Introduction to Set-Theoretic Methods for Social Sciences

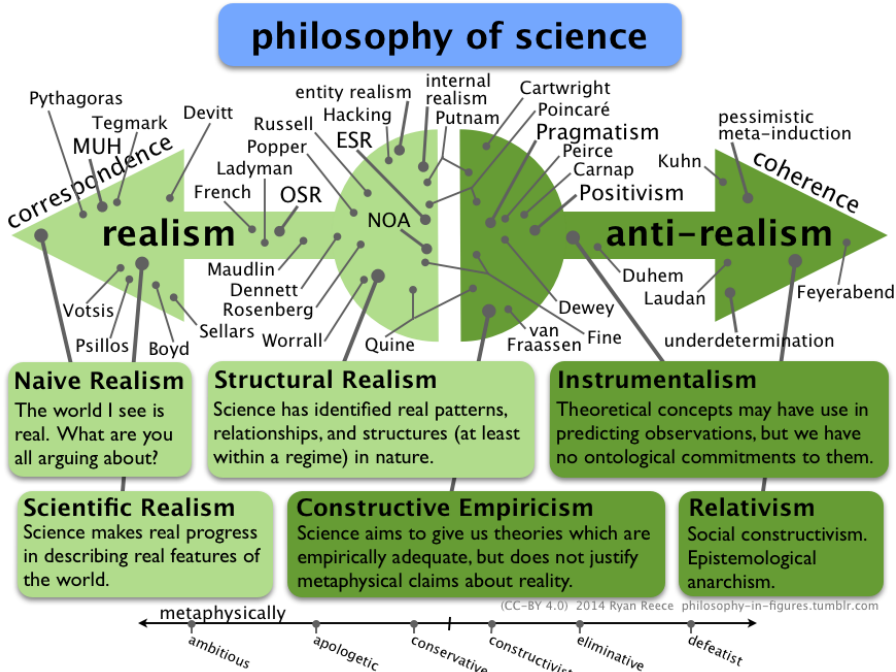
Juraj Medzihorsky

CEU CENTRAL EUROPEAN UNIVERSITY

UNIVERSITÀ DEGLI STUDI DI MILANO

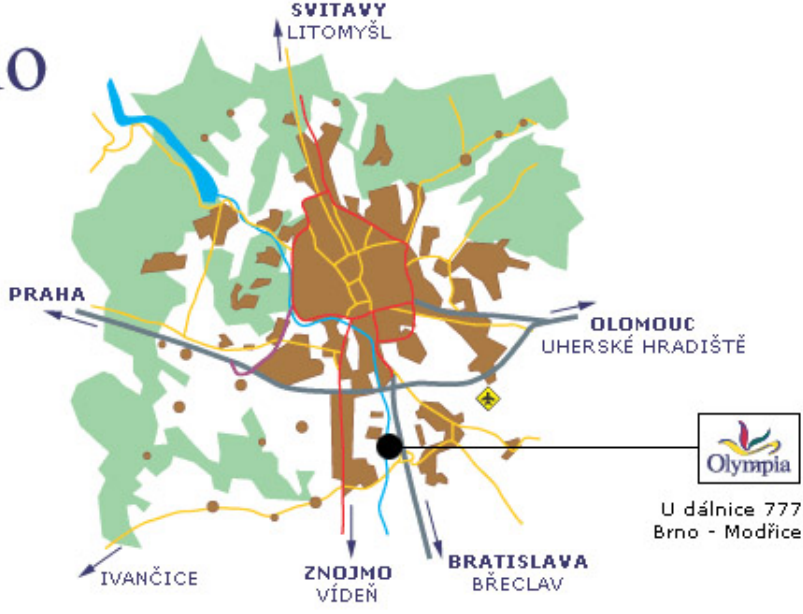
5 December 2014

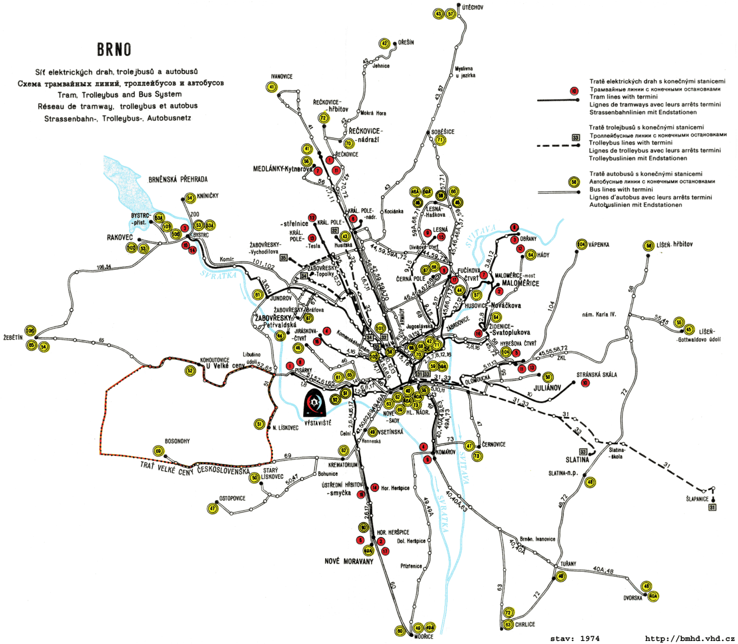
?

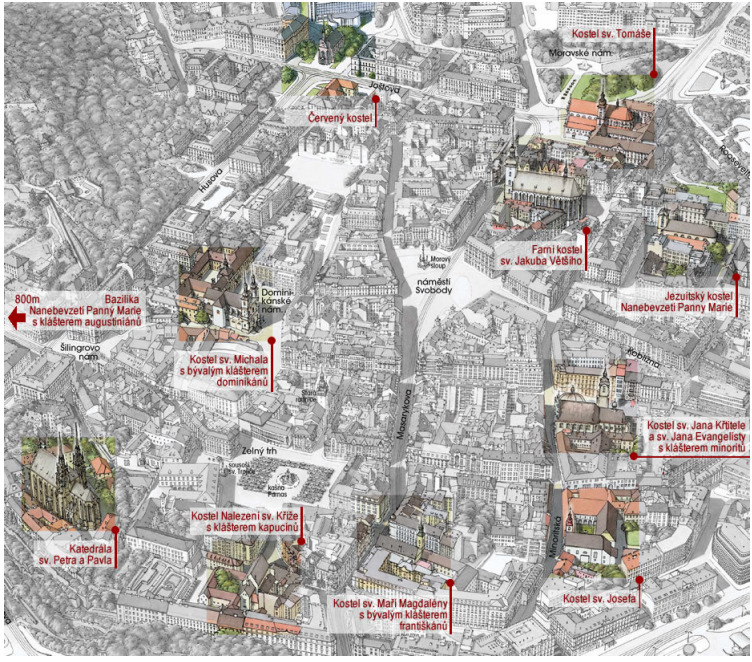

philosophy of science
Pythagoras
Tegmark
MUH
Devitt
correspondence
entity realism
Russell
Hacking
Popper
Ladyman
French
OSR
ESR
internal realism
Putnam
Cartwright
Poincaré
Pragmatism
Peirce
Carnap
Positivism
Kuhn
pessimistic meta-induction
coherence
realism
NOA
anti-realism
Maudlin
Dennett
Rosenberg
Sellars
Worrall
Quine
Votsis
Psillos
Boyd
Duhem
Laudan
Feyerabend
Dewey
van Fraassen
Fine
underdetermination
Naive Realism
The world I see is real. What are you all arguing about?
Structural Realism
Science has identified real patterns, relationships, and structures (at least within a regime) in nature.
Instrumentalism
Theoretical concepts may have use in predicting observations, but we have no ontological commitments to them.
Scientific Realism
Science makes real progress in describing real features of the world.
Constructive Empiricism
Science aims to give us theories which are empirically adequate, but does not justify metaphysical claims about reality.
Relativism
Social constructivism. Epistemological anarchism.
(CC-BY 4.0) 2014 Ryan Reece philosophy-in-figures.tumblr.com
metaphysically
ambitious
apologetic
conservative
constructivist
eliminative
defeatist

