

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

Je dostupná z [této adresy](#)

Verse 22. 7. 2019

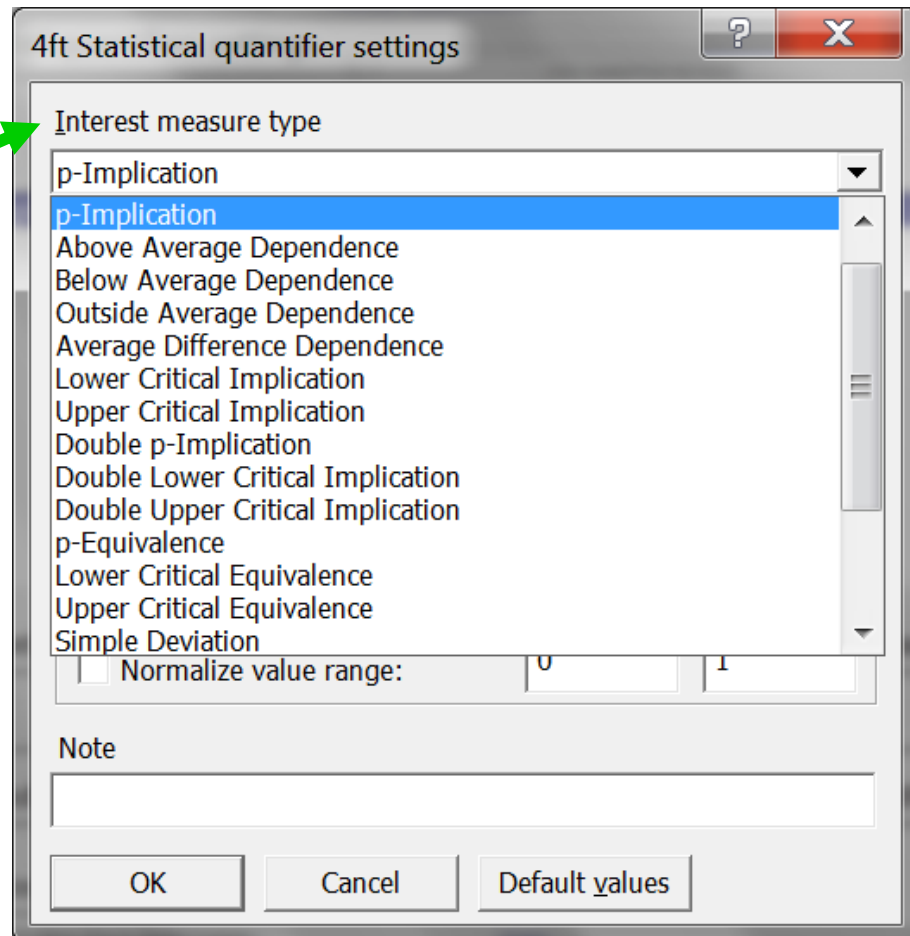
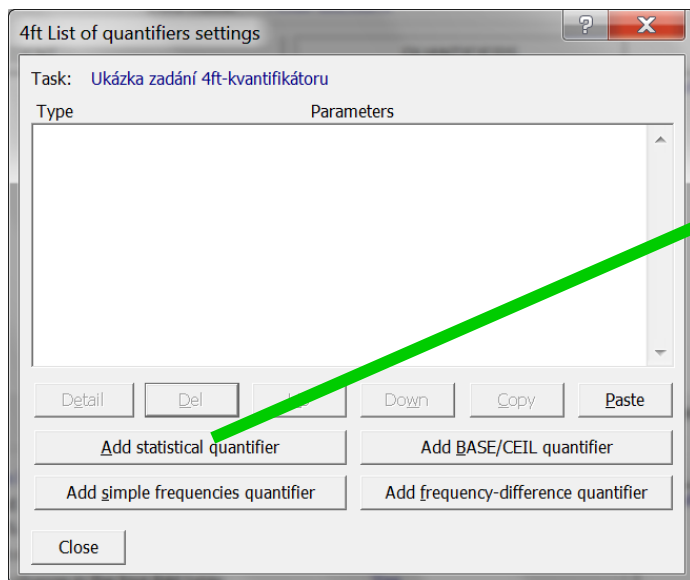
Zadávání statistických 4ft-kvantifikátorů pro proceduru 4ft-Miner

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

Start tlačítkem Add statistical quantifier



V menu **Interest measure type** se nabízí devatenáct 4ft-kvantifikátorů, z toho

- deset 4ft-kvantifikátorů založených na mírách asociace
- devět 4ft-kvantifikátorů založených na testech hypotéz.

