

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

Je dostupná z [této adresy](#)

Verse 28. 7. 2019

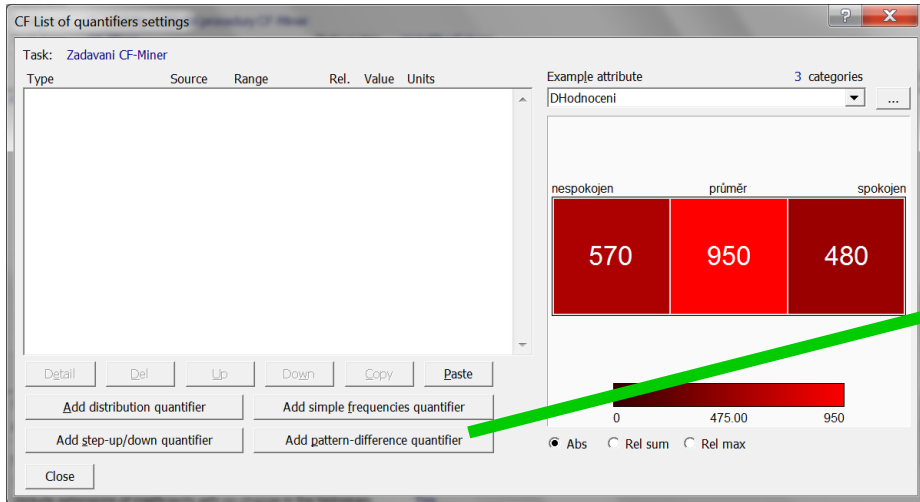
Definice pattern difference CF-kvantifikátorů pro proceduru CF-Miner

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

Start tlačítkem Add pattern difference quantifier



- Definice se provádí zadáním parametrů 1 - 6.
- Pro všechny parametry se nabízejí defaultní hodnoty.
- Podmínka kvantifikátoru je definována parametrem 6.

CF Pattern Difference quantifier settings

Interest measure type: PattDiffMax

Relation: Greater than or equal

Threshold value: 5

Threshold-value units: Absolute number

Source frequencies: Relative [%] to act condition

Category Range: From 0 To 100

Pattern used for comparison: Histogram of the same attribute on the whole matrix

CF: DHodnoceni: frequencies / PNoci(7) & MTeplota(neutrální)

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

Note: -

OK Cancel

Výchozí CF-tabulky

Pattern difference CF-kvantifikátor se pro CF-výraz $\approx A/\chi$ verifikuje na základě [rozšířené CF-tabulky CF\(A, \$\chi\$, M\)](#):

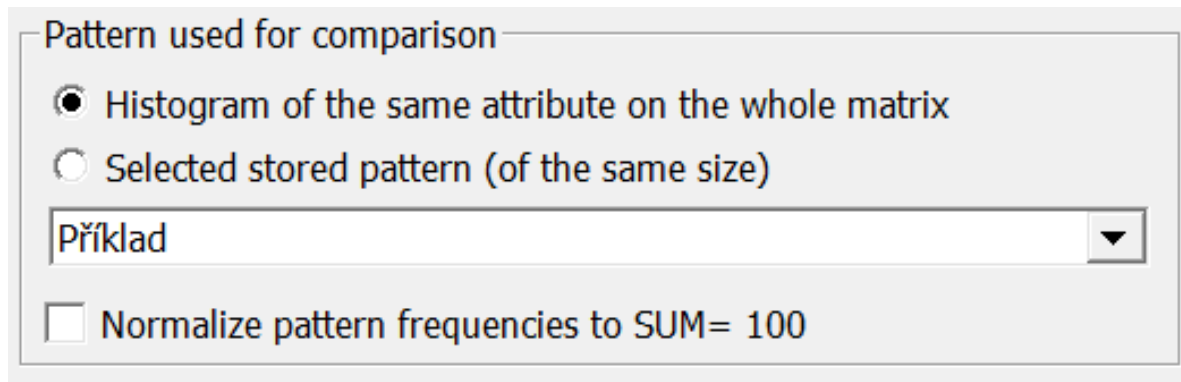
četnosti kategorií pro atribut A	a_1	...	a_p	...	a_q	...	a_K	Σ
absolutní četnosti v matici M/ χ	n_1	...	n_p	...	n_q	...	n_K	n
absolutní četnosti v matici M/ χ normalizované na 100	$w_1=100 n_1/n$...	$w_p=100 n_p/n$...	$w_q=100 n_q/n$...	$w_K=100 n_K/n$	100
relativní četnosti v matici M/ χ	$f_1=n_1/n$...	$f_p=n_p/n$...	$f_q=n_q/n$...	$f_K=n_K/n$	1
absolutní četnosti v matici M	m_1	...	m_p	...	m_q	...	m_K	n_T
relativní četnosti mezi M/ χ a M	$v_1= n_1/m_1$...	$v_p= n_p/m_p$...	$v_q= n_q/m_q$...	$v_K= n_K/m_K$	-

