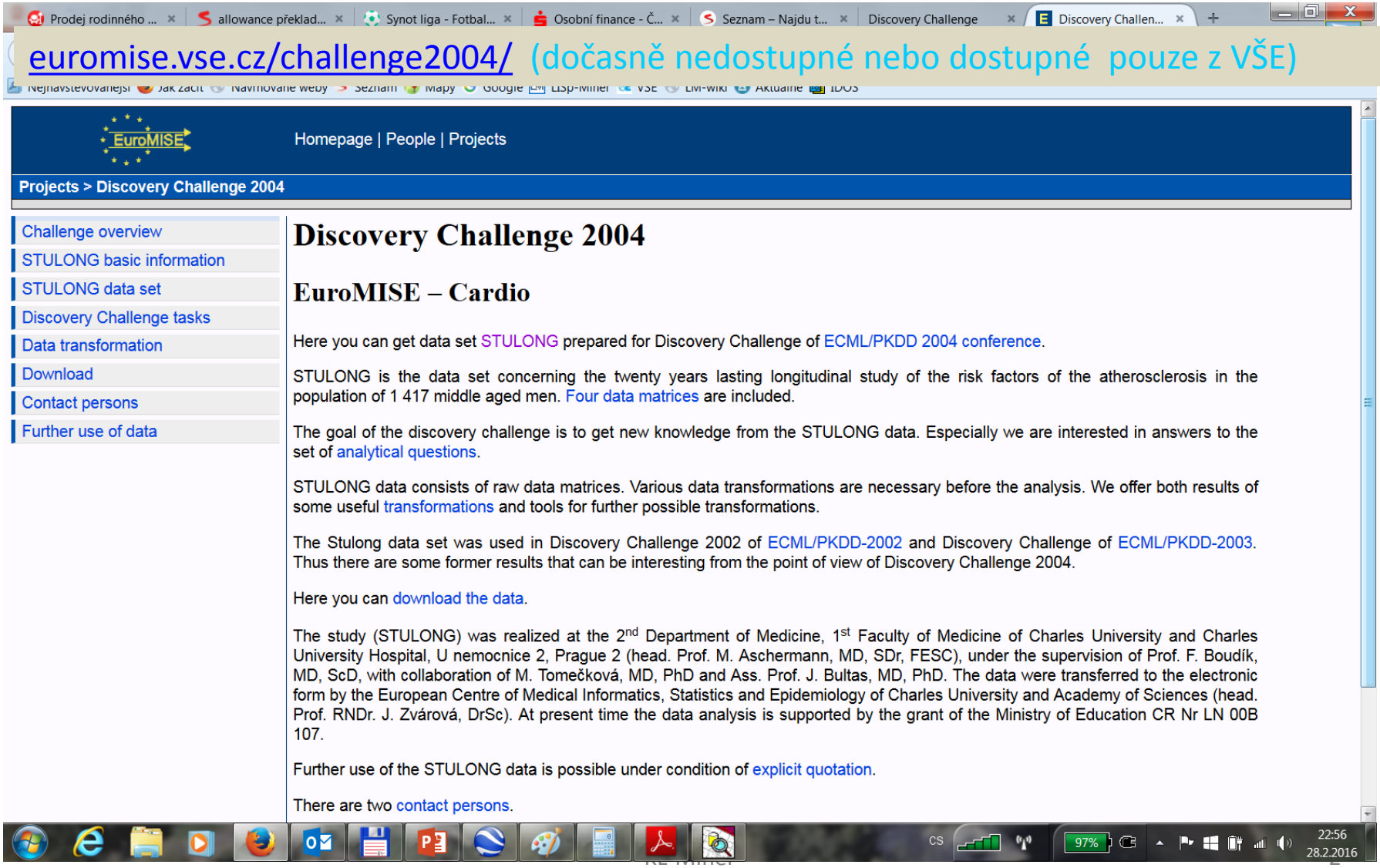
Typ úlohy: Pozitivní ordinální asociace

Data: Stulong

Problém: *Jaké ordinální závislosti platí mezi dostupnými ordinálními atributy ve skupinách pacientů definovaných pomocí osobních údajů, spotřeby alkoholu a anamnézy?*

Jan Rauch
Katedra informačního a znalostního inženýrství
Vysoká škola ekonomická v Praze

Aplikace procedury KL-Miner – příklad



Data STULONG

euromise.vse.cz/challenge2004/data/entry/ (dočasně nedostupné nebo dostupné pouze z VŠE)

EuroMISE

Homepage | People | Projects

Projects > Discovery Challenge 2004 > Data set > Entry

- A survey of attributes
- Identification data
- Social characteristics
- Activities
- Smoking
- Alcohol
- Sugar, coffee, tea
- Personal anamnesis
- Questionnaire A_2
- Physical exam.
- Biochemical exam.
- Risk factors

Entry examination – a survey of attributes

1 417 men have been examined during the entry examination. Values of 244 attributes have been surveyed with each patient. Values of 64 attributes are either codes or results of size measurements of different variables or results of transformations of the rest of the attributes. Values of all these 64 attributes are stored in the data matrix Entry. Attributes can be divided into groups according to the Table 1.

