

# Úvod do množinových metód pre spoločenské vedy

Juraj Medzihorský



MEB421 Vybrané metódy výzkumu v mezinárodních vztazích  
13. december 2012

# Program

Modelovanie v spoločenských vedách

Príčinnosť v spoločenských vedách

Čo je QCA

Postup QCA

Zdroje

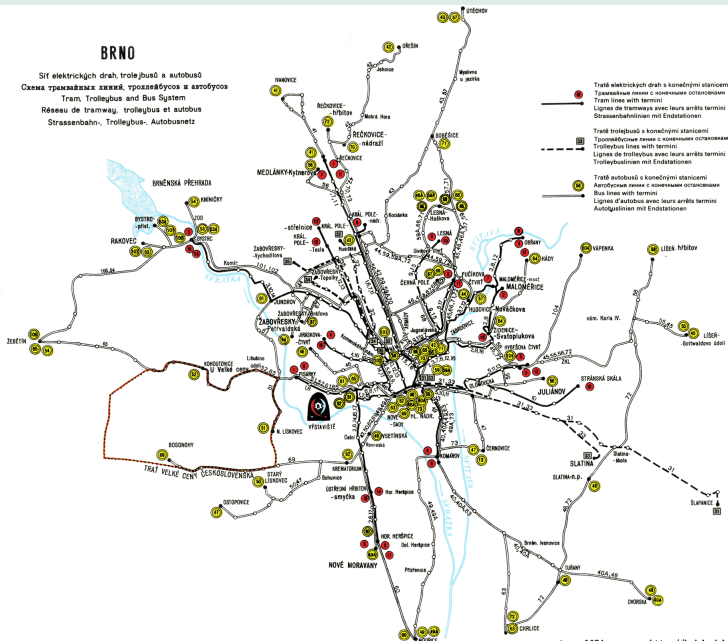


# Brno



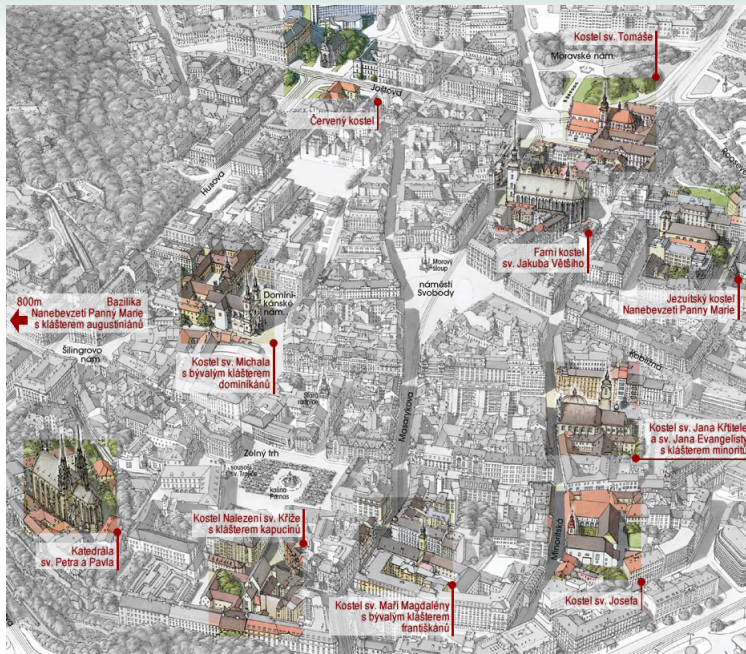
## BRNO

Síť elektrických drah, trolejbusů a autobusů  
 Схема трамвайных линий, троллейбусов и автобусов  
 Tram, Trolleybus and Bus System  
 Réseau de tramway, trolleybus et autobus  
 Strassenbahn-, Trolleybus-, Autobusnetz