Pokud je jako *Pattern used for comparison* vybrán Stored pattern, tak se využijí i jeho frekvence:

četnosti kategorií pro <i>Stored pattern</i>	A_1	...	a_p	...	a_q	...	a_K	Σ
absolutní četnosti	s_1	...	s_p	...	s_q	...	s_K	s
absolutní četnosti normalizované na 100	$u_1=100 s_1/s$...	$u_p=100 s_p/s$...	$u_q=100 s_q/s$...	$u_K=100 s_K/s$	100
relativní četnosti	$r_1=s_1/s$...	$r_p=s_p/s$...	$r_q=s_q/s$...	$r_K=s_K/s$	1
relativní četnosti mezi M/ χ a <i>Stored pattern</i>	$z_1= s_1/m_1$...	$z_p= s_p/m_p$...	$z_q= s_q/m_q$...	$z_K= s_K/m_K$	-

1 - Pattern used for comparison, výběr histogramů

Parametr *Pattern used for comparison* určuje se kterým histogramem bude vygenerovaný histogram atributu A na podmatici M/χ porovnáván.



The screenshot shows a dialog box titled "Pattern used for comparison". It contains two radio buttons: "Histogram of the same attribute on the whole matrix" (which is selected) and "Selected stored pattern (of the same size)". Below the radio buttons is a dropdown menu with the text "Příklad" and a downward arrow. At the bottom of the dialog is a checkbox labeled "Normalize pattern frequencies to SUM= 100", which is currently unchecked.

Platí:

- *Histogram of the same attribute on the whole matrix* znamená, že vygenerovaný histogram bude porovnáván s histogramem atributu A na celé matici M
- *Selected stored pattern (of the same size)* znamená, že vygenerovaný histogram bude porovnáván s vybraným, dříve uloženým histogramem atributu který má stejný počet kategorií jako atribut A.
- Volba *Normalize pattern frequencies to SUM = 100* znamená, že absolutní frekvence porovnávaných histogramů budou přepočítány na matici dat o 100 řádcích.

1 - Pattern used for comparison, porovnávané frekvence

Oba histogramy jsou porovnávány na základě frekvencí určených parametrem *Source frequencies*. Z těchto frekvencí se vypočítá míra podobnosti histogramů na základě parametru *Interest measure type*.

Frekvence charakterizující histogramy použité pro porovnání chápeme jako vektor a značíme takto:

- Frekvence histogramu atributu A na podmatici $M/\chi : \langle x_1, \dots, x_p, \dots, x_q, \dots, x_K \rangle$
- Frekvence histogramu vybraného parametrem *Pattern used for comparison* : $\langle y_1, \dots, y_p, \dots, y_q, \dots, y_K \rangle$

2 - Category Range

- Category Range určuje výseky vektorů frekvencí $\langle x_1, \dots, x_p, \dots, x_q, \dots, x_K \rangle$ a $\langle y_1, \dots, y_p, \dots, y_q, \dots, y_K \rangle$, na které budou aplikovány ostatní parametry.
- Provádí se způsobem popsaným [zde](#).
- Výseky vektorů jsou dány dvojicí $\langle p, q \rangle$ celých čísel splňujících $1 \leq p \leq q \leq K$.
- Výseky $\langle p, q \rangle$ obou vektorů zahrnují frekvence pro kategorie a_p, \dots, a_q .

