

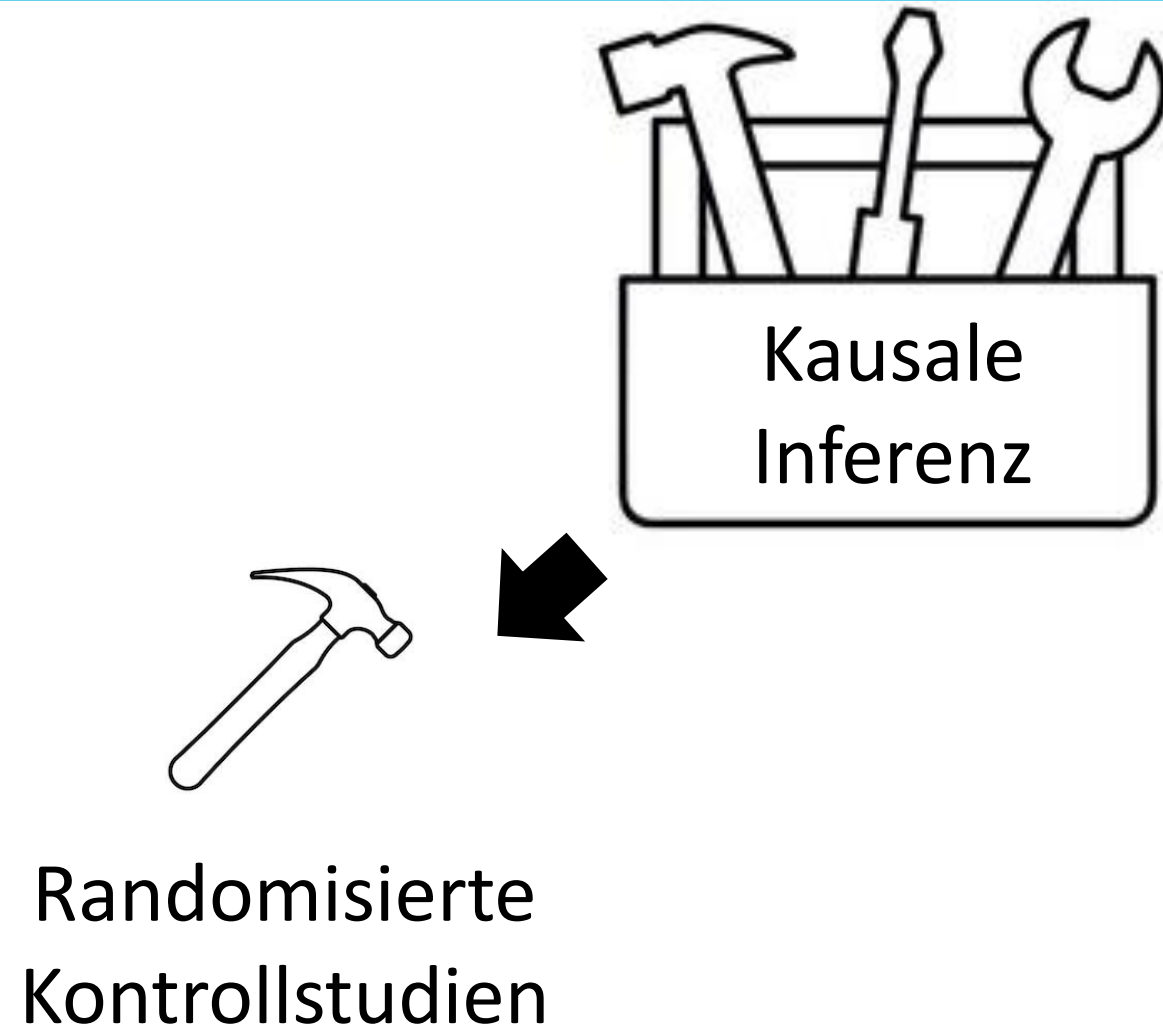
Evaluation der Wirksamkeit psychologischer Interventionen anhand von Regressions-Diskontinuitäts-Analysen

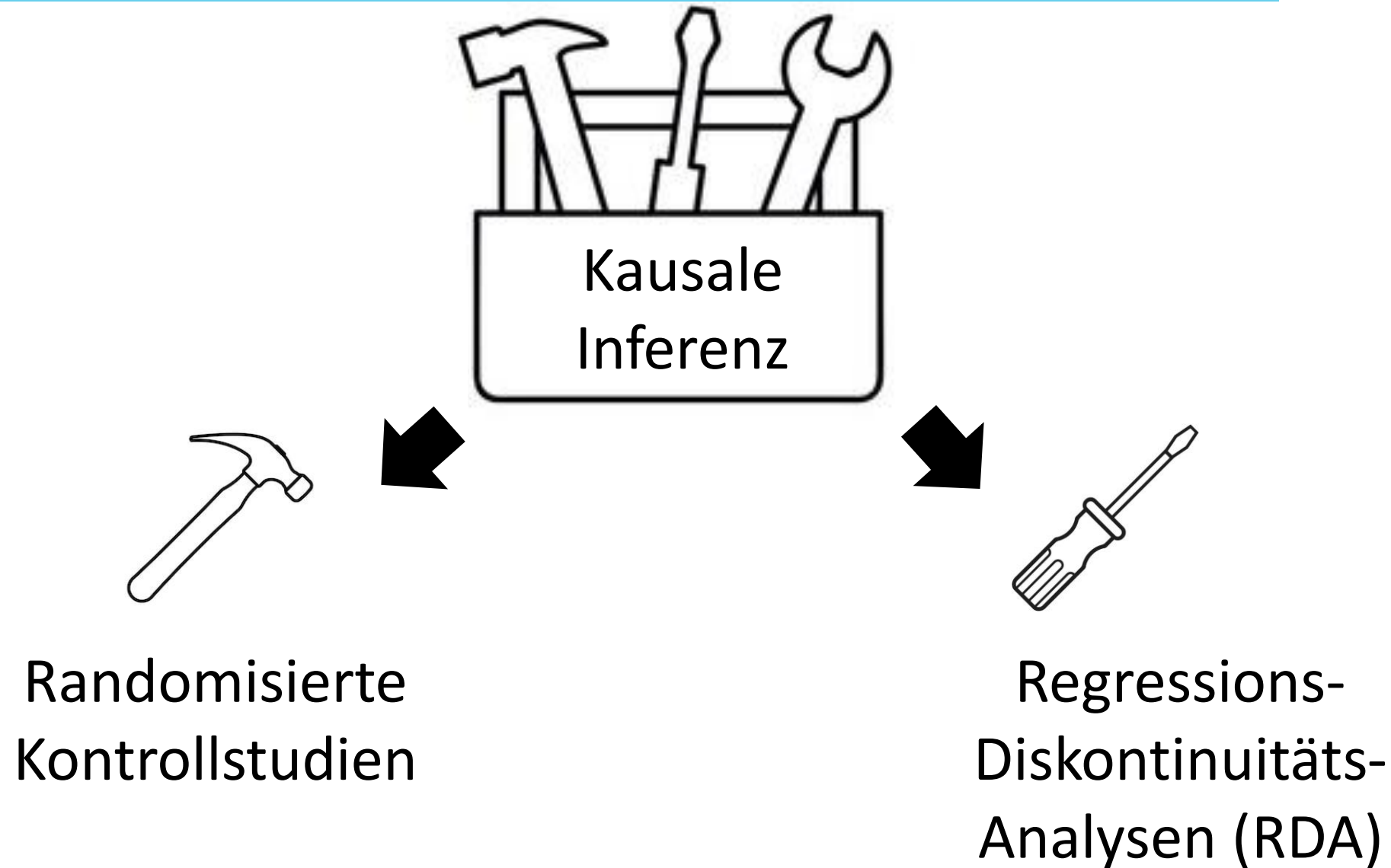
Alan Novaes Tump

Code & Daten



<https://github.com/alantump/RDA>





RDA kann angewendet werden wenn die Intervention durch eine Zuteilungsvariable bestimmt wird.

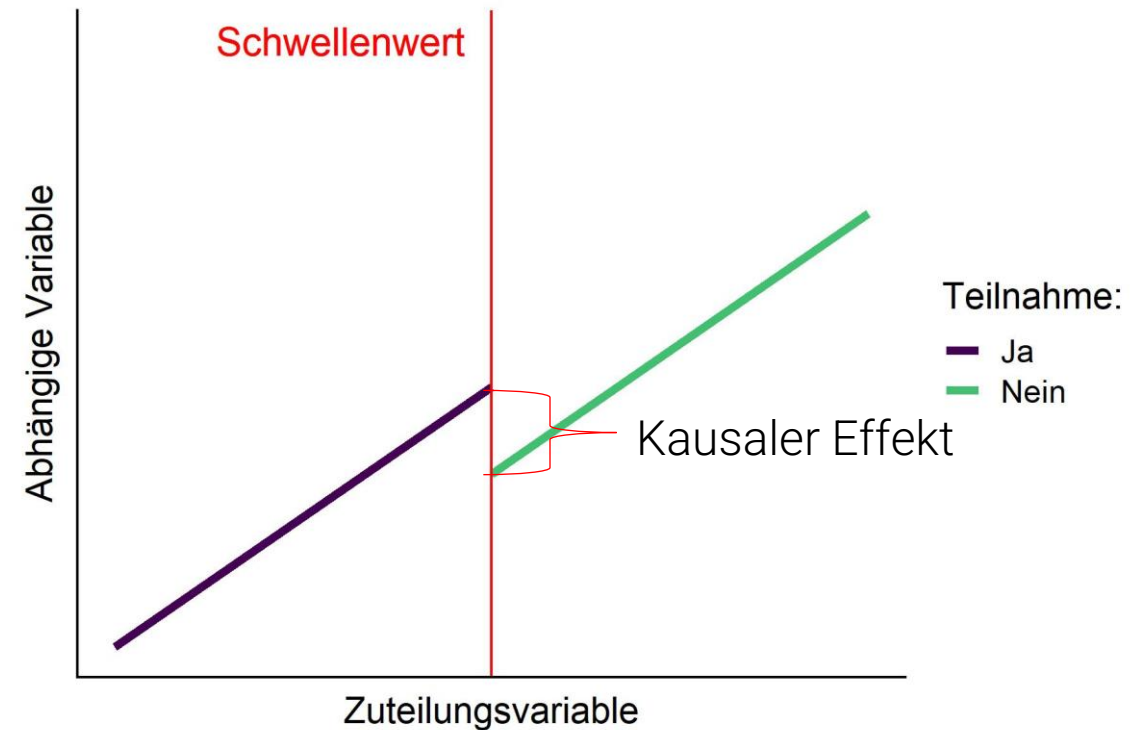
- z.B. Patienten unter einem Schwellenwert werden Behandelt

