Necessary Condition Analysis (NCA) in Three Steps: A Demonstration

Online Appendix to Dul, van der Laan, & Kuik, (2018). A statistical significance test for Necessary Condition Analysis. Organizational Research Methods.

Stefan Breet, Henk van Rhee, & Jan Dul
Rotterdam School of Management, Erasmus University

August 2018

Contents

A Step-by-Step Instruction **2**

1 Load the NCA R package **2**

2 Load the data that you want to analyze **2**
2.1 Load example data 2
2.2 Load your own data 3

3 Conduct a Necessary Condition Analysis **3**
3.1 Test a Necessary Condition Hypothesis 3
3.2 Analyze multiple necessary conditions 5
3.3 Check for Statistical Significance 6
3.4 Display the Bottleneck Table 6

More Information **7**

You can conduct a Necessary Condition Analysis and apply the statistical significance test in three steps:

1. Load the NCA R package
2. Load the data that you want to analyze
3. Use the `nca_analysis()` function to run the analysis

The following code block contains a demonstration of the three steps. You can copy-paste the code and use it to analyse your own data. The rest of this appendix contains a detailed description of the individual steps. More details can be found in the NCA Quick Start Guide.

```
#########################################################################################
## 1. Load the NCA R package
#########################################################################################

# Download and install the NCA package (delete the # before running the command)
# install.packages("NCA")

# Update the NCA package to the latest version (delete the # before running the command)
# update.packages("NCA")

# Load the NCA package into the workspace
library(NCA)

#########################################################################################
```

```
## 2. Load the data that you want to analyze
#########################################################################################

# Load the example data set
data(nca.example)

#########################################################################################
## 3. Use the `nca_analysis()` function to run the analysis
#########################################################################################

# Conduct the NCA analysis with the statistical significance test
# Define the conditions (X) and outcome (Y)
# Set the number of permutations to 500
model <- nca_analysis(data = nca.example,
                      x = c("Individualism", "Risk taking"),
                      y = "Innovation performance", test.rep = 500)

# Display the results
nca_output(model)
```

A Step-by-Step Instruction

1 Load the NCA R package

The NCA R package contains all the functions you need to conduct a Necessary Condition Analysis. You can download the package with the `install.packages()` function. We advise you to use the latest versions of the NCA package and the R software to ensure a proper analysis. Updating NCA to the latest version can be done with the `update.packages()` function.

```
# Install the NCA package
# install.packages("NCA") (delete the # before running the command)

# Update the NCA pacakge to the latest version
# update.packages("NCA") (delete the # before running the command)
```

When you have the (latest) NCA package installed on your computer, you can run the `library()` function to load it. You have to load the package every time you start a new R session.

```
# Activate the NCA package
library(NCA)
```

2 Load the data that you want to analyze

2.1 Load example data

We will use the `nca.example` data set for this demonstration. It is included in the NCA package and you can load this data set into your R session with the `data()` function.

```
# Load the example data set
data(nca.example)

# View the first lines of the data set
head(nca.example)
```

```
##           Individualism Risk taking Innovation performance
## Australia            90          84                   50.9
## Austria              55          65                   52.4
## Belgium              75          41                   75.1
## Canada               80          87                   81.4
## Czech Rep            58          61                   14.5
## Denmark              74         112                  116.3
```

The data consists of the innovative performance and cultural dimensions of 28 countries. The cultural dimensions are `Individualism` and `Risk taking` (Hofstede, 1980). The `Innovation performance` of the countries is measured by Gans and Stern's (2003) innovation index.

2.2 Load your own data

All the NCA functions that are demonstrated in this document can be applied to your own data sets as well. To import an existing data set into R, you can use a function that corresponds with its format or file type. For example, you can import a `.csv` file with the `read.csv()` function.

If your data is stored as an SPSS, SAS, or Stata file, we recommend you to use the Haven package. You can install this package with `install.packages("haven")` and activate it with `library("haven")`. The following functions can be used to import your data:

- `read_spss()` for `.sav` files
- `read_sas()` for `.sas7bdat` and `.sas7bcat` files
- `read_dta()` for `.dta` files

If your data is stored as an Excel (`.xlsx`) file, we recommend you to save it as a `.csv` file and import it with the `read.csv()` function.

3 Conduct a Necessary Condition Analysis

Our example data consists of information about cultural aspects of a country and its innovation performance. Suppose that we have a theory that states that `Individualism` and `Risk taking` each are necessary but not sufficient for a country's `Innovation performance`.

To test this theory, we formulate the following hypotheses:

- H1: `Individualism` is necessary but not sufficient for `Innovation performance`.
- H2: `Risk taking` is necessary but not sufficient for `Innovation performance`.

The `nca_analysis` function can be used to test these hypotheses.

