

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

Je dostupná z [této adresy](#)

Verse 20. 8. 2019

Typ úlohy: vysoký rozdíl konfidencí

Data: [Hotel](#)

Problém: *Jsou významné rozdíly mezi německy mluvícími zeměmi na straně jedné a jednotlivými slovanskými zeměmi ohledně vztahů mezi charakteristikami hosta (pohlaví, věk) a typickými parametry odpovědí v dotazníku pobytu?*

Jan Rauch

Katedra informačního a znalostního inženýrství

Vysoká škola ekonomická v Praze

© Jan Rauch

SD4ft-Miner – příklad analytické otázky

- **HOTEL:**

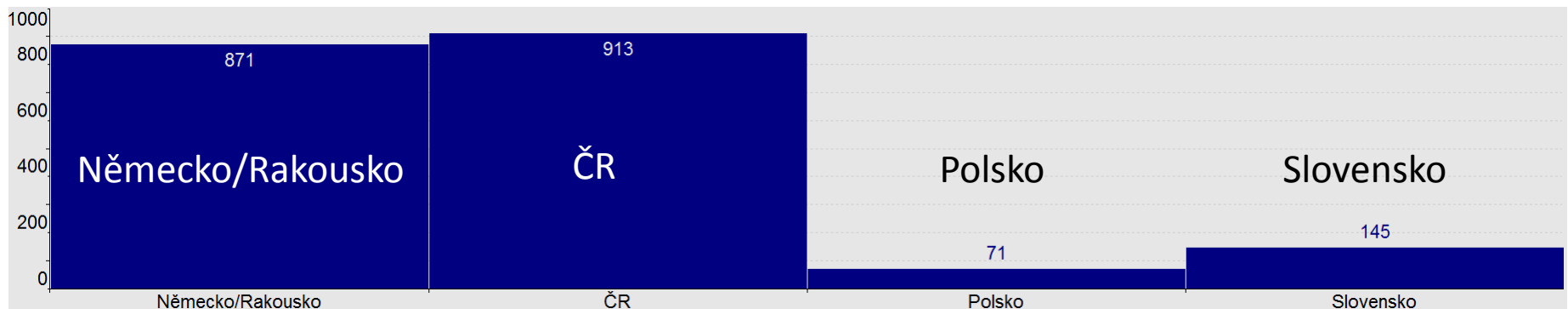
Jsou významné rozdíly mezi německy mluvícími zeměmi na straně jedné a jednotlivými slovanskými zeměmi ohledně vztahů mezi charakteristikami hosta (pohlaví, věk) a typickými parametry odpovědí v dotazníku pobytu?

Stát(Německo, Rakousko) x Stát(?) [Host \approx Dotazník]

Stát(Německo, Rakousko) x Stát(?)[Host ≈ Dotazník]

Jsou významné rozdíly mezi německy mluvícími zeměmi na straně jedné a jednotlivými slovanskými zeměmi ohledně vztahů mezi charakteristikami hosta (pohlaví, věk) a typickými parametry odpovědí v dotazníku pobytu?

Nový atribut: HStat_NS s následujícími kategoriemi:



Stát(Německo, Rakousko) x Stát (?)[Host ≈ Dotazník] – vstup

ANTECEDENT		QUANTIFIERS			SUCCEDENT		
Host	Con, 1 - 2	Type	Rel.	Value	Units	Dotazník	Con, 1 - 5
» HPohlavi (subset), 1 - 1	B, pos	a (BASE) FirstSet	>=	50.00	Abs	» DHodnoceni (seq), 1 - 2	B, pos
» HVek_ed10 (seq), 1 - 2	B, pos	a (BASE) SecondSet	>=	50.00	Abs	» DPersonal_ef3 (seq), 1 - 2	B, pos
		PIM DiffVal	>=	0.10	Abs	» DStrava_ef3 (seq), 1 - 2	B, pos
						» DUbytovani_ef3 (seq), 1 - 2	B, pos
						» DZabava_ef3 (seq), 1 - 2	B, pos
		$\frac{a_1}{a_1 + b_1} - \frac{a_2}{a_2 + b_2} \geq 0.1 \wedge a_1 \geq 50 \wedge a_2 \geq 50$					
Total length: 0 - 5 {1 - 2}						Total length: 1 - 5	
(1) FIRST SET		(2) SECOND SET			CONDITION		
Default Partial Cedent	Con, 1 - 1	Slovanské státy	Con, 1 - 1				
» HStat_NS(Německo/Rakou	B, pos	» HStat_NS#2 (subset), 1 - 1	B, pos				

Německo/Rakousko Slovanské státy Dotazník(*)