Problem:

- Keine Randomisierung
- Unterschiede zwischen Gruppen kann auf andere Variablen zurückgeführt werden

Grundidee:

- Individuen nah über und unter dem Schwellenwert sind sehr ähnlich
- Unterschiede bzw. Diskontinuität ist daher auf kausale Effekte zurückzuführen



Beispiele für Zuteilungsvariablen:

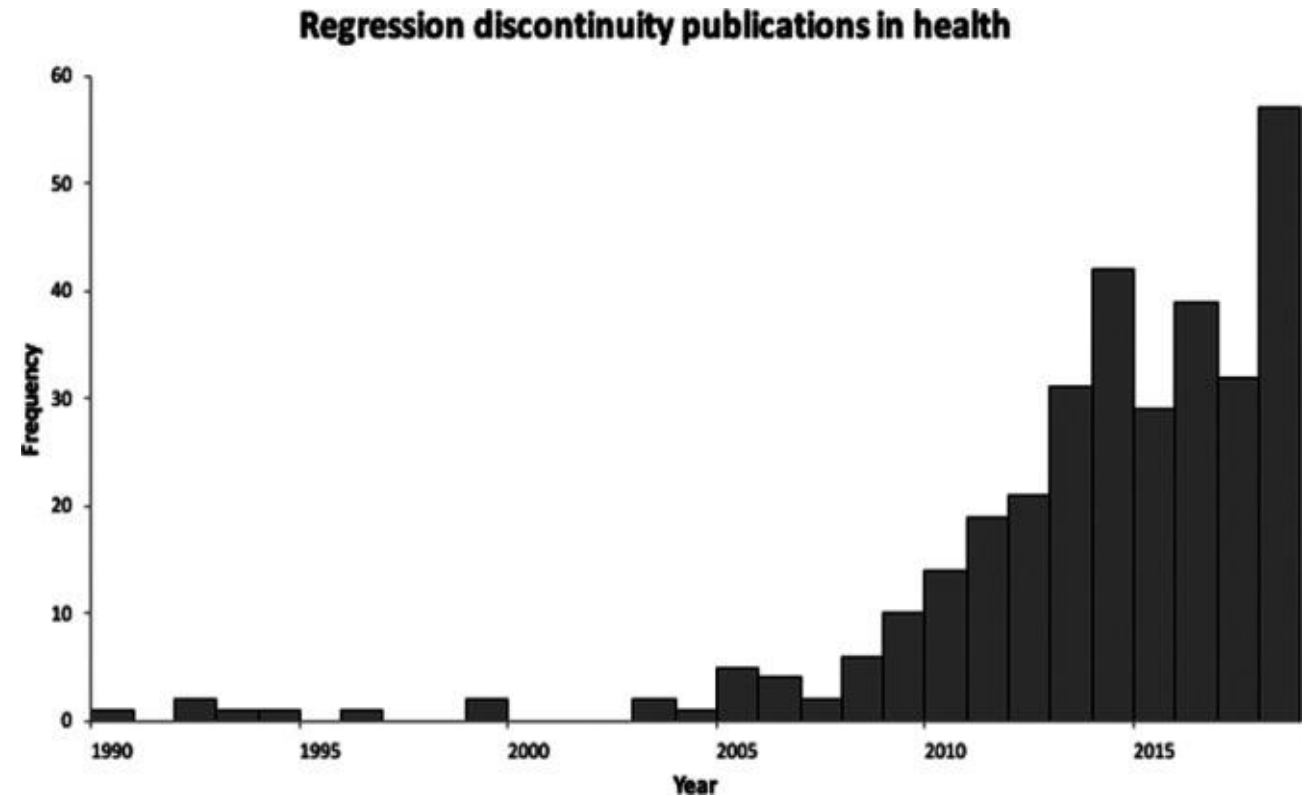
Alter (Schulbeginn, Rentenalter, Volljährigkeit)

Datum (Start oder Ende eines Gesetzes)

Sozioökonomische Faktoren (Einkommen, Wahlausgang, Klassengröße, Schulnoten)

Klinische Faktoren (Geburtsgewicht, Suchtgefahr, Stressindex)

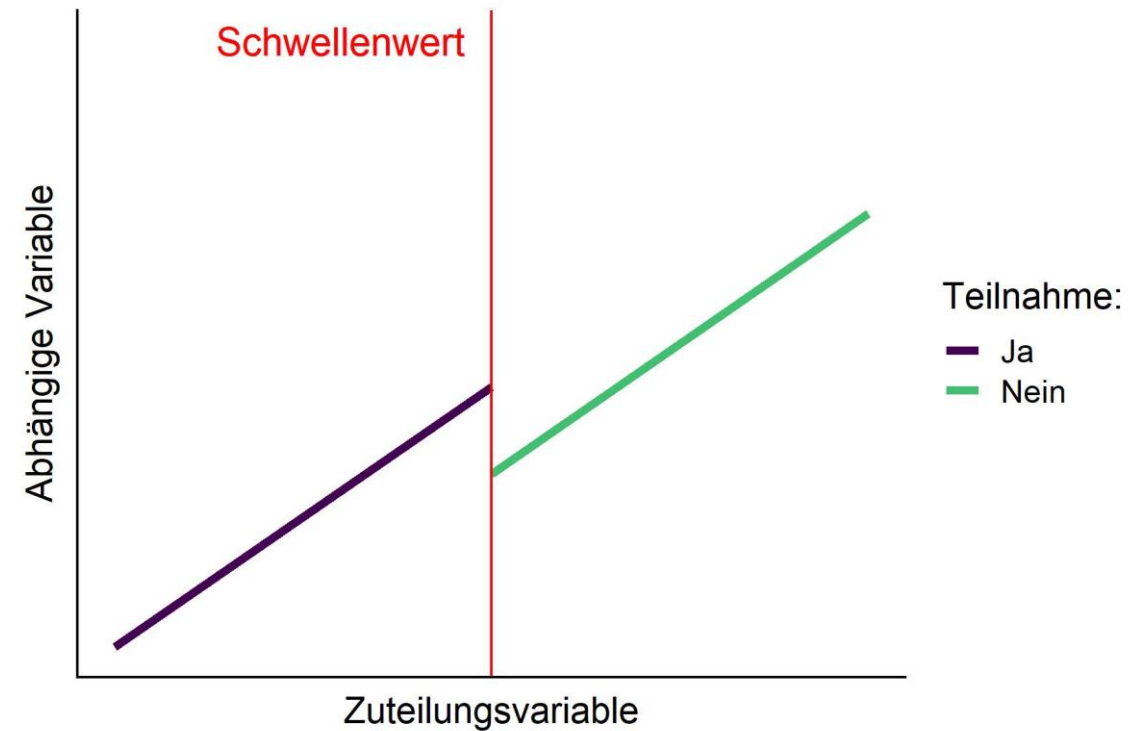
Geographische Faktoren (Distanz zur Grenze, Lengen- & Breitengrad)



(Boon et al., 2021)

Voraussetzung:

- Zuteilungsvariable ist kontinuierlich
- Schwellwert ist klar bestimmt
- Keine Selbstzuweisung
- Keine „Confounder“ (z.B. keine Kovariaten-Balance)



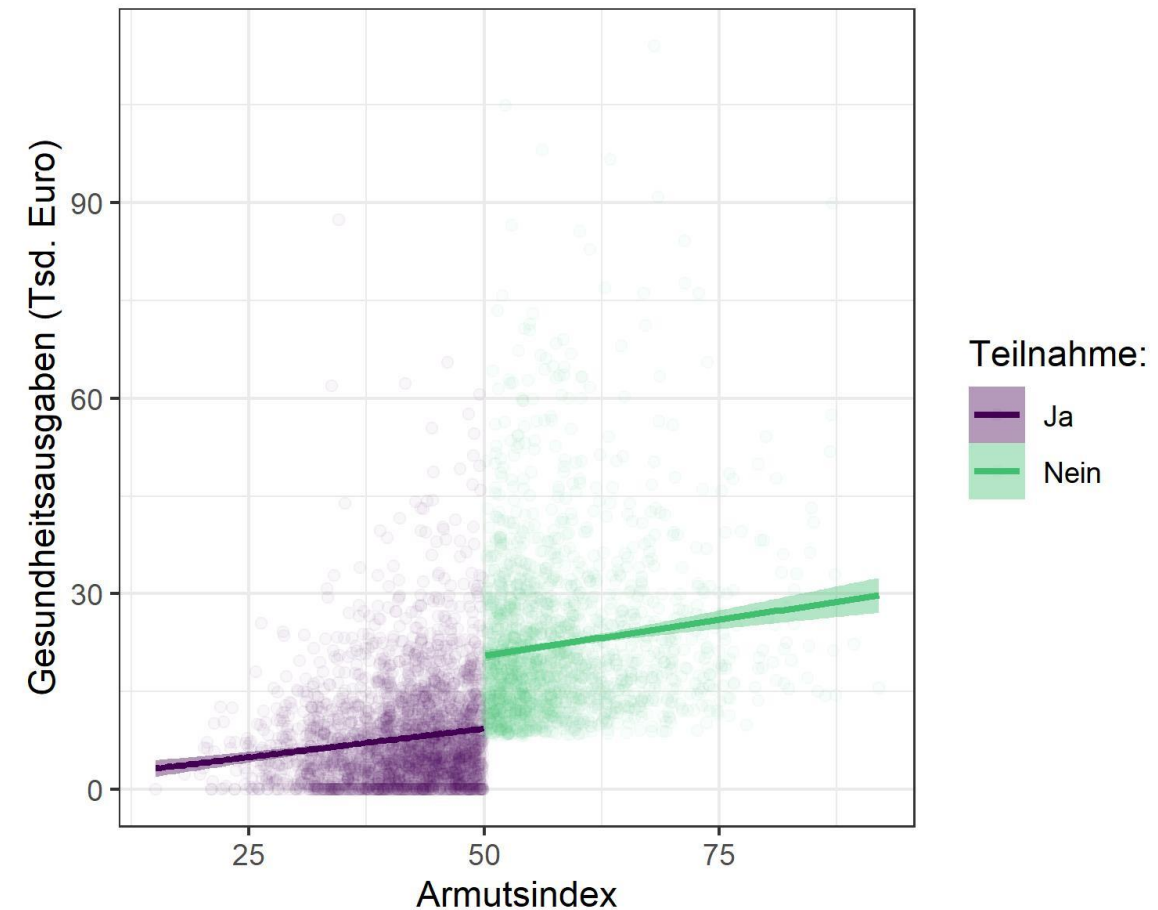
Fiktives Beispiel

Gesundheitsberatung für Einkommensschwache Haushalte

Frage: Reduziert die Gesundheitsberatung die Gesundheitsausgaben?



