

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

Je dostupná z [této adresy](#)

Verze 30. 7. 2019

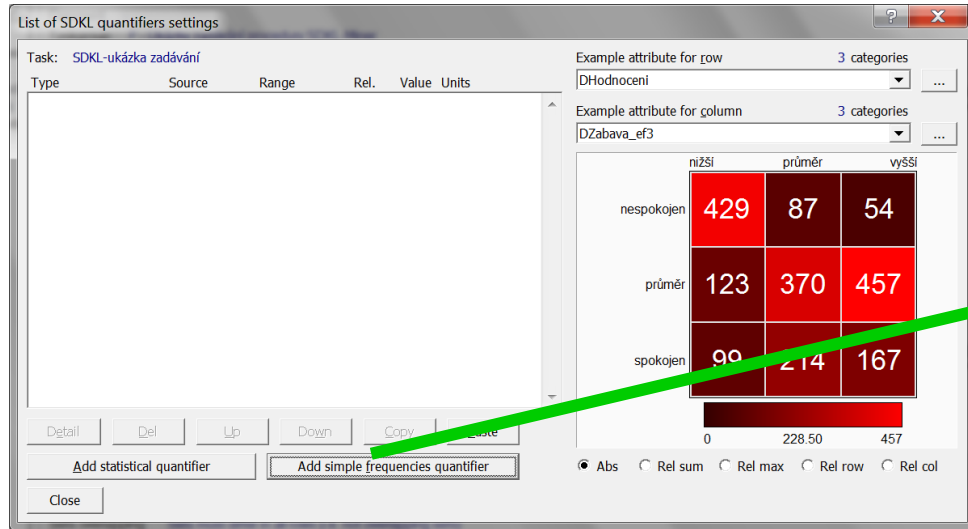
# Zadávání jednoduchých frekvenčních SDKL-kvantifikátorů pro proceduru SDKL-Miner

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

# Start tlačítkem Add simple frequencies quantifier



- Definice se provádí zadáním parametrů 1 - 5.
- Pro všechny parametry se nabízejí defaultní hodnoty.
- Podmínka kvantifikátoru je definována parametrem 5.

SDKL Simple frequencies quantifier settings

Interest measure type: Sum of frequencies (5)

Relation: Greater than or equal

Threshold value: 50

Threshold-value units: Relative [%] to act condition (4)

Source contingency table: Delta of absolute frequencies values

Quantifier test applied to the frequency table computed by subtracting corresponding frequencies

Category Range (1):  
Rows: From 0 To 100  
Columns: From 0 To 100  
Reset to All

Absolute category index  
 Relative range [%] to act number of categories

Primary IM Settings:  
 Set as primary IM  
 Normalize value range: From 0 To 1

Note: -

Buttons: OK, Cancel

# Výchozí KL-tabulky

Jednoduchý frekvenční SDKL-kvantifikátor se aplikuje na dvojici KL-tabulek  $\text{TKL}_\alpha = \text{KL}(\text{R,C,M}/\chi \wedge \alpha)$  a  $\text{TKL}_\beta = \text{KL}(\text{R,C,M}/\chi \wedge \beta)$ .

Tabulky zapisujeme i jako  $\text{TKL}_\alpha = \{n_{\alpha,i,j}\}_{i=1,\dots,K}^{j=1,\dots,L}$  a  $\text{TKL}_\beta = \{n_{\beta,i,j}\}_{i=1,\dots,K}^{j=1,\dots,L}$ . Pokud nehrozí nedorozumění, pak  $\text{TKL}_\alpha = \{n_{\alpha,i,j}\}$  a  $\text{TKL}_\beta = \{n_{\beta,i,j}\}$ .

