

Rosetta-for-OU

Vakgroep Methoden en Statistiek, Open Universiteit

2022-08-09

Contents

	10
1 Inhoudsopgave	11
I Basis informatie	13
2 Software	15
2.1 jamovi	15
2.2 R	19
2.3 SPSS	25
3 Datasets	27
3.1 pp15	27
3.2 Politiek_Media	29
3.3 statistiekangst	29
3.4 schoolsucces	29
3.5 sportcasus	30
3.6 contraProductiefGedrag	31
3.7 wiskunde	31
3.8 roken	32
4 Data laden	33
4.1 jamovi	33
4.2 R	33
4.3 SPSS	34

II	Analyses vooraf	37
5	Data screening	39
5.1	Intro	39
5.2	jamovi	39
5.3	R	41
5.4	SPSS	43
6	Ontbrekende waarden	47
6.1	Intro	47
6.2	jamovi	47
6.3	R	48
6.4	SPSS	48
7	Variabelen samenvoegen	49
7.1	Intro	49
7.2	jamovi	49
7.3	R	50
7.4	SPSS	51
8	Variabelen Standaardiseren	53
8.1	Intro	53
8.2	jamovi	53
8.3	R	54
8.4	SPSS	54
9	Hercoderen van variabelen	55
9.1	Intro	55
9.2	jamovi	56
9.3	R	57
9.4	SPSS	57

<i>CONTENTS</i>	5
10 Coefficient Alpha	59
10.1 Intro	59
10.2 jamovi	59
10.3 R	60
10.4 SPSS	60
III Analyses met twee variabelen	63
11 Kruistabellen	65
11.1 Intro	65
11.2 jamovi	65
11.3 R	66
11.4 SPSS	66
12 T-test	67
12.1 Intro	67
12.2 jamovi	67
12.3 R	68
12.4 SPSS	69
12.5 APA style report	69
13 Cohen's d	71
13.1 Intro	71
13.2 jamovi	71
13.3 R	72
13.4 SPSS	73
14 Correlaties	75
14.1 Intro	75
14.2 jamovi	75
14.3 R	76
14.4 SPSS	76

15 ANOVA	79
15.1 Intro	79
15.2 jamovi	80
15.3 R	81
15.4 SPSS	81
15.5 APA style report	82
16 Lineaire regressieanalyse	83
16.1 Intro	83
16.2 jamovi	83
16.3 R	84
16.4 SPSS	84
16.5 APA style report	84
17 Non-parametrische toetsen	87
17.1 Intro	87
17.2 jamovi	88
17.3 R	89
17.4 SPSS	90
IV Analyses met meer variabelen	91
18 Exploratieve Factoranalyse	93
18.1 Intro	93
18.2 jamovi	93
18.3 R	94
19 Factoriele ANOVA	99
19.1 Intro	99
19.2 jamovi	99
19.3 R	100
19.4 SPSS	101
19.5 APA style report [AANPASSEN]	101

<i>CONTENTS</i>	7
20 Covariantie analyse	103
20.1 Intro	103
20.2 jamovi	103
20.3 R	104
20.4 SPSS	105
20.5 APA style report [AANPASSEN]	105
21 Herhaalde metingen analyse	107
21.1 Intro	107
21.2 jamovi	107
21.3 R	108
21.4 SPSS	109
21.5 APA style report [AANPASSEN]	109
22 Multivariate Regressie	111
22.1 Intro	111
22.2 jamovi	111
22.3 R	113
22.4 SPSS	113
22.5 APA style report	113
23 Logistische Regressie	115
23.1 Intro	115
23.2 jamovi	115
23.3 R	117
23.4 SPSS	117
24 Moderatieanalyse	119
24.1 Intro	119
24.2 jamovi	120
24.3 R	122
24.4 SPSS	124