Table 1: Groups of the attributes in the entry examination

Groups of attributes	Number of attributes
identification data	2
social characteristics	6
physical activity	4
smoking	3
drinking of alcohol	
sugar, coffee, tea	
personal anamnesis	
questionnaire A_2	
physical examination	
biochemical examination	
risk faktors	5

Cíle:

- Příklad aplikace procedury KL-Miner na reálná data
- Ukázat význam dílčího cedentu – disjunkce

Print page PDF version

Projects > Discovery Challenge 2004 > Data set > Entry

Mail to: webmaster

Data STULONG

euromise.vse.cz/challenge2004/data/entry/ (dočasně nedostupné nebo dostupné pouze z VŠE)

Analytická otázka:
Jaké ordinální závislosti platí mezi dostupnými ordinálními atributy ve skupinách pacientů definovaných pomocí osobních údajů, spotřeby alkoholu a anamnézy?

Projects > Discovery Challen

- A survey of attributes
- Identification data
- Social characteristics
- Activities
- Smoking
- Alcohol
- Sugar, coffee, tea
- Personal anamnesis
- Questionnaire A_2
- Physical exam.
- Biochemical exam.
- Risk factors

1 417 men have been examined during the entry examination. Values of 244 attributes have been surveyed with each patient. Values of 64 attributes are either codes or results of size measurements of different variables or results of transformations of the rest of the

? KL: $\uparrow\uparrow_{0.7}$ STULONG / Osobní, Alkohol, Anamnéza

Table 1: Groups of the attributes in the entry examination

Groups of attributes	Number of attributes
identification data	2
social characteristics	6
physical activity	4
smoking	3
drinking of alcohol	9
sugar, coffee, tea	3
personal anamnesis	18
questionnaire A_2	3
physical examination	8
biochemical examination	3
risk faktors	5

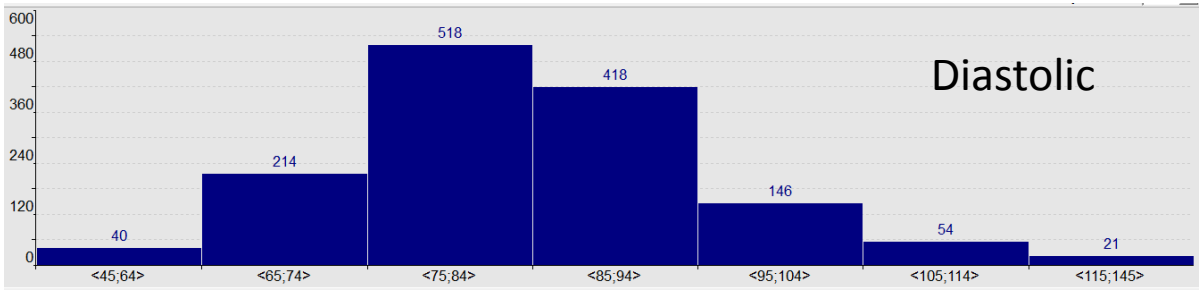
Print page PDF version

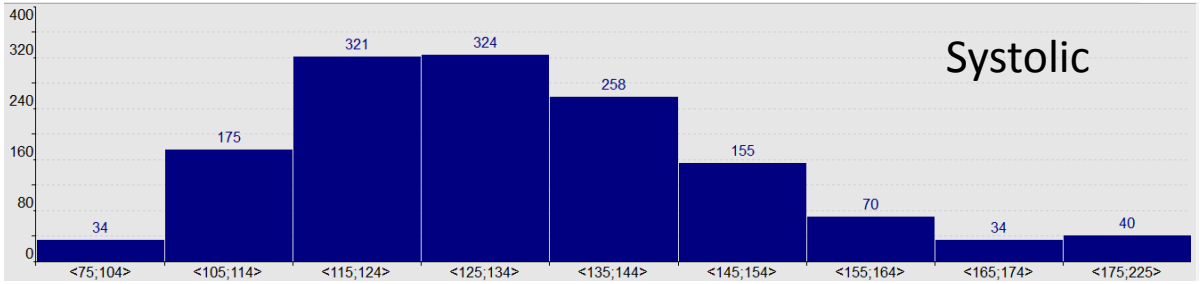
Projects > Discovery Challenge 2004 > Data set > Entry

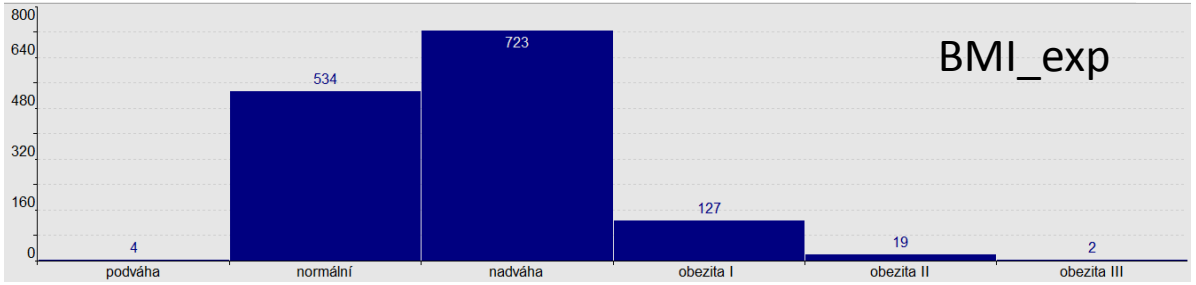
Mail to: webmaster

Ordinální atributy

Row attributes
- » Cholesterol(19 categories)
- » Triglyceridy(12 categories)
- » Cukr(8 categories)
- » Čaj(3 categories)
- » Káva(3 categories)
- » Doba kouření(4 categories)
- » Diastolic(7 categories)
- » Systolic(9 categories)
- » Bmi_exp(6 categories)
- » Výška_ed10(6 categories)
- » Triceps(12 categories)
- » Váha_ed5(11 categories)







...