$\text{M}/\chi \wedge \alpha$	$c_1$	...	$c_u$	...	$c_v$	...	$c_L$	$\Sigma_l$
$r_1$	$n_{\alpha,1,1}$	...	$n_{\alpha,1,u}$	...	$n_{\alpha,1,v}$	...	$n_{\alpha,1,L}$	$n_{\alpha,1,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_p$	$n_{\alpha,p,1}$	...	$n_{\alpha,p,u}$	...	$n_{\alpha,p,v}$	...	$n_{\alpha,p,L}$	$n_{\alpha,p,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_q$	$n_{\alpha,q,1}$	...	$n_{\alpha,q,u}$	...	$n_{\alpha,q,v}$	...	$n_{\alpha,q,L}$	$n_{\alpha,q,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_K$	$n_{\alpha,K,1}$	...	$n_{\alpha,K,u}$	...	$n_{\alpha,K,v}$	...	$n_{\alpha,K,L}$	$n_{\alpha,K,*}$
$\Sigma_k$	$n_{\alpha,*,1}$	...	$n_{\alpha,*,u}$	...	$n_{\alpha,*,v}$	...	$n_{\alpha,*,L}$	$n_\alpha$

$\text{M}/\chi \wedge \beta$	$c_1$	...	$c_u$	...	$c_v$	...	$c_L$	$\Sigma_l$
$r_1$	$n_{\beta,1,1}$	...	$n_{\beta,1,u}$	...	$n_{\beta,1,v}$	...	$n_{\beta,1,L}$	$n_{\beta,1,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_p$	$n_{\beta,p,1}$	...	$n_{\beta,p,u}$	...	$n_{\beta,p,v}$	...	$n_{\beta,p,L}$	$n_{\beta,p,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_q$	$n_{\beta,q,1}$	...	$n_{\beta,q,u}$	...	$n_{\beta,q,v}$	...	$n_{\beta,q,L}$	$n_{\beta,q,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_K$	$n_{\beta,K,1}$	...	$n_{\beta,K,u}$	...	$n_{\beta,K,v}$	...	$n_{\beta,K,L}$	$n_{\beta,K,*}$
$\Sigma_k$	$n_{\beta,*,1}$	...	$n_{\beta,*,u}$	...	$n_{\beta,*,v}$	...	$n_{\beta,*,L}$	$n_\beta$

Pro parametr 4 - *Threshold value unit* se využijí i KL-tabulky  $\text{KL}(\text{R,C,M}/\chi)$  a  $\text{KL}(\text{R,C,M})$ .

$\text{M}/\chi$	$c_1$	...	$c_L$	$\Sigma_l$
$r_1$	$n_{1,1}$	...	$n_{1,L}$	$n_{1,*}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
$r_K$	$n_{K,1}$	...	$n_{K,L}$	$n_{K,*}$
$\Sigma_k$	$n_{*,1}$	...	$n_{*,L}$	$n$

$\text{M}$	$c_1$	...	$c_L$	$\Sigma_l$
$r_1$	$m_{1,1}$	...	$m_{1,L}$	$m_{1,*}$
$\vdots$	$\vdots$		$\vdots$	$\vdots$
$r_K$	$m_{K,1}$	...	$m_{K,L}$	$m_{K,*}$
$\Sigma_k$	$m_{*,1}$	...	$m_{*,L}$	$n_T$

# 1 - Category Range

- Category Range určuje výseky KL-tabulek, na které budou aplikovány ostatní parametry.
- Provádí se způsobem popsaným [zde](#).
- Výsek KL-tabulky  $KL(R,C,M/?)$  je dán čtveřicí  $\langle p,q,u,v \rangle$  celých čísel splňujících  $1 \leq p \leq q \leq K$  a  $1 \leq u \leq v \leq L$
- Výsek  $\langle p,q,r,s \rangle$   $KL(R,C,M/?)$  zahrnuje frekvence pro kategorie  $r_p, \dots, r_q$  a  $c_u, \dots, c_v$