25 Moderatie-Mediatie Modellen	129
25.1 Intro	129
25.2 jamovi	130
25.3 R	131
25.4 SPSS	132
26 Multilevel analyse	135
26.1 Intro	135
26.2 jamovi	135
26.3 R	136
26.4 SPSS	137
27 Referenties	139

Rosetta for OU



Sectie Methoden en Statistiek

Open Universiteit



Chapter 1

Inhoudsopgave

1. *Basis Informatie*

- Software: Wat zijn de voor- en nadelen van de verschillende software pakketten? Hoe kun je R, jamovi, en SPSS installeren?
- Datasets: In dit boek gebruiken we meerdere datasets als voorbeeld voor de verschillende analyses. Hier bespreken we de structuur en de inhoud van de gebruikte datasets.
- Data laden: Het gebruiken van data vraagt om het laden (i.e., ‘activeren’) van een dataset. In dit hoofdstuk bespreken we hoe je dat voor de verschillende software pakketten aan kunt pakken.

2. *Analyses vooraf*

- Data screening: Voordat je analyses gaat uitvoeren is het belangrijk om een overzicht te krijgen van de beschikbare data. Denk hierbij aan descriptives (gemiddeldes; standaard deviates; standaard errors) en/of frequentie-tabellen.
- Ontbrekende waarden: In dit hoofdstuk bespreken we hoe je erachter komt of een variabelen ontbrekende waarden (i.e., ‘missing values’) heeft, hoe je hiermee kunt omgaan, en hoe je voor ontbrekende waarden kunt corrigeren.
- Variabelen samenvoegen: Een voorbeeld van het samenvoegen van variabelen is het berekenen van een gemiddelde score over meerdere variabelen/items. In dit hoofdstuk laten we zien hoe je dat kunt doen.
- Variabelen Standaardiseren: In dit hoofdstuk bespreken we hoe je variabelen kunt *standaardiseren* en *centraliseren*
- Variabelen Hercoderen: Hercoderen is het omschalen van variabelen (hoge waardes worden laag en omgekeerd) of het her-categoriseren van waardes. In dit hoofdstuk laten we zien hoe je dit handmatig of middels een berekening doet.

- Coefficient Alpha: Hoe je Cronbach's Alpha kunt berekenen. In jamovi en R worden McDonalds Omega ook automatisch gegeven.

3. *Analyses met twee variabelen*

- Kruistabellen Een kruistabel is een frequentie-overzicht van meerdere variabelen tegelijk. We bespreken ook hoe je **phi**, **Cramers' V**, en **Chi-kwardraat** berekent/toetst.
- T-test Het uitvoeren van onafhankelijke (independent) en gepaarde (paired) t-testen om verschil in gemiddelden te testen.
- Cohen's D Hoe groot het verschil is tussen de gemiddeldes van twee groepen.
- Correlaties De mate waarin twee continue verdeelde variabelen met elkaar samenhangen uitgedrukt in **Pearsons' s r**
- ANOVA Een variantieanalyse om twee of meer groepen met elkaar te vergelijken. In dit hoofdstuk komen de one-way ANOVA.
- Lineaire regressie Een regressie met een enkele predictor/voorspeller voor een afhankelijke variabele.
- Non-parametrische toetsen Alternatieve manieren om groepen te vergelijken wanneer niet aan de assumpties van 'traditionele' testen wordt voldaan. We bespreken **Wilcoxon's signed rank test**, **Wilcoxon's rank sum test**, **Mann-Whitney U test**, en de **Kruksal-Wallis test**.