3.1 Test a Necessary Condition Hypothesis

We first test whether `Individualism` is a necessary but not sufficient condition for `Innovation Performance`. Since this is the first model we test, we call the analysis `model.1`. We supply the function with the condition (X) and the outcome (Y) by using the corresponding variable names.

```
# Use the nca_analysis function to run the necessary condition analysis
# The condition (X) and outcome (Y) are supplied to the function by their names
# The analysis is stored as "model.1""
model.1 <- nca_analysis(data = nca.example,
                        x = "Individualism",
                        y = "Innovation performance")
```

Because we saved the analysis as `model.1`, we can view its results by calling the model name.

```r
# Display a short summary of the results (effect size):
model.1
```

```
##
## ------------------------------------------------------------------------------
## Effect size(s):
##                ce_fdh cr_fdh
## Individualism 0.416  0.307
## Risk taking   0.309  0.282
## ------------------------------------------------------------------------------
```

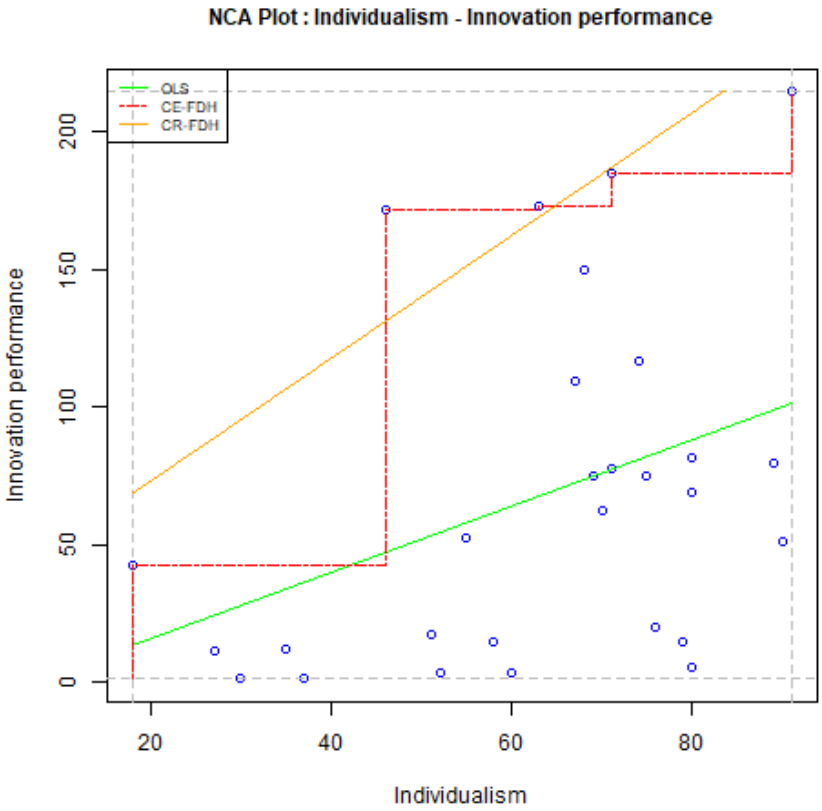
The displayed results consist of two effect sizes. The first one, `ce_fdh`, is based on a ceiling line that is drawn with a step function. It connects the highest values of the outcome (Y) for the values of the condition (X). The second effect size, `cr_fdh`, is based on a straight ceiling line that has been drawn through the points that are part of the step function. More information about the techniques can be found in the paper in Organizational Research Methods that describes the method (Dul, 2016).

A general rule of thumb qualifies effect sizes between 0.0 and 0.1 as a *small effect*, between 0.1 and 0.3 as a *medium effect*, and between 0.3 and 0.5 as a *large effect*. The effect sizes of our example can therefore be considered as large.

To display more detailed results, you can use the `nca_output()` function. For example, you can choose to display a model summary and a NCA plot.

```r
# Display a detailed summary and a plot
nca_output(model.1, summaries = TRUE, plots = TRUE)
```

```
##
## ------------------------------------------------------------------------------
## NCA Parameters : Individualism - Innovation performance
## ------------------------------------------------------------------------------
##
## Number of observations    28
## Scope                  15563.6
## Xmin                      18.0
## Xmax                      91.0
## Ymin                       1.2
## Ymax                     214.4
##
##                     ce_fdh   cr_fdh
## Ceiling zone     6466.800 4772.541
## Effect size         0.416    0.307
## # above             0        2
## c-accuracy         100%      92.9%
## Fit                100%      73.8%
##
## Slope                        2.230
## Intercept                   28.353
## Abs. ineff.      3000.300 6018.517
## Rel. ineff.        19.278   38.670
## Condition ineff.    0.000   10.383
## Outcome ineff.     19.278   31.565
```



We observe an empty space in the upper left corner, which indicates that `Individualim` is a necessary condition for `Innovation Performance`.