Motivation

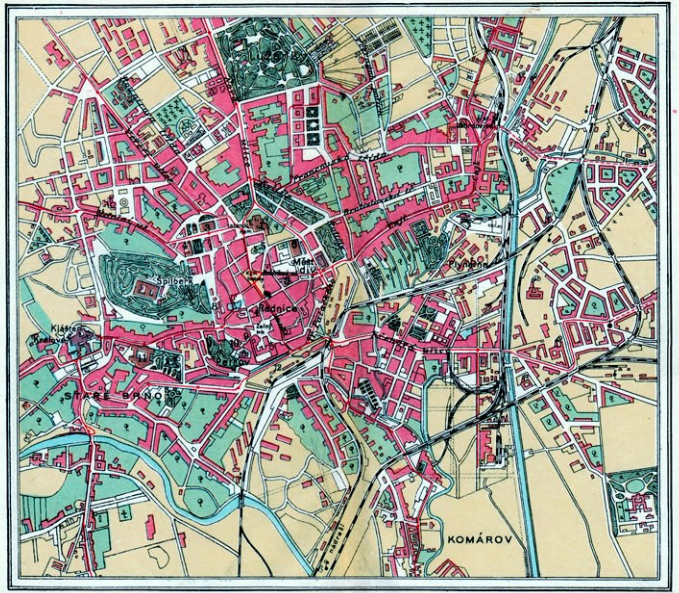
Brno


SVITAVY
LITOMYŠL
PRAHA
OLOMOUC
UHERSKÉ HRADIŠTĚ
Olympia
U dálnice 777
Brno - Modřice
IVANČICE
ZNOJMO
VÍDEŇ
BRATISLAVA
BŘECLAV


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
Tratě elektrických drah s konečnými stanicemi
Трамвайные линии с конечными остановками
Tram lines with termini
Lignes de tramways avec leurs arrêts termini
Strassenbahnlinien mit Endstationen
Tratě trolejbusů s konečnými stanicemi
Троллейбусные линии с конечными остановками
Trolleybus lines with termini
Lignes de trolleybus avec leurs arrêts termini
Trolleybuslinien mit Endstationen
Tratě autobusů s konečnými stanicemi
Автобусные линии с конечными остановками
Bus lines with termini
Lignes d'autobus avec leurs arrêts termini
Autobuslinien mit Endstationen
BRNĚNSKÁ PŘEHRADA
ÚTĚCHOV
OŘEŠÍN
IVANOVICE
ŘEČKOVICE-hřbitov
ŘEČKOVICE-nádraží
MEDLÁNKY-Kytnerova
SOBĚŠICE
KNÍNIČKY
BYSTRC
RAKOVEC
ŽEBĚTÍN
JUNDROV
ŽABOVŘESKY-Petrvaldská
KOHOUTOVICE U Velké ceny
PÍSAŘKY
VÝSTAVIŠTĚ
N. LÍSKOVEC
BOSONOHY
TRAŤ VELKÉ CENY ČESKOSLOVENSKA
STARÝ LÍSKOVEC
OSTOPOVICE
KREMATORIUM
ÚSTŘEDNÍ HŘBITOV-smyčka
HOR. HERŠPICE
NOVÉ MORAVANY
MODŘICE
HL. NÁDR.
KOMÁROV
ČERNOVICE
SLATINA
ŠLAPANICE
TUŘANY
DVORSKÁ
CHRLICE
JULIÁNOV
STRÁNSKÁ SKÁLA
ZIDENICE-Svatoplukova
HUSOVICE-Nováčkova
MALOMĚŘICE
OBŘANY
LESNÁ
ČERNÁ POLE
VÁPENKA
LÍŠEŇ hřbitov
LÍŠEŇ-Gottwaldovo údolí
SVRATKA
SVITAVA


Kostel sv. Tomáše
Moravské nám.
Jostova
Červený kostel
Husova
Farní kostel
sv. Jakuba Většího
Jezuitský kostel
Nanebevzetí Panny Marie
náměstí
Svobody
Dominikánské nám.
800m
Bazilika
Nanebevzetí Panny Marie
s klášterem augustiniánů
Šlingrovo nám.
Kostel sv. Michala
s bývalým klášterem
dominikánů
Masarykova
Zelný trh
Kostel sv. Jana Křtitele
a sv. Jana Evangelisty
s klášterem minoritů
Kostel Nalezení sv. Kříže
s klášterem kapucínů
Katedrála
sv. Petra a Pavla
Kostel sv. Mari Magdalény
s bývalým klášterem
františkánů
Kostel sv. Josefa

PLÁN MĚSTA BRNA.