4. *Analyses met meer variabelen*

- Exploratieve Factoranalyse Een manier om een grotere set items terug te brengen tot een kleiner aantal factoren.
- Factoriele ANOVA Een variantieanalyse voor het vergelijken van twee of meer nominale variabelen.
- Covariantie analyse Een uitbreiding op de one-way ANOVA of de factoriele ANOVA waarbij een covariant wordt meegenomen in de analyse.
- Herhaalde metingen analyse Een variantie analyse waarbij de afhankelijker variabelen meerdere keren is gemeten. Ook wel de Repeated-Measures ANOVA genoemd.
- Multivariate Regressie Een regressiemodel waarbij de afhankelijke variabele voorspelt wordt door meerdere predictoren.
- Logistische regressie Een regressiemodel voor dichotome/binaire uitkomstvariabelen (e.g., ja/nee).
- Moderatie analyse Een moderator is een variabele die het verband tussen twee variabelen beïnvloedt.
- Moderatie Mediatie Modellen Een combinatie van een mediatie- en moderatie model.
- Multilevel analyse Het analyseren van hiërarchische gegevens zoals kinderen binnen groepen of herhaalde metingen binnen proefpersonen.

Part I

Basis informatie

Chapter 2

Software

2.1 jamovi

Het jamovi-project is opgericht om een gratis en open statistisch platform te ontwikkelen dat intuïtief te gebruiken is en de nieuwste ontwikkelingen op het gebied van statistische methodologie kan bieden. De kern van de jamovi-filosofie is dat wetenschappelijke software “gemeenschapsgestuurd” moet zijn, zodat iedereen analyses kan ontwikkelen en publiceren, en deze beschikbaar kan maken voor een breed publiek. Het is beschikbaar voor Windows, macOS, Linux en ChromeOS via <https://jamovi.org>.

Een handige functie van jamovi is dat wanneer u een project opslaat, het de gegevens, analyses en uitvoer in hetzelfde bestand opslaat. Dit maakt samenwerking en begeleiding veel eenvoudiger. Een andere eigenschap is dat, in overeenstemming met de jamovi-filosofie, iedereen modules kan bijdragen. Dit betekent dat gebruikers extra functionaliteit kunnen installeren vanuit de jamovi-bibliotheek. Het betekent ook dat jamovi in de loop van de tijd zal blijven groeien.

Als u zojuist jamovi hebt geïnstalleerd en geopend, zult u iets zien dat lijkt op wat wordt weergegeven in Figuur 2.1. Dit boek gaat vergezeld van een jamovi-module die enkele van de analyses bevat waarnaar we verwijzen, evenals de datasets die we hebben opgesomd in hoofdstuk 3.

2.1.1 Installeren van jamovi modules

Om een nieuwe module in jamovi te installeren, klikt u op de knop met de blauwe “plus” met het label “Modules” in de rechterbovenhoek. Hierdoor wordt het menu geopend dat wordt weergegeven in Figuur 2.2.

