

Conjunctive Analysis of Case Configurations (CACC)

Timothy C. Hart ^{a*}, Asier Moneva ^{b, c}, and Miriam Esteve ^d

^a *Department of Criminology and Criminal Justice, University of Tampa, Tampa, Florida*

^b *Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, The Netherlands*

^c *Center of Expertise Cyber Security, The Hague University of Applied Sciences, The Hague, The Netherlands*

^d *Center of Operations Research (CIO), Miguel Hernández University, Elche, Spain*

* Corresponding author: thart@ut.edu

This is an unedited draft, not for citation, of a book chapter that has been accepted for publication in E. Groff and C. Haberman (Eds.), *The study of crime and place: A methods handbook* (Temple University Press, 2021). Temple University Press retains all rights and sub rights of this document. This pre-print is not published under a CC license.

Introduction

Most of the crime and place quantitative methods discussed in this handbook rely on what Ragin (1987, 1994) refers to as a variable-oriented approach to research. The goal of a variable-oriented approach is to study a small number of independent variables (e.g., the presence or absence of crime attractors and/or generators) across a very large number of cases and to identify a parsimonious set of causal variables that explain as much variation as possible in the dependent variable (e.g., crime hotspots). This is accomplished by constructing a generic representation of relationships, based on patterns observed across many cases (e.g., using correlations among variables) and reflects an additive-linear view of causation that depends upon strong homogenizing assumptions about cases which, in turn, make this approach insensitive to causal complexity. In short, the variable-oriented approach to research views causation as a contest between independent variables to explain variation in an outcome.

In contrast, the goal of a case-oriented approach is to examine many aspects of each observed case profile and to construct a representation of each individual from the interconnections among the aspects of these configurations. This approach views cases as configurations – combinations of aspects and conditions – and seeks to understand them at the level of specific instance. In this context, causation is viewed as conjunctural and plural (Schneider & Wagemann, 2012). In other words, causal conditions may combine in different and sometimes contradictory ways to generate the same outcome. A case-oriented approach, unlike a variable-oriented approach, views causation as often both conjunctural and multiple.

When applied to multiple cases or instances, the case-oriented strategy addresses a relatively small number of cases in an in-depth manner, paying attention to each case as an interpretable whole. In essence, this strategy is an extension of the single-case study to multiple cases with an eye toward configurations of similarities and differences. In-depth knowledge of cases provides the basis for constructing limited generalizations that hold for the cases studied. These generalizations may or may not have wider relevance.

Although much of the existing body of research on crime and place is based on a variable-oriented approach, our knowledge and understanding of the *situational context* of crime concentration – and other relevant outcomes – can be advanced through case-oriented approaches. Conjunctive analysis of case configurations (CACC) is one such technique, bridging the gap between the variable-oriented and other case-oriented methodologies.

Conjunctive Analysis of Case Configurations

Conceptual Aspects

Since Miethe and colleagues (2008) introduced CACC as an exploratory data analysis technique, researchers have used this case-oriented approach to improve our understanding of many different topics within the field of criminology, including:

- studies of offender trajectories across the life course (Doherty & Cwick, 2016),
- reporting crime to police (Rennison, Dragiewicz, & DeKeseredy, 2013),
- sentencing outcomes in the U.S. (Hart, Miethe, & Regoeczi, 2014) and in Australia (Lockwood, Hart, & Stewart, 2015),
- bullying victimization (Hart, Hart, & Miethe, 2013),
- risk factors and differential outcomes in carjacking (Miethe & Sousa, 2010), and
- acts of terrorism (DeLeeuw & Pridemore, 2018; Gruenewald, Klein, Drawve, Smith, & Ratcliff, 2019).

Furthermore, and more relevant to this manual, a growing number of studies have applied CACC to crime and place research, including:

- studies of bystander presence and intervention in nonfatal violent victimizations (Hart & Miethe, 2008),
- ship protection measures used to discourage maritime piracy (Bryant, Townsley, & Leclerc, 2014),
- micro-spatial crime hotspots and the influence of public bus stops on crime patterns (Hart & Miethe, 2014, 2015),
- the spatial concentration of drug activities (Hsu & Mogavero, 2020),
- the situational contexts of licensed alcohol premises and their influence on crime patterns (Newton, 2014),
- offending and victimization patterns in “virtual places” (Moneva, Miró-Llinares, & Hart, 2020), and
- the development of new spatial analytic methods (Summers & Caballero, 2017).

As mentioned previously, traditional data analysis techniques that are variable-oriented methods (e.g., regression or correlation analysis) focus on the explanatory power of predictor variables, or the associations between pairs of variables. What these quantitative methods cannot do is disentangle the complex causal relationships of combinations of variable attributes and how these causal recipes produce particular outcomes – especially when those attributes are considered in combination with one another simultaneously.

CACC was designed to overcome these limitations, by analyzing existing data and building composite profiles for a particular unit of analysis (e.g., victims, offenders, guardians, or crime

hotspots). The technique is used to understand the complex causal relationships that emerge when combinations of variable attributes are present or absent in a particular outcome. CACC has evolved over the past decade, from a simple exploratory data analysis technique (Miethe et al., 2008) to a method that can be used to formally test research hypotheses (Hart, 2019). Application of the CACC method is a simple and straightforward process that involves creating a truth table or data matrix, populating the truth table with observed data, preparing the table for analysis, and formally testing hypotheses related to situational clustering.

