

Hogeschool  van Arnhem en Nijmegen
 HAN University of Applied Sciences

Hello R!

An Introduction to R



Eghe Rice Osagie
 (eghe.osagie@han.nl)


22 nov. 2018

HAN

www.han.nl/masterscourses

MASTERS PROGRAMMES

1

Hogeschool  van Arnhem en Nijmegen
 HAN University of Applied Sciences

Who R we?

Eghe Osagie
 •Assistant lector (professor) at HAN University of AS
 •Lecturer Bachelor HRM, Master HRM, Master CE
 •Coordinator Minor HR Analytics
 •**Interests:** HR Analytics, Sustainability, HRM, Research methodology


Witek ten Hove
 •Instructor at HAN University of AS
 •Coordinator of MSL
 •**Interests:** Business Economics, Data Engineering, Data Mining, AI, Web Dev.

HAN

MASTERS PROGRAMMES

MASTERS PROGRAMMES

2

Hogeschool  van Arnhem en Nijmegen
 HAN University of Applied Sciences

Programma

1. Intro R
2. Practicum
3. Confirmatory factor analysis

HAN

MASTERS PROGRAMMES

MASTERS PROGRAMMES

3



Link naar alle docs:

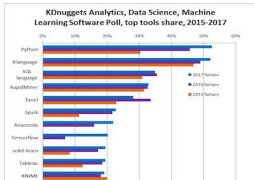

4



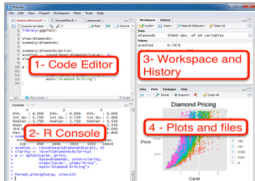
Intro R


5

R - software



- Ranking second as tool for data science (after Python)
- Upcoming tool in Social sciences!
- Rule of Thumb: Play with the R program before you work on anything professional and know your data !!







6

Characteristics R

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

- Created in: 1995 by Ross Ihaka & Robert Gentleman at the University of Auckland
- Free
- Computer language
- Windows, Mac, Linux
- and object oriented

- Extending software via 'packages'
- Each package is maintained and supported by the author, but not warranted (!)
- CRAN checks report any potential notes, warnings, and errors associated with a package
- Numerous Output options








Who can read this?

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

Command 1:

```
install.packages("threejs")
library(threejs)
data(ego)
graphjs(ego, bg="black")
```




Command 2:

```
HS.model <- ' Visual  == x1 + x2 + x3
Textual == x4 + x5 + x6
Speed == x7 + x8 + x9 '
```

Analysing with R

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

1. Install R
2. Install R-studio - of rstudio.cloud
3. Set working directory
4. Save workspace
5. Install packages and load them
6. Read tutorial
7. Amend commands



Exemplary Packages

Package	description
LAVAAN	Latent Variable Analysis (SEM,CFA)
AcousticNDLCoder	Coding Sound Files for Use with NDL
abd	The Analysis of Biological Data
RQDA	R-Based Qualitative Data Analysis
RSmartlyIQ	Loading Facebook and Instagram Advertising Data from 'Smartly.io'
qdap	Bridging the Gap Between Qualitative Data and Quantitative Analysis
gha	Qualitative Harmonic Analysis
quanteda	Quantitative Analysis of Textual Data

See for more packages:
https://cran.r-project.org/web/packages/available_packages_by_name.html

MAATWERK PROGRAMMEREN

10

Exemplary output

Copy paste command in R

```
install.packages("leaflet");
library(leaflet);

m <- leaflet() %>% addTiles() %>% #
Add default OpenStreetMap map tiles
addMarkers(lng=174.768, lat=-36.852,
label= "The birthplace of R",
labelOptions = labelOptions(noHide =
T));

m # Print the map
```

MAATWERK PROGRAMMEREN

11

Amending commands

Replace red.....

```
install.packages("leaflet");
library(leaflet);

m <- leaflet() %>% addTiles() %>% #
Add default OpenStreetMap map tiles
addMarkers(lng=174.768, lat=-36.852,
label= "The birthplace of R",
labelOptions = labelOptions(noHide =
T));

m # Print the map
```

....with green.

```
install.packages("leaflet");
library(leaflet);

m <- leaflet() %>%
addTiles() %>% # Add default
OpenStreetMap map tiles
addMarkers(lng= 5.949481,
lat=51.989683, label= "An introduction
to R", labelOptions =
labelOptions(noHide = T));

m # Print the map.
```

LET OP: Google maps toont eerst "Ling" en dan "L&P", dus net andersom invoeren.