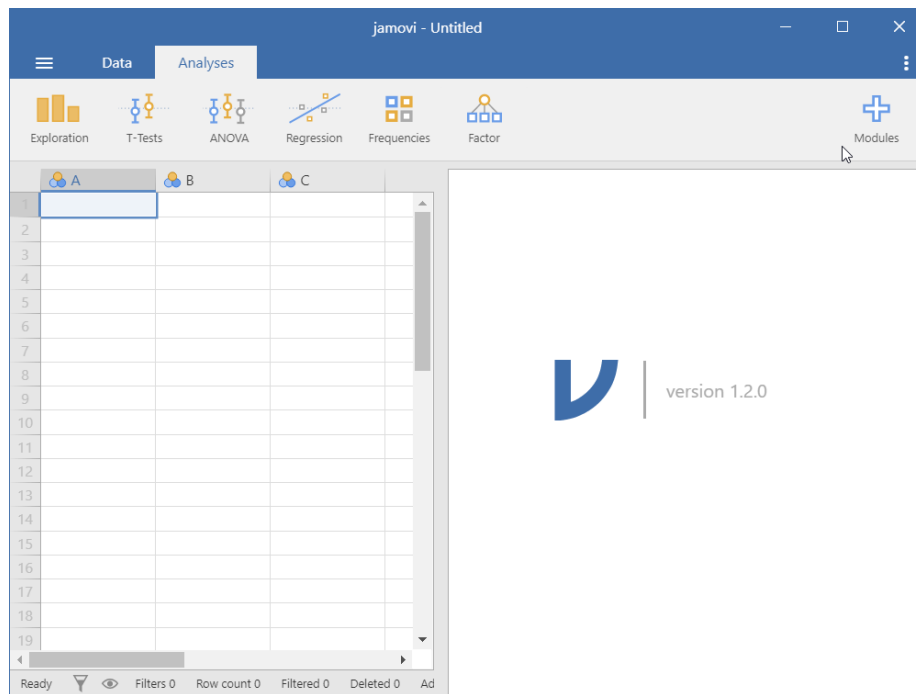


Figure 2.1: Installeren van jamovi.

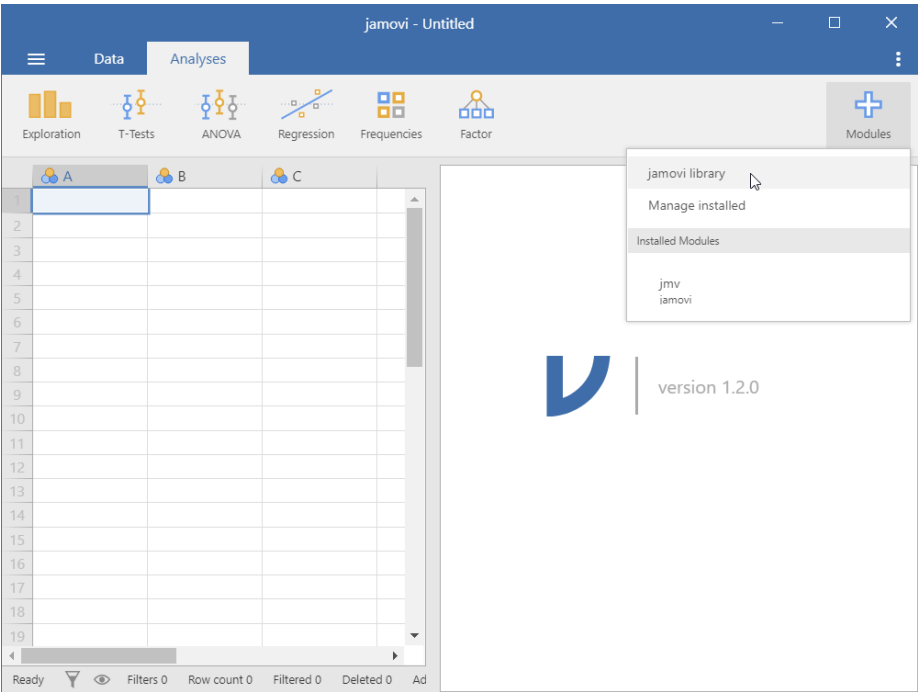


Figure 2.2: Het modules menu in jamovi.