Měřítko 1 : 25.000

Význačné budovy:

1 Zem. správa politická.
2 Národní divadlo.
3 Starý zemský dům.
4 Radnice.
5 Chrám sv. Jakuba.
6 Nejvyš. s. dvůr.
7 Ředitelství pošt a telegrafů.
8 Průmyslové museum.
9 Zemské museum.
10 Dóm.
11 Zem. nemocnice.
12 Státní nádraží.

“all models are false, but some are useful.”

George E. P. Box

Causality

Exercise #1

- Pick a cause & outcome relationship from social research.
- It can be an untested hypothesis.
- Write it down with a single sentence.

Exercise #1: Evaluation

Who wrote down a relationship where both the cause and the outcome can be either present or missing?

Exercise #1: Evaluation

What about these:

1. The outcome is where the cause is.

   The cause is sufficient for the outcome.

Exercise #1: Evaluation

What about these:

1. The outcome is where the cause is.

   The cause is sufficient for the outcome.
2. The outcome is only where the cause is.

   The cause is necessary for the outcome.

Exercise #1: Evaluation

What about these:

1. The outcome is where the cause is.

   The cause is sufficient for the outcome.
2. The outcome is only where the cause is.

   The cause is necessary for the outcome.
3. The outcome is only and always where the cause is.

Exercise #1: Evaluation

- Did you use the term **condition**?

Exercise #1: Evaluation

- Did you use the term **condition**?
- What about **treatment**?

Exercise #1: Evaluation

- Did you use the term **condition**?
- What about **treatment**?
- What's the difference?

Causes & Conditions

- **Treatment**–an intervention.
- **Cause**–manipulable at least in principle.

Causes & Conditions

- Having three angles is a condition for being a triangle.
- Did having three angles caused an object to be a triangle?

Causes & Conditions

- Having three angles is a condition for being a triangle.
- Did having three angles caused an object to be a triangle?
- No.

Causes & Conditions

- Cause is a ‘difference maker.’

Causes & Conditions

- Cause is a ‘difference maker.’
- Counterfactual reasoning.

Causes & Conditions

- Cause is a 'difference maker.'
- Counterfactual reasoning.
- What's the main problem with counterfactuals?

Causes & Conditions

- Cause is a 'difference maker.'
- Counterfactual reasoning.
- What's the main problem with counterfactuals?
- What can we do about it?

Causes & Effects

- Causes of effects.
- Effects of causes.

Causes & Effects

- Causes of effects.
- Effects of causes.
- Do you have examples?

Correlation & Causation

- What does it mean that ‘correlation does not imply causation’?

Independence & Association

- What is **association**?

Independence & Association

- What is **association**?
- What kinds of association do you know?

Independence & Association

- What is **association**?
- What kinds of association do you know?
- Can you improve the statement ‘correlation does not imply causation’?

Independence & Association

- What is **independence**?

Independence & Association

- What is **independence**?
- What about **lack of association**?

Independence & Association

- What is **independence**?
- What about **lack of association**?
- What are the observable implications of no association?

2-by-2 Tables

		X	
		No	Yes
Y	Yes	**a**	**b**
	No	**c**	**d**

Lack of Association

- What do you need to see to say that X and Y are **not** associated?

Lack of Association

- What do you need to see to say that X and Y are **not** associated?
- What do you need to see to say that X and Y are associated?