3.2 Analyze multiple necessary conditions

Rather than repeating the analysis for `Risk taking` as a necessary condition for `Innovation performance`, we can analyze both necessary conditions in one analysis with the concatenate ("combine") function `c("condition1", "condition2", ...)`. We store the new model as `model.2`.

```
# Supply the two conditions (X) as names with the combine function
model.2 <- nca_analysis(data = nca.example,
                        x = c("Individualism", "Risk taking"),
                        y = "Innovation performance")
# Display the results
model.2
```

```
##
## ------------------------------------------------------------------------------
## Effect size(s):
##                ce_fdh cr_fdh
## Individualism 0.416  0.307
## Risk taking   0.309  0.282
## ------------------------------------------------------------------------------
```

3.3 Check for Statistical Significance

Any effect size we observe could be the result of random chance. We can use the statistical significance test that is part of the `nca_analysis` function to test whether this were the case. The test resamples the data to create a range of samples (permutations) in which the condition (X) and the outcome (Y) are unrelated. The outcome of the test is the probability that we observe our results if this is the case. The probability is represented by the p value. The more the p value of the test approaches zero, the more unlikely it is that the observers effect size is caused by random chance. See Dul, van der Laan, & Kuik, 2018 for more information about the statistical significance test for NCA.

To conduct the test, we supply the *number of permutations* to the `nca_analysis()` function via the `test.rep` argument. We recommend using at least 10,000 permutations if you run the test on your own data set. Increasing the number of permutations, however, increases the processing time as well. In this demonstration we will therefore use only 500 permutations.

```
# Conduct the necessary condition analysis with the permutation test
model.3 <- nca_analysis(data = nca.example,
                        x = c("Individualism", "Risk taking"),
                        y = "Innovation performance", test.rep = 500)
```

```
# Display the results
model.3
```

```
##
## ---------------------------------------------------------------------------
## Effect size(s):
##                ce_fdh p     cr_fdh p
## Individualism 0.416  0.082 0.307  0.170
## Risk taking   0.309  0.112 0.282  0.092
## ---------------------------------------------------------------------------
```

The p values of the effect sizes are relatively large ($p > 0.05$), suggesting that the probability that the observed effect size is due to random chance is considerable. For example, the chance that individualism is not a necessary condition for innovation performance is approximately 8 percent for `ce_fdh` and 17 percent for `cr_fdh`. We therefore do not find support for our two hypotheses.

3.4 Display the Bottleneck Table

The bottleneck table shows which level of the condition (X) is necessary for which level of the outcome (Y). You can display the bottleneck table via the `bottlenecks` argument in the `nca_output()` function. In the bottleneck table NN means 'not necessary'. The X and Y values displayed in the bottleneck table are percentages of the range of X and Y, respectively. This means that 0 = smallest X,Y value; 100 = largest X,Y value, 50 = middle X,Y value. With the `bottleneck.x` and `bottleneck.y` arguments the values can be expressed as percentages of maximum, actual values or percentiles.

```
# Show the bottleneck table
nca_output(model.3, bottlenecks = TRUE, summaries = FALSE)
```

```
##
## ---------------------------------------------------------------------------
## Bottleneck CE-FDH (cutoff = 0)
## Y Innovation performance (percentage.range)
## 1 Individualism          (percentage.range)
## 2 Risk taking            (percentage.range)
## ---------------------------------------------------------------------------
## Y        1     2
## 0       NN    NN
## 10      NN    20.2
```

```
## 20      38.4  20.2
## 30      38.4  20.2
## 40      38.4  22.5
## 50      38.4  22.5
## 60      38.4  22.5
## 70      38.4  22.5
## 80      61.6  59.6
## 90      100.0 74.2
## 100     100.0 74.2
##
##
## ------------------------------------------------------------------------------
## Bottleneck CR-FDH (cutoff = 0)
## Y Innovation performance (percentage.range)
## 1 Individualism          (percentage.range)
## 2 Risk taking            (percentage.range)
## ------------------------------------------------------------------------------
## Y        1    2
## 0       NN   NN
## 10      NN   NN
## 20      NN   NN
## 30      NN   8.0
## 40      11.0 17.1
## 50      24.1 26.2
## 60      37.2 35.2
## 70      50.3 44.3
## 80      63.4 53.4
## 90      76.5 62.4
## 100     89.6 71.5
##
```

More Information

If you have questions about the functions in the R package, you can access the help documentation by adding a question mark before a function. For example, if you want to know more about the `nca_analysis()` function, you can type `?nca_analysis`.

More information about NCA can be found on http://www.erim.nl/nca. If you have any questions about the method or the R package, feel free to contact us by email (breet@rsm.nl, vanrhee@rsm.nl, jdul@rsm.nl).