Osobní údaje

Basic parameters

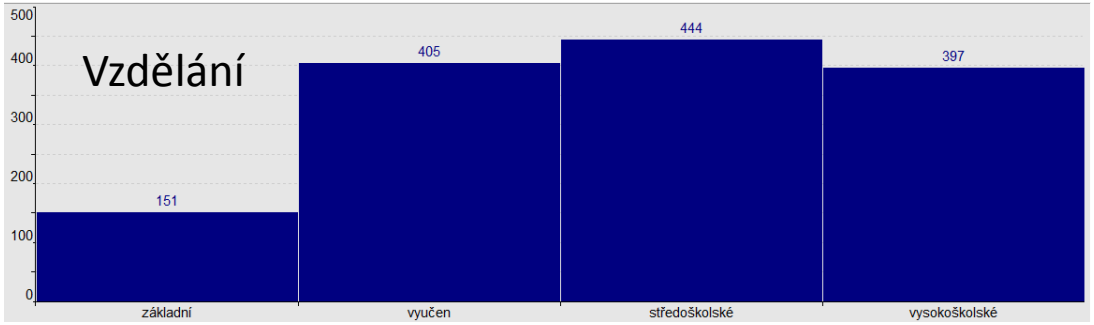
Name: Osobní údaje

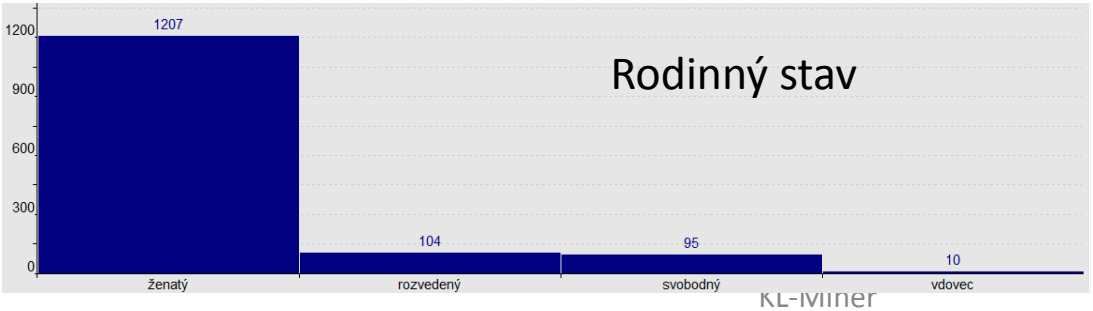
Min. length: 0 Max. length: 2 Literals boolean operation type: Conjunction

Comment: -

Literals Settings

Underlying attribute	Categories	X-cat	Coefficient type	Length	+/-	B/R
Vzdělání	4	Yes	Subsets	1 - 1	pos	Basic
Rodinný stav	4	Yes	Subsets	1 - 1	pos	Basic





Spotřeba alkoholu

Basic parameters

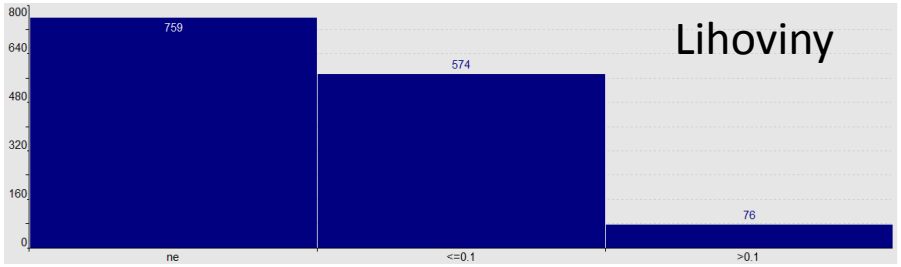
Name: Alkohol

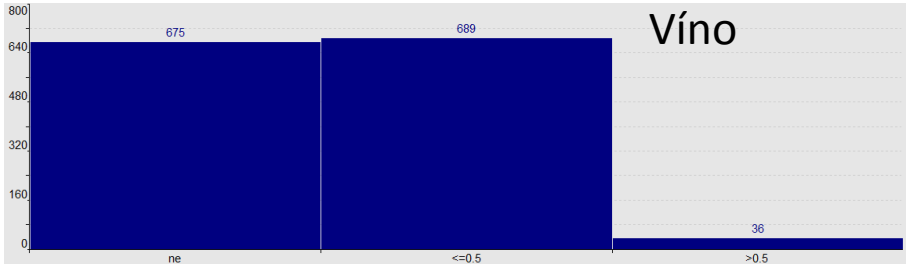
Min. length: 0 Max. length: 3 Literals boolean operation type: Conjunction