Sufficiency

		Condition	
		No	Yes
Outcome	Yes	✓	✓
	No	✓	

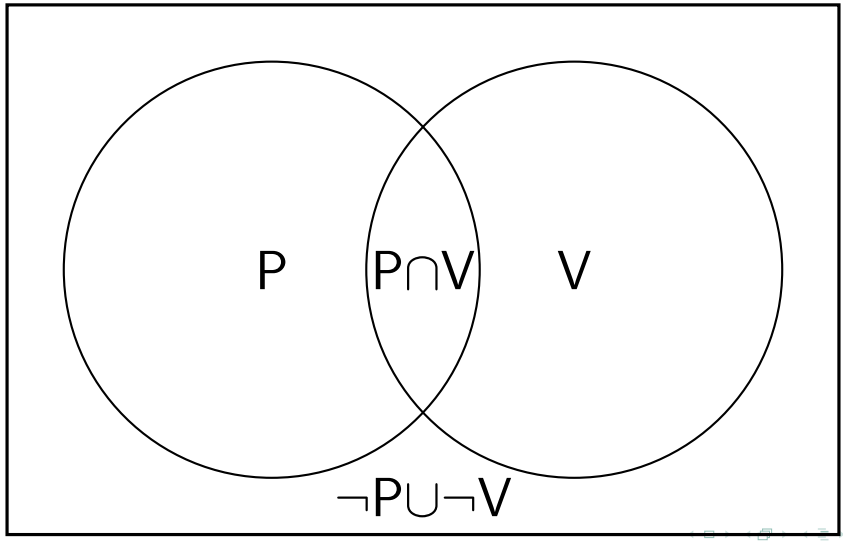
Necessity

		Condition	
		No	Yes
Outcome	Yes		✓
	No	✓	✓

Necessity & sufficiency

		Condition	
		No	Yes
Outcome	Yes		✓
	No	✓	

Set-Theoretic Formulation



Set-Theoretic Formulation

$$
\begin{aligned}
&\textit{Sufficiency:} && C \subseteq O, \\
&\textit{Necessity:} && C \supseteq O, \\
&\textit{Equality:} && C = O.
\end{aligned}
$$

Set-Theoretic Formulation

$$
\begin{aligned}
&\textit{Sufficiency:} && \varnothing = (C \cap \neg O),\\
&\textit{Necessity:} && \varnothing = (\neg C \cap O),\\
&\textit{Equality:} && \varnothing = [(C \cap \neg O) \cup (\neg C \cap O)].
\end{aligned}
$$

Exercise #2

Write down three hypotheses about a cause and an effect

- Cause and effect are both binary.
- One is about necessity, one about sufficiency, and one about both.
- Define what does not exist if the hypothesis is true.

Exercise #2: Evaluation

- How did it go?

Causal Heterogeneity

- We often think that the same effect can have different causes in different cases.
- We can call this causal **heterogeneity** or **substitutability**.

Causal Heterogeneity

- We often think that the same effect can have different causes in different cases.
- We can call this causal **heterogeneity** or **substitutability**.
- Do you have any examples?

Causal Complexity

- Sometimes we are interested in a set of factors that influence their effects.
- E.g., X only causes Y where Z is present.

Causal Complexity & Heterogeneity

- Sometimes we think both might be present.
- E.g., A & B or by C & D both cause Y.

Causal Complexity & Heterogeneity

- Sometimes we think both might be present.
- E.g., A & B or by C & D both cause Y.
- In this example A, B, C, and D are insufficient, but necessary parts of substitutable configurations, a.k.a., INUS conditions.

Exercise #3

Write down two hypotheses.

- The first is about causal heterogeneity.
- The second is about causal complexity.

Exercise #3: Evaluation

- What about INUS conditions?

Exercise #3: Evaluation

- What about INUS conditions?
- Does anyone have causes that are necessary only in a combination?

Exercise #3: Evaluation

- What about INUS conditions?
- Does anyone have causes that are necessary only in a combination?
- Why?

Set-Theoretic Methods

There are several methods based on set theory. The differ in

- *Logic used*: Boolean logic, fuzzy logic, discrete multivalue logic.

Set-Theoretic Methods

There are several methods based on set theory. The differ in

- *Logic used*: Boolean logic, fuzzy logic, discrete multivalue logic.
- *Handling of uncertainty*: deterministic, frequentist, Bayesian.

Set-Theoretic Methods

There are several methods based on set theory. The differ in

- *Logic used*: Boolean logic, fuzzy logic, discrete multivalue logic.
- *Handling of uncertainty*: deterministic, frequentist, Bayesian.
- *Declared goals*: purely descriptive or also causal analysis.