Stát(Německo, Rakousko) x Stát (?) [Host ≈ Dotazník] – výstup

4 vteřiny
449 565 verifikací
84 vztahů

Task run

Start: 30.10.2015 14:45:25 Total time: 0h 0m 4s
Number of verifications: 449565
Number of hypotheses: 84 Mode: Standard

Add group

Del group

Edit group

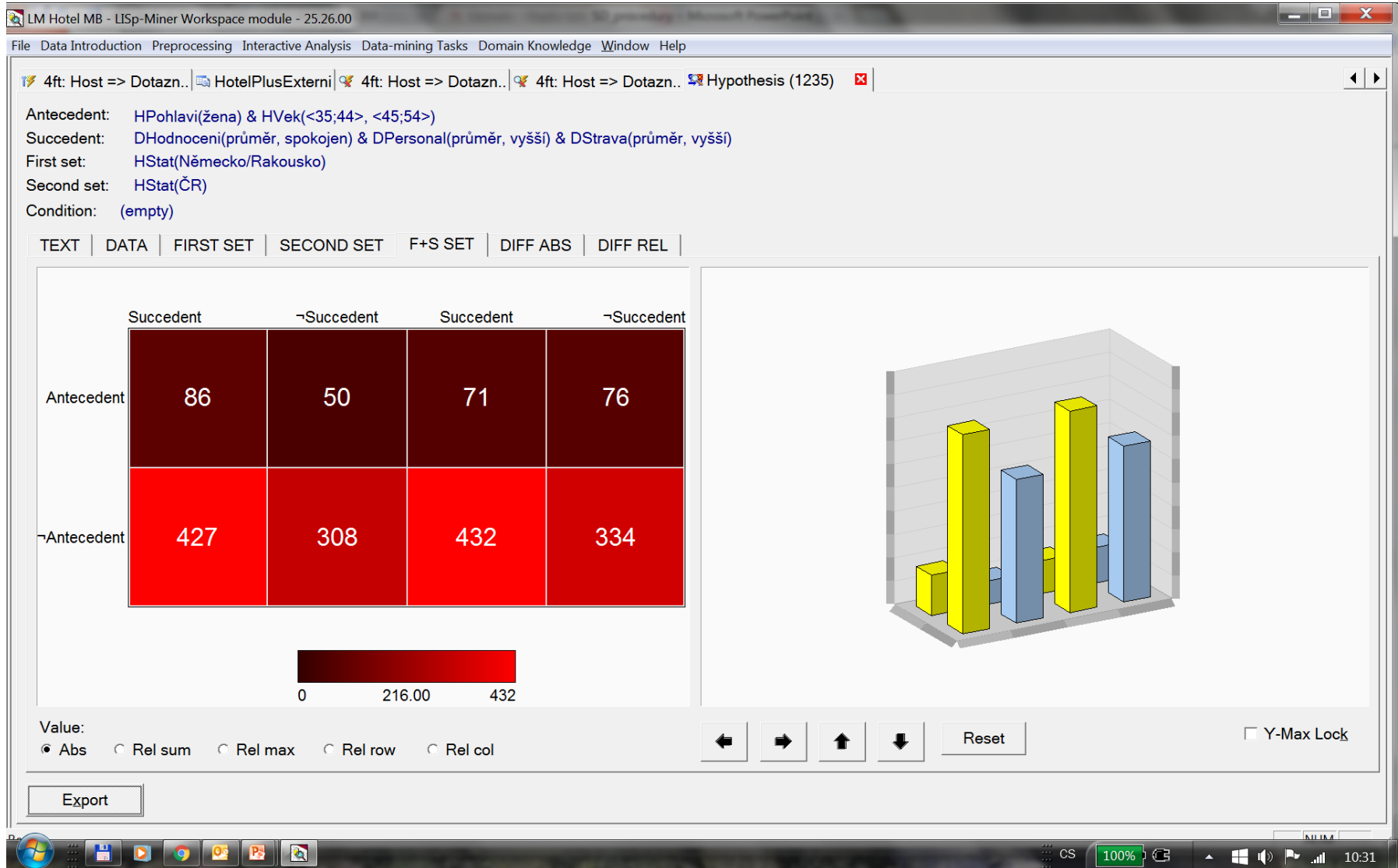
Actual group of hypotheses: All hypotheses

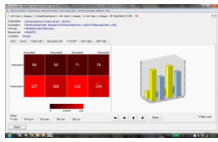
Hypotheses in group: 84 Shown hypotheses: 84 Highlighted: 0