Creating a truth table

A truth table is constructed from an outcome variable and predictor variables contained in an existing data file. The truth table is comprised of all theoretically observable profiles created from the predictor variables' attributes. Each row in the truth table, also referred to as a CACC data matrix, represents all possible combinations of variable attributes associated with a particular outcome (i.e., the theoretical population of outcomes); columns constitute the specific variables used in the analysis.

Populating the truth table

Next, each observation contained in the existing data file is aggregated into one of the rows that defines the truth table. The truth table is then analyzed to determine the frequency with which a specific profile is observed and how often the outcome variable is observed for each case profile.

Preparing the truth table for analysis

Decision rules for defining dominant case configurations are applied to the truth table after it is constructed and populated. The resulting truth table contains one row for each dominant case configuration that is empirically observed in the original existing data file, the number of times it was observed, and the proportion of times it yielded the outcome of interest (i.e., the dependent variable).

Testing hypotheses

Observed patterns among the dominant case configurations contained in the truth table can be analyzed and formally tested. Recent methods have been developed to determine whether there is significant clustering of the data aggregated to a truth table's case configurations; and if there is, quantifying the magnitude clustering (Esteve, Moneva, & Hart, 2019; Hart, 2019). Techniques for examining the main effects that each predictor variable has on the outcome can also easily be determined using these recently developed methods (Esteve et al., 2019).

Data Requirements

An existing data file containing observations that are associated with theoretically-relevant predictor variables and one outcome variable is needed to conduct CACC. Predictor variables must be measured at either the nominal or ordinal level, whereas the outcome variable must be a dichotomous variable, coded 0 and 1. The number of observations contained in the existing data file should also be considered when selecting the number of predictor variables to use in the analysis because the total number of case configurations will define a truth table will be a function of the number of predictor variables used in the analysis and the number of attributes associated with them.

For example, if five predictor variables are selected and two of them are have three categories and the other three predictors have four categories, then there will be 576 profiles that define the truth table (i.e., $3^2 \times 4^3 = 576$). However, if the existing data file used in the analysis contains only 300 observations, then aggregating them to a truth table with nearly 600 profiles may lead to incorrect conclusions about dominant case configurations and/or situational clustering. It is recommended, therefore, that the number of observable profiles contained in a truth table not exceed the number of observations contained in an existing data file used in the analysis.

Limitations of CACC

As with any methodology, CACC is not without certain limitations. One of the most common critiques levied against this technique is that it can only be applied to nominal- or ordinal-level measures. However, this limitation can easily be overcome through the use of fuzzy sets to define group membership (Ragin, 2000, 2006). With fuzzy sets, membership scores can range from 0.0 (non-membership) to 1.0 (full membership) and cases can vary in the degree to which they satisfy membership criteria. Miethe, Troshynski, and Hart (2017) recently demonstrated the use of set-theoretic methods in CACC to predict national-level imprisonment rates based on predictors typical used in variable-oriented models. This study provides an excellent example of how this perceived limitation of CACC can be easily overcome.

Another shortcoming believed to be associated with CACC is that findings are dataset-specific and cross-study comparisons of findings are not possible. However, Hart (2019) recently introduced a formal test that can be used to determine whether data aggregated to a truth table cluster among dominant case configurations more so than expected and a standardized measure of the magnitude of situational clustering when it is detected. The formal test and standardized measure of situational clustering allow researchers to make cross-study comparisons of CACC findings.

Finally, some critics of CACC suggest findings from configural analysis cannot be generalized, because the unit of analysis is changed from the original sample element to case configurations. However, this is not a shortcoming of CACC; rather, it is a characteristic of most case-oriented methods and how these techniques define populations. Most variable-oriented

techniques involve the analysis of data drawn from samples, which are intended to represent “given” populations. However, Ragin (2013) argues that relying on analysis based on given populations can be problematic because given populations can include irrelevant cases. He further suggests that irrelevant cases often correspond to low scores on relevant causal conditions. In other words, they inflate correlations and paradoxically can appear to confirm theory, despite being irrelevant.

Findings from CACC are not generalizable, because like most case-oriented methods populations are not given; they are constructed. Specifically, populations are built from *candidates* for a qualitative outcome of interest. These candidates include both positive and relevant negative cases, which can be identified once their candidacy for the outcome is established (Ragin, 2013). Using constructed populations that include both positive and relevant negative cases that explain a qualitative outcome, instead of given populations, represent one of the fundamental differences between variable- and case-oriented approaches to data analysis. Therefore, to say that the latter’s inability to generalize findings is a limitation is a red herring because it fails to understand how populations are defined and how this definition impacts generalizability.

Step-by-step Example

CACC can be performed with any software that researchers commonly use to analyze data. This step-by-step example uses R version 4.0.0 (R Core Team, 2020), RStudio version 1.2.5 (RStudio Team, 2019)¹, and the *cacc* R package² recently developed by Esteve and her colleagues (2019). This package automates the construction of a truth table, conducts a formal test of situational clustering, calculates a Situational Clustering Index (SCI), and produces a box plot for each predictor variable the user wants to analyze. These boxplots show the contextual variability associated with each independent variable and the main effect they have on the dependent variable. The demonstration that follows uses data provided by the editors of this handbook.