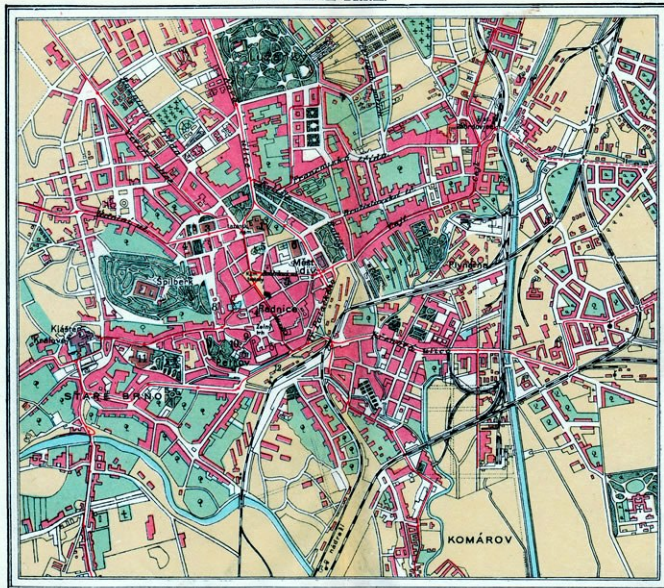
stav: 1974

<http://bmhd.vhd.cz>



# PLÁN MĚSTA BRNA.

Čís. 28.



Měřítko 1:25,000

## Význačné budovy:

Reprodukce Voj. zeměp. ústavu v Praze.

- 1 Zem. správa politická.
- 2 Národní divadlo.

- 4 Radnice
- 5 Chrám sv. Jakuba.

- 7 Ředitelství pošty a telegrafů
- 8 Průmyslové museum.
- 9 Zemské museum.
- 10 Dóm.
- 11 Zem. nemocnice.
- 12 Státní nádraží

MEB421

# Modelovanie v spoločenských vedách

“[V]šetky modely sú nesprávne, ale niektoré sú užitočné.”  
George E. P. Box



## Cvičenie 1: zadanie

- Vyberte si vzťah medzi príčinou a výsledkom zo spoločenských vied.
- Môže to byť aj neoverená hypotéza.
- Zapíšte ho jednou vetou.

## Cvičenie 1: vyhodnotenie

Kto zapísal vzťah, kde príčina a výsledok môžu byť len prítomné alebo chýbajúce?

## Cvičenie 1: vyhodnotenie

Kto zapísal vzťah takto?:

1. Kde je príčina P tam je aj výsledok V.

Pre výsledok V stačí príčina P.

## Cvičenie 1: vyhodnotenie

Kto zapísal vzťah takto?:

1. Kde je príčina P tam je aj výsledok V.  
Pre výsledok V stačí príčina P.
2. Bez príčiny P nie je výsledok V možný.  
Pre výsledok V je potrebná príčina P.

## Cvičenie 1: vyhodnotenie

Kto zapísal vzťah takto?:

1. Kde je príčina P tam je aj výsledok V.  
Pre výsledok V stačí príčina P.
2. Bez príčiny P nie je výsledok V možný.  
Pre výsledok V je potrebná príčina P.
3. Výsledok V je všade tam a len tam kde je príčina P.

# Tabuľkový zápis príčinnosti

Všeobecne			
		Príčina	
		Nie	Áno
Výsledok	Áno	<b>a</b>	<b>b</b>
	Nie	<b>c</b>	<b>d</b>