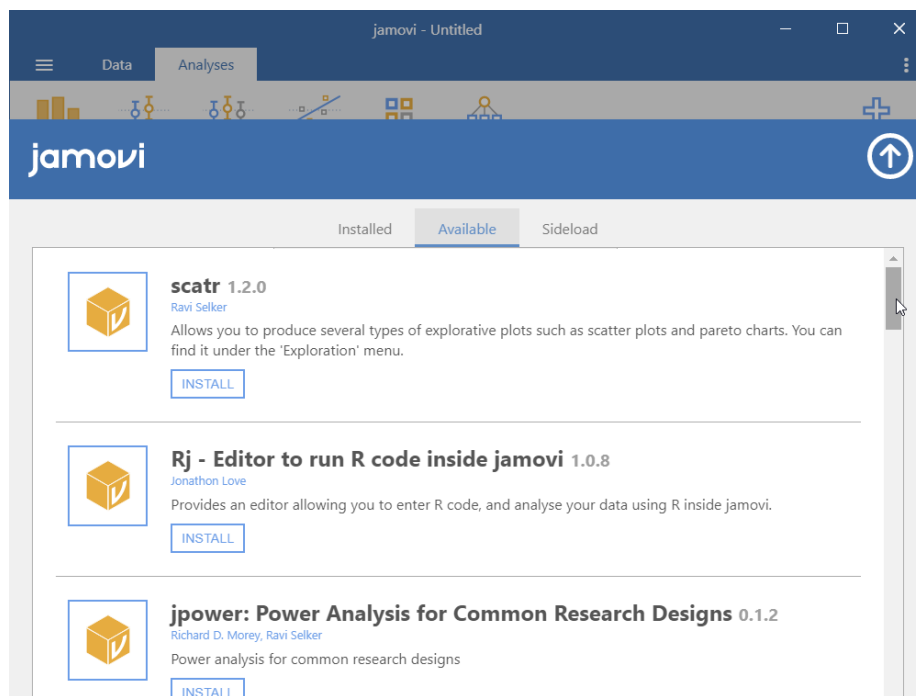


Figure 2.3: De jamovi library.

Selecteer in dit menu “jamovi-bibliotheek”, waarmee de jamovi-bibliotheek wordt geopend met een overzicht van alle beschikbare modules, zoals weergegeven in Figuur 2.3. Scroll naar beneden naar de “rosetta”-module, die “Parallel Use of Statistical Packages in Teaching” wordt genoemd en klik op de overeenkomstige “Installeren”-knop om de installatie te starten. Zodra de module is geïnstalleerd, verschijnt deze in de menubalk zoals weergegeven in Figuur 2.4.

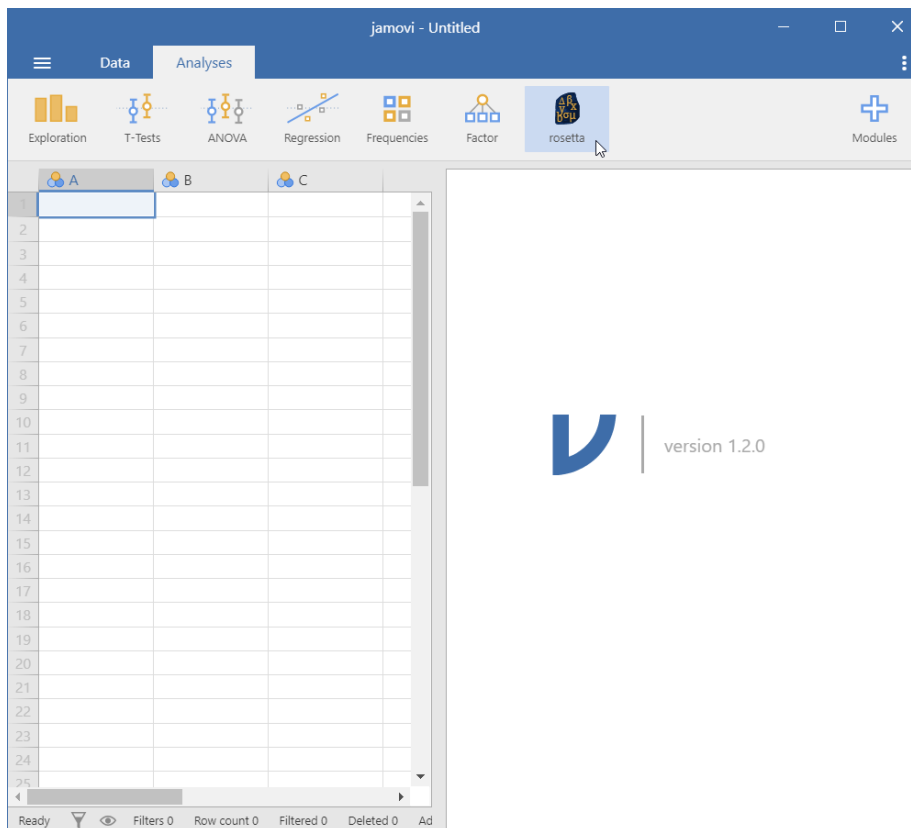


Figure 2.4: De jamovi library.