Category Range

From
To

Rows

Columns

Absolute category index

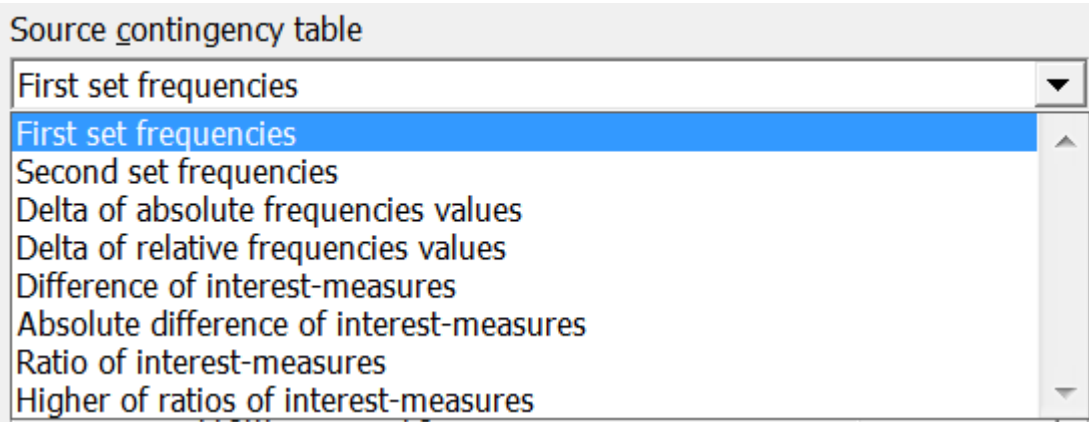
Relative range [%] to act number of categories

$M/\chi \wedge \alpha$	$c_1$	...	$c_u$	...	$c_v$	...	$c_L$	$\Sigma_l$
$r_1$	$n_{\alpha,1,1}$	...	$n_{\alpha,1,u}$	...	$n_{\alpha,1,v}$	...	$n_{\alpha,1,L}$	$n_{\alpha,1,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_p$	$n_{\alpha,p,1}$	...	$n_{\alpha,p,u}$	...	$n_{\alpha,p,v}$	...	$n_{\alpha,p,L}$	$n_{\alpha,p,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_q$	$n_{\alpha,q,1}$	...	$n_{\alpha,q,u}$	...	$n_{\alpha,q,v}$	...	$n_{\alpha,q,L}$	$n_{\alpha,q,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_K$	$n_{\alpha,K,1}$	...	$n_{\alpha,K,u}$	...	$n_{\alpha,K,v}$	...	$n_{\alpha,K,L}$	$n_{\alpha,K,*}$
$\Sigma_k$	$n_{\alpha,*,1}$	...	$n_{\alpha,*,u}$	...	$n_{\alpha,*,v}$	...	$n_{\alpha,*,L}$	$n_\alpha$

$M/\chi \wedge \beta$	$c_1$	...	$c_u$	...	$c_v$	...	$c_L$	$\Sigma_l$
$r_1$	$n_{\beta,1,1}$	...	$n_{\beta,1,u}$	...	$n_{\beta,1,v}$	...	$n_{\beta,1,L}$	$n_{\beta,1,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_p$	$n_{\beta,p,1}$	...	$n_{\beta,p,u}$	...	$n_{\beta,p,v}$	...	$n_{\beta,p,L}$	$n_{\beta,p,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_q$	$n_{\beta,q,1}$	...	$n_{\beta,q,u}$	...	$n_{\beta,q,v}$	...	$n_{\beta,q,L}$	$n_{\beta,q,*}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$r_K$	$n_{\beta,K,1}$	...	$n_{\beta,K,u}$	...	$n_{\beta,K,v}$	...	$n_{\beta,K,L}$	$n_{\beta,K,*}$
$\Sigma_k$	$n_{\beta,*,1}$	...	$n_{\beta,*,u}$	...	$n_{\beta,*,v}$	...	$n_{\beta,*,L}$	$n_\beta$

## 2 - Source contingency table

Vybírá se jeden z možností nabízených v menu *Source contingency table*. Ta určuje, jakým způsobem získána míra zajímavosti  $IM$  pro porovnání v podmínce SDKL-kvantifikátoru.