# Tabuľkový zápis príčinnosti

		Dostatočnosť	
		Príčina	
		Nie	Áno
Výsledok	Áno	✓	✓
	Nie	✓	

## Tabuľkový zápis príčinnosti

		Nevyhnutnosť	
		Príčina	
		Nie	Áno
Výsledok	Áno		✓
	Nie	✓	✓

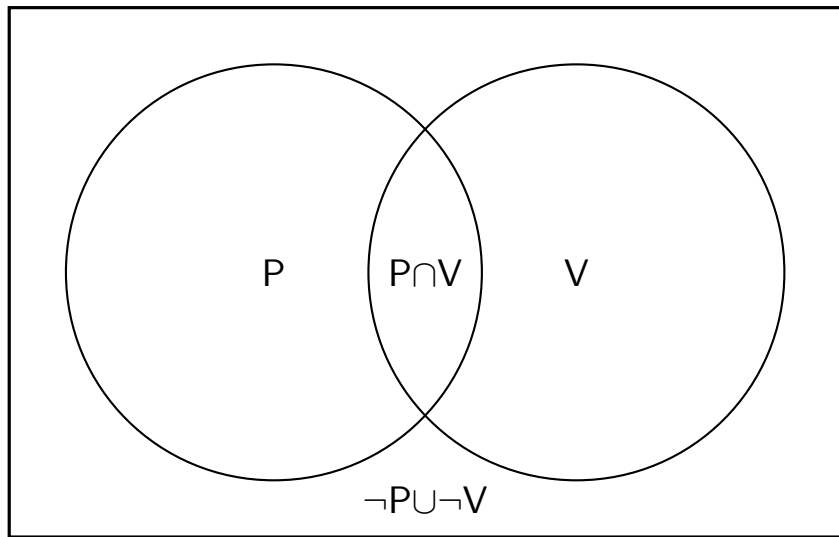


## Tabuľkový zápis príčinnosti

Nevyhnutnosť a dostatočnosť

		Príčina	
		Nie	Áno
Výsledok	Áno		✓
	Nie	✓	

## Množinové zobrazenie príčinnosti



## Množinový zápis podmienenosti

Ak  $\{x \in C | \mathbf{P}(x)\}$  a  $\{x \in O | \mathbf{V}(x)\}$ , tak

*dostatočnosť:*  $C \subseteq O$ ,

*nevyhnutnosť:*  $C \supseteq O$ ,

*rovnocennosť:*  $C = O$ .

A teda

*dostatočnosť:*  $\emptyset = (C \cap \neg O)$ ,

*nevyhnutnosť:*  $\emptyset = (\neg C \cap O)$ ,

*rovnocennosť:*  $\emptyset = [(C \cap \neg O) \cup (\neg C \cap O)]$ .

## Cvičenie 2: zadanie

Zapíšte spoločenskovednú hypotézu o vzťahu medzi výsledkom a príčinou.

- Výsledok a príčina môžu byť len prítomné, alebo chýbať.
- V prvom je príčina dostatočná pre výsledok, v druhom nevyhnutná, a v treťom dostatočná a nevyhnutná súčasne.
- Zadefinujte, čo neexistuje, ak je hypotéza pravdivá.

## Rôznorodá príčinnosť

- Často predpokladáme, že rovnaký výsledok môže mať v rôznych prípadoch rôzne príčiny.
- V takom prípade môžeme hovoriť o *rôznorodosti*, *nahraditeľnosti*, či *zastupiteľnosti* príčin.

## Zložitá príčinnosť

- Často nás zaujíma vzťah medzi výsledkom a *súborom* príčin, ktoré vzájomne ovplyvňujú svoj účinok.
- Napr. takejto formy:  $P_1$  zapríčiňuje  $V$  iba v prítomnosti  $P_2$ .

## Zložitá príčinnosť

- Často nás zaujíma vzťah medzi výsledkom a *súborom* príčin, ktoré vzájomne ovplyvňujú svoj účinok.
- Napr. takejto formy:  $P_1$  zapríčiňuje  $V$  iba v prítomnosti  $P_2$ .
- Niekedy v spojení s príčinnou rôznorodosťou:  $V$  vzniká ak sú súčasne prítomné  $P_1$  a  $P_2$  alebo  $P_3$  a  $P_4$ .
- V tom prípade sú  $P_1$  až  $P_4$  nedostačujúce nevyhnutné súčasti zastupiteľných dostačujúcich konfigurácií, teda *INUS* podmienky.

## Cvičenie 3: zadanie

Zapíšte dve hypotézy o vzťahu medzi výsledkom a príčinami.