Category Range

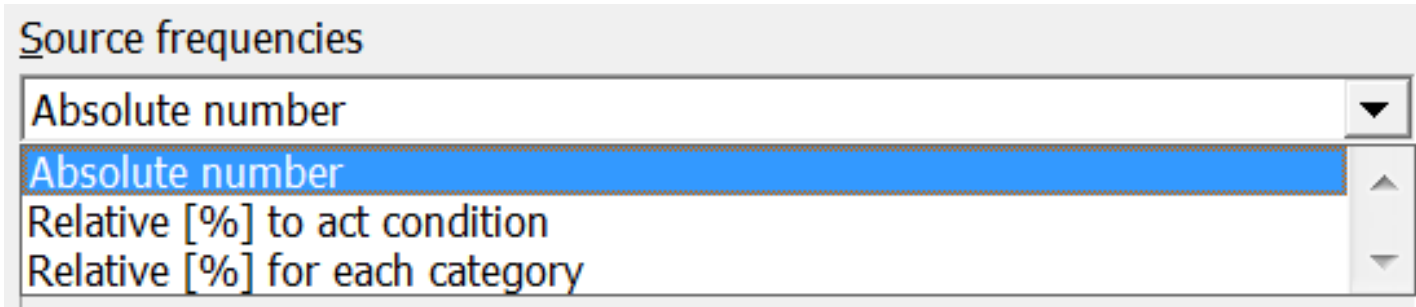
From To

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

$$\langle x_1, \dots, x_p, \dots, x_q, \dots, x_K \rangle$$
$$\langle y_1, \dots, y_p, \dots, y_q, \dots, y_K \rangle$$

3 - Source frequencies

Vybírá se jedna z voleb nabízených v menu *Source frequencies*.

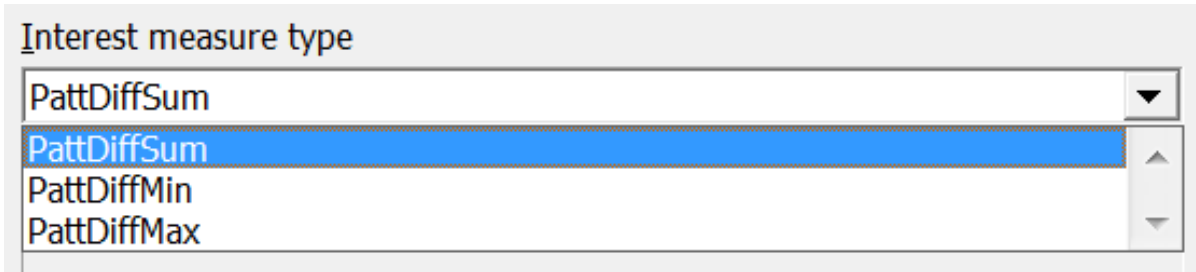


Na základě této volby se určí vektory $\langle x_p, \dots, x_q \rangle$ a $\langle y_p, \dots, y_q \rangle$ frekvencí použité při verifikaci. Platí:

- pro *Absolute number*
 - pokud není zaškrtnuta volba *Normalize pattern frequencies to SUM = 100*:
 $\langle x_p, \dots, x_q \rangle = \langle n_p, \dots, n_q \rangle$ a $\langle y_p, \dots, y_q \rangle = \langle s_p, \dots, s_q \rangle$, jedná se tedy o [histogram absolutních frekvencí](#)
 - pokud je zaškrtnuta volba *Normalize pattern frequencies to SUM = 100*:
 $\langle x_p, \dots, x_q \rangle = \langle w_p, \dots, w_q \rangle$ a $\langle y_p, \dots, y_q \rangle = \langle u_p, \dots, u_q \rangle$, jedná se tedy o modifikaci histogramu absolutních frekvencí
- pro *Relative [%] to act condition* $\langle x_p, \dots, x_q \rangle = \langle f_p, \dots, f_q \rangle$ a $\langle y_p, \dots, y_q \rangle = \langle r_p, \dots, r_q \rangle$, jedná se tedy o [histogramy podílu kategorií na podmínce](#)
- pro *Relative [%] for each category* $\langle x_p, \dots, x_q \rangle = \langle f_p, \dots, f_q \rangle$ a $\langle y_p, \dots, y_q \rangle = \langle z_p, \dots, z_q \rangle$, jedná se tedy o [histogramy podílu podmínky na kategoriích](#)