Door de “rosetta”-module te installeren, worden de voorbeelddatasets die in dit boek worden gebruikt, beschikbaar. Hoofdstuk 4 legt uit hoe deze datasets (of andere datasets) in jamovi geladen kunnen worden.

2.2 R

R is een zeer krachtig hulpmiddel voor statistische analyses. Omdat het gemakkelijk uitbreidbaar is via pakketten, is het uitgegroeid tot een tool die

u met vrijwel alles kan helpen: er zijn pakketten voor analyse op meerdere niveaus, het modelleren van structurele vergelijkingen, maar ook voor het visualiseren van geografische kaarten, het weergeven van driedimensionale objecten, het uitvoeren van tekst mining, werken met big data, directe interactie met databases, geavanceerde datavisualisatie, kwalitatieve analyse, en ga zo maar door.

R kan worden gedownload vanaf <https://cloud.r-project.org/>. Wanneer u R downloadt, krijgt u allen het statistische pakket zelf. Er zijn veel manieren om het werken met R prettiger en efficiënter te maken, zoals het populaire RStudio.

2.2.1 R Studio

Een populaire omgeving is RStudio. RStudio kan worden gedownload van <https://rstudio.com/products/rstudio/> (kies RStudio Desktop). Zie Figuur 2.5 voor een illustratie van hoe RStudio er uitziet.

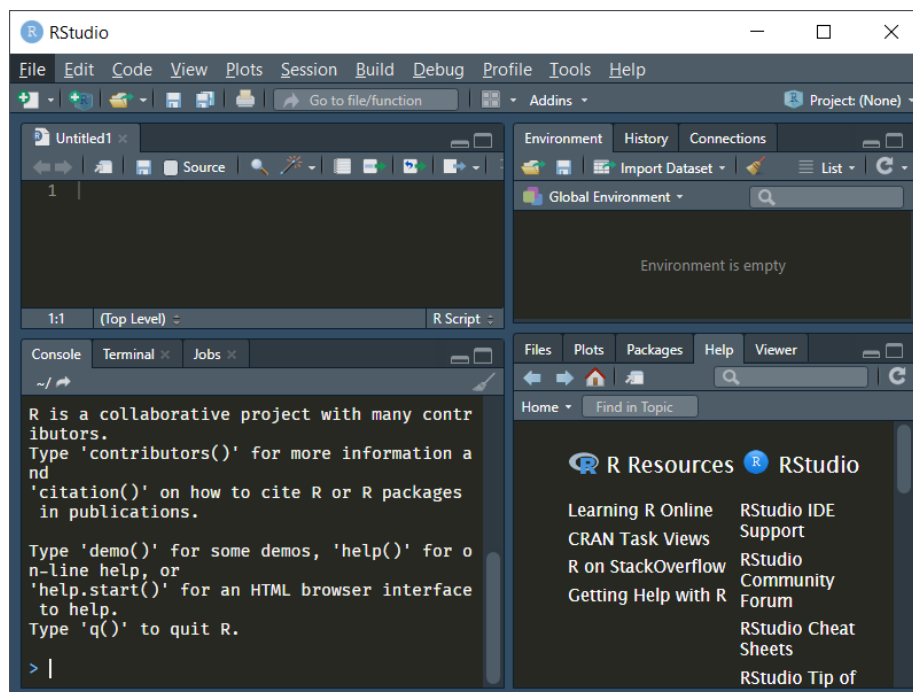


Figure 2.5: Voorbeeld RStudio.

RStudio is een interface die vier deelvensters met tabbladen gebruikt om interacties met R te vergemakkelijken. Het deelvenster linksboven bevat het analysescript of meerdere scripts waaraan u werkt. Het paneel linksonder bevat de R-console. Het paneel rechtsboven bevat een overzicht van alle datasets en

andere objecten die u hebt geladen, evenals een geschiedenis van de opdrachten die u aan R hebt verstrekt. Ten slotte biedt het paneel rechtsonder eenvoudig toegang tot een overzicht van alle bestanden in uw project, de plots die je hebt gemaakt, de R-pakketten die je hebt geïnstalleerd en geladen, en de online handleiding van R en je pakketten.