Data Set

The data set included for this exercise contain publicly available census data for Washington, DC. Information in this file include sociodemographic, block, and crime incident measures. The original data were transformed, however, to meet the CACC data requirements discussed previously. Specifically, we recoded the variables *race_het*, *p_pov*, *med_hh_inc*, *p_move*, *gas*, *restaurant*, and *pawn*; and used the variables *Shape_Area*, *tot_pop*, and *crime2018* to create a dichotomized outcome variable. Details of these transformations are provided below in Step 3.

An Application of the Conjunctive Analysis of Case Configurations (CACC)

Step 1: Start R and load the cacc package

The first step in conducting CACC in R is to open RStudio and download the `cacc` R package file from the book resources. Once downloaded, double-click that file and R Studio will open. If you open the `.R` file provided with the book correctly, your screen should look something like this:

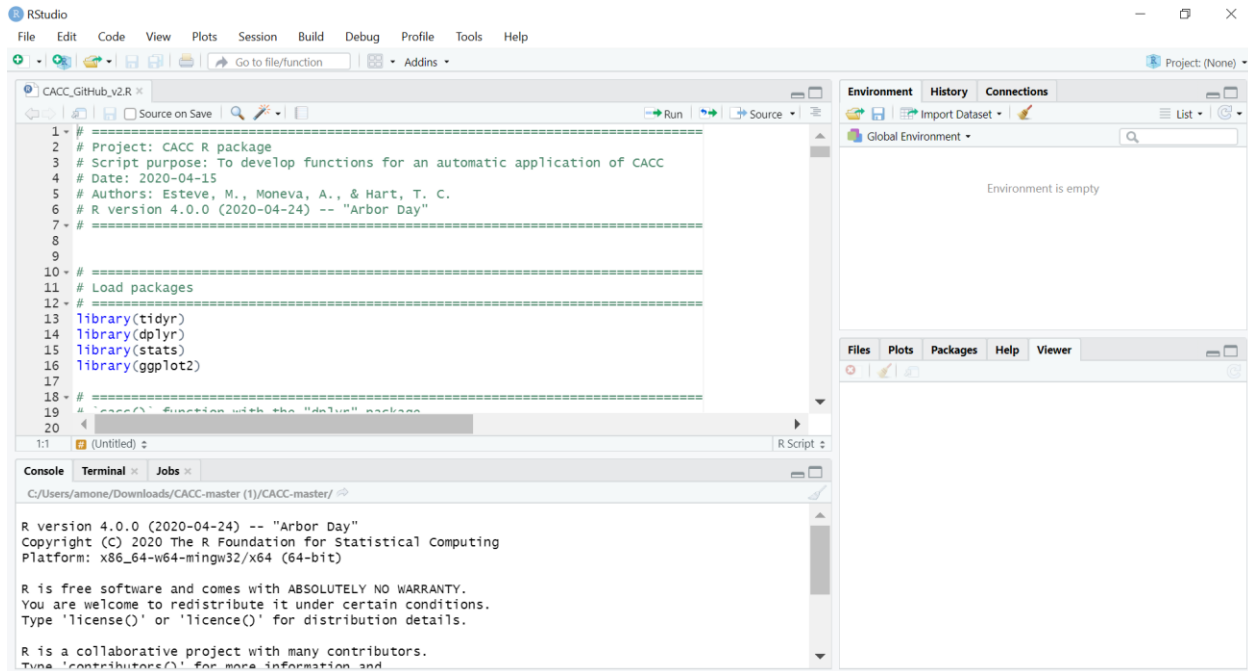


Figure 1. RStudio interface with the script for the cacc package

The R Studio window is divided into four sections. The top-left tab is where the R code used in the CACC will appear; it is the code contained in the `.R` file. In this section, users can make changes to this code if needed. The code will not run until the user clicks the word "Run", located in the top-right portion of this section or uses the appropriate commands. After the CACC code is run, output (including any warnings) will appear in the bottom-left tab of the R Studio window. The environment section is located in the top-right tab, and the bottom-right tab lists the packages, libraries, and graphical output (i.e., graphs, figures, maps, etc.).

Before continuing, all the contents of the script must be run in order to prepare all the functionalities of the package. To do this, click on the "Run" button and then on the "Run All" option in the drop-down menu. Alternatively, simply execute the command `Ctrl + Alt + R`. All `cacc` functions should now appear on the top-right tab (i.e., `cacc()`, `cacc_xsq()`, `sci()`, `gg_lorenz_curve()`, `main_effect()`, among others).