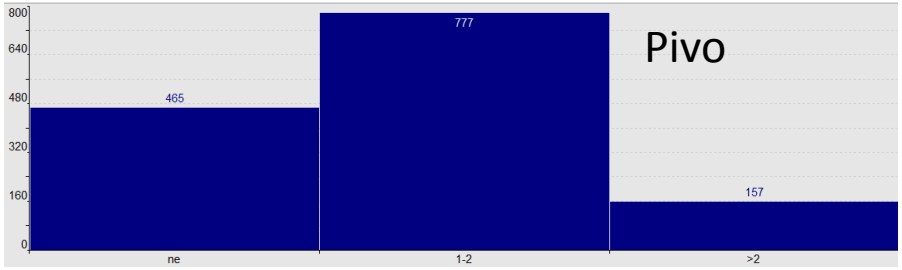
Comment: -

Literals Settings

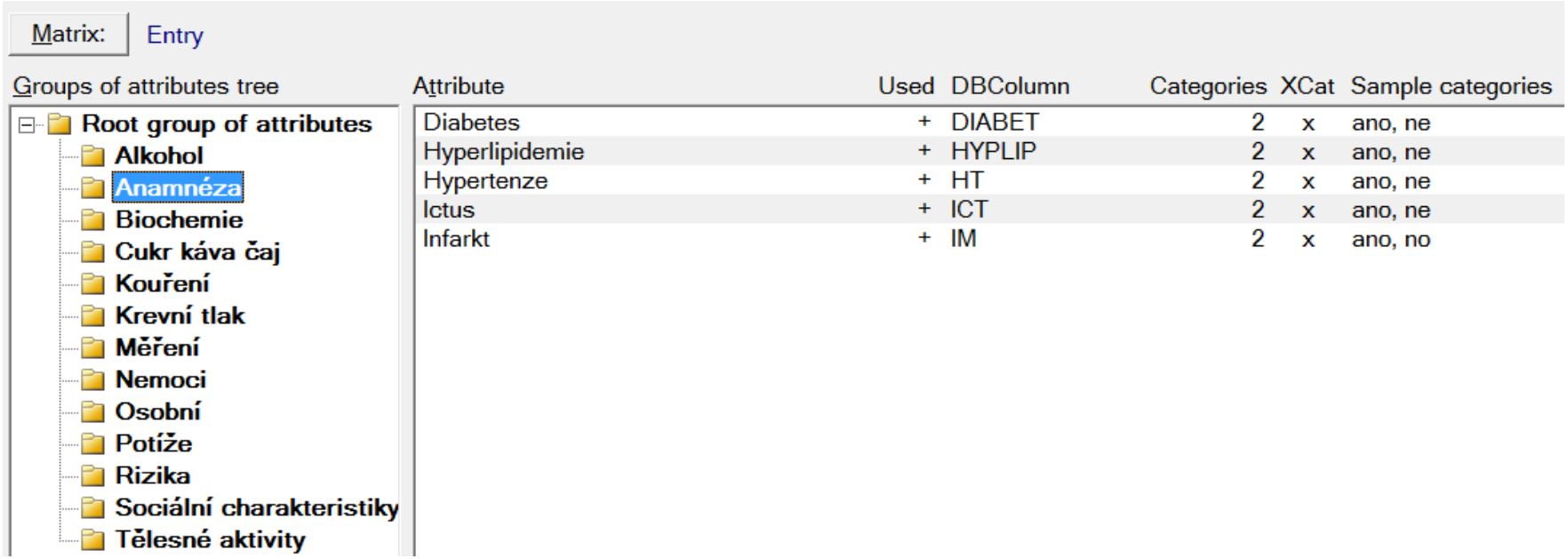
Underlying attribute	Categories	X-cat	Coefficient type	Length	+/-	B/R
Lihoviny	3	Yes	Subsets	1 - 1	pos	Basic
Pivo	3	Yes	Subsets	1 - 1	pos	Basic
Víno	3	Yes	Subsets	1 - 1	pos	Basic







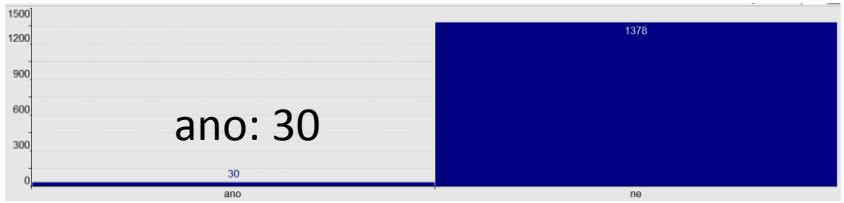
Anamnéza



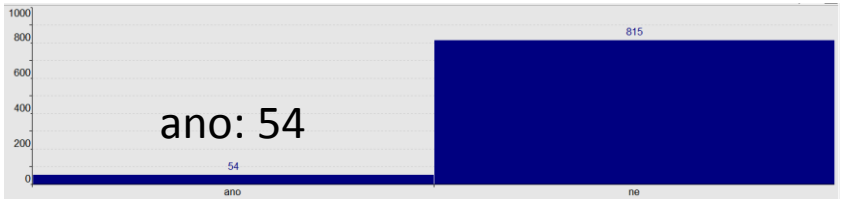
Skupina *Anamnéza* – u atributů nás zajímá kategorie *ano*