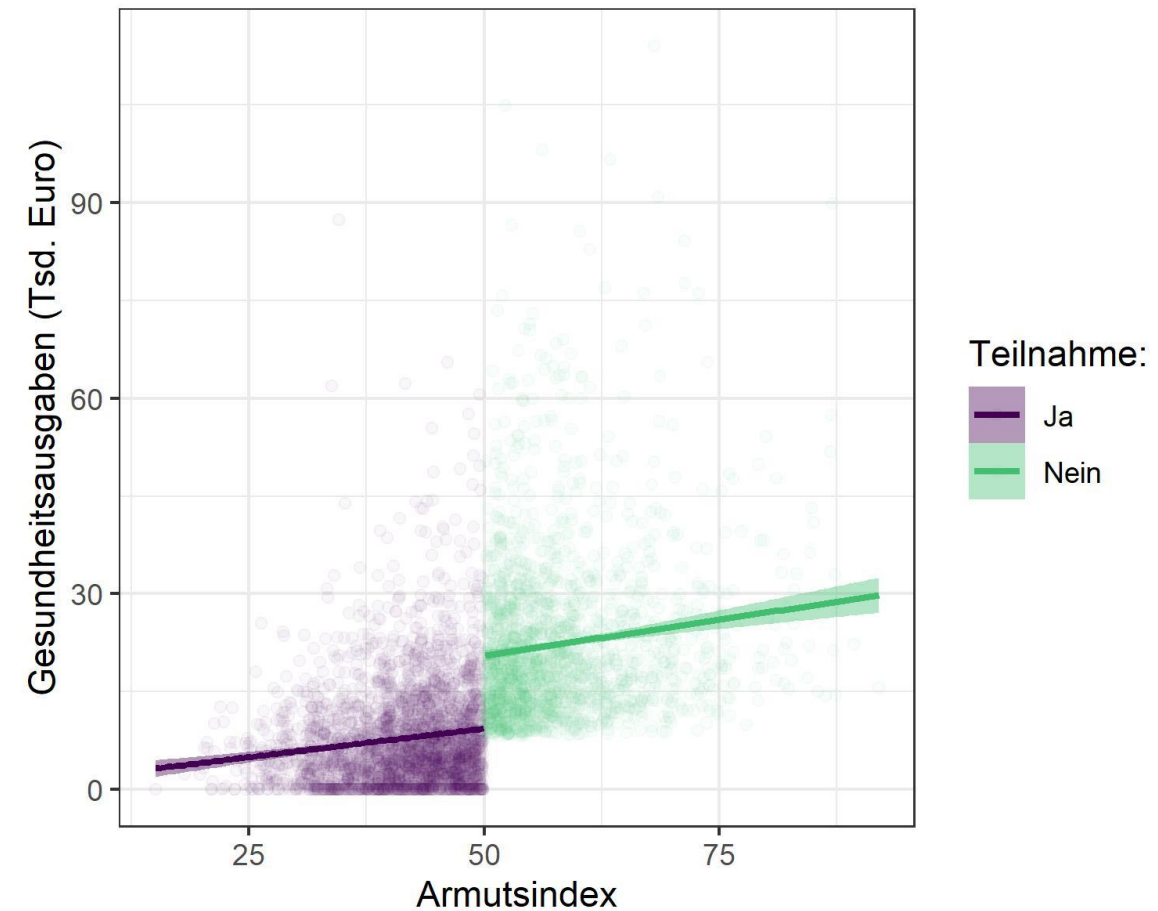
Code und Daten



Qualitätsanalyse:

Voraussetzung:

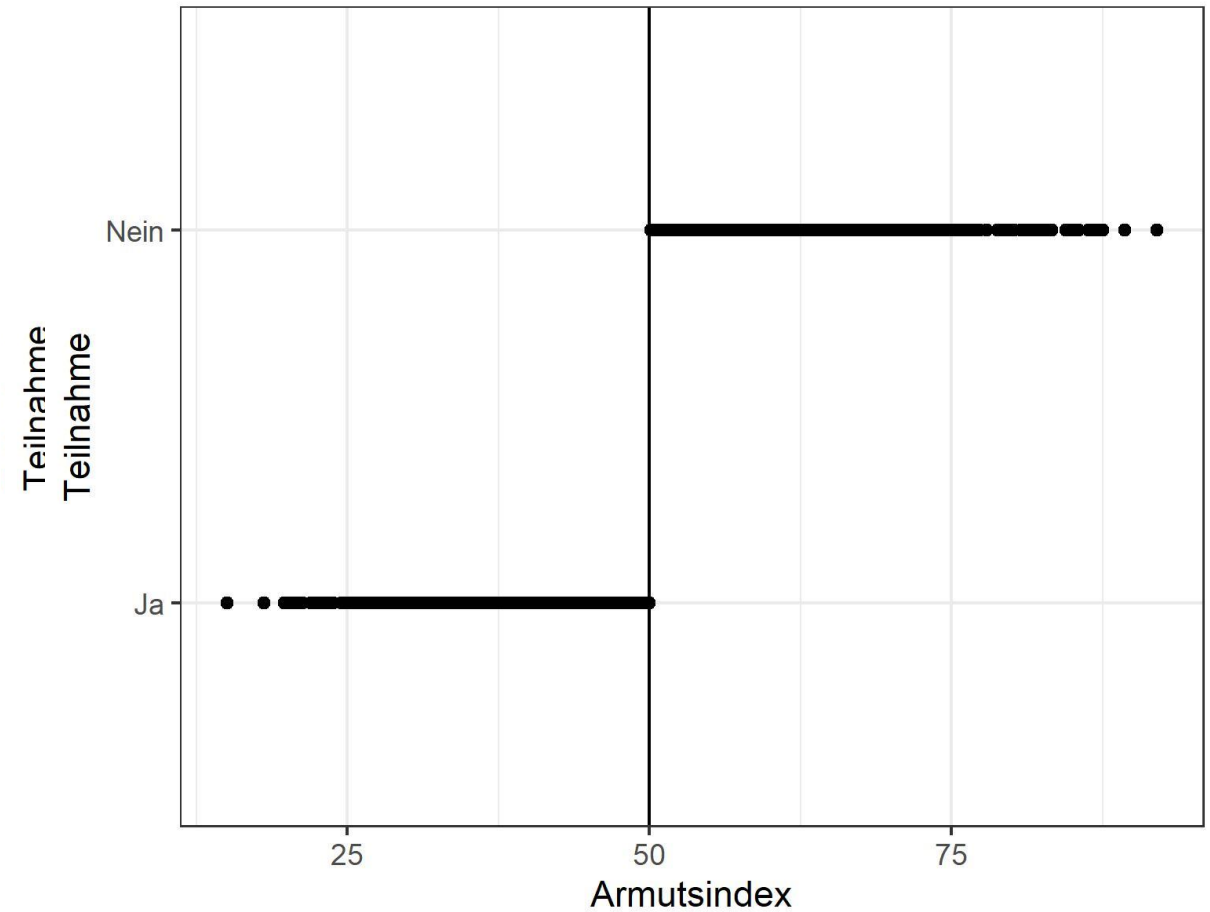
- Zuteilungsvariable ist kontinuierlich ✓



Qualitätsanalyse:

Voraussetzung:

- Zuteilungsvariable ist kontinuierlich ✓
- Schwellwert ist klar bestimmt ✓



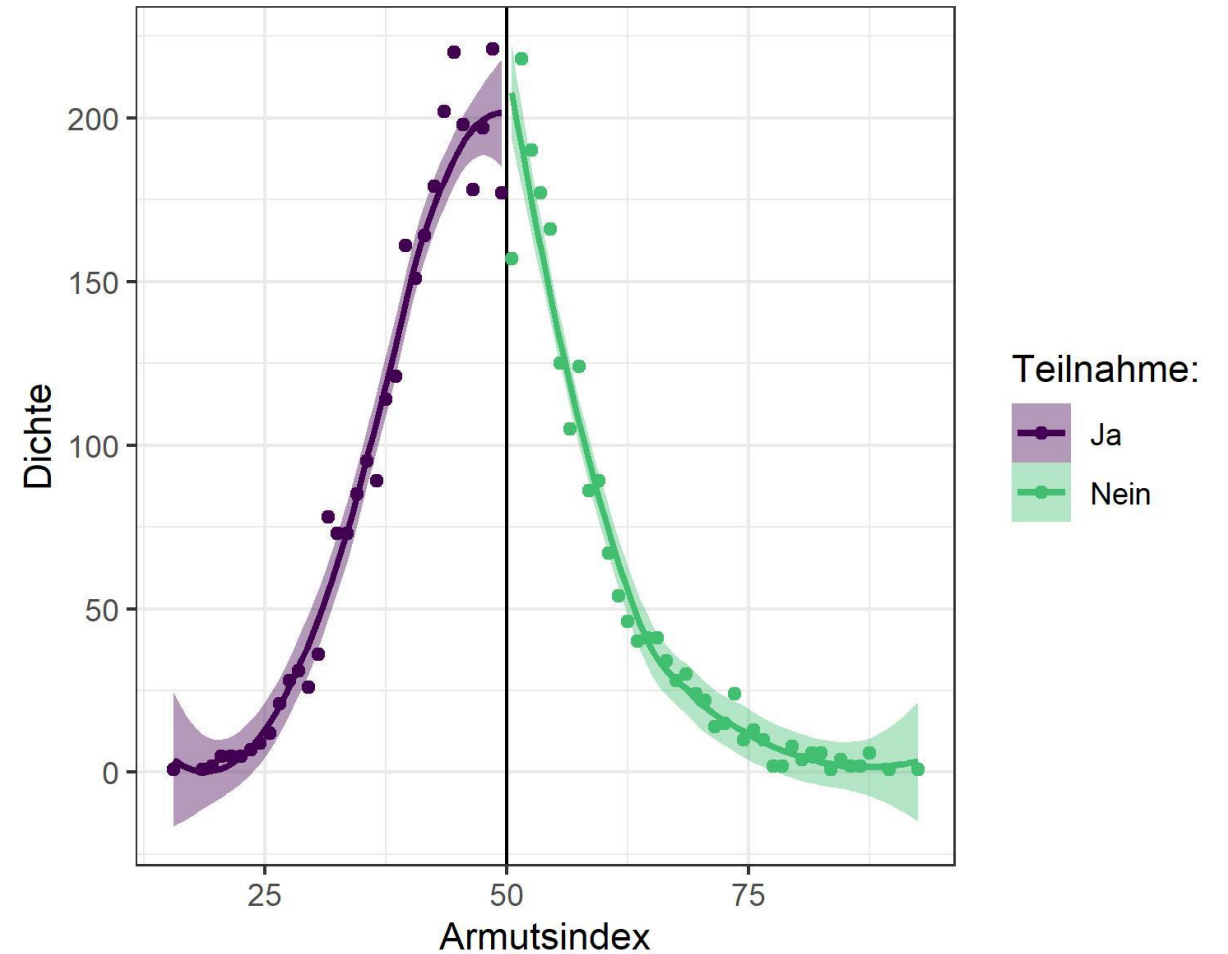
Qualitätsanalyse:

Voraussetzung:

- Zuteilungsvariable ist kontinuierlich ✓
- Schwellwert ist klar bestimmt ✓
- Keine Selbstzuweisung ✓

Siehe auch:

<https://cran.r-project.org/web/packages/rddensity/rddensity.pdf>

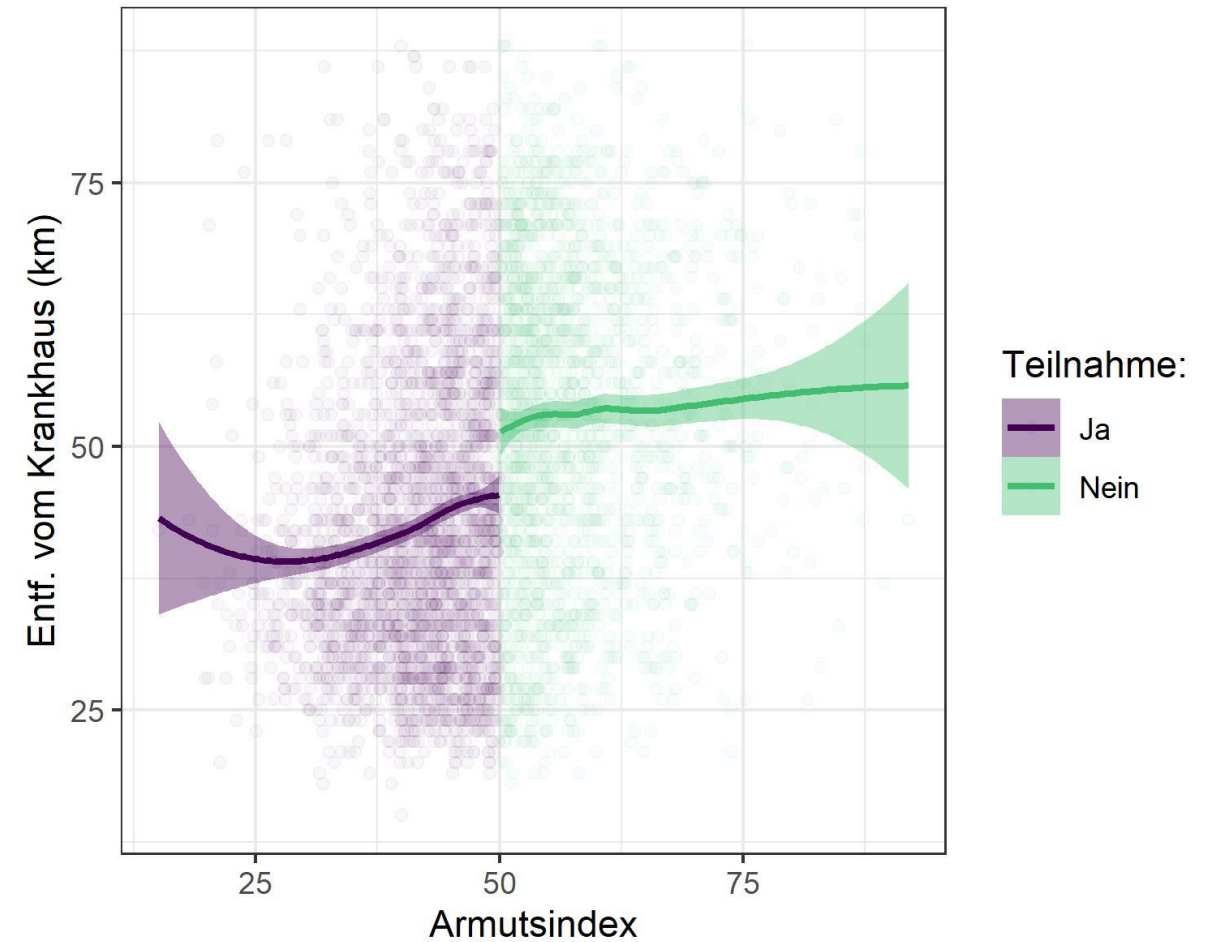


Qualitätsanalyse:

Voraussetzung:

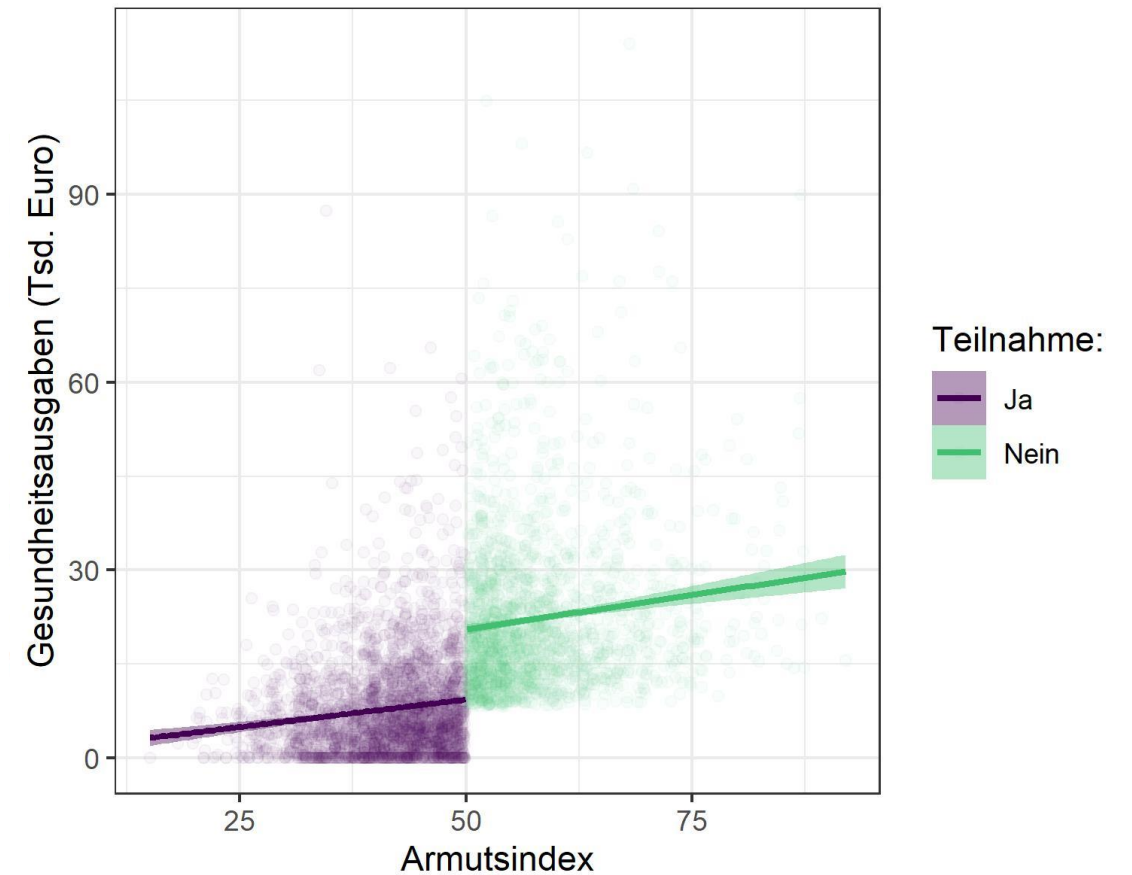
- Zuteilungsvariable ist kontinuierlich ✓
- Schwellwert ist klar bestimmt ✓
- Keine Selbstzuweisung ✓
- Keine „Confounder“

-> Kontinuitätstest



Regressionsmodell

Ziel: Denn Effekt von der Intervention am Schwellenwert schätzen.