Step 2: Load packages and import data

The next step involves installing or loading additional packages for tidying, transforming, and visualizing data. In this example, we use three R packages that are part of the tidyverse metapackage (Wickham et al., 2019): `tidyr` (Wickham & Henry, 2020), `dplyr` (Wickham, François, Henry, & Müller, 2020), and `ggplot2` (Wickham, 2016). In addition, we use `sf` (Pebesma, 2018). If you do not normally use R, you may not have these packages installed. You can install them from the RStudio interface (see the Packages button on the bottom right tab in Figure 1) or you can run this line of code instead.

```
# Install packages  
install.packages(c("tidyr", "dplyr", "ggplot2", "sf"))
```

The `cacc` script is already prepared for loading these packages. However, in this case, you also need `sf` to import the shapefile with the crime data from Washington. When the packages are loaded, the console displays some information about the versions of the attached packages and possible conflicts between functions.

```
# Load packages  
library(tidyr)  
library(dplyr)  
library(ggplot2)  
library(sf)
```

Once the packages are loaded, the data can be imported into R by indicating the file path with the following function. Note that you must specify your own path. We named our dataset `df`. When data are imported, the console displays information about the type of data it contains. Because `sf` is a package designed to handle geocoded data, the description provides details on such data.

```
# Read the Washington shapefile  
df <- st_read(dsn = "path")
```

Step 3: Create the truth table

After the data and the functions are imported, you can proceed with the analysis. CACC can be performed on any dataset that meets the data requirements described above, regardless of its dimensions; however, it is recommended to reduce its size to only include the variables of greatest interest and that are theoretically relevant. In this example, we prepare a truth table comprised of seven predictor variables and one outcome variable that define the situational context of crime in Washington, DC.

Fortunately, the dataset provided is already tidy (Wickham, 2014), so we just need to transform it to meet the data requirements described previously in Step 1. Data transformation involves creating new variables and recode existing ones (Grolemund & Wickham, 2016).

First, we create the outcome variable `CrimeUpQTL` from the `CrimeRate` variable. The `CrimeRate` variable is a somewhat convoluted variable specific to this exercise, representing the 2018 crime rate controlling for population density. Here are the five steps we took to create this variable.

```
# Create the outcome variable:
df <- df %>%
  as_tibble() %>%
  mutate(
    # 1. Transform the area to square miles;
    SquareMiles = Shape_Area * 0.000000386102,
    # 2. Calculate a density measure for the number of people in a Census
block group, per square mile;
    PplPerMiles = tot_pop / SquareMiles,
    # 3. Obtain the crime rate by the population density;
    CrimeRate = crime2018 / PplPerMiles,
    # 4. Standardize the variable by multiplying it by '10,000';
    CrimeRate = CrimeRate * 10000,
    # 5. Dichotomize the outcome variable depending on whether the value is
in the upper quartile or not.
    CrimeUpQTL = if_else(
      condition = CrimeRate > quantile(
        x = CrimeRate,
        prob = 0.75
      ),
      true = 1,
      false = 0
    )
  )
)
```

Second, it is necessary to transform the continuous numeric variables (i.e., `race_het`, `p_pov`, `med_hh_inc`, and `p_move`) and the integer numeric predictors (i.e., `gas`, `restaurant`, and `pawn`) into categorical variables. You can do this in many ways, but we recommend following a justified criterion. For instance, we created three categories for the continuous variables according to their distribution using quartile breaks. Integer variables, however, were dichotomized based on whether their respective fields indicated the presence or absence of these activity nodes.

```
# The continuous numeric variables are categorized depending on their
distribution (i.e., lower quartile, IQR, and upper quartile).
df <- df %>%
  mutate(
    # Variable 'race_het':
    RaceHetQTL = if_else(
      condition = race_het <= quantile(
        x = race_het,
        prob = 0.25
      ),
      true = 1,
```

```

false = if_else(
  condition = race_het > quantile(
    x = race_het,
    prob = 0.75
  ),
  true = 3,
  false = 2
)
),
# Variable 'p_pov':
PctPvtQTL = if_else(
  condition = p_pov <= quantile(
    x = p_pov,
    prob = 0.25
  ),
  true = 1,
  false = if_else(
    condition = p_pov > quantile(
      x = p_pov,
      prob = 0.75
    ),
    true = 3,
    false = 2
  )
),
# Variable 'med_hh_inc':
MedHhIncQTL = if_else(
  condition = med_hh_inc <= quantile(
    x = med_hh_inc,
    prob = 0.25
  ),
  true = 1,
  false = if_else(
    condition = med_hh_inc > quantile(
      x = med_hh_inc,
      prob = 0.75
    ),
    true = 3,
    false = 2
  )
),
# Variable 'p_move':
PctMvQTL = if_else(
  condition = p_move <= quantile(
    x = p_move,
    prob = 0.25
  ),
  true = 1,
  false = if_else(

```

```

    condition = p_move > quantile(
      x = p_move,
      prob = 0.75
    ),
    true = 3,
    false = 2
  )
),

```

Then, the integer numeric variables are dichotomized depending on their presence or absence.

```

# Variable 'gas':
Gas = if_else(
  condition = gas == 0,
  true = 0,
  false = 1
),
# Variable 'restaurant':
Restaurant = if_else(
  condition = restaurant == 0,
  true = 0,
  false = 1
),
# Variable 'pawn':
Pawn = if_else(
  condition = pawn == 0,
  true = 0,
  false = 1
)
)

```

Finally, the truth table is created by combining the seven predictors and the outcome variable into a data frame. Make sure the outcome variable occupies the rightmost column of the data frame. This way, the `cacc()` function will detect it automatically. To verify that all variables are correctly categorized, we can summarize their basic descriptive statistics. As shown below, the data transformation was successful.