Delete hypotheses

Nr.	Id	Df-Conf	1:Conf	2:Conf	Hypothesis
1	27	0.149	0.632	0.483	HPohlavi(žena) & HVek(<35;54>) >+< DHodnoceni(průměr, spokojen) & DPersonal(>=průměr) & DStrava(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
2	38	0.149	0.632	0.483	HPohlavi(žena) & HVek(<35;54>) >+< DPersonal(>=průměr) & DStrava(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
3	13	0.149	0.825	0.675	HPohlavi(muž) & HVek(<45;64>) >+< DUbytovani(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
4	30	0.144	0.566	0.422	HPohlavi(žena) & HVek(<35;54>) >+< DHodnoceni(průměr, spokojen) & DPersonal(>=průměr) & DStrava(>=průměr) & DZabava(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
5	41	0.144	0.566	0.422	HPohlavi(žena) & HVek(<35;54>) >+< DPersonal(>=průměr) & DStrava(>=průměr) & DZabava(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
6	50	0.144	0.557	0.412	HVek(<=24) >+< DHodnoceni(průměr, spokojen) & DStrava(>=průměr) & DUbytovani(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
7	15	0.138	0.699	0.560	HPohlavi(žena) & HVek(<=24) >+< DHodnoceni(průměr, spokojen) : HStat(Německo/Rakousko) × HStat(ČR)
8	55	0.135	0.547	0.412	HVek(<=24) >+< DStrava(>=průměr) & DZabava(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
9	35	0.135	0.618	0.483	HPohlavi(žena) & HVek(<35;54>) >+< DHodnoceni(průměr, spokojen) & DStrava(>=průměr) & DZabava(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
10	45	0.133	0.680	0.546	HPohlavi(žena) & HVek(>=65) >+< DHodnoceni(nespokojen, průměr) & DPersonal(<=průměr) & DStrava(<=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
11	47	0.133	0.680	0.546	HPohlavi(žena) & HVek(>=65) >+< DPersonal(<=průměr) & DStrava(<=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
12	25	0.133	0.500	0.367	HPohlavi(žena) & HVek(<35;54>) >+< DHodnoceni(průměr) & DZabava(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
13	49	0.132	0.613	0.481	HVek(<=24) >+< DHodnoceni(průměr, spokojen) & DStrava(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
14	14	0.132	0.649	0.518	HPohlavi(muž) & HVek(<45;64>) >+< DUbytovani(>=průměr) & DZabava(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
15	24	0.132	0.485	0.354	HPohlavi(žena) & HVek(<35;54>) >+< DHodnoceni(průměr) & DPersonal(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)
16	22	0.132	0.485	0.354	HPohlavi(žena) & HVek(<35;54>) >+< DHodnoceni(nespokojen, průměr) & DPersonal(>=průměr) : HStat(Německo/Rakousko) × HStat(ČR)

Stát(Německo, Rakousko) x Stát (?)[Host ≈ Dotazník] – detail





Stát(Německo, Rakousko) x Stát(?)[Host \approx Dotazník] – detail

HPohlaví(žena) \wedge HVek(35;54) = žena \wedge (35;54)

OK = DHodnoceni(průměr, spokojen) \wedge DPersonal(průměr, vyšší) \wedge DStrava(průměr, vyšší)

Německo/Rakousko	OK	\neg OK
žena \wedge (35;54)	86	50
\neg (žena \wedge (35;54))	427	308

ČR	OK	\neg OK
žena \wedge (35;54)	71	76
\neg (žena \wedge (35;54))	432	334

$$\frac{86}{86+50} = 0.632$$

$$\frac{86}{86+50} - \frac{71}{71+76} = 0.149$$

$$\frac{71}{71+76} = 0.483$$

Rozdíl konfidencí pravidla žena \wedge (35;54) \approx OK mezi Německo/Rakousko a ČR je 0.149

Německo/Rakousko x ČR: žena \wedge (35;54) $\Rightarrow_{0.149,86,71}$ OK

Poznámka

Místo

Německo/Rakousko	OK	¬OK	ČR	OK	¬OK
žena \wedge $\langle 35;54 \rangle$	86	50	žena \wedge $\langle 35;54 \rangle$	71	76
\neg (žena \wedge $\langle 35;54 \rangle$)	427	308	\neg (žena \wedge $\langle 35;54 \rangle$)	432	334

by mělo být

HotelPlusExterni/HStat(Německo/Rakousko)	OK	¬OK
žena \wedge $\langle 35;54 \rangle$	86	50
\neg (žena \wedge $\langle 35;54 \rangle$)	427	308

HotelPlusExterni/HStat(ČR)	OK	¬OK
žena \wedge $\langle 35;54 \rangle$	71	76
\neg (žena \wedge $\langle 35;54 \rangle$)	432	334

Pokud nehrozí nebezpečí nedorozumění, používáme kratší zápis.