

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

Je dostupná z [této adresy](#)

Verse 28. 7. 2019

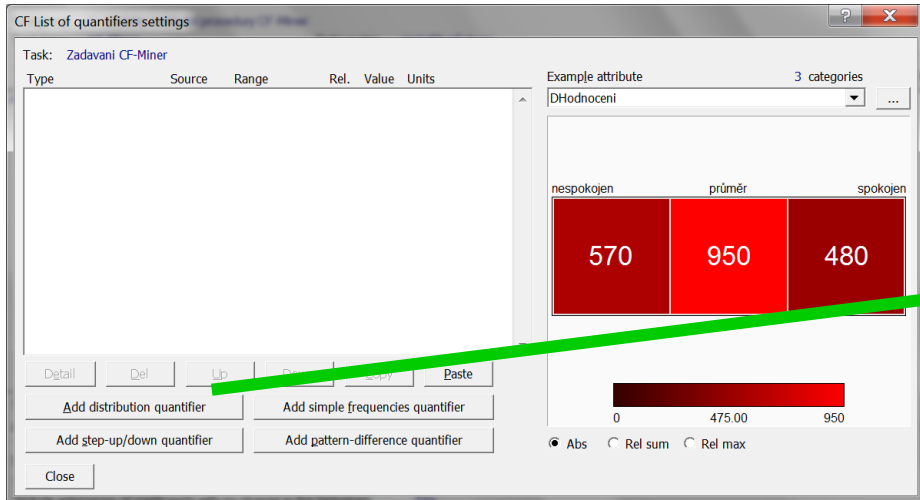
# Definice step-up/down CF-kvantifikátorů pro proceduru CF-Miner

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

# Start tlačítkem Add step-up/down quantifier



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

**1** Category Range

From: 0 To: 100

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

**2** Source frequencies

Absolute number

Absolute frequencies. Frequencies in histogram are left as is.

**3** Step type

Steps-up

Number of steps-up in frequency of adjecant given range in histogram

**4** Step Properties

Minimal step size (frequency difference): 1

Minimal-step-size units: Absolute number

Change in frequency as an absolute number. Minimal step size is left as given.

**5** Step-count units

Relative [%] to act range

Relative to number of categories in the currently selected range-1. Count value is multiplied by ('number of categories in current range minus 1')/100

**6** Relation

Greater than or equal

Step count threshold value: 80

Consecutive steps (the longest sequence)

# Výchozí CF-tabulka

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

četnosti kategorií pro atribut A	$a_1$	...	$a_p$	...	$a_q$	...	$a_K$	$\Sigma$
absolutní četnosti v matici M/ $\chi$	$n_1$	...	$n_p$	...	$n_q$	...	$n_K$	$n$
relativní četnosti v matici M/ $\chi$	$f_1=n_1/n$	...	$f_p=n_p/n$	...	$f_q=n_q/n$	...	$f_K=n_K/n$	1
absolutní četnosti v matici M	$m_1$	...	$m_p$	...	$m_q$	...	$m_K$	$n_T$
relativní četnosti mezi M/ $\chi$ a M	$v_1= n_1/m_1$	...	$v_p= n_p/m_p$	...	$v_q= n_q/m_q$	...	$v_K= n_K/m_K$	-

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

# 1 - Category Range

- Category Range určuje výsek CF-tabulky, na který budou aplikovány ostatní parametry.
- Provádí se způsobem popsáním [zde](#).
- Výsek CF-tabulky týkající se  $K$  kategorií je dvojice  $\langle p, q \rangle$  celých čísel splňujících  $1 \leq p \leq q \leq K$ .
- Výsek  $\langle p, q \rangle$  CF-tabulky zahrnuje frekvence pro kategorie  $a_p, \dots, a_q$ .