QCA

Q Qualitative

C Comparative

A Analysis

What is QCA?

- A set of techniques built on set theory.
- Designed for comparative analysis.
- Uses the similarity of necessary and sufficient causes with necessity and sufficiency of conditions in logic.

What is QCA?

- Different versions based on different kinds of logic–Boolean in Crisp Set QCA, fuzzy in Fuzzy Set QCA, discrete multivalue in Multivalue QCA.
- Some specialized variants–multilevel QCA, time-series QCA.

What is QCA?

- Typically used with medium-sized datasets, increasingly for large ones as well.
- A.k.a., medium-N.
- More observations handled well, more conditions (variables) not so much.

Is this QCA?

- Different algorithm–Coincidence Analysis (CNA) (author says it's not QCA).
- Classification and regression trees–similar goals, otherwise quite different.
- Boolean regression–in computer science.

Is this QCA?

- Different algorithm–Coincidence Analysis (CNA) (author says it's not QCA).
- Classification and regression trees–similar goals, otherwise quite different.
- Boolean regression–in computer science.
- Well, let's not call it QCA now.

What QCA isn't I.

- Completely separated from quantitative methods–set theory is everywhere, QCA uses numbers.
- Indivisible package–parts can be modified, substituted with other methods.

What QCA isn't II.

- A cure for small datasets.
- Lighter on assumptions than other methods.
- A unified consensually defined approach.

Research Process in QCA

Research Process in QCA

1. Get the data.
2. Prepare the truth table.
3. Evaluate the models.

1. Obtaining the Data

Data for QCA

- Numerically coded set memberships.
- Crisp set QCA–0s and 1s.
- Fuzzy set QCA–from 0 to 1.
- Multivalue QCA–discrete numbers.

Data for QCA

- There are some publicly available datasets prepared for QCA.
- In most research contexts one needs to prepare the data for QCA.

1.2 Encoding information as set memberships

How does this

P
-1.83
20.64
100.43
16.13
-73.89
4.57

or this

P
little
weak
small
strong
medium
none

or this

become this

P
0
1
0
1
1
1

Calibration

- The quality of the analysis depends on it.
- Many procedures available.
- Some originate in classical quantitative analysis, some developed in the context of QCA.

Calibration

- Calibration procedures vary in the degree of formalization.
- One end: *black box*.
- The other end: a formula/algorithm.

2. Truth tables

Truth Tables

- It is necessary to process the *table of cases* into a *truth table*.

Example table of cases

V	A	B	C	Country
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

Example truth table

V	A	B	C	Count
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
1	1	0	1	11
1	1	1	1	2

Table of Cases

- Each row one case (observation, unit).
- Columns contain set memberships or identifying information.

Truth Table

- Each row is a combination of conditions.
- K conditions means 2^K rows.
- Each row can either be in the outcome set, in its complement, or unknown.
- Usually also a column with the number of cases and another with the list of cases from the row.

Processing the Truth Table

- Necessity is evaluated for single conditions.
- Simplification to obtain sufficient combinations. *Method of prime implicants*, i.e., Quine–McCluskey algorithm.

Limited Diversity

- Not all configurations of conditions are observed.
- Assumptions about the unobserved configurations are required.

Limited Diversity

- Impossible vs. unobserved.

Limited Diversity

- Impossible vs. unobserved.
- Example of an impossible combination?

Limited Diversity

- Impossible vs. unobserved.
- Example of an impossible combination?
- Example of an unobserved combination?

2.2 Simplification

From a table of cases

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

to a truth table

V	A	B	C	Count
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

Referential Conventions

QCA often uses simplified referential conventions

- Sets are denoted with uppercase: $A, B, \ldots, Z$.
- Their complements with lowercase: A^c is a.
- Set intersection without a sign: $A \cap B$ is AB.
- Set union as $+$: $A \cup B$ is $A + B$.

Simplification of Configurations

We observe three combinations

$$\mathrm{AbC} + \mathrm{ABc} + \mathrm{ABC},$$

which can be simplified into two

$$\mathrm{AC} + \mathrm{AB}$$

this is called **Boolean minimization**.