2.2.2 Packages

R wordt geleverd met veel basisfuncties, ook wel Base R genoemd. Onderzoekers maken georganiseerde verzamelingen van R-functies, ook wel pakketten genoemd, om het gebruik van verschillende statistische methodologieën en analyses te vergemakkelijken. Deze pakketten zijn gratis en leveren bruikbare output op voor gebruikers. Iedereen kan een R-pakket maken. Het is gebruikelijk om de pakketten die u gebruikt te citeren in het verwijzingsgedeelte van uw onderzoek.

Pakketten bevinden zich doorgaans op twee plaatsen: het Comprehensive R Archive Network, afgekort CRAN, dat enige kwaliteitscontrole op pakketten uitvoert. Standaard kijkt R alleen in de CRAN-repositories wanneer u hem vertelt een pakket te installeren. Veel pakketontwikkelaars bieden ook pakketten aan in Git-opslagplaatsen. Deze zijn vaak minder gepolijst dan CRAN-pakketten, maar ook meer cutting edge. Hieronder leggen we uit hoe u “niet-CRAN-pakketten” installeert (zie Paragraaf 2.2.2.2).

Een pakket dat u misschien wilt installeren als u dit boek gebruikt, is het rosetta-pakket. We hebben dit gemaakt om parallel gebruik van en overgang tussen verschillende statistische pakketten te vergemakkelijken. Het bevat een aantal functies die zijn ontworpen om zich op dezelfde manier te gedragen als hun tegenhangers in andere softwarepakketten. Dit betekent dat we ernaar streven om vergelijkbare namen te gebruiken voor de functies en argumenten, vergelijkbare standaardinstellingen en vergelijkbare uitvoer. We zullen dit pakket als voorbeeld gebruiken in de volgende secties.

2.2.2.1 Installeren van CRAN packages

Om een pakket te installeren, kunt u de volgende code gebruiken:

```
install.packages('rosetta');
```

Vervang “rosetta” door de naam van het pakket dat u wilt installeren. De naam van het pakket is hoofdlettergevoelig en moet tussen aanhalingstekens staan. Zodra het pakket is geïnstalleerd, hoeft u het nooit meer te installeren (hoewel u mogelijk wordt gevraagd om het bij te werken).

Als u RStudio gebruikt, kunt u ook pakketten installeren met behulp van het tabblad Pakketten in het rechteronderdeelvenster van RStudio door op “Pakketten” en “Installeren” te klikken en de naam in te voeren van het pakket dat u wilt installeren (zie Figuur 2.6).

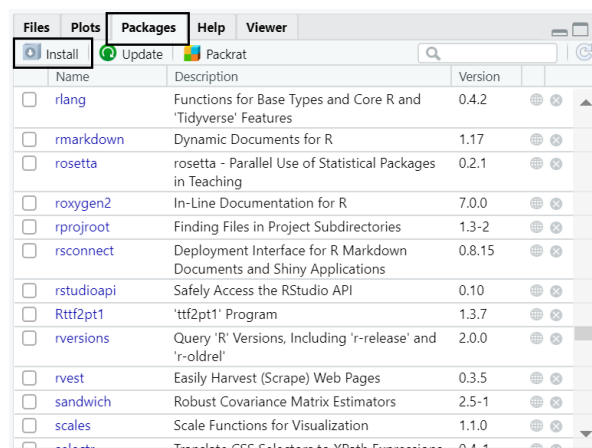


Figure 2.6: How to install package using the RStudio GUI.