Select the seven predictor variables and the outcome variable.

```

df <- df %>%
  select(
    RaceHetQTL,
    PctPvtQTL,
    MedHhIncQTL,
    PctMvQTL,
    Gas,
    Restaurant,
    Pawn,
    CrimeUpQTL
  )

```

```
# Explore the truth table
```

```
summary(df)
```

```
##      RaceHetQTL      PctPvtQTL      MedHhIncQTL      PctMvQTL      Gas
## Min.      :1.00    Min.      :1.00    Min.      :1.00    Min.      :1.00    Min.      :0.0000
## 1st Qu.:1.25    1st Qu.:1.25    1st Qu.:1.25    1st Qu.:1.25    1st Qu.:0.0000
## Median :2.00    Median :2.00    Median :2.00    Median :2.00    Median :0.0000
## Mean   :2.00    Mean   :2.00    Mean   :2.00    Mean   :2.00    Mean   :0.1489
## 3rd Qu.:2.75    3rd Qu.:2.75    3rd Qu.:2.75    3rd Qu.:2.75    3rd Qu.:0.0000
## Max.   :3.00    Max.   :3.00    Max.   :3.00    Max.   :3.00    Max.   :1.0000
##      Restaurant      Pawn      CrimeUpQTL
## Min.      :0.0000    Min.      :0.00000    Min.      :0.0000
## 1st Qu.:0.0000    1st Qu.:0.00000    1st Qu.:0.0000
## Median :1.0000    Median :0.00000    Median :0.0000
## Mean   :0.5444    Mean   :0.01111    Mean   :0.2511
## 3rd Qu.:1.0000    3rd Qu.:0.00000    3rd Qu.:0.7500
## Max.   :1.0000    Max.   :1.00000    Max.   :1.0000
```

Step 4: Apply the cacc functions

Now that the cacc functions and the truth table are ready, you can conduct CACC. The first function that will be used is `cacc()`, which automates the processes of “populating the truth table” and “preparing the truth table for analysis”. This function takes the truth table created in the previous step as input and produces a CACC matrix.

The resulting CACC matrix shows the dominant profiles observed in the existing Washington, DC data file, sorted by the likelihood of observing the outcome. Note that since the data have less than 1,000 observations, the function has set the threshold of dominant profiles at five or more (see Hart, 2014; Hart, Rennison, & Miethe, 2017). With this parameter, this CACC reveals 31 dominant profiles in the existing data, as indicated by the table dimensions below. You can store the CACC matrix as a separate data frame for further examination and analysis.

```
# Conduct a CACC
```

```
cacc_matrix <- cacc(df)
```

```
# Examine the CACC matrix
```

```
print(
```

```
  cacc_matrix,
```

```
  # Display the top five rows only
```

```
  n = 5,
```

```
  # Display all columns
```

```
  width = Inf
```

```
)
```

```
## # A tibble: 31 x 9
```

```
##   RaceHetQTL PctPvtQTL MedHhIncQTL PctMvQTL Gas Restaurant Pawn N_Break p
##   <dbl>      <dbl>      <dbl>      <dbl> <dbl> <dbl> <dbl> <int> <dbl>
```

```
## 1      1      3      1      2      0      1      0      10 0.5
## 2      2      2      3      3      0      1      0      10 0.5
## 3      2      1      3      1      0      0      0       7 0.429
## 4      3      2      2      2      1      1      0       8 0.375
## 5      3      2      3      2      0      1      0       8 0.375
## # ... with 26 more rows
```

Suppose you want to know whether the observations in your data are disproportionately concentrated in a few dominant profiles (i.e., they cluster among dominant case configurations more than expected). To test this hypothesis related to the situational clustering of crime, you can now apply formal statistical tests on the CACC matrix, as indicated in the “testing hypotheses” section. In this demonstration, we introduce the functions `cacc_xsq()` and `sci()` to apply the two techniques developed by Hart (2019) and the function `gg_lorenz_curve()` to visualize the SCI results.

First, results of a Chi-square goodness-of-fit test indicate that the observations in our data ($N = 450$) are significantly concentrated among a few dominant profiles that define the CACC matrix ($X^2(30, N = 269) = 66.47, p < .001$). Second, the SCI shows that the magnitude of the clustering is moderately strong ($SCI = .24$), as illustrated in Figure 2. In other words, crime is moderately concentrated within a few dominant situational profiles that define crime incident locations.

```
# Chi-square test of independence
cacc_xsq(cacc_matrix)

##   X_Square df          p
## 1  66.4684 30 0.0001418771

# SCI
sci(cacc_matrix)

## [1] "The dataset has been prepared."

## [1] 0.2357597

# Plot the SCI results
gg_lorenz_curve(cacc_matrix)

## [1] "The dataset has been prepared."
```