Pro zbývající čtyři nabízené módy se vybraná míra zajímavosti aplikuje zvlášť na části SDCF tabulky pro atributy  $\alpha$  a  $\beta$ . Poté se výsledky dále zpracují.

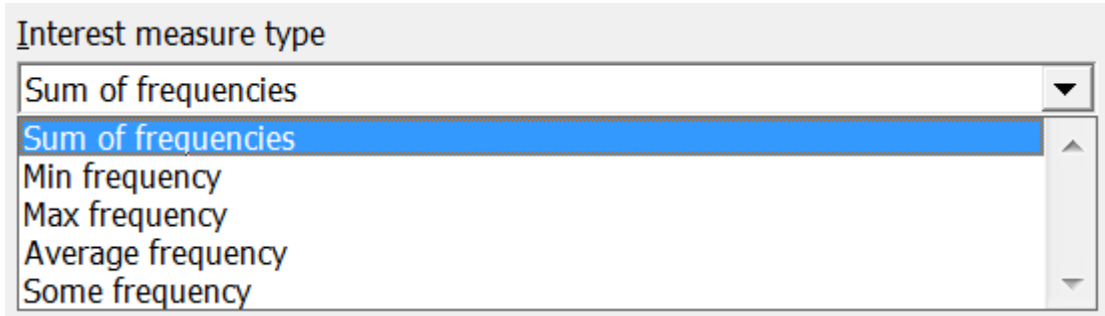
Pro první čtyři volby se nejprve vytvoří výsek KL-tabulky

$\{z_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v}$  takto:

- pro *First set frequencies*  $\{z_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v} = \{n_{\alpha,i,j}\}_{i=p,\dots,q}^{j=u,\dots,v}$
- pro *Second set frequencies*  $\{z_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v} = \{n_{\beta,i,j}\}_{i=p,\dots,q}^{j=u,\dots,v}$
- pro *Delta of absolute frequencies values*  $\{z_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v} = \{|n_{\alpha,i,j} - n_{\beta,i,j}|\}_{i=p,\dots,q}^{j=u,\dots,v}$
- pro *Delta of relative frequencies values*  $\{z_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v} = \left| \frac{n_{\alpha,i,j}}{n_{\alpha}} - \frac{n_{\beta,i,j}}{n_{\beta}} \right|_{i=p,\dots,q}^{j=u,\dots,v}$

## 3 - Interest measure type (A)

Vybírá se jedna z měr zajímavosti nabízených v menu *Interest measure type*.



Pro výpočet hodnoty  $IM$  pro první čtyři volby v Source contingency table

- *First set frequencies*
- *Second set frequencies*
- *Delta of absolute frequencies values,*
- *Delta of relative frequencies values*

se použije výsek KL-tabulky  $\{z_{i,j}\}_{i=p,\dots,q}^{j=u,\dots,v}$ .

Na základě Interest measure type se určí hodnota  $IM$  pro porovnání s prahem dle relace zadané parametrem 5. Platí:

- pro *Sum of frequencies* se spočítá hodnota  $IM = \sum_{i=p}^q \sum_{j=u}^v z_{i,j}$
- pro *Min frequency* se spočítá hodnota  $IM = \min_{i,j} \{z_{i,j} | i = p, \dots, q; j = u, \dots, v\}$
- pro *Max frequency* se spočítá hodnota  $IM = \max_{i,j} \{z_{i,j} | i = p, \dots, q; j = u, \dots, v\}$
- pro *Average frequency* se spočítá  $IM = \frac{\sum_{i=p}^q \sum_{j=u}^v z_{i,j}}{(q-p+1)(v-u+1)}$
- pro *Some frequency* se zjišťuje, zda alespoň jedno  $z_{i,j}$  pro  $i = p, \dots, q; j = u, \dots, v$  splňuje zadanou nerovnost.