Anamnéza – frekvence kategorie *ano*

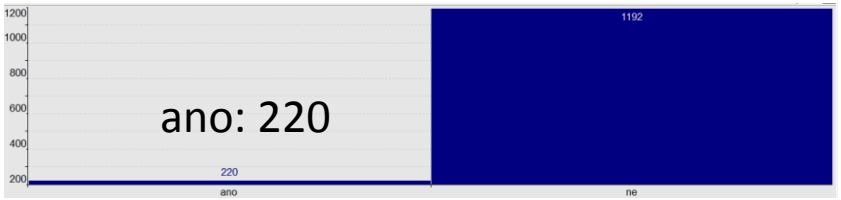
Diabetes



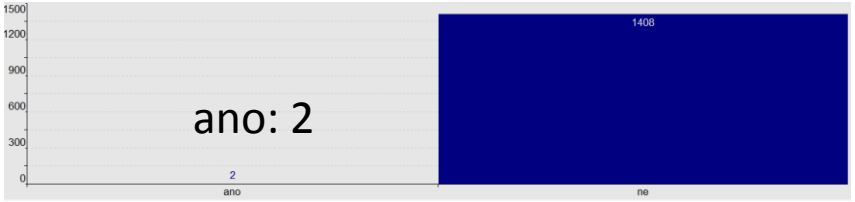
Hyperlipidemie



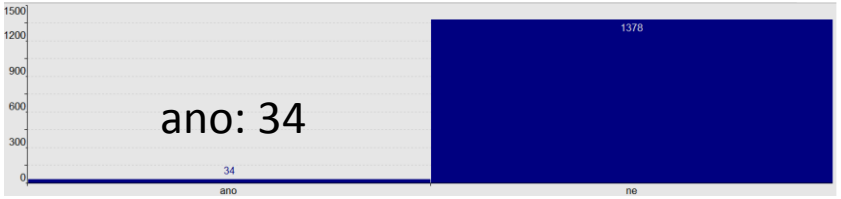
Hypertenze



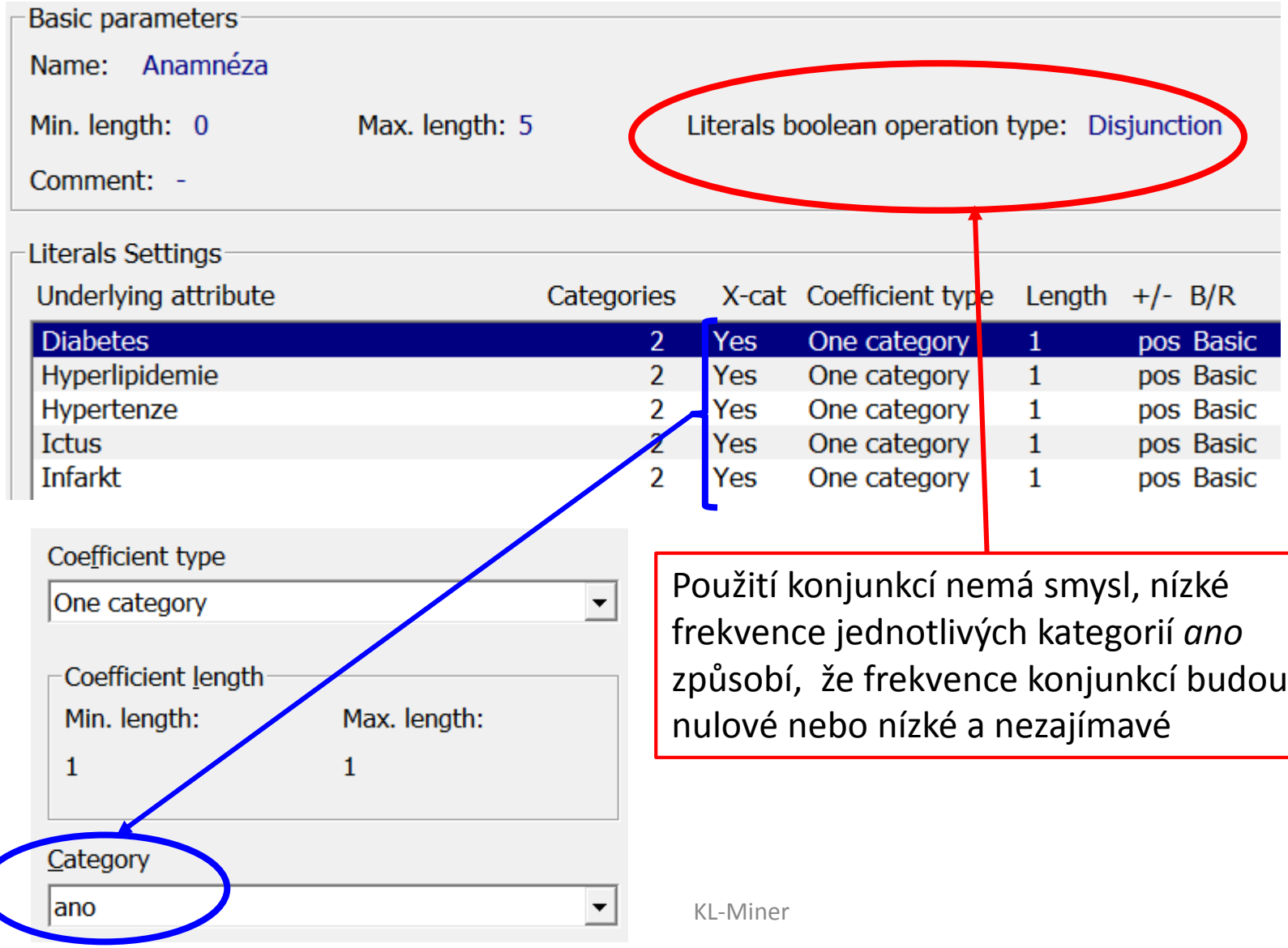
Ictus



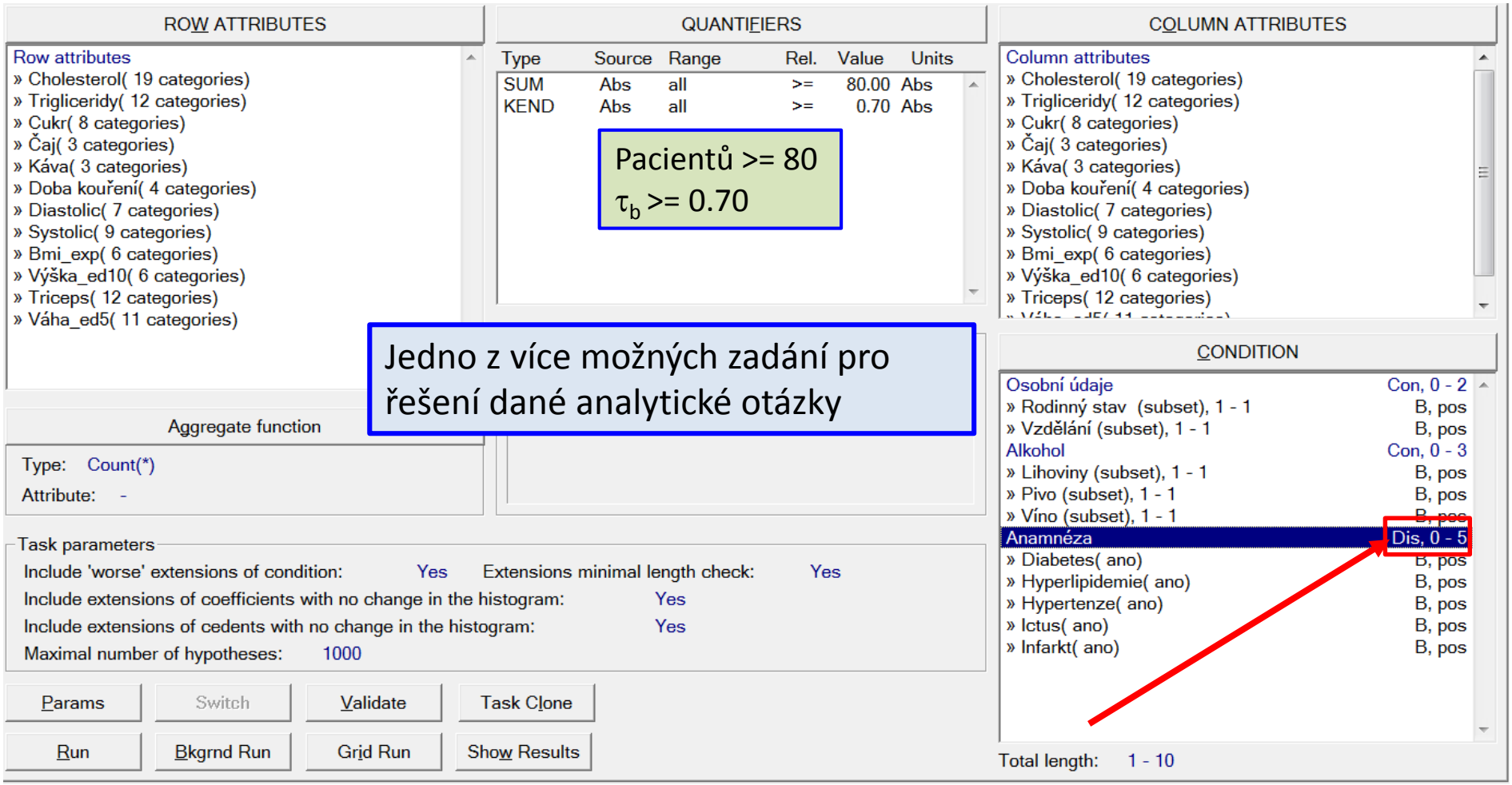
Infarkt



Anamnéza – zadání dílčího cedentu



? KL: $\uparrow\uparrow_{0.7}$ STULONG / Osobní, Alkohol, Anamnéza (1)



? KL: $\uparrow\uparrow_{0.7}$ STULONG / Osobní, Alkohol, Anamnéza (1) – výstup

Task run
Start: 1.10.2016 12:12:07 Total time: 0h 0m 34s
Number of verifications: 697884
Number of hypotheses: 216 Mode: Standard

Add group | Del group | Edit group

Actual group of hypotheses: All hypotheses
Hypotheses in group: 216 Shown hypotheses: 216 Highlighted: 0

Delete hypotheses

Nr.	Id	KEND	Hypothesis
1	147	0.770	**Diastolic** × **Systolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*) | **Infarkt**(*ano*)]
2	148	0.770	**Systolic** × **Diastolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*) | **Infarkt**(*ano*)]
3	149	0.770	**Diastolic** × **Systolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Infarkt**(*ano*)]
4	150	0.770	**Systolic** × **Diastolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Infarkt**(*ano*)]
5	145	0.767	**Diastolic** × **Systolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*)]
6	146	0.767	**Systolic** × **Diastolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*)]
7	143	0.766	**Diastolic** × **Systolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*)]
8	144	0.766	**Systolic** × **Diastolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*)]