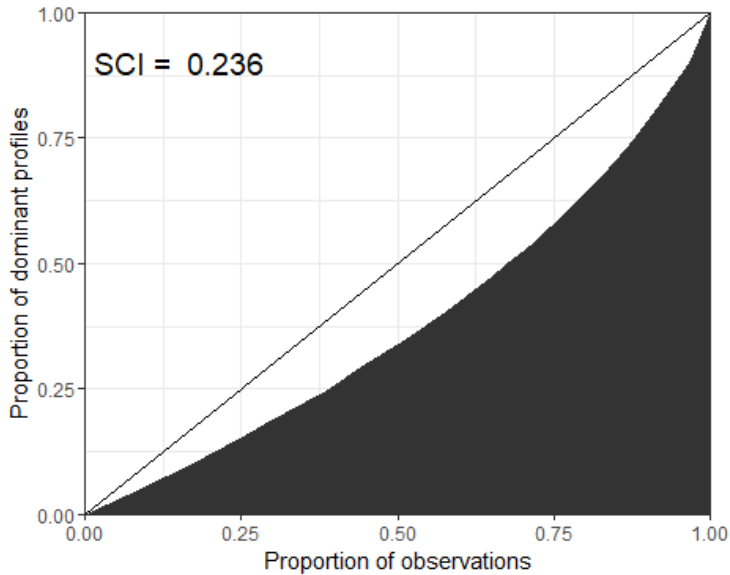


Figure 2. Lorenz Curve for the magnitude of the situational clustering

Further examination of the CACC matrix reveals that the main effect that each predictor variable has on the outcome varies greatly. For example, profiles 1 and 24 are identical, except for the `Restaurant` variable. At least one restaurant is present in profile 1, but absent in profile 24 – all other factors are held constant across both. Isolating the main effect that restaurants have on the outcome in this way shows that their presence is associated with a much higher likelihood of crime ($p_1 = .50$) than when they are absent ($p_{24} = .12$). This illustrates what is known as a “context-specific interaction effect” (Miethe et al., 2008, p. 235), and helps provide insight into the complex causal relationships between predictors and outcomes that traditional analytic techniques struggle to disentangle.

```
# Examine the differences between dominant profiles 1 and 24
cacc_matrix[c(1, 24), ]

## # A tibble: 2 x 9
##   RaceHetQTL PctPvtQTL MedHhIncQTL PctMvQTL   Gas Restaurant Pawn N_Break    p
##   <dbl>      <dbl>      <dbl>      <dbl> <dbl>   <dbl> <dbl> <int> <dbl>
## 1         1         3         1         2     0       1     0     10 0.5
## 2         1         3         1         2     0       0     0     26 0.115
```

We were able to identify these effects between specific pairs of profiles easily in the current CACC matrix because it contains only 31 dominant profiles. However, examining larger truth tables for all sources of contextual variability would be an arduous task and prone to error. In order to automate this process, `cacc` provides the dedicated function `main_effect()` for identifying and visually displaying these sources of contextual variability (Hart et al., 2017).

When executed, this function scans all existing values in each variable and lets the user choose which ones to include in the contextual variability analysis as the reference value. To do

this, you must select a value for each variable (i.e., the reference value) and enter it directly into the console. Any variable can be purposely omitted by entering the value -1. If the CACC matrix contains any constants (i.e., a variable with only one observed attribute), the function omits it because no effect can be evaluated. For example, in this example, Pawn is constant and is therefore excluded from this analysis.

For the current demonstration, we chose the highest value associated with each predictor variable: PctPvtQTL = 3 (i.e., upper quartile); Restaurant = 1 (i.e., present); Gas = 1 (i.e., present). After these values were selected, the function performs the analysis and displays the results as a series of box plots (see Figure 3). Box plots are an effective way to illustrate the degree to which the likelihood of observing an outcome varies when the main effect of a predictor variable is isolated from the influence of all other factors used in the analysis.

For example, when all other predictors are held constant, the likelihood of a crime occurring when at least one restaurant is present in the environment is increased by about 50% in certain contexts, but decreased by about 15% in others. The average likelihood of crime occurring in the environment with a restaurant present is about 17%. In addition to the box plots, the function returns a numerical summary of the distribution of the selected variables in the console.

```
# Scan the CACC matrix for potential sources of contextual variability  
main_effect(cacc_matrix)
```

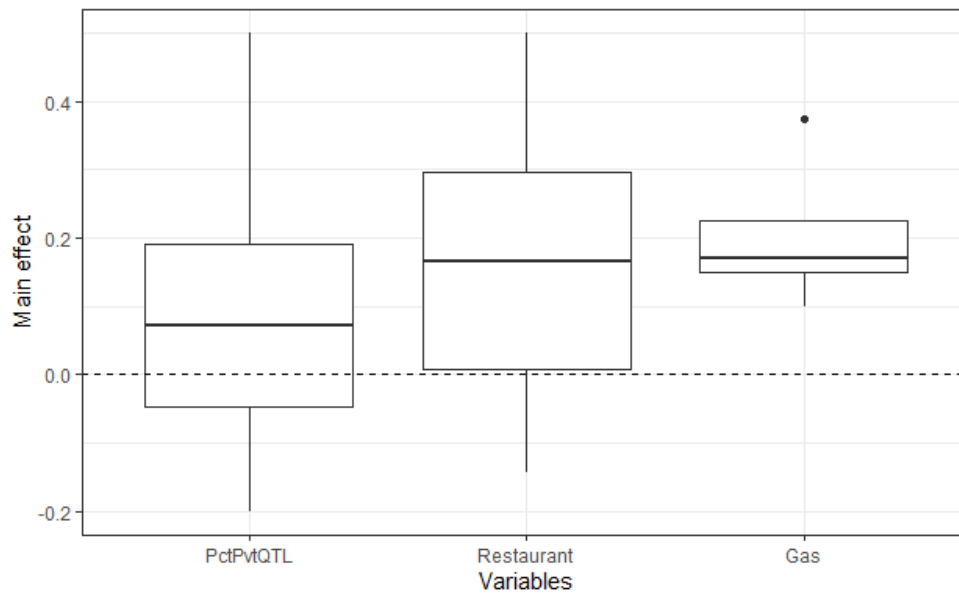


Figure 3. Main effects of the highest variable values.