Regressionsmodell

$$y_i = \beta_0 + \beta_1(Z_i - c) + \beta_2T_i + \varepsilon_i$$

Abhängige variable = y_i

Unabhängige Zuteilungsvariable = Z_i

Teilnahme = T_i

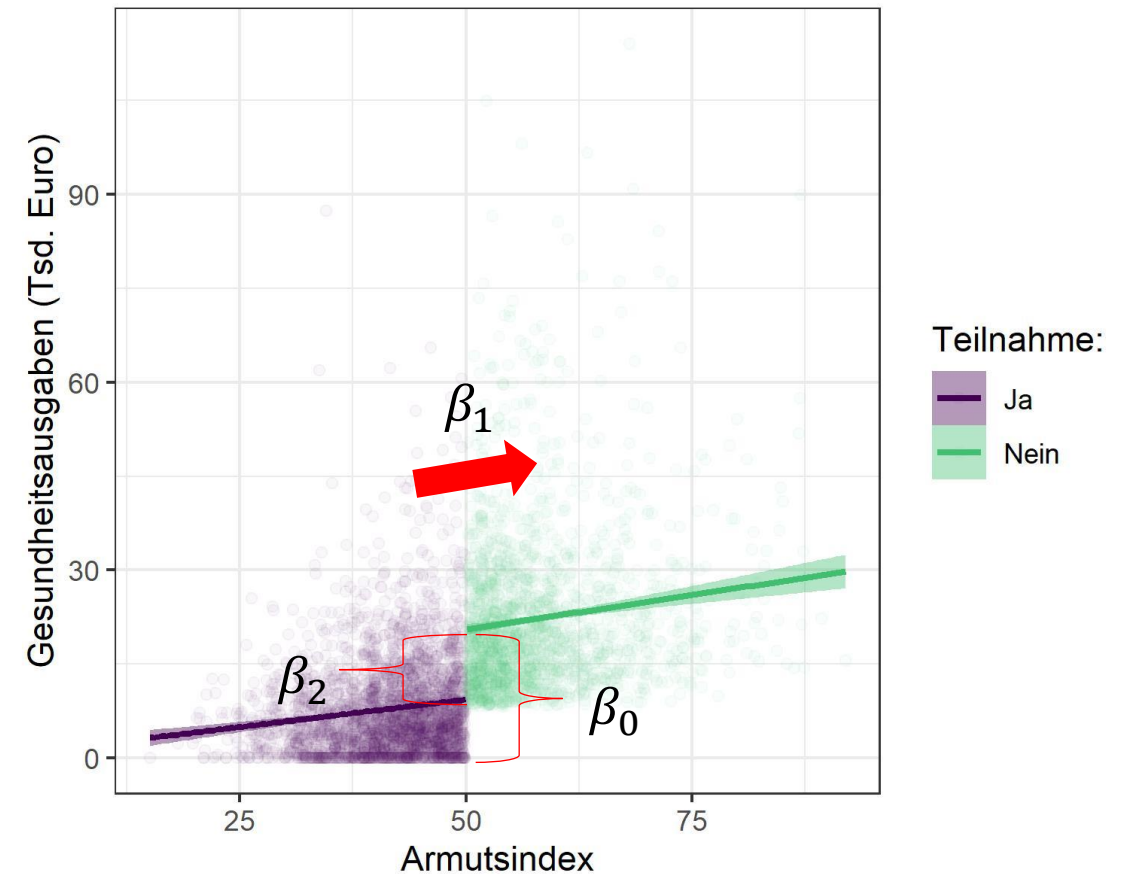
Schwellenwert = c

Intercept (y-Achsenabschnitt) = β_0

Steigung (slope) = β_1

Behandlungseffekt = β_2

Statistische Streuung = ε_i

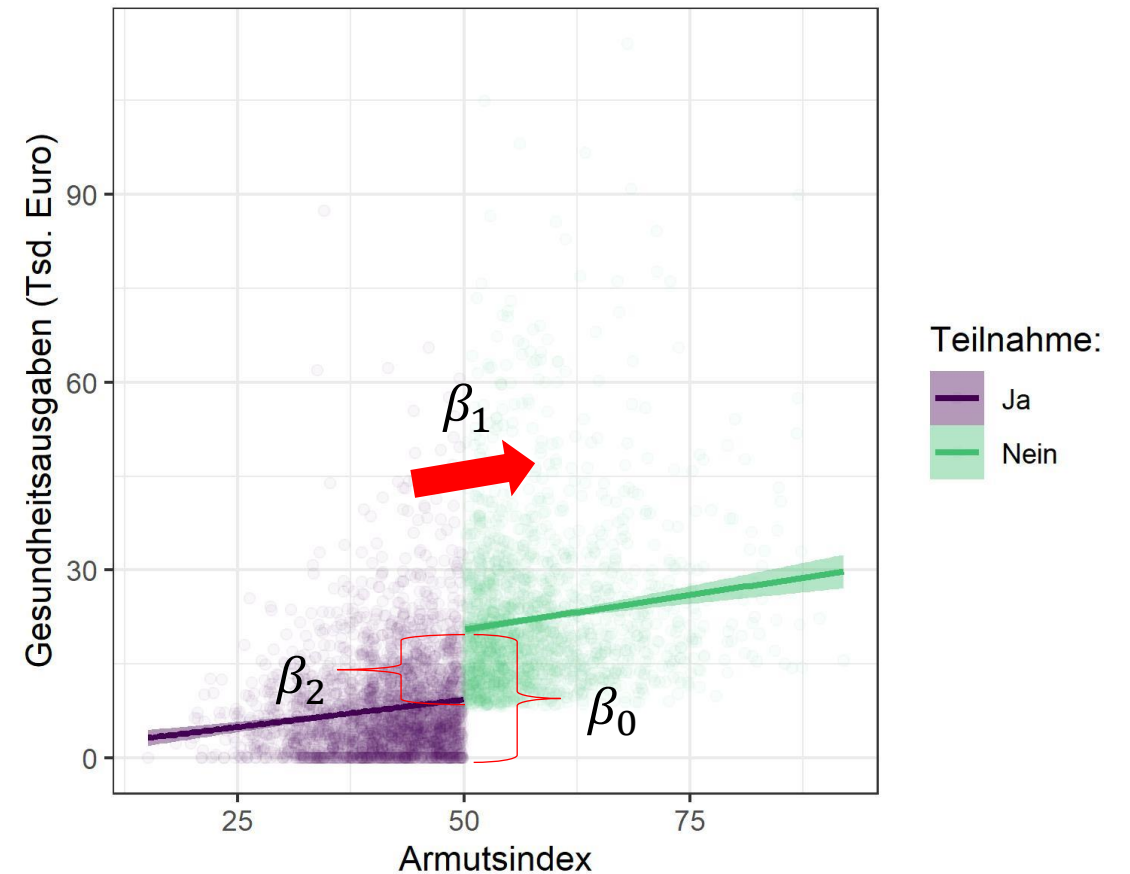


Regressionsmodel

$$y_i = \beta_0 + \beta_1(Z_i - c) + \beta_2T_i + \varepsilon_i$$

```
summary(
  lm(
    gesundh_ausgaben ~ I(armuts_index - 50) + teilnahme,
    data = df_eval))
```

```
##
## Call:
## lm(formula = gesundh_ausgaben ~ I(armuts_index - 50) + teilnahme,
##     data = df_eval)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.352  -6.400  -2.294   3.412   89.726
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    20.73949    0.28502  72.765  <2e-16 ***
## I(armuts_index - 50)  0.19700    0.02185   9.016  <2e-16 ***
## teilnahme      -11.19090    0.46618 -24.005  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.1 on 4957 degrees of freedom
## Multiple R-squared:  0.3379, Adjusted R-squared:  0.3376
## F-statistic: 1265 on 2 and 4957 DF, p-value: < 2.2e-16
```

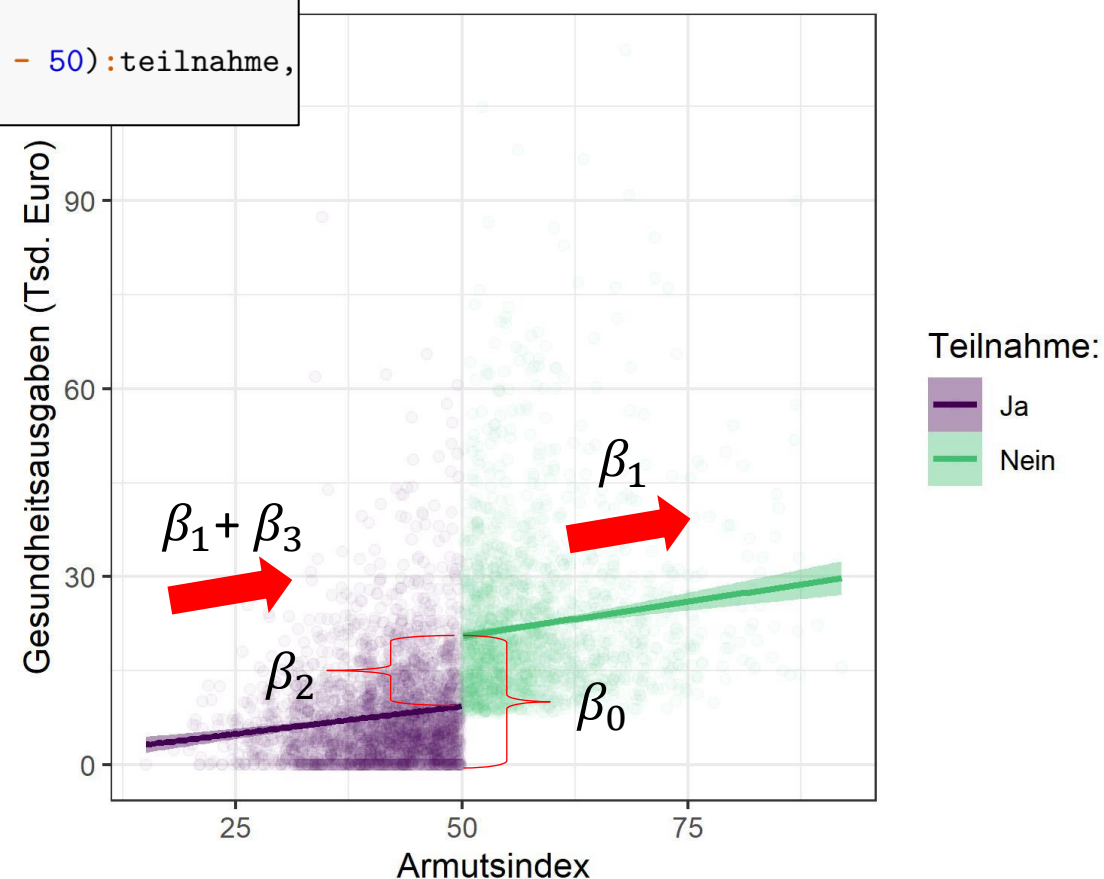


Regressionsmodel mit unterschiedlicher Steigung

$$y_i = \beta_0 + \beta_1(Z_i - c) + \beta_2T_i + \beta_3T_i(Z_i - c) + \varepsilon_i$$

```
summary(
  lm(
    gesundh_ausgaben ~ I(armuts_index - 50) + teilnahme + I(armuts_index - 50):teilmahme,
    data = df_eval))
```

```
## Call:
## lm(formula = gesundh_ausgaben ~ I(armuts_index - 50) + teilnahme +
##     I(armuts_index - 50):teilmahme, data = df_eval)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -16.816  -6.449  -2.293   3.440  89.489
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      20.55449    0.33763  60.878 < 2e-16 ***
## I(armuts_index - 50)    0.22028    0.03156   6.979 3.37e-12 ***
## teilnahme         -11.19171    0.46618 -24.007 < 2e-16 ***
## I(armuts_index - 50):teilmahme -0.04471    0.04374  -1.022  0.307
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.1 on 4956 degrees of freedom
## Multiple R-squared:  0.338, Adjusted R-squared:  0.3376
## F-statistic: 843.5 on 3 and 4956 DF, p-value: < 2.2e-16
```



Unscharfe RD-Analyse

Was ist wenn die Grenze nicht scharf trennt?

- Personen die nicht Teilnehmen obwohl sie können oder andersrum.

Sehr häufiges Problem

Schätzung der Diskontinuität beschreibt nur die Wirkung des Überschreitens des Schwellenwerts -> „Intention-to-treat“

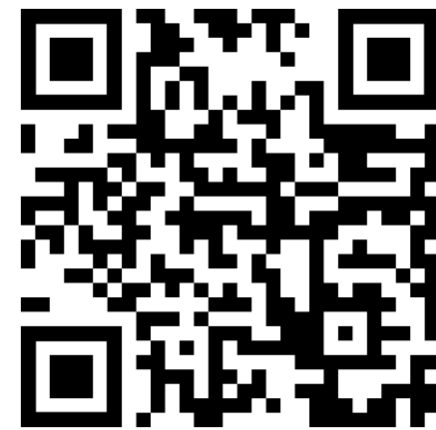
Lösung:

Analyse mit Instrumentvariablen schätzt die Wirkung für Individuen, deren Teilnahme durch die Trennvariable beeinflusst wurde. (siehe: Imbens & Limieux.2008)

Zusammenfassung

RDA erlaubt kausale Inferenz (ohne Randomisierung) wenn die Zuteilung abhängig von einem Schwellenwert ist.