- V prvej je príčinnosť rôznorodá.
- V druhej je príčinnosť zložitá.



## Cvičenie 3: vyhodnotenie

- Zapísal niekto *INUS* podmienky?

## Cvičenie 3: vyhodnotenie

- Zapísal niekto *INUS* podmienky?
- Zapísal niekto podmienky, ktoré su nevhnutné v kombinácii, ale nie samostatne?

## Cvičenie 3: vyhodnotenie

- Zapísal niekto *INUS* podmienky?
- Zapísal niekto podmienky, ktoré su nevhnutné v kombinácii, ale nie samostatne?
- Prečo?

# Množiny a spoločenskovedné metódy

Existuje niekoľko spoločenskovedných metód stavajúcich na teórii množín. Líšia sa:

- *Druhom logiky*, na ktorej stavajú: Booleova logika, fuzzy logika, diskretná viachodnotová logika.

# Množiny a spoločenskovedné metódy

Existuje niekoľko spoločenskovedných metód stavajúcich na teórii množín. Líšia sa:

- *Druhom logiky*, na ktorej stavajú: Booleova logika, fuzzy logika, diskretná viachodnotová logika.
- *Chápaním neistoty*: deterministické, frekventistické, Bayesovské.

# Množiny a spoločenskovedné metódy

Existuje niekoľko spoločenskovedných metód stavajúcich na teórii množín. Líšia sa:

- *Druhom logiky*, na ktorej stavajú: Booleova logika, fuzzy logika, diskretná viachodnotová logika.
- *Chápaním neistoty*: deterministické, frekventistické, Bayesovské.
- *Deklarovaným cieľom*: popisná alebo príčinná analýza.

# QCA

**Qualitative Comparative Analysis** možno preložiť do čestiny doslovne ako **Jakostní srovnávací rozbor**, medzinárodne sa používa skratka QCA.

## Čím QCA je

- Súbor techník stavajúcich na teórii množín, na ktorom je možné postaviť komparatívnu analýzu.
- Využíva podobnosť nevyhnutnej a dostatočnej *príčiny* s pojmami nevyhnutnej a dostatočnej *podmienky* v logike.
- Štandardne sa používa na stredne veľký počet prípadov, rádovo desiatky. Možno ju použiť aj s väčším počtom prípadov.
- Má niekoľko verzií postavených na rôznych druhoch logiky. Najpoužívanější je *crisp set* QCA postavená na Booleovskej logike. Dalšie varianty sú *fuzzy set* QCA a *multi value* QCA.
- Niekoľko špecializovaných variant ako *viacúrovňová* QCA a *časová* QCA.



## Čím QCA nie je

- Nie je úplne oddelená od iných kvantitatívnych metód. Teória množín patrí medzi základné časti modernej matematiky.
- Nie je nedeliteľným balíkom. Zložky štandardnej QCA procedúry možno nahradiť alebo doplniť inými metódami.
- Nie je všeliekom na nedostatok dát.
- Rovnako ako iné metódy kladie náročné predpoklady na modely a dáta.
- Nemá jednotnú všeobecne prijatú podobu a výklad.

# Postup analýzy

1. Získanie dát.
2. Spracovanie tabuľky pravdivosti.
3. Vyhodnotenie modelov.

# 1. Získanie dát

- Potrebné sú dáta kde je členstvo prípadov v množinách číselne kódované.
- V prípade Booleovskej QCA (csQCA) 1 značí, že daný prípad je členom danej množiny a 0, že jej členom nie je.
- V prípade fuzzy QCA (fsQCA) je okrem týchto dvoch hodnôt možné aj čiastočné členstvo, ktoré sa zaznačuje číslami z intervalu  $(0, 1)$ .
- V prípade viachodnotovej diskkrétnej QCA (mvQCA) sú možné rôzne hodnoty zaznačované celými číslami.