MAATWERK PROGRAMMEREN

12

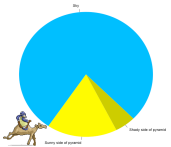
Exemplary output

Copy paste command in R


```
pie(c(a=78, b=17, c=5), init.angle = 315, col = c("deepskyblue", "yellow", "yellow3"), border = FALSE, radius = 1.0)
```

Copy paste command in R

```
install.packages("threejs");
library(threejs);
data(ego);
graphs(ego, bg="red")
```




More examples:
<https://github.com/witusi/hellor/blob/master/hellor.Rmd>



MASTERS PROGRAMMES

13

Practicum R



MASTERS PROGRAMMES

14

Practicum


First some initial information

Go to: [_witusi.github.io/WorkshopSI/](https://witusi.github.io/WorkshopSI/)

Perform the following exercises:


- Voorbereiding
- Basis R

Remaining exercises can be performed at home



MASTERS PROGRAMMES

15



Hello R!


This presentation can be found online:

witusi.github.io/hello/hello.html
press F for fullscreen


For the Workshop R (Dutch) go to:

witusi.github.io/WorkshopR/

Workshop documents can be found here (docs folder): <https://github.com/witusi/hello>




16



Who can read this?


Command 1:

- install.packages("threejs")
- library(threejs)
- data(ego)
- graphjs(ego, bg="black")



Command 2:

```
HS.model <- ' Visual  =~ x1 + x2 + x3
Textual =~ x4 + x5 + x6
Speed =~ x7 + x8 + x9 '
```



17



Confirmatory Factor Analysis (Dutch)



18

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

CFA

Doel confirmatory factor analysis:
bevestiging krijgen voor van te voren bepaald model/structuur

HAN

MASTERS PROGRAMME

19

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

CFA model:

Kenmerken:

- **NIET** elke manifeste variabele een lading op elke factor
- **WEL** relatie tussen de componenten
- **WEL** meetfouten

HAN

MASTERS PROGRAMME

20

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

CFA model:

De asterixen verwijzen naar de te schatten parameters

Parameters = die delen van het model die nog onbekend zijn voor de onderzoeker, en dus berekend moeten worden

Hier:

- meetfouten,
- factorladingen,
- correlaties tussen factoren,
- variantie van factoren,
- ...

HAN

MASTERS PROGRAMME

21




Belangrijke bergippen in CFA

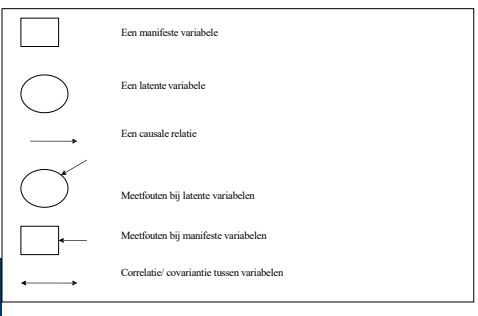
- **Rondje** = niet direct gemeten (latente var. [f])
- **Vierkant** = direct gemeten (manifeste var./indicator/item [x])
- **ind.** = indicator[x]
- **→** = impact van 1 variabele/factor op een andere variabele/factor
- **↔** = covariantie of correlatie tussen variabelen/factoren.
- **Meetmodel** = relatie tussen latente variabelen en indicatoren
- **Structuurmodel** = relaties tussen latente variabelen
- **EXO** = Exogene construct/factor (pijlje exit)
- **e** = meetfout




22




Notatie voor tekenen van modellen

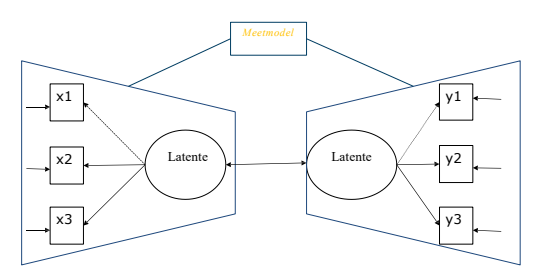





23



Voorbeeld model CFA





24

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

Belangrijke Commands LAVAAN

Formule type	Operator	Betekenis
• Definitie van latente variabele	$\equiv \sim$	Is measured by/ Is gemeten door
• regressie	\sim	Is regressed on
• (residu) (co)variantie	$\sim \sim$	Is correlated with/ gecorreleerd met
• intercept	~ 1	intercept
	f	Latente variabele
	y	Afhankelijke var
	x	Onafhank. Var/observed variable/indicator
	$cfa()$	Voer een CFA analyse uit. Met help("cfa"), krijg je uitleg over de functie
	$sem()$	Voer een SEM analyse uit. Met help("sem"), krijg je uitleg over de functie
	$Growth()$	Voer een Growth curve analyse uit. Met help("growth"), krijg je uitleg over de functie