9	135	0.764	**Diastolic** × **Systolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Diabetes**(*ano*) | **Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*) |
10	136	0.764	**Systolic** × **Diastolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Diabetes**(*ano*) | **Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*) |
11	137	0.763	**Diastolic** × **Systolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Diabetes**(*ano*) | **Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Infarkt**(*ano*)
12	138	0.763	**Systolic** × **Diastolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Diabetes**(*ano*) | **Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Infarkt**(*ano*)
13	53	0.761	**Diastolic** × **Systolic** / **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*) | **Infarkt**(*ano*)]
14	54	0.761	**Systolic** × **Diastolic** / **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*) | **Infarkt**(*ano*)]
15	51	0.761	**Diastolic** × **Systolic** / **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*)]
16	52	0.761	**Systolic** × **Diastolic** / **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*)]
17	55	0.761	**Diastolic** × **Systolic** / **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Infarkt**(*ano*)]
18	56	0.761	**Systolic** × **Diastolic** / **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Infarkt**(*ano*)]
19	133	0.760	**Diastolic** × **Systolic** / **Rodinný stav** (*ženatý*) & **Lihoviny**(*ne*) & **Víno**(*ne*) & [**Diabetes**(*ano*) | **Hyperlipidemie**(*ano*) | **Hypertenze**(*ano*) | **Ictus**(*ano*)]