Summary

Most crime and place studies use traditional variable-oriented data analysis techniques to answer research questions. Although these methods provide a parsimonious and well understood way to

answer important questions, they are not designed to provide us with an in-depth understanding of the context within which outcomes of interest occur. For example, simple linear regression involves “fitting” a Best Linear Unbiased Estimator (i.e., a BLUE line) to pairs of observations and calculating the slope of the line to determine whether the predictor and outcome variables are linearly associated. The formula for the BLUE line’s slope (m) is calculated by applying the following equation:

$$m = r \left(\frac{s_y}{s_x} \right)$$

where r is the correlation between the independent variable (x) and the dependent variable (y), and s_x and s_y are their standard deviations, respectively. The slope of the BLUE line can then be used to determine the y -intercept, which is based on the average values of both x and y .

Single measures of central tendency and dispersion are at the heart of most traditional methods, like regression analysis. Reducing observed data to a single measure is an efficient way to understand and describe large samples. Using them to model associates and make predictions, however, can result in a loss of information and understanding of how “context matters”. CACC is designed to overcome this limitation and should be used to complement these more popular approaches to data analysis.

With the application of the newly developed `cacc` package, in this chapter, we demonstrated how configural analysis can be more easily applied to both exploratory and confirmatory crime and place research. The package allows researchers to simultaneously focus on specificity and multiple causality that underlies most qualitative research and the variable-oriented search for general patterns across contexts that is at the heart of quantitative studies. By assuming maximum causal complexity and using basic methods to visually identify patterns within case configurations, CACC provides a simple way of addressing both of these concerns. Even for crime and place researchers who prefer more variable-oriented quantitative methods, CACC can augment and inform their substantive analysis by identifying possible problems with multicollinearity and low-cell frequencies. In short, the simplicity of CACC provides a succinct and visually appealing way to explore and confirm the nature of crime concentration, diversity, and complex causal patterns among multiple categorical variables.

Funding

This work was supported by the Spanish Ministry of Science, Innovation and Universities under Grant FPU16/01671; and under Grant FPU17/05365.

Endnotes

¹ R must be downloaded and installed before the CACC package can be run. R for Windows can be downloaded from <http://cran.r-project.org/bin/windows/base/> with Mac and Linux versions also available. Once R is installed, it is recommended that R Studio be installed, which can be downloaded from <https://www.rstudio.com/products/rstudio/download/>. Although R Studio is not necessary, it is an interface that is easier to use than the default R Console.

² The CACC R package is available for download at <https://github.com/MiriamEsteve/CACC>

References

- Bryant, W., Townsley, M., & Leclerc, B. (2014). Preventing maritime pirate attacks: A conjunctive analysis of the effectiveness of ship protection measures recommended by the international maritime organisation. *Journal of Transportation Security*, 7(1), 69–82. <https://doi.org/10.1007/s12198-013-0130-2>
- DeLeeuw, J. G., & Pridemore, W. A. (2018). The Threat from Within: A Conjunctive Analysis of Domestic Terrorism Incidents in the United States, United Kingdom, and Ireland. *Perspectives on Terrorism*, 12(4), 26–54.
- Doherty, E. E., & Cwick, J. M. (2016). Unpacking the Complexity of Life Events and Desistance: An Application of Conjunctive Analysis of Case Configurations to Developmental and Life Course Criminology. *Journal of Developmental and Life-Course Criminology*, 2(1), 45–63. <https://doi.org/10.1007/s40865-015-0023-0>
- Esteve, M., Moneva, A., & Hart, T. C. (2019). CACC R package: Conjunctive Analysis of Case Configurations (Version 1.0.3). <https://doi.org/10.5281/zenodo.3472967>
- Grolemund, G., & Wickham, H. (2016). *R for Data Science*. Retrieved from <https://proquest.safaribooksonline.com/9781491910382>
- Gruenewald, J., Klein, B. R., Drawve, G., Smith, B. L., & Ratcliff, K. (2019). Suspicious preoperational activities and law enforcement interdiction of terrorist plots. *Policing: An International Journal*, 42(1), 89–107. <https://doi.org/10.1108/PIJPSM-08-2018-0125>
- Hart, T. C. (2014). Conjunctive Analysis of Case Configurations. Retrieved from JDIBrief Series website: <http://www.ucl.ac.uk/jdibrief/analysis/CACC>
- Hart, T. C. (2019). Identifying Situational Clustering and Quantifying Its Magnitude in Dominant Case Configurations: New Methods for Conjunctive Analysis. *Crime & Delinquency*, 66(1), 143–159. <https://doi.org/10.1177/0011128719866123>
- Hart, T. C., Hart, J. L., & Miethe, T. D. (2013). Situational Context of Student Bullying Victimization and Reporting Behavior: A Conjunctive Analysis of Case Configurations. *Justice Research and Policy*, 15(2), 43–73. <https://doi.org/10.3818/JRP.15.2.2013.43>
- Hart, T. C., & Miethe, T. D. (2008). Exploring Bystander Presence and Intervention in Nonfatal Violent Victimization: When Does Helping Really Help? *Violence and Victims*, 23(5), 637–651. <https://doi.org/10.1891/0886-6708.23.5.637>
- Hart, T. C., & Miethe, T. D. (2014). Street robbery and public bus stops: A case study of activity nodes and situational risk. *Security Journal*, 27(2), 180–193. <https://doi.org/10.1057/sj.2014.5>
- Hart, T. C., & Miethe, T. D. (2015). Configural Behavior Settings of Crime Event Locations: Toward an Alternative Conceptualization of Criminogenic Microenvironments. *Journal of*