25

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

Voorbeeld met LAVAAN in R

- Bepaal model(len)**
 $Visual \sim x1 + x2 + x3$
 $Textual \sim x4 + x5 + x6$
 $Speed \sim x7 + x8 + x9$
 Wat staat hier: Latent variable \sim indicator1 + indicator2 + indicator3
- Specificeer model(len) in R**
 $HS.model <- 'Visual \sim x1 + x2 + x3$
 $Textual \sim x4 + x5 + x6$
 $Speed \sim x7 + x8 + x9'$
- Fit model(len) in R**
 $fit1 <- cfa(HS.model, data = HolzingerSwineford1939)$
 Bijv.: $fitM1 <- cfa(HS.model, data = HolzingerSwineford1939)$
- Lees Fit indices afvergelijk ze**
 $summary(..., fit.measures = TRUE)$
 Bijv.: $summary(fit1, fit.measures = TRUE)$
 In het geval van niet normal verdeelde data en $n=200+$ (estimator MLM test = SB):
 $fitM1 <- cfa(HS.model, data = HolzingerSwineford1939, estimator = "MLM", test = "satorra.bentler")$
 Mocht je specifieke fit indices willen opvragen:
 $fitMeasures(fitM1, c("df.scaled", "rmsea", "gfi", "agfi", "nnfi", "chisq", "chisq.scaled", "df.scaled"))$
 Voor de modificatie indices:
 $modificationIndices(fitM1)$
 Rood = aanpassen aan eigen variabelen/items/namen

26

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

Voorbeeld met LAVAAN in R

- Bepaal model(len)**
 $Visual \sim x1 + x2 + x3$
 $Textual \sim x4 + x5 + x6$
 $Speed \sim x7 + x8 + x9$
 Wat staat hier: Latent variable \sim indicator1 + indicator2 + indicator3
- Specificeer model(len) in R**
 $HS.model <- 'Visual \sim x1 + x2 + x3$
 $Textual \sim x4 + x5 + x6$
 $Speed \sim x7 + x8 + x9'$
- Fit model(len) in R**
 $fit1 <- cfa(..., data = ...)$
 Bijv.: $fitM1 <- cfa(HS.model, data = HolzingerSwineford1939)$
- Lees Fit indices afvergelijk ze**
 $summary(..., fit.measures = TRUE)$
 Bijv.: $summary(fit1, fit.measures = TRUE)$
- Bepaal beste model nadat je meerdere modellen hebt gefit**
 $anova(fit1, fit3)$
 Kijk naar AIC waarde....Jagere AIC of chi kwadaat is beter model
 Rood = aanpassen aan eigen variabelen/items/namen

27

Hogeschool van Arnhem en Nijmegen
 HAN University of Applied Sciences

Fit indices

Fit indices	Thresholds (cut-offs)
• Relative Chi square (Chi-square/df; cmin/df)	< 2 ^a of < 3 = good ^b (soms is < 5 ook toegelaten ^c)
• p value of the model	> .05
• RMSEA	< .05=good; .05-.10=moderate; > .10=bad ^b
• CFI	> .95=great; > .90 traditional; > .80 sommige gevallen toelaatbaar ^bstreven > .93 ^d
• GFI	> .90 ^d ...liefst > .95 ^b
• (N)NFI	> .90 ^d ...of > .95 ^c
• AGFI	> .80 ^b

HAN

a = Ullman(2001), b = Hu & Bentler (1999), c = Schumacker & Lomax (2004), d = Byrne (1994)

MAATWERK PROGRAMMEREN
28

Hogeschool van Arnhem en Nijmegen
 HAN University of Applied Sciences

Modification indices (MI) & Standardized residuals covar (SRC)

Aanpassen model : doe je bij geen goede fit. Theoretische onderbouwing belangrijk!!