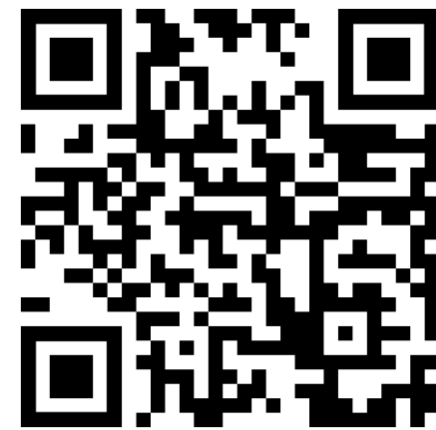
- Vorteil:
 - Einsetzbar wenn Randomisierung aus ethischen oder praktischen Gründen nicht durchführbar sind.
 - Kann in nicht experimentellen Kontexten benutzt werden
 - Praxisnah weil mit relevanter Experimentalgruppe
- Nachteil:
 - Weniger Robust als randomisierte Kontrollstudien da stärkere Annahmen gemacht werden müssen
 - Beschreibt nur Wirkung am Schwellenwert



Code & Daten

Lernquiz

- Entwickeln Sie ein hypothetisches Forschungsdesign für eine RDA in einem Bereich Ihrer Wahl (z.B. Bildung, Gesundheit, Politik).
- Diskutieren sie die Vor- und Nachteile einer RDA in diesem Forschungsdesign.
- Malen sie entweder in R oder mit einem Stift wie der erwartete Zusammenhang zwischen Zuteilungsvariable und abhängige Variable aussehen könnte. Diskutieren sie die Annahmen die sie dafür machen mussten.
- Beschreiben Sie, wie Sie sicherstellen würden, dass die Voraussetzungen für eine RDA in Ihrem Forschungsdesign erfüllt sind.
 - Ist die Zuteilungsvariable geeignet für eine RDA?
 - Überlegen sie erstmal ob der Versuchsaufbau Anlass gibt zu bezweifeln das alle Annahmen zutreffen. Überlegen sie danach wie sie die Annahmen testen würden.



Code & Daten



Code & Daten

Weiterführende Literatur

- Lee & Lemieux (2010). Regression Discontinuity Designs in Economics. *Journal of economic literature*, 48(2), 281-355.
- Boon, M. H., Craig, P., Thomson, H., Campbell, M., & Moore, L. (2021). Regression discontinuity designs in health: a systematic review. *Epidemiology (Cambridge, Mass.)*, 32(1), 87.
- Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2), 615-635.
- Stigler, M. (2014). *RDDtools: an R package for Regression Discontinuity Design*. GitHub. <https://github.com/MatthieuStigler/RDDtools>

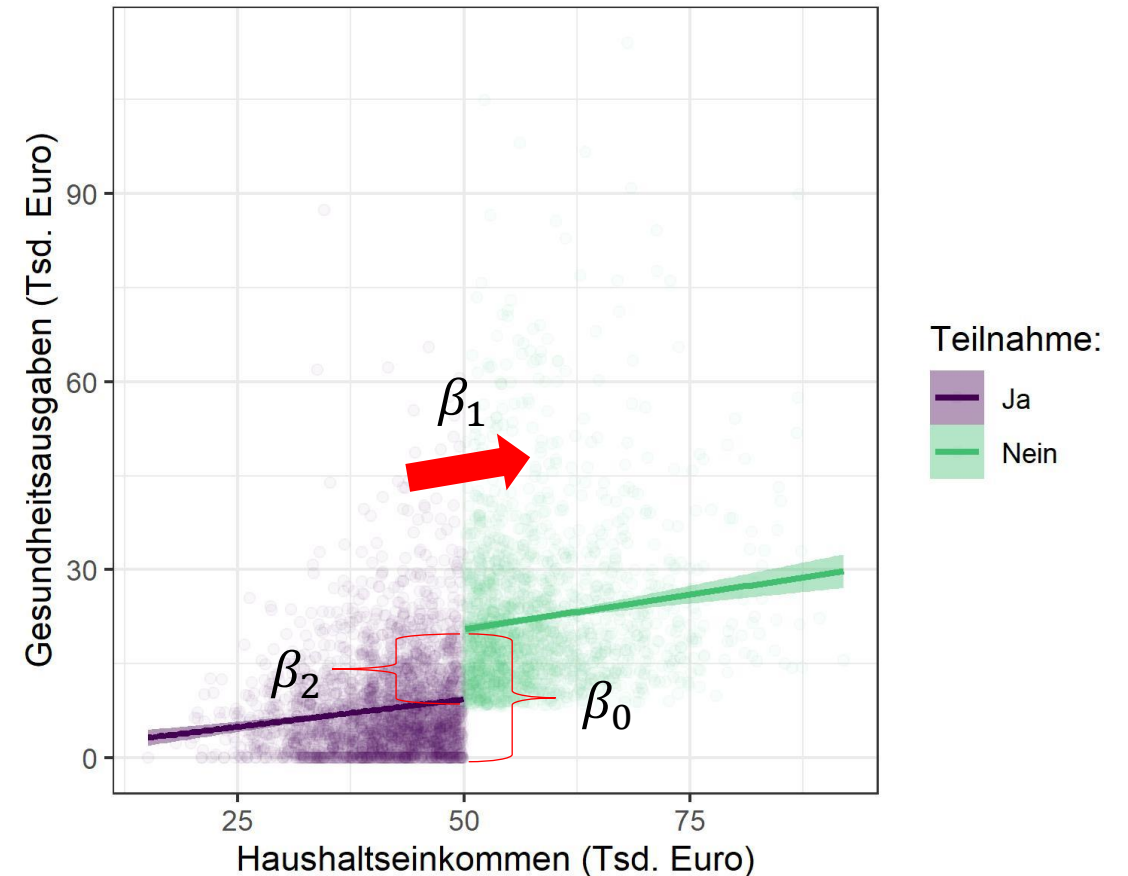
Regressionsmodell (bayesianisch)

$$y_i = \beta_0 + \beta_1(Z_i - c) + \beta_2T_i + \varepsilon_i$$

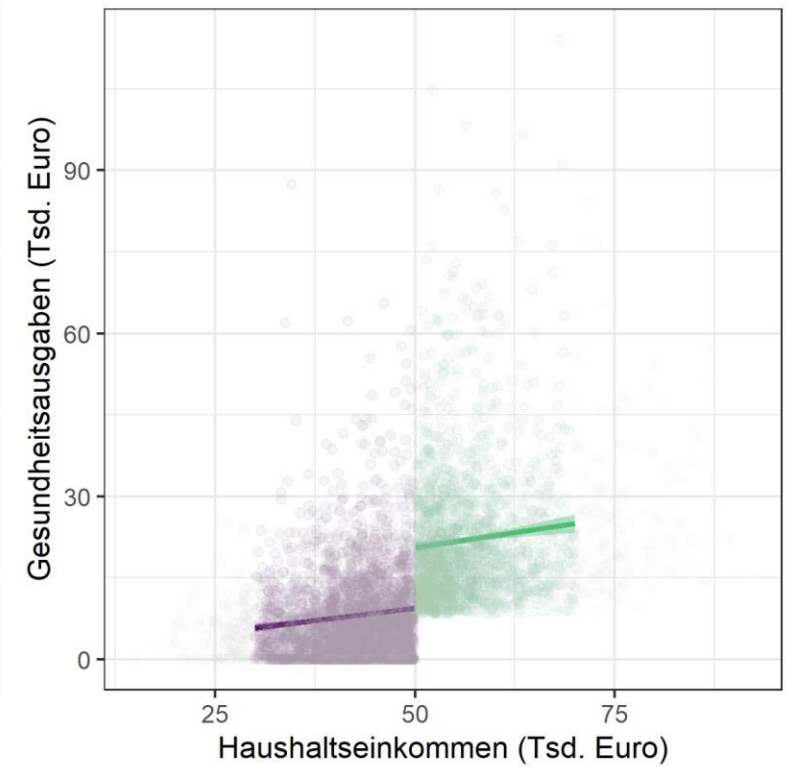
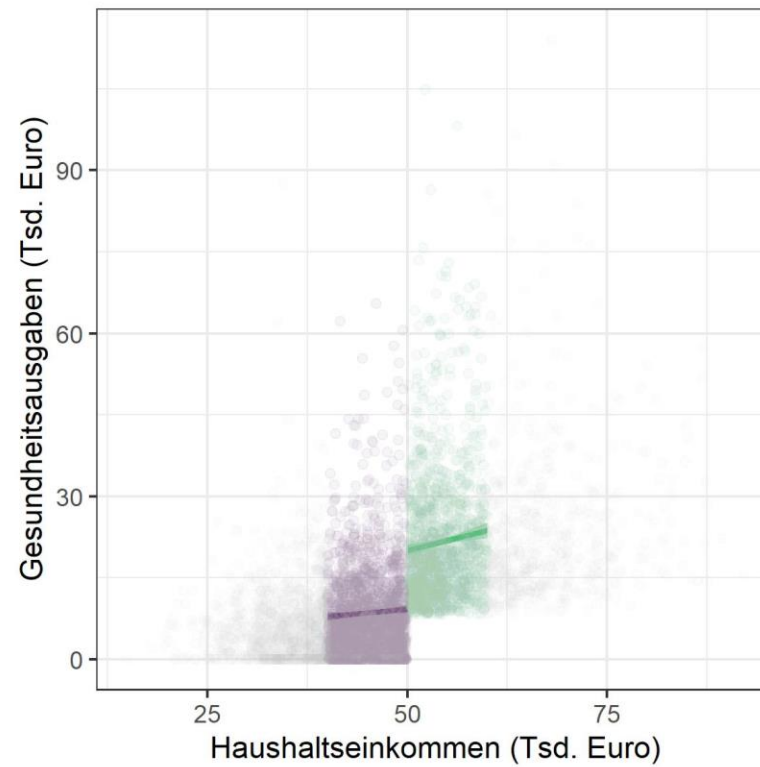
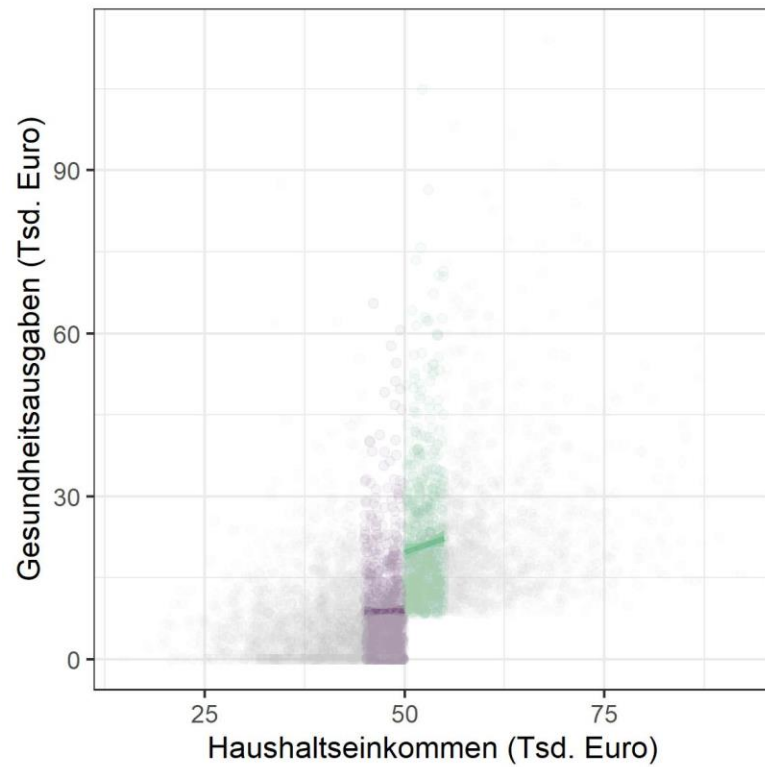
```
library(brms)
```

```
brm( gesundh_ausgaben ~ I(einkommen - 50) + teilnahme,
     data = df_eval)
```

```
## Family: gaussian
## Links: mu = identity; sigma = identity
## Formula: gesundh_ausgaben ~ I(einkommen - 50) + teilnahme
## Data: df_eval (Number of observations: 4960)
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
## total post-warmup draws = 4000
##
## Population-Level Effects:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## Intercept          20.76     0.29   20.21   21.32 1.00   2685   2306
## IeinkommenM50         0.20     0.02    0.15    0.24 1.00   2359   2437
## teilnahme          -11.22     0.47  -12.18  -10.30 1.00   2283   2103
##
## Family Specific Parameters:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma          10.10     0.10    9.90   10.30 1.00   3824   2559
##
## Draws were sampled using sampling(NUTS). For each parameter, Bulk_ESS
## and Tail_ESS are effective sample size measures, and Rhat is the potential
## scale reduction factor on split chains (at convergence, Rhat = 1).
```