- Research in Crime and Delinquency*, 52(3), 373–402.
<https://doi.org/10.1177/0022427814566639>
- Hart, T. C., Miethe, T. D., & Regoeczi, W. C. (2014). Contextualizing sentencing disparities: Using conjunctive analysis of case configurations to identify patterns of variability. *Criminal Justice Studies*, 27(4), 344–361. <https://doi.org/10.1080/1478601X.2014.947031>
- Hart, T. C., Rennison, C. M., & Miethe, T. D. (2017). Identifying Patterns of Situational Clustering and Contextual Variability in Criminological Data: An Overview of Conjunctive Analysis of Case Configurations. *Journal of Contemporary Criminal Justice*, 33(2), 112–120. <https://doi.org/10.1177/1043986216689746>
- Hsu, K. H., & Mogavero, M. C. (2020, November). *Examining concentrations of drug activities: A conjunctive analysis of situational characteristics*. Presented at the 76th Annual Meeting of the American Society of Criminology, Washington, DC.
- Lockwood, K., Hart, T. C., & Stewart, A. (2015). First Nations Peoples and Judicial Sentencing: Main Effects and the Impact of Contextual Variability. *British Journal of Criminology*, 55(4), 769–789. <https://doi.org/10.1093/bjc/azv005>
- Miethe, T. D., Hart, T. C., & Regoeczi, W. C. (2008). The Conjunctive Analysis of Case Configurations: An Exploratory Method for Discrete Multivariate Analyses of Crime Data. *Journal of Quantitative Criminology*, 24(2), 227–241. <https://doi.org/10.1007/s10940-008-9044-8>
- Miethe, T. D., & Sousa, W. H. (2010). Carjacking and its consequences: A situational analysis of risk factors for differential outcomes. *Security Journal*, 23(4), 241–258. <https://doi.org/10.1057/sj.2008.19>
- Miethe, T. D., Troshynski, E. I., & Hart, T. C. (2017). Social Conditions and Cross-National Imprisonment Rates: Using Set-Theoretic Methods for Theory Testing and Identifying Deviant Cases. *Journal of Contemporary Criminal Justice*, 33(2), 152–172. <https://doi.org/10.1177/1043986216688815>
- Moneva, A., Miró-Llinares, F., & Hart, T. C. (2020). Hunter or Prey? Exploring the Situational Profiles that Define Repeated Online Harassment Victims and Offenders. *Deviant Behavior*, 1–16. <https://doi.org/10.1080/01639625.2020.1746135>
- Newton, A. (2014, June). *Activity nodes and licensed premises: Risky mixes and risky facilities*. Presented at the Environmental Criminology and Crime Analysis Symposium, Kerkrade.
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439. <https://doi.org/10.32614/RJ-2018-009>
- R Core Team. (2020). R: A Language and Environment for Statistical Computing (Version 4.0.0). Retrieved from <https://www.R-project.org/>

- Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.
- Ragin, C. C. (1994). *Constructing social research: The unity and diversity of method*. Thousand Oaks, CA: Pine Forge Press.
- Ragin, C. C. (2000). *Fuzzy-set social science*. Chicago: University of Chicago Press.
- Ragin, C. C. (2006). Set Relations in Social Research: Evaluating Their Consistency and Coverage. *Political Analysis*, 14(3), 291–310. <https://doi.org/10.1093/pan/mpj019>
- Ragin, C. C. (2013). New Directions in the Logic of Social Inquiry. *Political Research Quarterly*, 66(1), 171–174.
- Rennison, C. M., Dragiewicz, M., & DeKeseredy, W. S. (2013). Context Matters: Violence Against Women and Reporting to Police in Rural, Suburban and Urban Areas. *American Journal of Criminal Justice*, 38(1), 141–159. <https://doi.org/10.1007/s12103-012-9164-4>
- RStudio Team. (2019). RStudio: Integrated Development Environment for R (Version 1.2.5). Retrieved from <http://www.rstudio.com/>
- Schneider, C. Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge: Cambridge University Press.
- Summers, L., & Caballero, M. (2017). Spatial conjunctive analysis of (crime) case configurations: Using Monte Carlo methods for significance testing. *Applied Geography*, 84, 55–63. <https://doi.org/10.1016/j.apgeog.2017.05.002>
- Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, 59(10), 1–23.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (Second edition). Cham: Springer.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., & Müller, K. (2020). dplyr: A Grammar of Data Manipulation (Version 0.8.5). Retrieved from <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., & Henry, L. (2020). tidyr: Tidy Messy Data (Version 1.0.2). Retrieved from <https://CRAN.R-project.org/package=tidyr>