Definice statistických 4ft-kvantifikátorů - zásady

- Pro asociační pravidla $\varphi \approx \psi$ se používají frekvence a, b, c, d z tabulky $4ft(\varphi, \psi, M)$
- Pro podmíněná asociačních pravidel $\varphi \approx \psi / \chi$ se používají frekvence a, b, c, d z tabulky $4ft(\varphi, \psi, M / \chi)$

M	ψ	$\neg\psi$	
φ	a	b	r
$\neg\varphi$	c	d	s
	k	l	n

$4ft(\varphi, \psi, M)$

M/ χ	ψ	$\neg\psi$	
φ	a	b	r
$\neg\varphi$	c	d	s
	k	l	n

$4ft(\varphi, \psi, M / \chi)$

Krom těchto frekvencí se používají i součty $r = a + b$, $k = a + c$ a $n = a + b + c + d$

Dále se používají:

- parametr ρ
- parametr α (zadáva se Alpha)
- parametr δ (zadáva se Delta)

4ft-kvantifikátory založené na mírách asociace

- Definice je dána podmínkou na frekvence a, b, c, d a na jeden parametr.
- Tučně je uveden název 4ft-kvantifikátoru použitý v proceduře 4ft-Miner.
- Pro některé 4ft-kvantifikátory jsou uvedeny i další názvy.

Název	Podmínka
Support podpora	$\frac{a}{a+b+c+d} \geq p \wedge a + b + c + d > 0$
<i>p</i>-Implication konfidence spolehlivost	$\frac{a}{a+b} \geq p \wedge a + b > 0$
Above Average Dependence AA-kvantifikátor	$\frac{a}{a+b} \geq (1 + p) \frac{a+c}{a+b+c+d} \wedge a + b > 0$
Below Average Dependence BA-kvantifikátor	$\frac{a}{a+b} \leq (1 - p) \frac{a+c}{a+b+c+d} \wedge a + b > 0$
Outside Average Dependence	$\frac{a}{a+b} \geq (1 + p) \frac{a+c}{a+b+c+d} \vee$ $\frac{a}{a+b} \leq (1 - p) \frac{a+c}{a+b+c+d} \wedge a + b > 0$
Average Difference Dependence	$\frac{\frac{a}{a+b}}{\frac{a+c}{a+b+c+d}} - 1 \geq p \wedge a + b > 0$
Double <i>p</i>-implication Jaccardův kvantifikátor	$\frac{a}{a+b+c} \geq p \wedge a + b + c > 0$
<i>p</i>-Equivalence přesnost	$\frac{a+d}{a+b+c+d} \geq p \wedge a + b + c + d > 0$
E-quantifier	$\max\{\frac{b}{a+b}, \frac{c}{c+d}\} \leq \delta \wedge a + b > 0 \wedge c + d > 0$
Paraconsistent separation	$(1 + p)a \leq b + c$

4ft-kvantifikátory založené na statistických testech hypotéz

- Jsou uvedeny názvy použité v proceduře 4ft-Miner
- Definice je dána podmínkou na frekvence a, b, c, d a na jeden nebo dva parametry.
- χ^2_α je $1 - 2\alpha$ kvantil χ^2 rozdělení s jedním stupněm volnosti.

Název	Podmínka
Lower Critical Implication	$\sum_{i=a}^{a+b} \binom{a+b}{i} p^i (1-p)^{a+b-i} \leq \alpha$
Upper Critical Implication	$\sum_{i=0}^a \binom{a+b}{i} p^i (1-p)^{a+b-i} > \alpha$
Double Lower Critical Implication	$\sum_{i=a}^{a+b+c} \binom{a+b+c}{i} p^i (1-p)^{a+b+c-i} \leq \alpha$
Double Upper Critical Implication	$\sum_{i=0}^a \binom{a+b+c}{i} p^i (1-p)^{a+b+c-i} > \alpha$
Lower Critical Equivalence	$\sum_{i=a+d}^n \binom{n}{i} p^i (1-p)^{n-i} \leq \alpha$
Upper Critical Equivalence	$\sum_{i=0}^{a+d} \binom{n}{i} p^i (1-p)^{n-i} > \alpha$
Simple Deviation	$ad > e^\delta bc$
Fisher Quantifier	$\sum_{i=a}^{\min(r,k)} \frac{\binom{k}{i} \binom{n-k}{r-i}}{\binom{n}{r}} \leq \alpha \wedge ad > bc$
Chi-Square Quantifier	$\frac{(ad-bc)^2}{rk(n-k)(n-r)} n \geq \chi^2_\alpha \wedge ad > bc$