# 1. Získanie dát

Vďaka stúpajúcej používanosti QCA sú dostupné viaceré dátové súbory pripravené pre QCA. Vo väčšine prípadov však používateľ musí vytvoriť vlastný dátový súbor.

# Premena informácií na členstvá v množinách

Ako sa z niečoho takéhoto:

$P$
-1.83
20.64
100.43
16.13
-73.89
4.57

# Premena informácií na členstvá v množinách

alebo takéhoto:

---

 $P$ 

---

veľmi

slabo

málo

silne

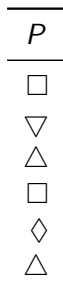
stredne

vôbec

---

# Premena informácií na členstvá v množinách

či takéhto:



# Premena informácií na členstvá v množinách

stane:

$$\begin{array}{c} \hline P \\ \hline 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \\ \hline \end{array}$$



## Premena informácií na členstvá v množinách

- Tzv. *kalibrácia*. Na jej kvalite závisí kvalita celej analýzy.
- Množstvo procedúr na výber. Mnohé boli vyvinuté na zber klasických číselných dát, iné priamo pre QCA.
- V praxi rôzny stupeň formalizácie a prehľadnosti: od 'čiernej skrinky' kódovania po výpočet členstva v množinách pomocou vzorcov z dostupných údajov.

## 2. Spracovanie tabuľky pravdivosti

Analýza vyžaduje spracovanie *tabuľky prípadov* do *tabuľky pravdivosti*.

# Tabuľky prípadov a pravdivosti

Príklad tabuľky prípadov:

$V$	$A$	$B$	$C$	Krajina
0	0	0	0	GAB
1	1	1	1	DJI
1	0	0	1	KEN
0	0	1	0	UGD
1	1	1	0	MAR
1	1	1	0	CON
1	1	1	1	SAR

# Tabuľky prípadov a pravdivosti

Príklad tabuľky pravdivosti:

<i>V</i>	<i>A</i>	<i>B</i>	<i>C</i>	Počet
0	0	0	0	3
0	0	0	1	8
0	0	1	0	10
1	1	0	0	7
1	1	1	0	2
0	0	1	1	9
1	1	0	1	11
1	1	1	1	2

# Tabuľky prípadov a pravdivosti

V tabuľke prípadov je

- každý riadok je jeden prípad a
- každý stĺpec označuje členstvá v jednej množine, alebo obsahuje identifikátor prípadu.

V tabuľke pravdivosti je

- každý riadok jedna kombinácia hodnôt členstiev v množinách, teda ak máme  $K$  podmienok, máme  $2^K$  riadkov;
- každý riadok môže alebo byť členom výsledku, alebo nečlenom, alebo mať túto hodnotu prázdnu;
- zvyčajne každý riadok obsahuje informáciu o počte prípadov, ktoré do neho patria a/alebo ich zoznam.

## Spracovanie tabuľky pravdivosti

- Nevyhnutnosť sa vyhodnocuje pre samostatné podmienky.
- Zjednodušenie s cieľom získať dostačujúce konfigurácie.  
*Metódou hlavných implikantov*, tzv. Quine–McCluskey algoritmus.
- Problém *obmedzenej rozmanitosti*: nie všetky konfigurácie podmienok sú pozorované. Potreba výberu predpokladov o nepozorovaných konfiguráciách.

# Zjednodušovanie konfigurácií

Máme tabuľku prípadov:

$V$	$A$	$B$	$C$
0	0	0	0
1	1	1	1
1	0	0	1
0	0	1	0
1	1	1	0
1	1	1	0
1	1	1	1

## Zjednodušovanie konfigurácií

Dáta si prevedieme do *tabuľky pravdivosti*:

V	A	B	C	Počet
0	0	0	0	1
?	0	0	1	0
0	0	1	0	1
?	1	0	0	0
1	1	1	0	2
?	0	1	1	0
1	1	0	1	1
1	1	1	1	2



# Zjednodušovanie konfigurácií

V QCA sa často používa zjednodušený zápis:

- Množiny sa zapisujú veľkými písmenami:  $A, B, \dots, Z$ .
- Množinový doplnok sa zapisuje malým písmenom:  $A^c$  je  $a$ .
- Množinový prienik sa vynecháva:  $A \cap B$  je  $AB$ .
- Množinové spojenie sa zapisuje ako  $+$ :  $A \cup B$  je  $A + B$ .