4 - Interest measure type

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

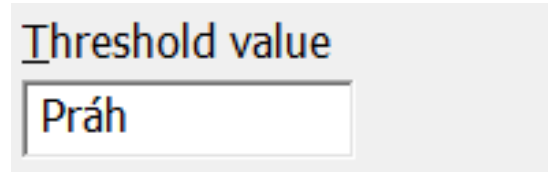
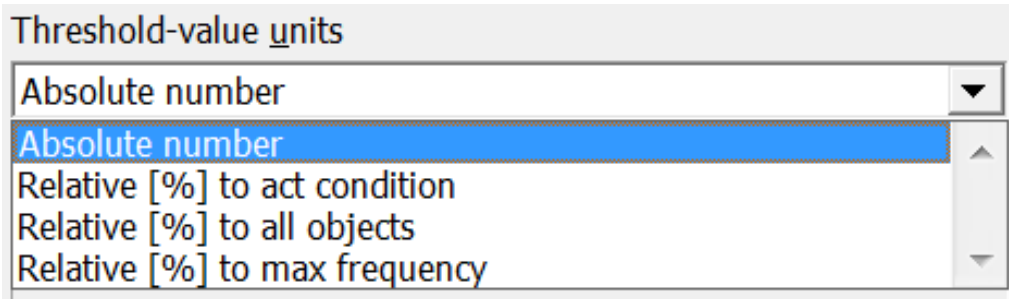


Na základě této volby se určí hodnota IM pro porovnání s prahem dle relace zadané parametrem 6. Platí:

- pro *PattDiffSum* se spočítá hodnota $IM = \sum_{i=p}^q |x_i - y_i|$
- pro *PattDiffMin* se spočítá hodnota $IM = \min_i \{|x_i - y_i| \mid i = p, \dots, q\}$
- pro *PattDiffMax* se spočítá hodnota $IM = \max_i \{|x_i - y_i| \mid i = p, \dots, q\}$

5 - Threshold-value units

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

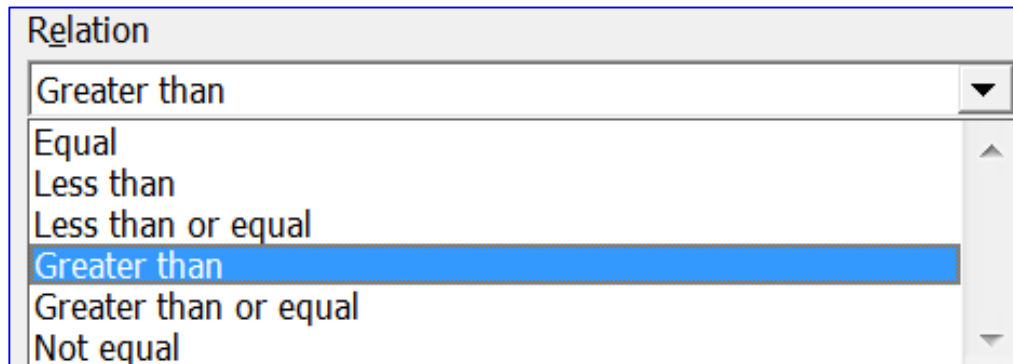


Na základě této volby se určí jednotky, ve kterých je hodnota *Práh* zadaná v poli *Threshold value*. To znamená, že se před porovnáním dle parametru 5 zadaná hodnota *Práh* vynásobí vhodným číslem dle volby v menu. Výslednou hodnotu pro porovnání značíme *PráhAkt*. Platí:

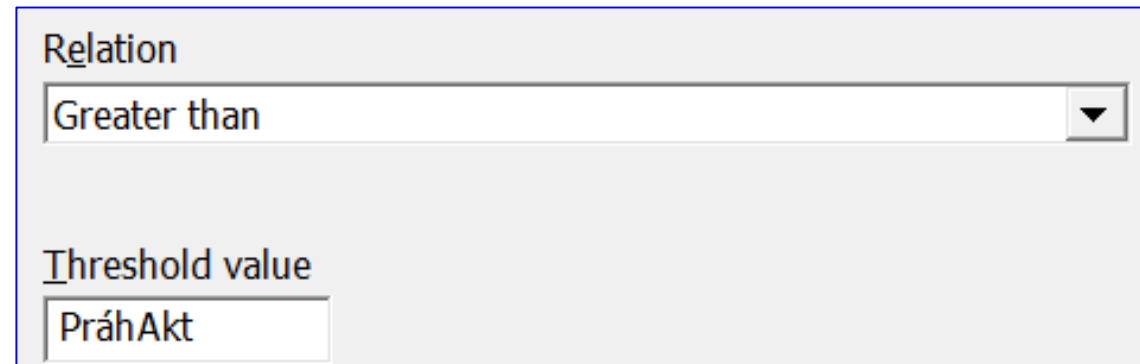
- pro *Absolute number* $PráhAkt = Práh$
- pro *Relative [%] to act condition* $PráhAkt = \frac{100}{n} * Práh$
- pro *Relative [%] to all objects* $PráhAkt = \frac{100}{n_T} * Práh$
- pro *Relative [%] to all objects* $PráhAkt = \frac{100}{Max} * Práh$ kde $Max = \max_i \{n_i | i = 1, \dots, K\}$

6 - Relation x Threshold value

Na základě volby v nabídce *Relation* se vybere relace, která se použije pro porovnání hodnoty *IM* vypočtené dle parametru 4 - *Inte*nterest measure type s hodnotou *PráhAkt* vypočtenou dle parametru 5 - *Threshold-value units*.



A screenshot of a software interface showing a dropdown menu for the 'Relation' parameter. The menu is open, displaying several options: 'Greater than', 'Equal', 'Less than', 'Less than or equal', 'Greater than', and 'Not equal'. The 'Greater than' option is highlighted in blue, indicating it is the selected value.

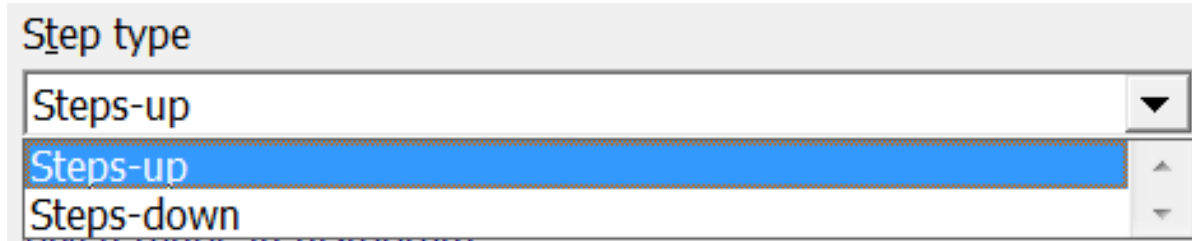


A screenshot of a software interface showing the 'Relation' parameter set to 'Greater than' and the 'Threshold value' parameter set to 'PráhAkt'. The 'Relation' dropdown is closed, and the 'Threshold value' text box contains the text 'PráhAkt'.