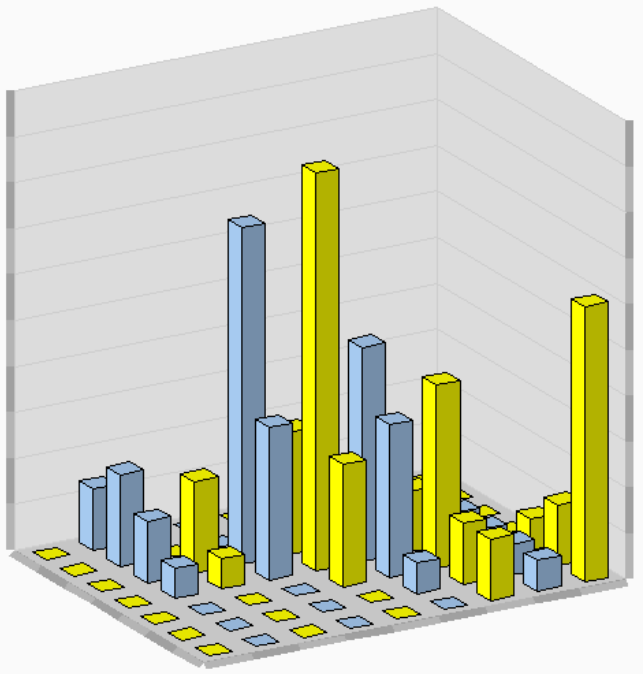
Detailní výstup nejsilnějšího vztahu

Diastolic $\uparrow\uparrow$ Systolic / Rodinný stav(ženatý) $\wedge$ Lihoviny(ne) $\wedge$ Víno(ne) $\wedge$
$\wedge$ (Hyperlipidemie(ano) $\vee$ Hypertenze(ano) $\vee$ Ictus(ano) $\vee$ Infarkt(ano))
SUM = 90, $\tau_b = 0.77$

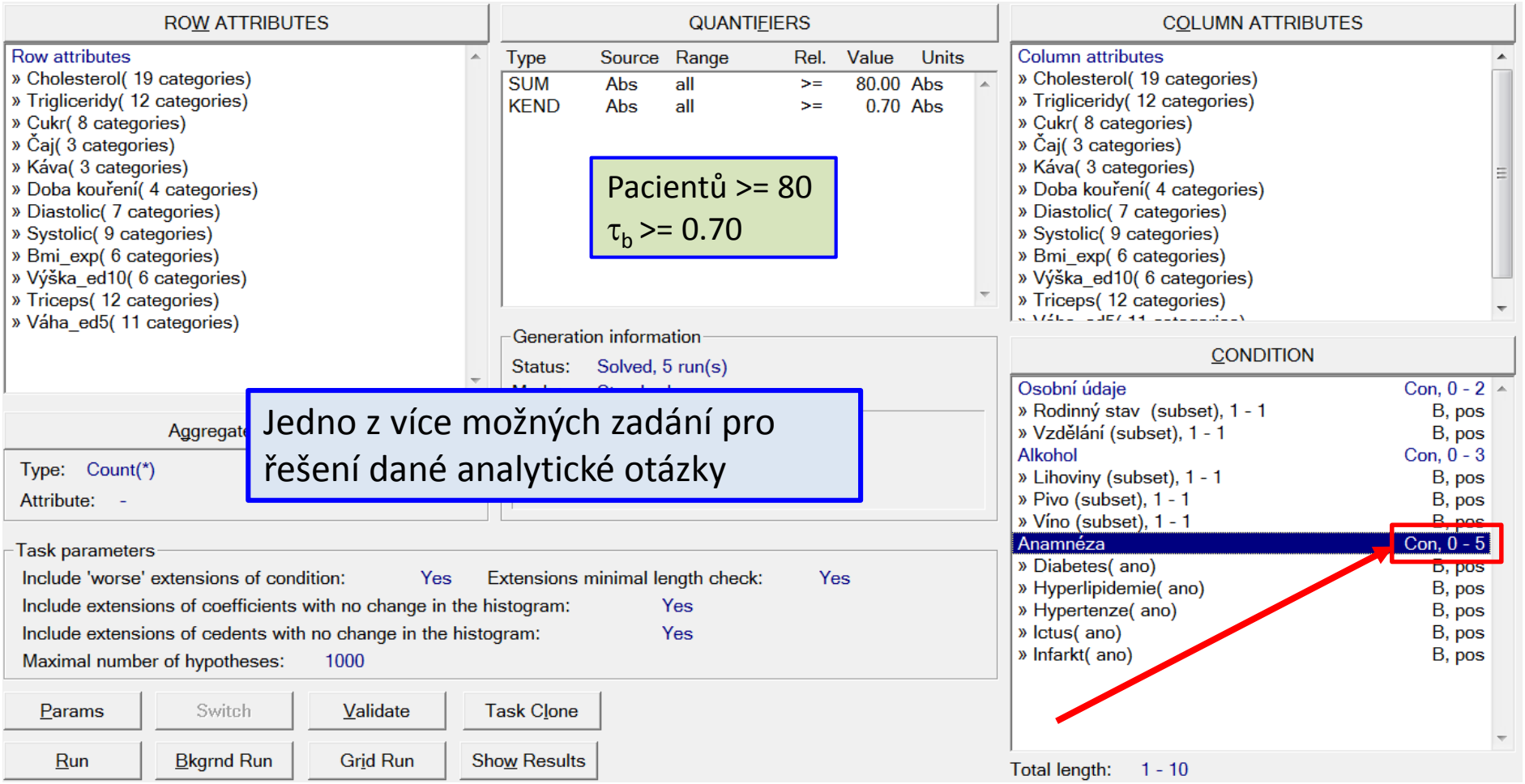
Attributes: Diastolic × Systolic: frequencies
Condition: Rodinný stav (ženatý) & Lihoviny(ne) & Víno(ne) & [Hyperlipidemie(ano) | Hypertenze(ano) | Ictus(ano) | Infarkt(ano)]