Bandbreite

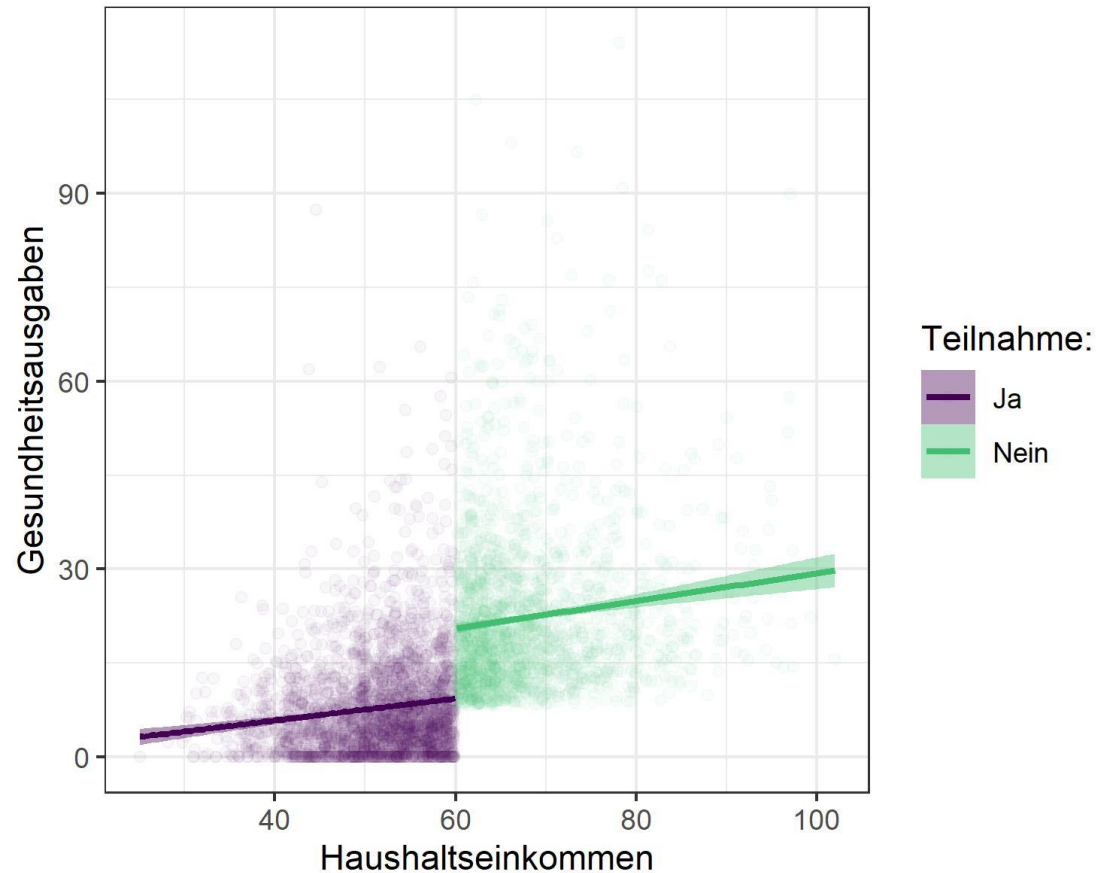


Out

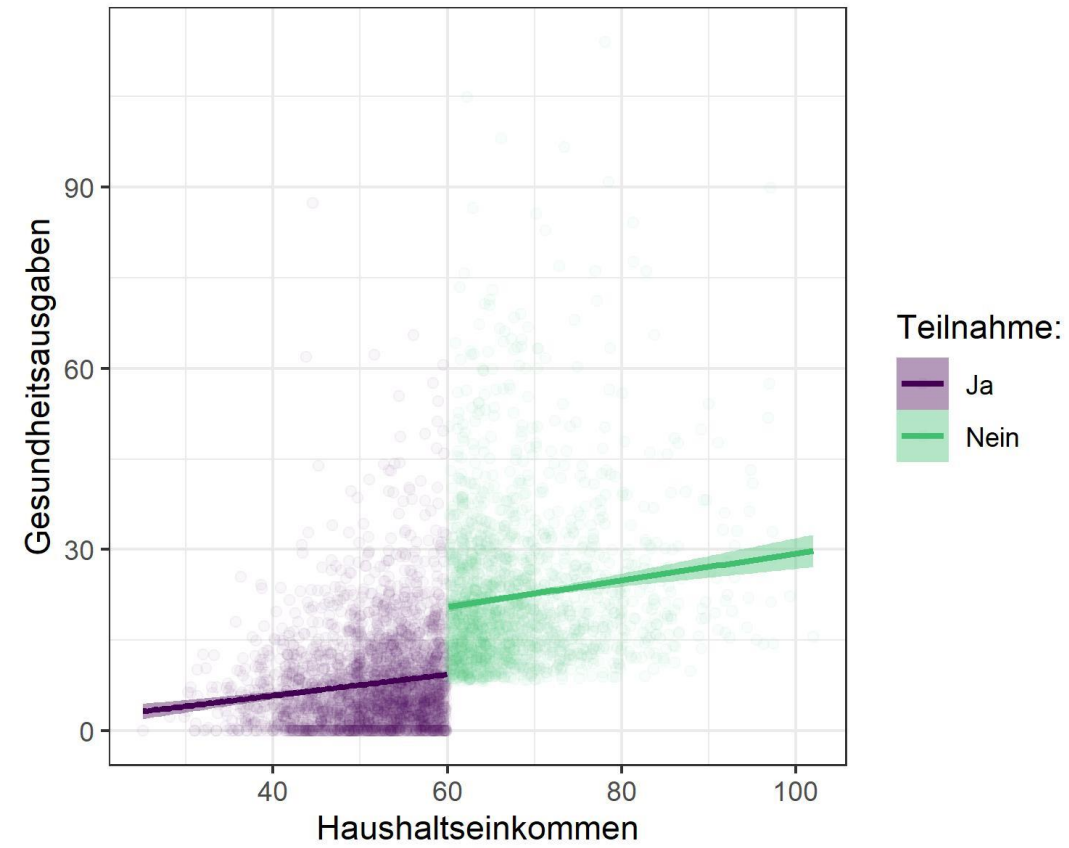
Qualitäts analyse:

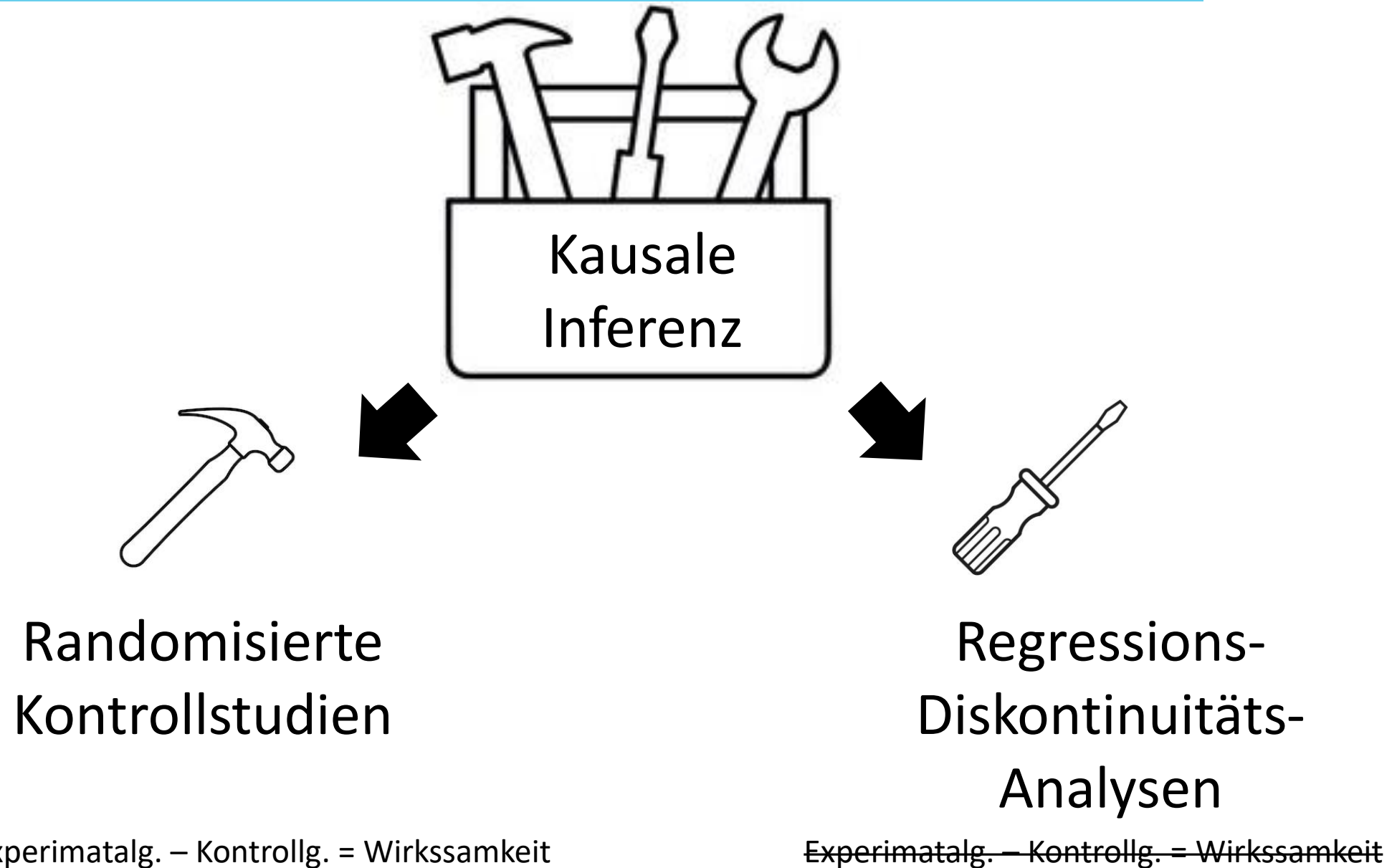
Voraussetzung:

- Teilungsvariable ist kontinuierlich
- Grenze ist klar bestimmt
- Keine Selbstzuweisung
- Keine „Confounder“ (z.B. andere Interventionen mit ähnlicher Grenze)



Qualitäts analyse:

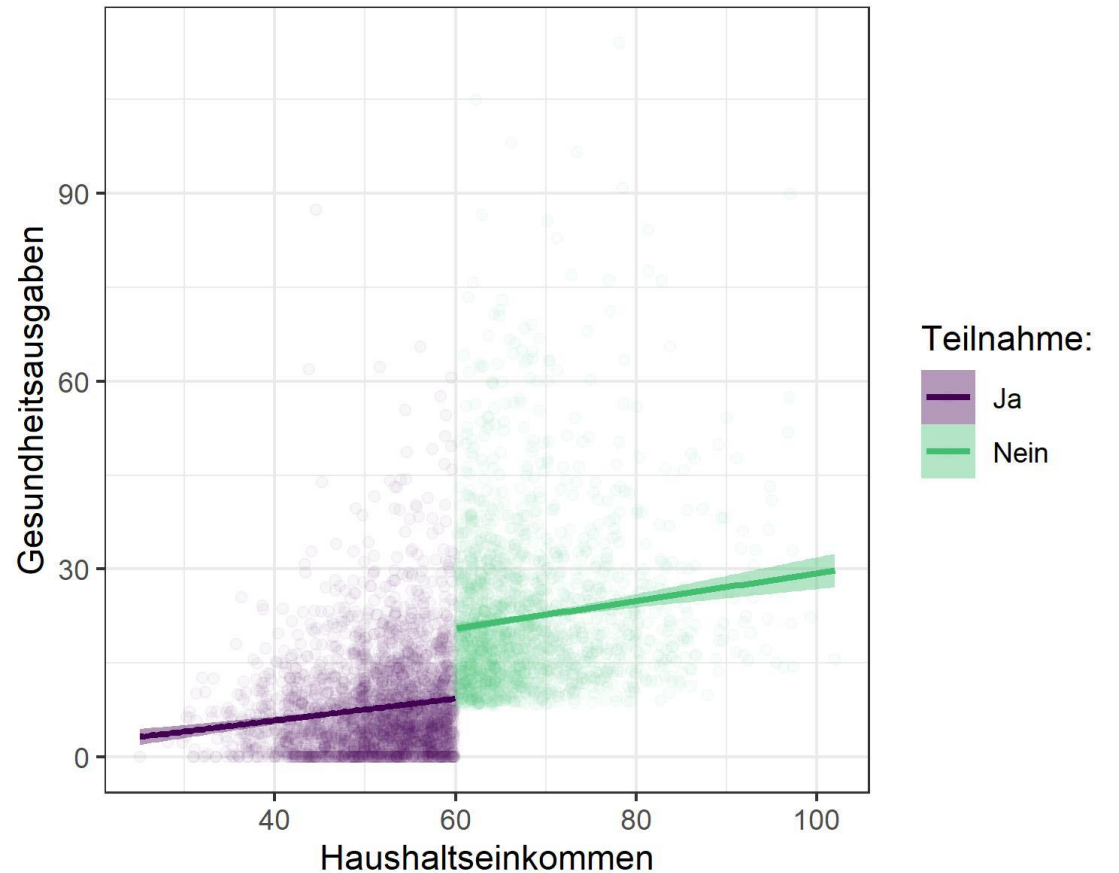




Qualitätsanalyse:

Voraussetzung:

- Teilungsvariable ist kontinuierlich
- Grenze ist klar bestimmt
- Keine Selbstzuweisung
- Keine „Confounder“ (z.B. weitere Zuteilungsvariable oder andere Interventionen mit ähnlichem Schwellwert)



```
#For data wrangling  
library(dplyr)  
  
#For plotting  
library(ggplot2)  
library(cowplot)  
  
#Loading the data:  
df_eval <- read.csv( "evaluation.csv")
```

- Vortest?? Erlaubt eine Schätzung der Zuteilsvariable auf gemessenen Werten ohne Intervention
- Test auf veränderte Slopes??
- Key assumptions: threshold, z is continuous. Continuity (without intervention the red line would not jump)

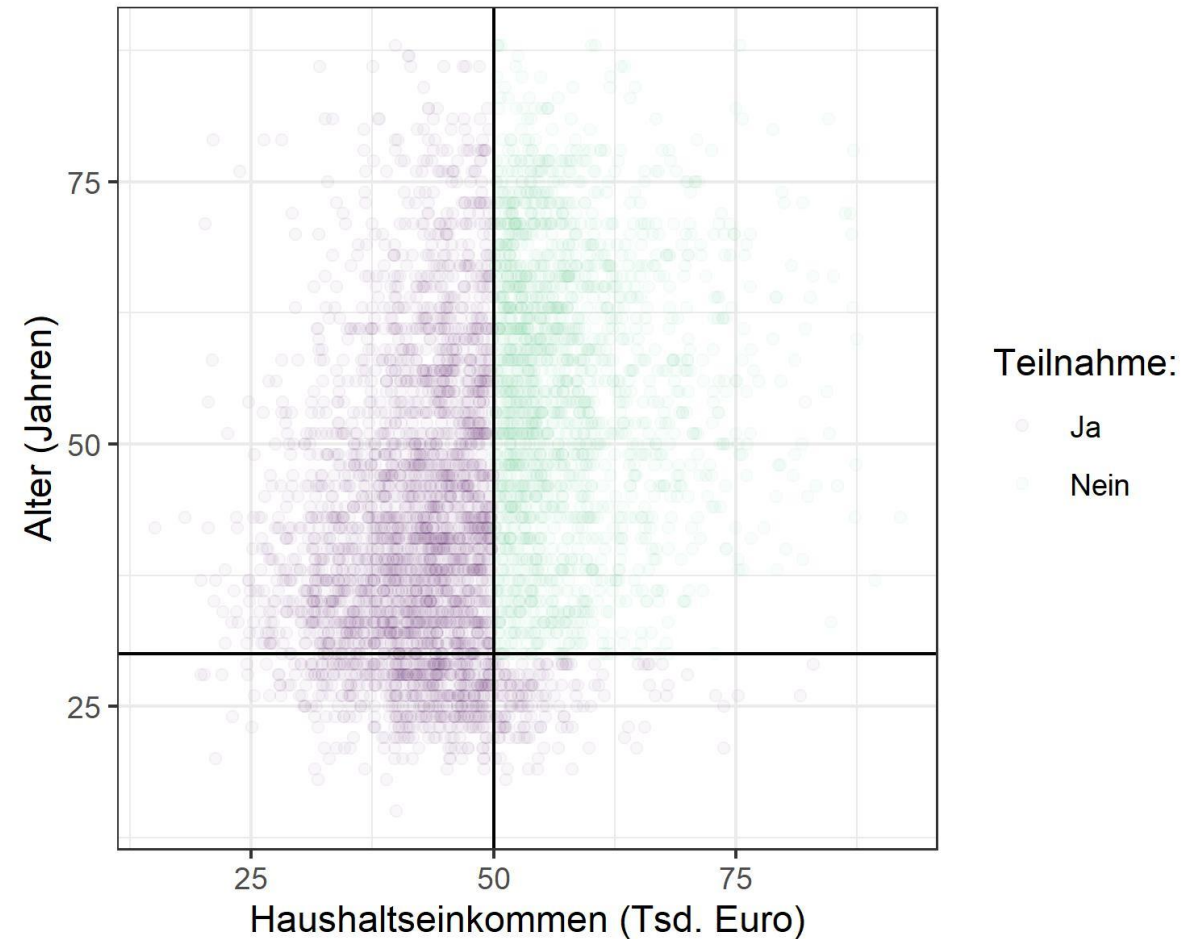
- Überblick über RDA
- Fiktives Fallbeispiel
- Qualitätsanalyse
- Res

Qualitätsanalyse:

Voraussetzung:

- Teilungsvariable ist kontinuierlich ✓
- Schwellwert ist klar bestir ✓
- Keine Selbstzuweisung ✓
- Keine „Confounder“

-> Problem: zwei Zuteilungsvariablen



Qualitätsanalyse:

Voraussetzung:

- Teilungsvariable ist kontinuierlich
- Schwellwert ist klar bestimmt
- Keine Selbstzuweisung
- Keine „Confounder“ (z.B. andere Interventionen mit ähnlichem Schwellwert)