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

3 - Step type

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

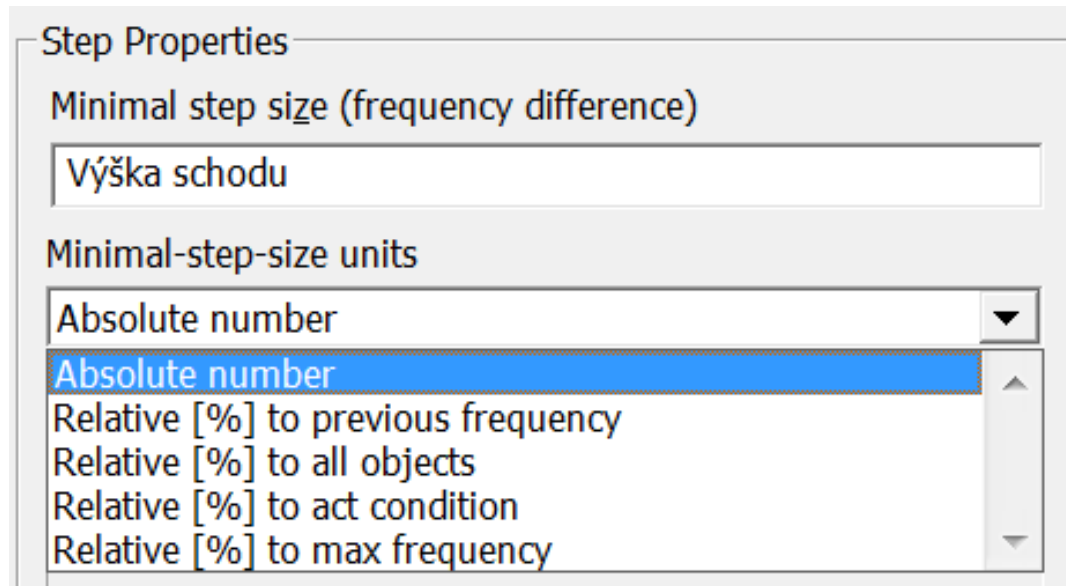


Podrobnosti o nabízených typech schodů jsou zde:

- [Steps-up](#)
- [Step-down](#) .

4 - Step properties

Zadáva se minimální výška schodu v poli *Minimal step size (frequency difference)*
a vybírá se jedna z voleb nabízených v menu *Minimal-step-size units*



Step Properties

Minimal step size (frequency difference)

Výška schodu

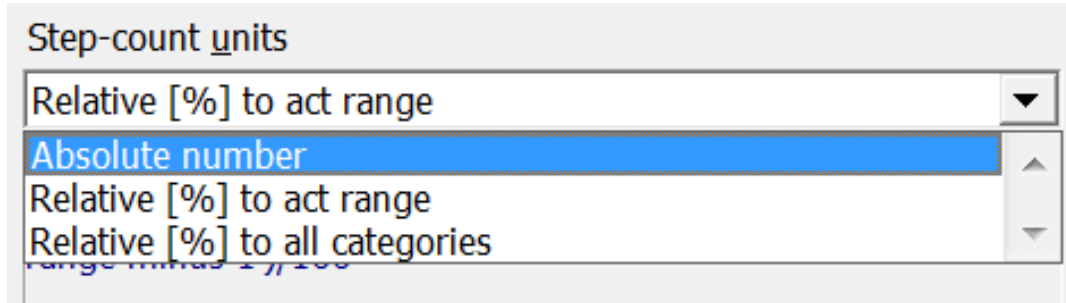
Minimal-step-size units

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

Podrobnosti jsou [zde](#).

5 - Step-count units

Vybírá se jedna z voleb nabízených v menu *Step-count units*



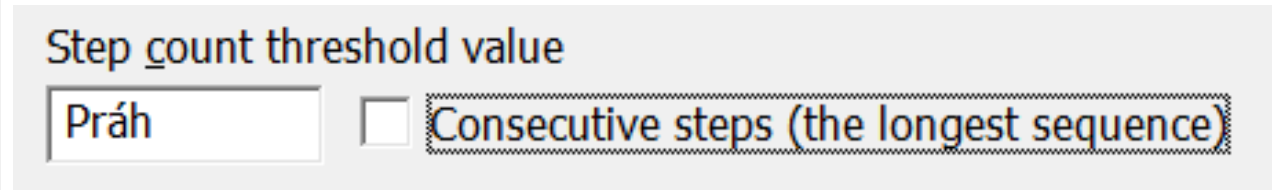
Step-count units

Relative [%] to act range

Absolute number

Relative [%] to act range

Relative [%] to all categories



Step count threshold value

Práh Consecutive steps (the longest sequence)