### 3 - Interest measure type (B)

Při výpočtu hodnoty  $IM$  pro zbývající čtyři operační módy se nejprve se vypočtou hodnoty  $IM_\alpha$  a  $IM_\beta$  pro vybraný *Interest measure type*:

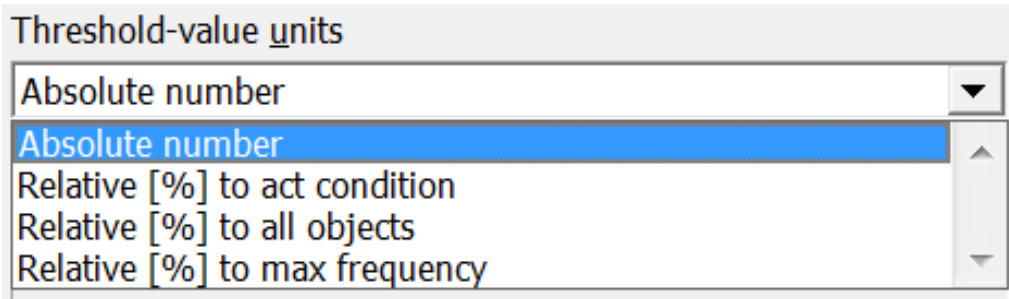
Interest measure type	Hodnota $IM_\alpha$	Hodnota $IM_\beta$
Sum of frequencies	$\sum_{i=p}^q \sum_{j=u}^v n_{\alpha,i,j}$	$\sum_{i=p}^q \sum_{j=u}^v n_{\beta,i,j}$
Min frequency	$\min_{i,j} \{n_{\alpha,i,j}   i = p, \dots, q; j = u, \dots, v\}$	$\min_{i,j} \{n_{\beta,i,j}   i = p, \dots, q; j = u, \dots, v\}$
Max frequency	$\max_{i,j} \{n_{\alpha,i,j}   i = p, \dots, q; j = u, \dots, v\}$	$\max_{i,j} \{n_{\beta,i,j}   i = p, \dots, q; j = u, \dots, v\}$
Average frequency	$\frac{\sum_{i=p}^q \sum_{j=u}^v n_{\alpha,i,j}}{(q-p+1)(v-u+1)}$	$\frac{\sum_{i=p}^q \sum_{j=u}^v n_{\beta,i,j}}{(q-p+1)(v-u+1)}$
Some frequency	Výpočet není k dispozici, není rozumná definice	

Poté se vypočte hodnota  $IM$  v závislosti na vybraném *Operation mode*:

Operation mode	$IM$
Difference of interest-measures	$IM_\alpha - IM_\beta$
Absolute difference of interest-measures	$ IM_\alpha - IM_\beta $
Ratio of interest-measures	$\frac{IM_\alpha}{IM_\beta}$
Higher of ratios of interest-measures	$\max\left\{\frac{IM_\alpha}{IM_\beta}, \frac{IM_\beta}{IM_\alpha}\right\}$

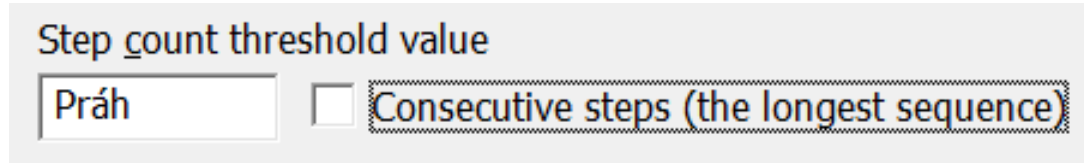
## 4 - Threshold-value units

Vybírá se jedna z voleb nabízených v menu *Threshold-value units*



Threshold-value units

- Absolute number
- Absolute number
- Relative [%] to act condition
- Relative [%] to all objects
- Relative [%] to max frequency



Step count threshold value

Práh  Consecutive steps (the longest sequence)