## Zjednodušovanie konfigurácií

Zaujímá nás aké konfigurácie podmieňujú V. Súbežne s V pozorujeme tri kombinácie podmienok:

$$AbC + ABc + ABC,$$

Čo môžeme zjednodušiť na:

$$AC + AB$$

keďže V pozorujeme s AB ako v prípadoch, ktoré sú členmi C aj tými, ktoré nimi nie sú.

## Cvičenie 4: zadanie

1. Zjednodušte

$$abc + Abc + aBc + ABc + AbC.$$

2. Rozpíšte

$$aC + Bc.$$

## Cvičenie 4: vyhodnotenie

1. Zjednodušené  $abc + Abc + aBc + ABc + AbC$  je  $c + Ab$ .
2. Rozpísané  $aC + Bc$  je  $abC + aBC + aBc + ABc$ .

## Zjednodušovanie konfigurácií

Niekedy nepozorujeme ani jeden prípad s danou konfiguráciou podmienok. V tom prípade je daný riadok v tabuľke pravidiel prázdný:

<i>V</i>	<i>A</i>	<i>B</i>	<i>C</i>	Počet
?	1	0	0	0

Môžeme:

1. *Konzervatívne* neprijatť žiadne predpoklady.
2. Rozhodnúť sa na základe iných informácií.
3. Predpokladať to, čo povedie k *najjednoduchšiemu riešeniu*.

## Zjednodušovanie konfigurácií

Niekedy narážame v dátach na tzv. protichodné riadky:

<i>V</i>	<i>A</i>	<i>B</i>	<i>C</i>	Id.
0	1	1	1	DJI
1	1	1	1	IND

V takom prípade sa musíme rozhodnúť, či danú konfiguráciu zaradíme do analýzy, a v prípade, že áno, ako ju zakódujeme v tabuľke pravdivosti.

### 3. Vyhodnotenie modelov

- *Súhlas* (*consistency*) – miera zhody modelu s dátami.
- *Pokrytie* (*coverage*) – miera pokrytia relevantnej časti pozorovaní.
- Klasické miery súhlasu a pokrytia sa zakladajú na podieloch určitej kategórie prípadov na určitej všeobecnejšej kategórii. Ak sú vyjadrené v % môžu nadobúdať akúkoľvek hodnotu na intervale  $[0, 100]$  a ak podielom, na intervale  $[0, 1]$ . Vyššie hodnoty ukazujú vyššiu mieru súhlasu, resp. pokrytia.

## Miera súhlasu dát s nevyhnutnosťou

		Príčina	
		Nie	Áno
Výsledok	Áno	$a = 0$	$b$
	Nie	$c$	$d$



## Miera súhlasu dát s nevyhnutnosťou

Klasická miera súhlasu dát s nevyhnutnosťou sa vyjadruje v % prípadov, ktoré majú danú podmienku aj výsledok. Čím väčší podiel takýchto prípadov, tým lepšie pre podmienku. Ktorá z týchto dvoch možností je vhodnejšia?

$$\frac{b}{a + b} \quad (1)$$

$$\frac{b + c + d}{a + b + c + d} \quad (2)$$

## Miera súhlasu dát s nevyhnutnosťou

Klasická miera súhlasu s nevyhnutnosťou podmienky je podiel prípadov s danou podmienkou aj výsledkom na celkovom počte prípadov s výsledkom:

$$S_{P \supseteq V} = \frac{b}{a + b}.$$

# Miera súhlasu dát s dostatočnosťou

		Príčina	
		Nie	Áno
Výsledok	Áno	$a$	$b$
	Nie	$c$	$d = 0$

## Miera súhlasu dát s dostatočnosťou

Klasická mierka súhlasu dát s dostatočnosťou podmienky (alebo konfigurácie) je podiel prípadov s výsledkom a danou podmienkou na celkovom počte prípadov s danou podmienkou:

$$S_{P \subseteq V} = \frac{b}{b + d}.$$

## Miera pokrytia nevyhnutnosťou

		Príčina	
		Nie	Áno
Výsledok	Áno	<i>a</i>	<b>b</b>
	Nie	<i>c</i>	<b>d</b>

## Miera pokrytia nevyhnutnosťou

Klasická miera pokrytia nevyhnutnosťou je podiel prípadov s výsledkom aj danou podmienkou na celkovom počte prípadov s danou podmienkou.

$$P_{P \supseteq V} = \frac{b}{b + d}.$$

## Miera pokrytia nevyhnutnosťou

*Triviálna nevyhnutná podmienka* je taká nevyhnutná podmienka, ktorá pokrýva všetky prípady s daným výsledkom, ale súčasne aj ďaleko väčší počet prípadov bez daného výsledku.

## Miera pokrytia dostatočnosťou

		Príčina	
		Nie	Áno
Výsledok	Áno	<b>a</b>	<b>b</b>
	Nie	<i>c</i>	<i>d</i>



## Miera pokrytia dostatočnosťou

Klasická miera je podiel prípadov s podmienkou (alebo konfiguráciou) a výsledkom na celkovom počte prípadov s výsledkom:

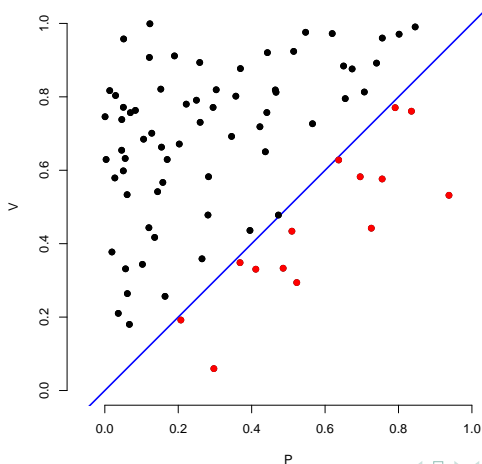
$$P_{P \subseteq V} = \frac{b}{a + b}.$$

## Miera pokrytia dostatočnosťou

*Jedinečné pokrytie* dostatočnou podmienkou (alebo konfiguráciou) je podiel pokrytých prípadov, ktoré nie sú pokryté žiadnou inou dostatočnou podmienkou na celkovom počte prípadov s výsledkom.

# Vyhodnotenie modelov v fsQCA

Vo fuzzy set QCA sa zohľadňujú všetky prípady.



# Softvér

Viacere voľne dostupné balíky:

- `fs/QCA`,
- `Tosmana`,
- knižnice `QCA` , `QCA3` a `SetMethods` pre jazyk `R`,
- `KirqST`.

Tiež

- knižnica `fuzzy` pre komerčný balík `Stata`.

# Komunity užívateľov

- Stránka [www.compassss.org](http://www.compassss.org).
- E-mailové list [QUAL-COMPARE](#) a [QCA-NET](#).

## Vybraná bibliografia

Ragin, Charles C. (1987). *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. University of California Press 1987.

Ragin, Charles C. (2000). *Fuzzy-Set Social Science*. University of Chicago Press.

Ragin, Charles C. (2008). *Redesigning Social Inquiry: Fuzzy Sets and Beyond*. University of Chicago Press.

Goertz, Gary and James Mahoney. (2012). *A Tale of Two Cultures: Qualitative and Quantitative Research in the Social Sciences*. Princeton University Press.

Schneider, Carsten.Q. and Claudius Wagemann. (2012). *Set-Theoretic Methods for the Social Sciences: A Guide to Qualitative Comparative Analysis*. Cambridge University Press.