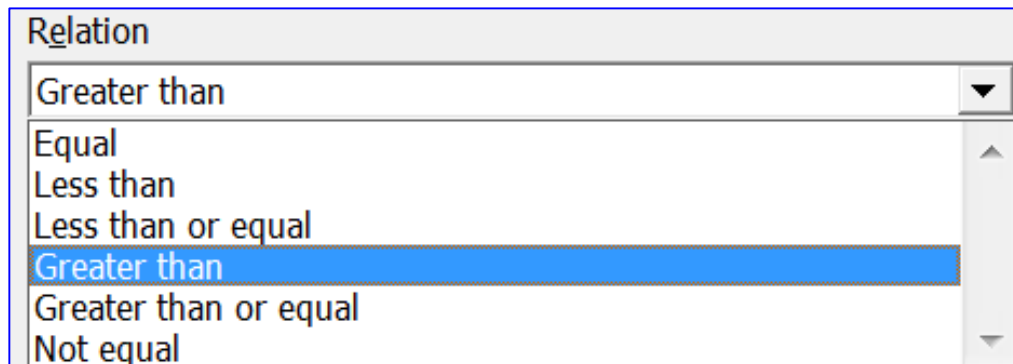
Na základě této volby se určí jednotky, ve kterých je hodnota *Práh* zadaná v poli *Step-count threshold value*. To znamená, že se před porovnáváním dle parametru 5 zadaná hodnota *Práh* vynásobí vhodným číslem dle volby v menu. Výslednou hodnotu pro porovnání značíme *PráhAkt*. Platí:

- pro *Absolute number* $PráhAkt = Práh$
- pro *Relative [%] to act range* $PráhAkt = \frac{100}{p-q} * Práh$
- pro *Relative [%] to all categories* $PráhAkt = \frac{100}{K-1} * Práh$

6 - Relation x Threshold value

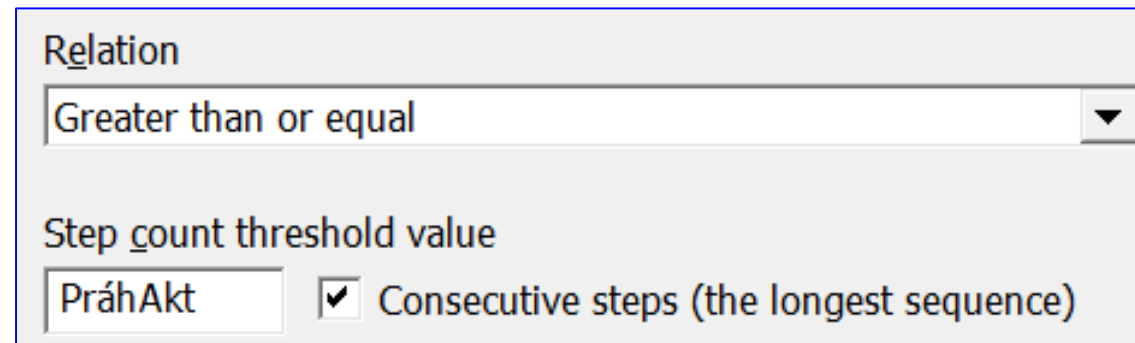
Na základě volby v nabídce *Relation* se vybere relace, která se použije pro porovnání počtu schodů zadaného typu s hodnotou *PráhAkt* vypočtenou dle parametru 5 - [Step-count units](#).

Zaškrtnutím volby *Consecutive steps (the longest sequence)* je možno požadovat že schody vybraného typu následují bezprostředně za sebou.



Relation

- Greater than
- Equal
- Less than
- Less than or equal
- Greater than
- Greater than or equal
- Not equal



Relation

Greater than or equal

Step count threshold value

PráhAkt Consecutive steps (the longest sequence)

Platnost vybrané relace je považována za podmínku definující CF-kvantifikátor zadaný parametry 1 až 6.