Category Range

From  To  Reset to All

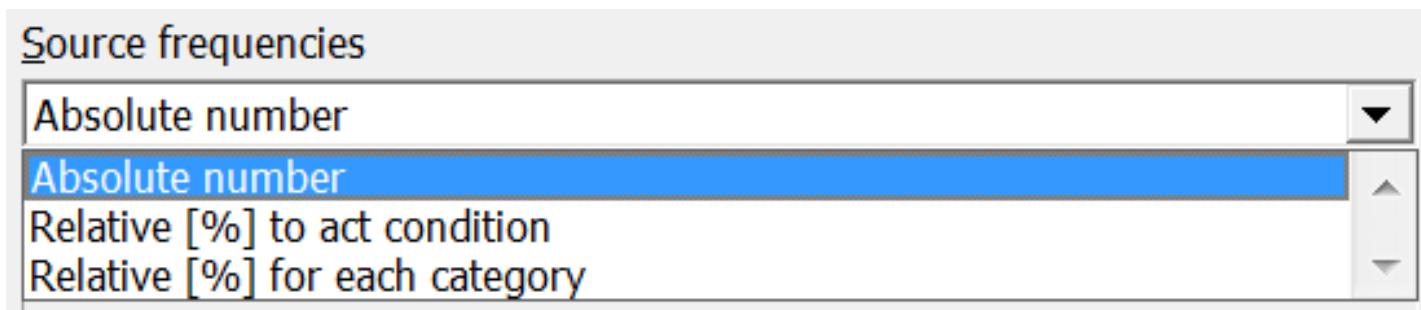
Absolute category index

Relative range [%] to act number of categories

četnosti kategorií pro atribut A	$a_1$	...	$a_p$	...	$a_q$	...	$a_K$	$\Sigma$
absolutní četnosti v matici $M/\chi$	$n_1$	...	$n_p$	...	$n_q$	...	$n_K$	$n$
relativní četnosti v matici $M/\chi$	$f_1 = n_1/n$	...	$f_p = n_p/n$	...	$f_q = n_q/n$	...	$f_K = n_K/n$	$1$
absolutní četnosti v matici M	$m_1$	...	$m_p$	...	$m_q$	...	$m_K$	$n_T$
relativní četnosti mezi $M/\chi$ a M	$v_1 = n_1/m_1$	...	$v_p = n_p/m_p$	...	$v_q = n_q/m_q$	...	$v_K = n_K/m_K$	-

## 2 - Source frequencies

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



Na základě této volby se určí vektor  $\langle w_p, \dots, w_q \rangle$  frekvencí použitý při verifikaci. Platí:

- pro *Absolute number*  $\langle w_p, \dots, w_q \rangle = \langle n_p, \dots, n_q \rangle$ , jedná se tedy o [histogram absolutních frekvencí](#)
- pro *Relative [%] to act condition*  $\langle w_p, \dots, w_q \rangle = \langle f_p, \dots, f_q \rangle$ , jedná se tedy o [histogram podílu kategorií na podmínce](#)
- pro *Relative [%] for each category*  $\langle w_p, \dots, w_q \rangle = \langle v_p, \dots, v_q \rangle$ , jedná se tedy o [histogram podílu podmínky na kategoriích](#)

## 3 - Step type

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

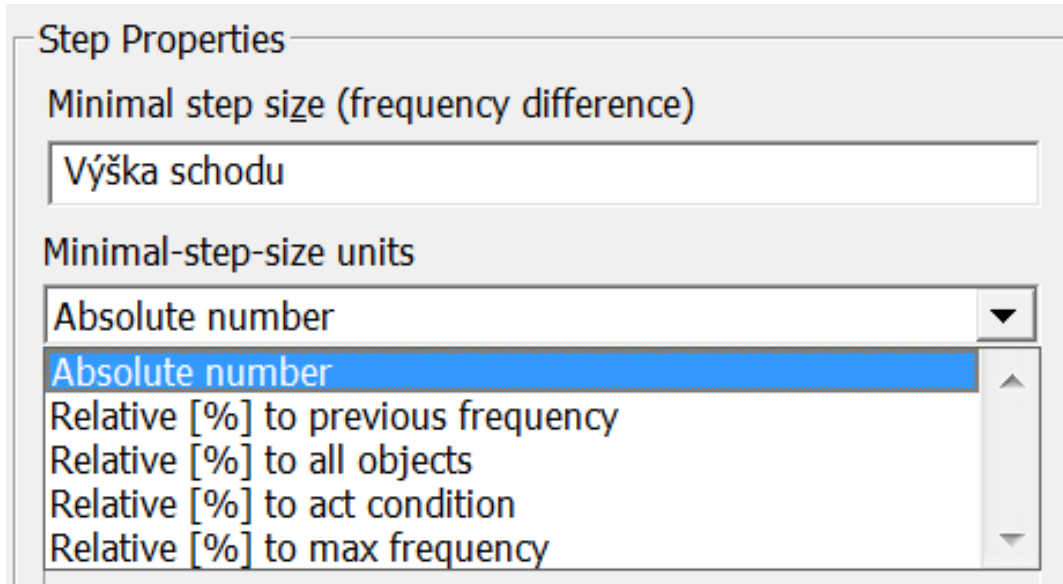


Podrobnosti o nabízených typech schodů jsou zde:

- [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

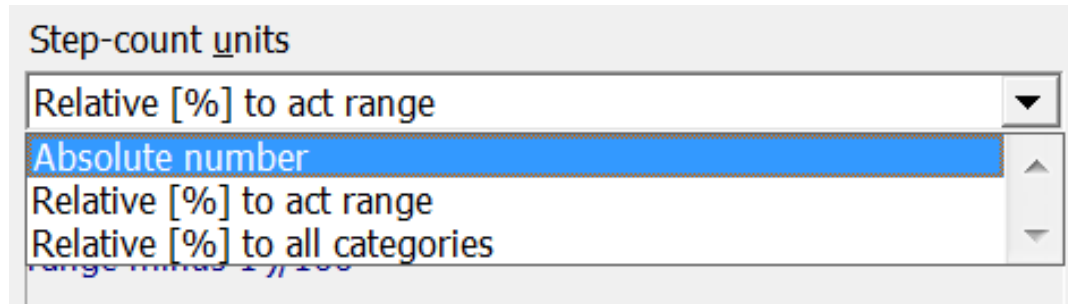
Minimal-step-size units

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

Podrobnosti jsou [zde](#).

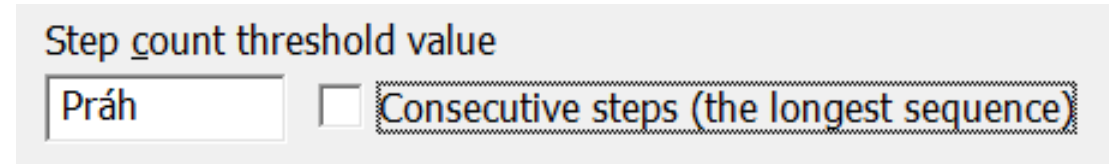
## 5 - Step-count units

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:

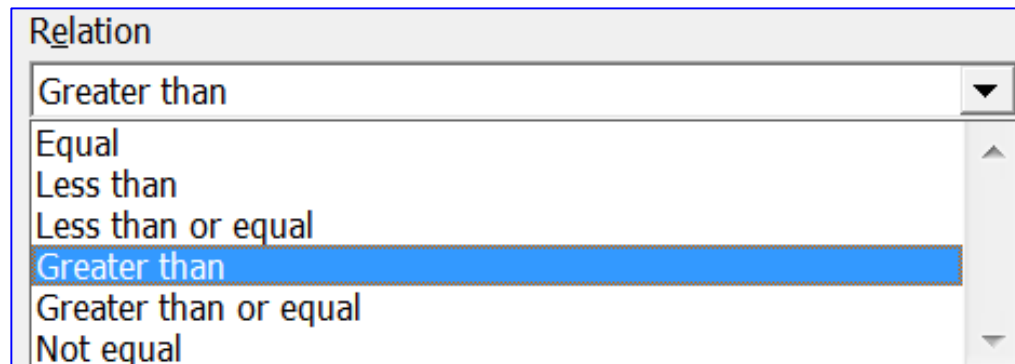
- pro *Absolute number*  $PráhAkt = Práh$
- pro *Relative [%] to act range*  $PráhAkt = \frac{100}{p-q} * Práh$
- pro *Relative [%] to all categories*  $PráhAkt = \frac{100}{K-1} * Práh$



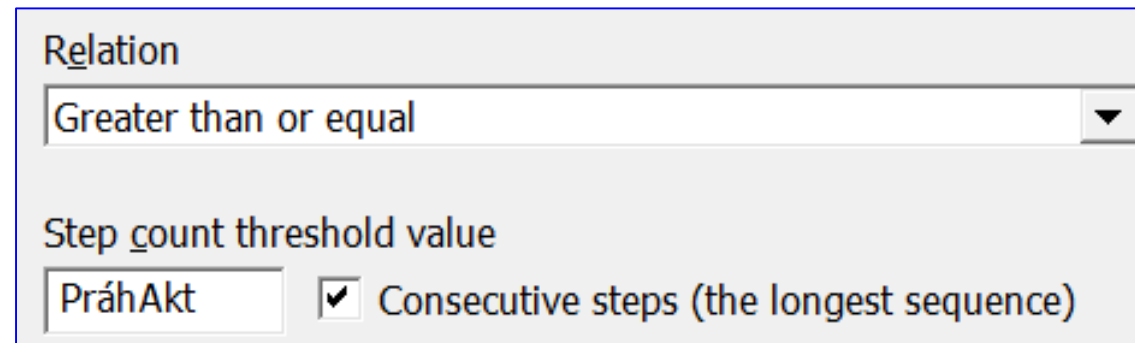
## 6 - Relation x Threshold value

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, aby schody vybraného typu následovaly bezprostředně za sebou.



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



A screenshot of a software interface showing a dropdown menu titled "Relation" with "Greater than or equal" selected. Below the menu is a text input field containing "PráhAkt" and a checked checkbox labeled "Consecutive steps (the longest sequence)".

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