Exercise #4

1. Simplify
$$abc + Abc + aBc + ABc + AbC.$$
2. Expand
$$aC + Bc.$$

Exercise #4: Solution

1. Simplified $abc + Abc + aBc + ABc + AbC$ is
$$c + Ab.$$
2. Expanded $aC + Bc$ is
$$abC + aBC + aBc + ABc.$$

Assumptions in Simplification

Sometimes a truth table row is without observations.

V	A	B	C	Count
?	1	0	0	0

Assumptions in Simplification

When a truth table row is without observations, we can

1. Be *conservative* and not assume anything.
2. Use other information to decide.
3. Assume whatever it takes to get the simplest solution.

Assumptions in Simplification

Often we observe 'contradictory' rows in the data:

V	A	B	C	Id.
0	1	1	1	DJI
1	1	1	1	IND

Assumptions in Simplification

With contradictory rows in the table of cases we can

1. Exclude the configuration from the analysis.
2. Decide how to code it–is it in the outcome set or not?

3. Model Evaluation

Model Evaluation

Two quantities

- *Consistency* – how much does the model agree with the data?
- *Coverage* – to what extent does the model cover the relevant observations?

Model Evaluation

- Consistency and coverage look on the shares of some categories of cases in some broader categories.
- Thus values are from $[0, 100]$ in % and from $[0, 1]$ if expressed as fractions.
- Higher values are better–more consistency or more coverage.

Consistency of Necessity

		Condition	
		No	Yes
Outcome	Yes	$a = 0$	b
	No	c	d

Consistency of Necessity

Which of these do you prefer?

$$\frac{b}{a+b} \tag{1}$$

$$\frac{b+c+d}{a+b+c+d} \tag{2}$$

Consistency of Necessity

The classical measure is the fraction of cases with the outcome of interest and the inspected condition in the set of all cases with the outcome.

Consistency of Necessity

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

Consistency of Sufficiency

		Condition	
		No	Yes
Outcome	Yes	a	b
	No	c	$d = 0$

Consistency of Sufficiency

The classical measure is the fraction of cases with the outcome of interest with the inspected condition (or configuration) in the set of all cases with the condition (or configuration).

Consistency of Sufficiency

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

Coverage of Necessity

		Condition	
		No	Yes
Outcome	Yes	a	**b**
	No	c	**d**

Coverage of Necessity

The classical measure is the fraction of cases with the outcome of interest and the inspected condition in the set of all cases with the inspected condition.

Coverage of Necessity

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

Coverage of Necessity

A trivial necessary condition covers all cases with the outcome of interest, but also many more cases without the outcome.

Coverage of Sufficiency

		Condition	
		No	Yes
Outcome	Yes	**a**	**b**
	No	*c*	*d*

Coverage of Sufficiency

The classical measure is the share of cases with the condition (or configuration) and the outcome in the set of all cases with the outcome.

Coverage of Sufficiency

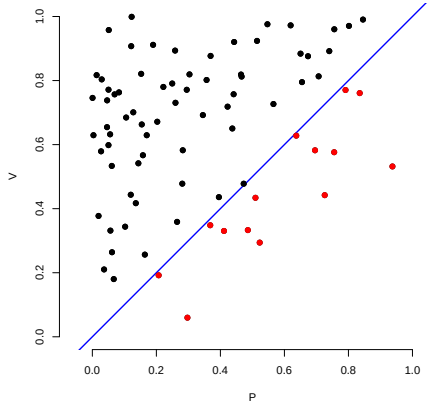
$$P_{P \subseteq V} = \frac{b}{a+b}.$$

Coverage of Sufficiency

Unique coverage of a sufficient condition (or configuration) is the share of the covered cases that are not covered by any other sufficient condition or configuration among all the cases with the outcome.

Model Evaluation in fsQCA

All cases are taken into account.



Software

Free:

- `fs/QCA`
- `Tosmana`
- libraries `QCA`, `QCA3`, and `SetMethods` for `R`,
- `KirqST`.

Also

- library `fuzzy` for commercial software `Stata`.

User Communities

- Website `www.compasss.org`.
- E-mail lists `QUAL-COMPARE` and `QCA-NET`.

Selected Bibliography

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.