

Tato prezentace je součástí wiki-prezentace [Metoda GUHA a systém LISp-Miner](#)

Je dostupná z [této adresy](#)

Verze 25. 7. 2019

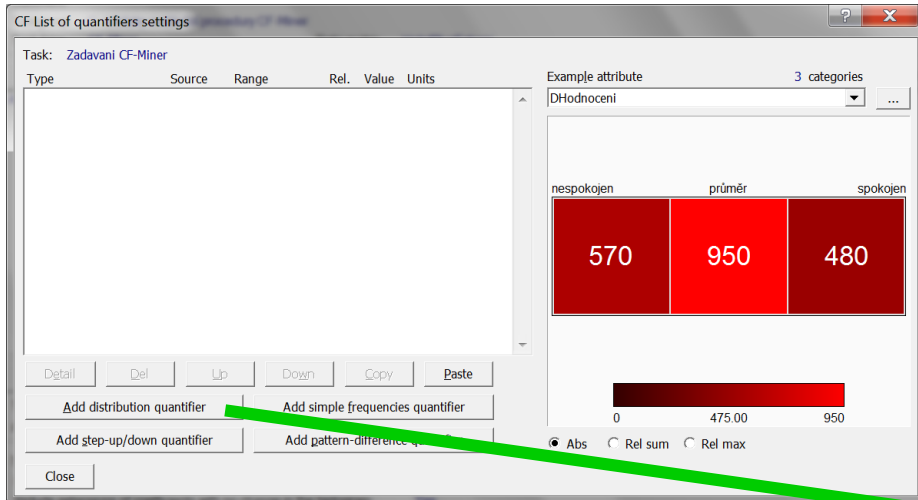
Zadávání distribučních CF-kvantifikátorů pro proceduru CF-Miner

Jan Rauch

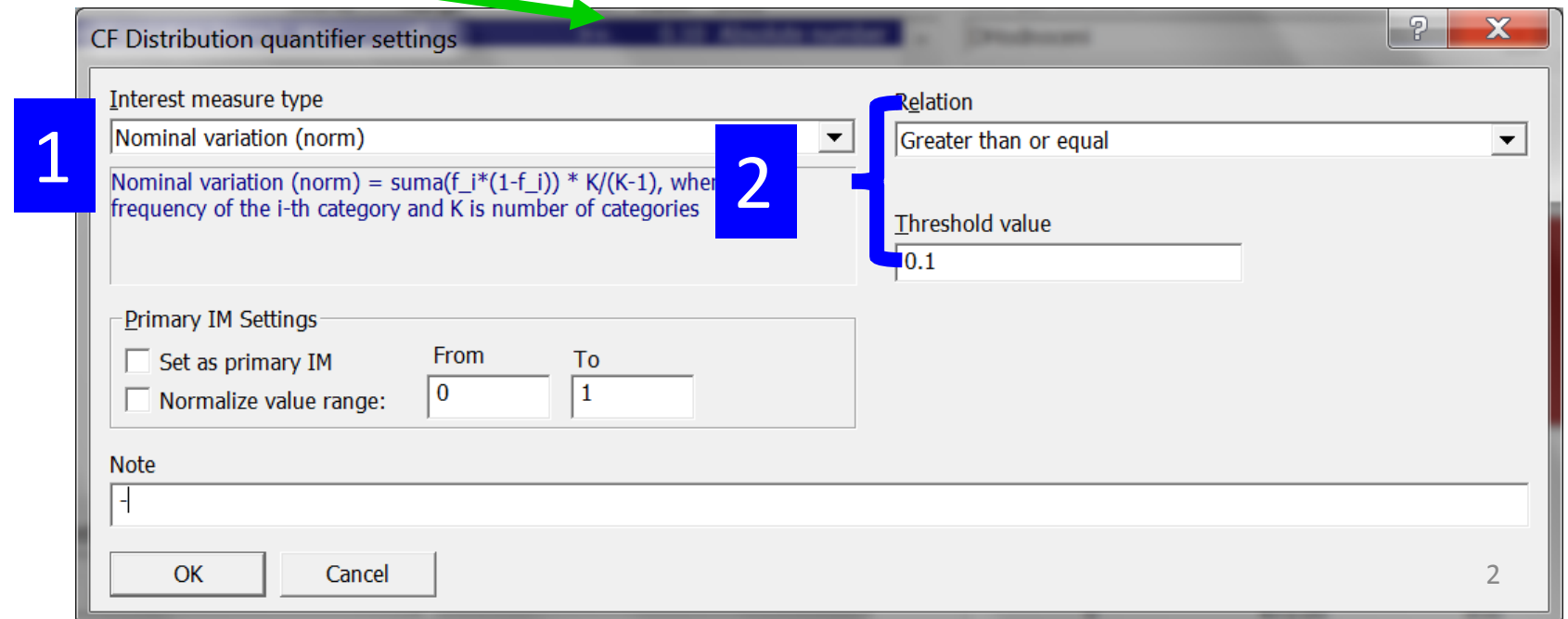
Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

Start tlačítkem Add simple frequencies quantifier



Zadání kvantifikátoru se provádí ve dvou krocích.