TEXT | DATA | FREQUENCIES

	<75;104>	<115;124>	<135;144>	<175;225>
<45;64>	0	2	0	0	0	0	0	0	0
<65;74>	0	3	0	0	0	0	0	0	0
<75;84>	0	2	3	11	4	0	0	0	0
<85;94>	0	1	1	5	13	7	2	1	0
<95;104>	0	0	0	0	4	5	6	1	1
<105;114>	0	0	0	0	0	1	2	1	2
<115;145>	0	0	0	0	0	0	2	1	9



? KL: $\uparrow\uparrow_{0.7}$ STULONG / Osobní, Alkohol, Anamnéza (2)



? KL: $\uparrow\uparrow_{0.7}$ STULONG / Osobní, Alkohol, Anamnéza (2) – výstup

Task run
Start: 1.10.2016 12:07:08 Total time: 0h 0m 4s
Number of verifications: 41580
Number of hypotheses: 70 Mode: Standard

Add group | Del group | Edit group

Actual group of hypotheses: All hypotheses
Hypotheses in group: 70 Shown hypotheses: 70 Highlighted: 0

Delete hypotheses

Nr.	Id	KEND	Hypothesis
1	27	0.755	**Diastolic** × **Systolic** / **Vzdělání**(*vyučen*) & **Lihoviny**(*<=0.1*) & **Víno**(*<=0.5*)
2	28	0.755	**Systolic** × **Diastolic** / **Vzdělání**(*vyučen*) & **Lihoviny**(*<=0.1*) & **Víno**(*<=0.5*)
3	57	0.748	**Diastolic** × **Systolic** / **Rodinný stav** (*ženatý*) & **Vzdělání**(*vyučen*) & **Lihoviny**(*<=0.1*)
4	58	0.748	**Systolic** × **Diastolic** / **Rodinný stav** (*ženatý*) & **Vzdělání**(*vyučen*) & **Lihoviny**(*<=0.1*)
5	3	0.746	**Diastolic** × **Systolic** / **Lihoviny**(*ne*) & **Víno**(*ne*) & **Hypertenze**(*ano*)
6	4	0.746	**Systolic** × **Diastolic** / **Lihoviny**(*ne*) & **Víno**(*ne*) & **Hypertenze**(*ano*)
7	63	0.744	**Diastolic** × **Systolic** / **Rodinný stav** (*ženatý*) & **Vzdělání**(*vyučen*) & **Lihoviny**(*<=0.1*) & **Víno**(*<=0.5*)
8	64	0.744	**Systolic** × **Diastolic** / **Rodinný stav** (*ženatý*) & **Vzdělání**(*vyučen*) & **Lihoviny**(*<=0.1*) & **Víno**(*<=0.5*)
9	17		
10	18		
11	15		
12	16		
13	49		
14	50		
15	47		
16	48	0.729	**Systolic** × **Diastolic** / **Rodinný stav** (*ženatý*) & **Vzdělání**(*vysokoškolské*) & **Pivo**(*1-2*) & **Víno**(*<=0.5*)
17	9	0.727	**Bmi_exp** × **Weight** / **Vzdělání**(*středoškolské*) & **Lihoviny**(*ne*) & **Pivo**(*1-2*)
18	10	0.727	**Weight** × **Bmi_exp** / **Vzdělání**(*středoškolské*) & **Lihoviny**(*ne*) & **Pivo**(*1-2*)
19	53	0.727	**Diastolic** × **Systolic** / **Rodinný stav** (*ženatý*) & **Vzdělání**(*vysokoškolské*) & **Lihoviny**(*<=0.1*) & **Pivo**(*1-2*) & **Víno**(*<=0.5*)

V případě použití konjunkce u dílčího cedentu Anamnéza je výstupem 70 vztahů na rozdíl od použití disjunkce, kdy je výstupem 216 vztahů.

Pouze dva vztahy se týkají dílčího cedentu Anamnéza – atribut Hypertenze(ano).