- **Theorie**
- **MI**
 - Error van verschillende constructen mogen niet correleren
 - Error mag niet correleren met latente of observeerd constructen
- **SRC**
 - Error van verschillende constructen mogen niet correleren
 - Error mag niet correleren met latente of observeerd constructen

HAN

MAATWERK PROGRAMMEREN

29

Hogeschool van Arnhem en Nijmegen
 HAN University of Applied Sciences

MI rules

Acceptable

Inacceptable

Inacceptable

Inacceptable

HAN

MAATWERK PROGRAMMEREN

30




CFA samengevat

- *CFA om model-fit te schatten*: past model bij de data?
=> fit indices: χ^2 , GFI, AGFI, NNFI, CFI, RMSEA
- *CFA om modellen onderling te vergelijken*: kijk naar AIC waarde, lagere waarde dan past model beter bij data
- *En hoe het model interpreteren?* => interpreteren van parameterschattingen




31




Practicum

- **CFA** → open Tutorial LAVAAN, perform excises on p. 4-8


Remaining exercises can be performed at home



32



Questions?



33

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

Confirmatory Factor Analysis (Dutch)

HAN
MASTERS PROGRAMMES

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

CFA

Doel confirmatory factor analysis:
bevestiging krijgen voor van te voren bepaald
model/structuur

HAN
MASTERS PROGRAMMES

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

CFA model:

kenmerken:
NIET elke manifeste
variabele een lading op elke
factor
WEL relatie tussen de
componenten
WEL meetfouten

HAN
MASTERS PROGRAMMES

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

CFA model:

De asterixen verwijzen naar de te schatten parameters

Parameters = die delen van het model die nog onbekend zijn voor de onderzoeker, en dus berekend moeten worden

Hier:

- meetfouten,
- factorladingen,
- correlaties tussen factoren,
- variantie van factoren,
- ...

HAN
MASTER'S PROGRAMMES

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

Belangrijke bergippen in CFA

- Rondje = niet direct gemeten (latente var. [f])
- Vierkant = direct gemeten (manifeste var./indicator/item [x])
- ind. = indicator[x]
- → = impact van 1 variabele/factor op een andere variabele/factor
- ↔ = covariantie of correlatie tussen variabelen/factoren.
- Meetmodel = relatie tussen latente variabelen en indicatoren
- Structuurmodel = relaties tussen latente variabelen
- EXO = Exogene construct/factor (pijlje exit)
- e = meetfout

HAN
MASTER'S PROGRAMMES

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

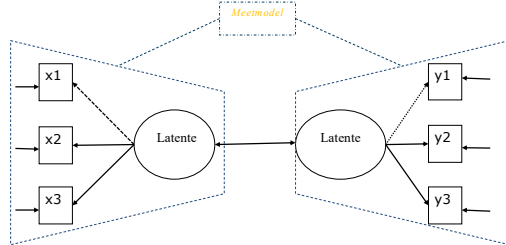
Notatie voor tekenen van modellen

- Een manifeste variabele
- Een latente variabele
- Een causale relatie
- Meetfouten bij latente variabelen
- Meetfouten bij manifeste variabelen
- Correlatie/ covariantie tussen variabelen

HAN
MASTER'S PROGRAMMES

Voorbeeld model CFA

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences



HAN

MASTERS PROGRAMMES

Belangrijke Commands LAVAAN

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

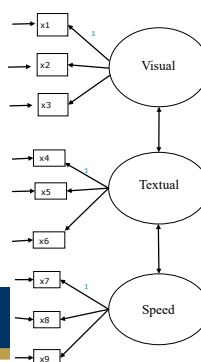
Formule type	Operator	Betekenis
• Definitie van latente variabele	\sim	Is measured by/ Is gemeten door
• regressie	\sim	Is regressed on
• (residu) (co)variantie	$\sim\sim$	Is correlated with/ gecorreleerd met
• intercept	~ 1	intercept
	f	Latente variabele
	y	Afhankelijke var
	x	Onafhank. Var/observed variable/indicator
	<code>cfa()</code>	Voer een CFA analyse uit. Met help("cfa"), krijg je uitleg over de functie
	<code>sem()</code>	Voer een SEM analyse uit. Met help("sem"), krijg je uitleg over de functie
	<code>Growth()</code>	Voer een Growth curve analyse uit. Met help("growth"), krijg je uitleg over de functie

HAN

MASTERS PROGRAMMES

Voorbeeld met LAVAAN in R

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences



1. Begaal model(len)
Visual \sim x1 + x2 + x3
Textual \sim x4 + x5 + x6
Speed \sim x7 + x8 + x9

Wat staat hier: Latent variable \sim indicator1 + indicator2 + indicator3

2. Specificeer model(len) in R
H5.model <- ' Visual \sim x1 + x2 + x3
Textual \sim x4 + x5 + x6
Speed \sim x7 + x8 + x9 '