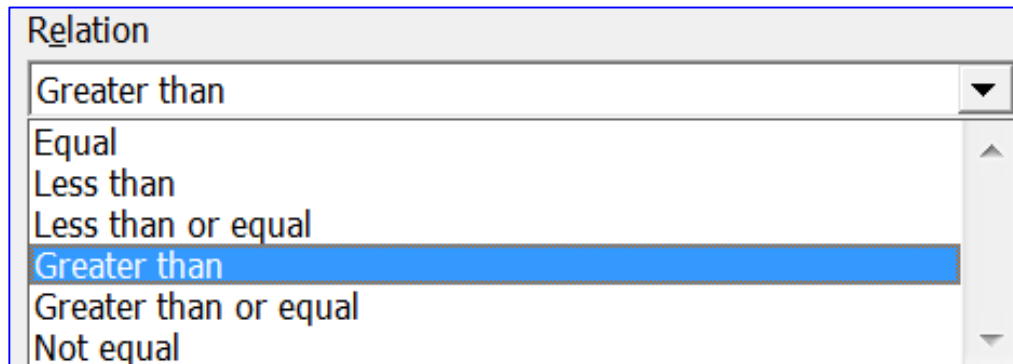
Na základě této volby se určí jednotky, ve kterých je hodnota *Práh* zadaná v poli *Threshold value*. To znamená, že se před porovnáním dle parametru 5 zadaná hodnota *Práh* vynásobí vhodným číslem dle volby v menu. Výslednou hodnotu pro porovnání značíme *PráhAkt*. Platí:

- pro *Absolute number*  $PráhAkt = Práh$
- pro *Relative [%] to act condition*  $PráhAkt = \frac{100}{n} * Práh$
- pro *Relative [%] to all objects*  $PráhAkt = \frac{100}{n_T} * Práh$
- pro *Relative [%] to max frequency*  $PráhAkt = \frac{100}{Max} * Práh$  kde  $Max = \max_{i,j} \{m_{i,j} | i = 1, \dots, K, j = 1, \dots, L\}$

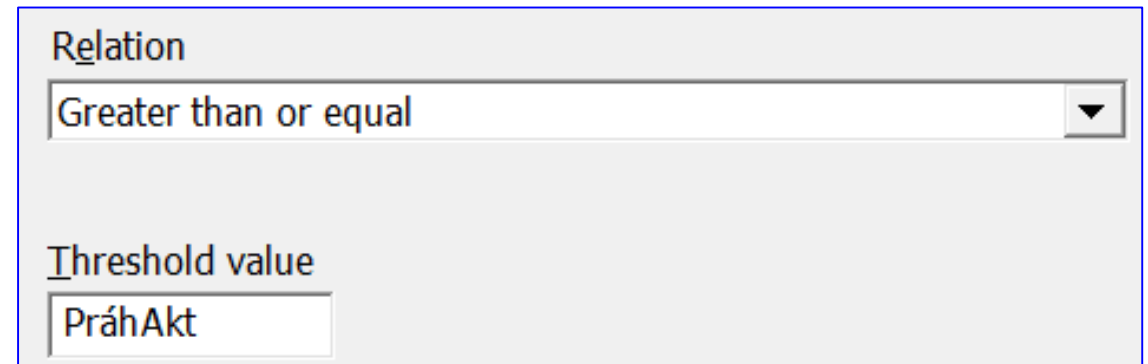


## 5 - Relation x Threshold value

Na základě volby v nabídce *Relation* se vybere relace, která se použije pro porovnání hodnoty *IM* vypočtené dle parametru 3 - *Inte*nterest measure type s hodnotou *PráhAkt* vypočítanou dle parametru 4 - *Threshold-value units* .



A screenshot of a software interface showing a dropdown menu labeled 'Relation'. The menu is open, displaying a list of comparison operators: 'Greater than', 'Equal', 'Less than', 'Less than or equal', 'Greater than', and 'Not equal'. The second 'Greater than' option is highlighted with a blue background. The first 'Greater than' option is currently selected and displayed in the dropdown box above the list.



A screenshot of a software interface showing a dropdown menu labeled 'Relation' with 'Greater than or equal' selected. Below the dropdown is a text input field labeled 'Threshold value' containing the text 'PráhAkt'.

Platnost vybrané relace je považována za podmínku definující SDKL-kvantifikátor zadaný parametry 1 až 5.