Výchozí CF-tabulka

Distribuční CF-kvantifikátor se pro CF-výraz $\approx A/\chi$ verifikuje na matici dat M na základě CF-tabulky $CF(A, \chi, M)$:

četnosti kategorií pro atribut A	a_1	...	a_K	Σ
absolutní četnosti v matici M/χ	n_1	...	n_K	n
relativní četnosti v matici M/χ	$f_1=n_1/n$...	$f_K=n_K/n$	1

Výraz $\approx A$ týkající se rozložení histogramu na celé matici dat chápeme jako $\approx A/True$.

Krok 1 - výběr míry zajímavosti

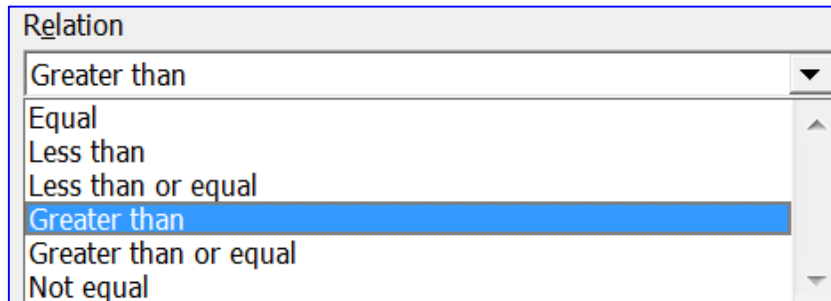
Vybere se jedna z měr zajímavosti nabízených v menu *Interest measure type* a přiřadí se jí hodnota dle tabulky

Interest measure type
Asymetry
Variation ratio
Nominal variation (norm)
Discrete ordinary variation (norm)
Median-category index (absolute)
Median-category index (relative)
Arithmetic average
Geometric average
Variance
Standard deviation
Skewness
Asymetry

Interest measure type	Hodnota IM	Poznámka
Variation ratio	$1 - \max\{f_1, \dots, f_K\}$	
Nominal variation (norm)	$\frac{K}{K-1} \sum_{i=1}^K f_i(1 - f_i)$	
Discrete ordinary variation (norm)	$\frac{2}{K-1} \sum_{i=1}^K F_i(1 - F_i)$	$F_i = \sum_{j=1}^i f_j, i = 1, \dots, K$
Median category index (absolute)	absolutní index mediánu	$\in \langle 1, K \rangle$
Median category index (relative)	relativní index mediánu	$\in (0, 1)$
Arithmetic average	$\sum_{i=1}^K f_i a_i$	pouze pro kardinální atributy
Geometric average	$\prod_{i=1}^K a_i^{f_i}$	pouze pro kardinální atributy
Variance	$\sum_{i=1}^K f_i (a_i - AvgA)^2$	$AvgA = \sum_{i=1}^K f_i a_i$
Standard deviation	$\sqrt{\sum_{i=1}^K f_i (a_i - AvgA)^2}$	dále značíme jako <i>StdDev</i>
Skewness	$\frac{\sum_{i=1}^K f_i (a_i - AvgA)^3}{StdDev^3}$	
Asymetry	$\frac{N_2 - N_1}{n}$	$N_1 = \sum n_i$ pro $a_i < AvgA$ $N_2 = \sum n_i$ pro $a_i > AvgA$

Krok 2 - porovnání vypočtené hodnoty s prahem

Na základě volby v nabídce *Relation* se vybere relace která bude použita pro porovnání vypočtené hodnoty *IM* s hodnotou zadanou v poli *Threshold value*.



Relation

Greater than

Equal

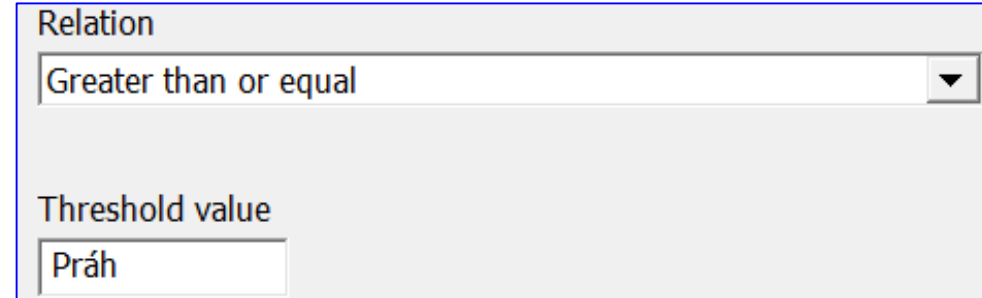
Less than

Less than or equal

Greater than

Greater than or equal

Not equal



Relation

Greater than or equal

Threshold value

Práh

Platnost vybrané relace mezi hodnotou *IM* a hodnotou zadanou v poli *Threshold value* je považována za podmínku definující CF-kvantifikátor zadaný v krocích 1 až 2.