3. Fit model(len) in R
...<- cfa (... , data = ...)

Rijv.: fitM1 <- cfa (H5.model, data = HobingerSwineford1999)

4. Lees fit indices af/vergelijk ze
summary (... , fit.measures = TRUE)

Rijv.: summary (fitM1, fit.measures = TRUE)

In het geval van niet normal verdeelde data en n>200+ (estimator MLM test = 58)

fitM1 <- cfa(H5.model, data = HobingerSwineford1999, estimator = "MLM", test = "satorra.bentler")

Mocht je specifieke fitmaken willen opvragen:

fitMeasures(fitM1, c("chi.scaled", "rmsea", "gfi", "agfi", "nfi", "chiq", "chiq.scaled", "df.scaled"))

Voor de modificatie indices:
modindices(fitM1)

Rood = aanpassen aan eigen variabelen/items/namen

HAN

MASTERS PROGRAMMES

Voorbeeld met LAVAAN in R

1. Bepaal model(len)
`Visual =~ x1 + x2 + x3`
`Textual =~ x4 + x5 + x6`
`Speed =~ x7 + x8 + x9`

Wat staat hier? Latent variable =~ indicator1 + indicator2 + indicator3

2. Specificeer model(len) in R
`H5.model <- 'Visual =~ x1 + x2 + x3`
`Textual =~ x4 + x5 + x6`
`Speed =~ x7 + x8 + x9'`

3. Fit model(len) in R
`fit1 <- cfa (H5.model, data = HolzingerSwineford1939)`

4. Lees fit indices af/vergelijk ze
`summary (fit1, fit.measures = TRUE)`

5. Bepaal beste model nadat je meerdere modellen hebt gefit
`anova(fit1, fit3)`

Kijk naar AIC waarde... lagere AIC of chi kwadaat is beter model

Rood = aanpassen aan eigen variabelen/items/namen

HAN
 MASTER'S PROGRAMMES

Fit indices

Fit indices	Thresholds (cut-offs)
• Relative Chi square (Chi-square/df; χ^2/df)	< 2 ^a of < 3 = good ^b (soms is < 5 ook toegelaten ^c)
• p value of the model	> .05
• RMSEA	< .05=good; .05-.10=moderate; > .10=bad ^b
• CFI	> .95=great; > .90 traditional; > .80 sommige gevallen toelaatbaar ^bstreven > .93 ^d
• GFI	> .90 ^d ... liefst > .95 ^b
• (N)NFI	> .90 ^d ... of > .95 ^c
• AGFI	> .80 ^b

a = Ullman(2001). b = Hu & Bentler (1999). c = Schumacker & Lomax (2004). d = Byrne (1994)

HAN
 MASTER'S PROGRAMMES

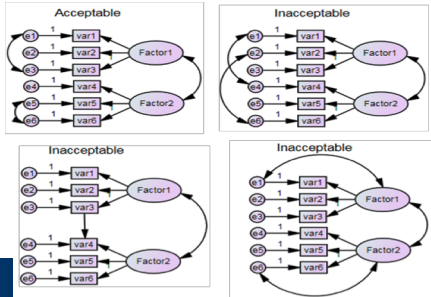
Modification indices (MI) & Standardized residuals covar (SRC)

Aanpassen model : doe je bij geen goede fit. Theoretische onderbouwing belangrijk!!

- **Theorie**
- **MI**
 - Error van verschillende constructen mogen niet correleren
 - Error mag niet correleren met latente of observerd constructen
- **SRC**
 - Error van verschillende constructen mogen niet correleren
 - Error mag niet correleren met latente of observerd constructen

HAN
 MASTER'S PROGRAMMES

MI rules



CFA samengevat

CFA om model-fit te schatten: past model bij de data? => fit indices: χ^2 , GFI, AGFI, NNFI, CFI, RMSEA

CFA om modellen onderling te vergelijken: kijk naar AIC waarde, lagere waarde dan past model beter bij data

En hoe het model interpreteren? => interpreteren van parameterschattingen

Practicum

- CFA → open Tutorial LAVAAN, perform excises on p. 4-8

Remaining exercises can be performed at home

Hogeschool van Arnhem en Nijmegen
HAN University of Applied Sciences

Questions?

HAN

MASTERS PROGRAMMES