2.2.2.2 Installeren van niet-CRAN packages

Als je een pakket wilt installeren dat niet op CRAN staat, of een nieuwere versie van een pakket dan de versie op CRAN, kun je pakketten rechtstreeks vanuit Git repositories installeren. Dit vereist het pakket `remotes`. Eenmaal geïnstalleerd, kunt u de functies `remotes::install_gitlab()` en `remotes::install_github()` gebruiken om pakketten te installeren vanuit repositories die worden gehost door respectievelijk GitLab en GitHub.

Om dit te doen, zoek eerst de URL naar de repository (deze heeft meestal de vorm van “https://gitlab.com/xxx/yyy” of “https://github.com/xxx/yyy”), en onthoud het deel achter de domeinnaam (dus “xxx/yyy”). Geef dit als argument door aan de functie `remotes::install_gitlab()` of `remotes::install_github()`.

Om bijvoorbeeld de huidige ontwikkelversie van `rosetta` te installeren (die meer mogelijkheden heeft maar meer fouten kan bevatten), gebruik je de volgende twee commando’s (je kunt de eerste overslaan als je het `remotes` pakket al geïnstalleerd hebt):

```
install.packages('remotes');
remotes::install_gitlab('r-packages/rosetta');
```

2.2.2.3 Functies gebruiken van packages

Zodra een pakket (package) is geïnstalleerd, kunt u alle functies in het pakket uitvoeren door de functienaam vooraf te laten gaan met `pakket::`. Als u bijvoorbeeld het pakket `rosetta` hebt geïnstalleerd, kunt u alle functies in het Rosetta-boek uitvoeren door te beginnen met `rosetta::`. Merk op dat als een functieaanroep begint met `somethingelse::`, je het pakket `somethingelse` moet installeren. Geavanceerde datavisualisaties vereisen bijvoorbeeld het pakket `ggplot2` en functies van `ggplot` beginnen allemaal met `ggplot2::`. Het `rosetta` pakket gebruikt ook het `ggplot2` pakket op de achtergrond, dus je hoeft dat niet apart te installeren.

Als u een functie probeert te gebruiken maar u vergeet R te vertellen uit welk pakket het de functie moet halen, kan R u een foutmelding geven. Dit gebeurt niet altijd omdat sommige functies vanuit `base R` (het basis pakket) al beschikbaar zijn (bijv. `nrow()` voor het bepalen van het aantal rijen in een dataset.).

2.2.2.4 Het laden van packages in het geheugen

Soms wil je R vertellen dat hij altijd naar objecten (zoals functies) in een bepaald pakket moet zoeken. In dat geval kunt u het pakket tijdens uw sessie aan het “zoekpad” van R koppelen. U kunt zich voorstellen als het “laden” van het pakket. Hoewel u een pakket maar één keer hoeft te installeren, moet u die opdracht elke keer dat u een nieuwe R-sessie opent, als u het wilt laden, herhalen.

U kunt een pakket laden (d.w.z. het aan het zoekpad van R koppelen) met dit commando:

```
library("rosetta")
```

Als u RStudio gebruikt om met R te werken, kunt u pakketten laden door het tabblad ‘Packages’ in het rechteronderdeelvenster te openen en op het pakket te klikken dat u wilt laden (zie Figuur 2.7). Als het pakket niet in uw lijst met pakketten voorkomt, is dat omdat het niet is geïnstalleerd (zie Paragraaf 2.7).

Merk op dat het loont om er een gewoonte van te maken om altijd expliciet de pakketnaam te specificeren met behulp van de `::` operator. Omdat R geen unieke functienamen afdwingt tussen pakketten, hebben veel pakketten functies met dezelfde namen (zoals `recode()` in de `Hmisc`, `car` en `rosetta` pakketten en `filter` in beide basis R en `dplyr`). Als u meerdere pakketten laadt en een dergelijke functie gebruikt, is het niet altijd intuïtief welke versie van een functie u “krijgt”.

Hierdoor is foutopsporing veel efficiënter als u expliciet de pakketten specificeert samen met de functies in elke functieaanroep. Aangezien wanneer men met R

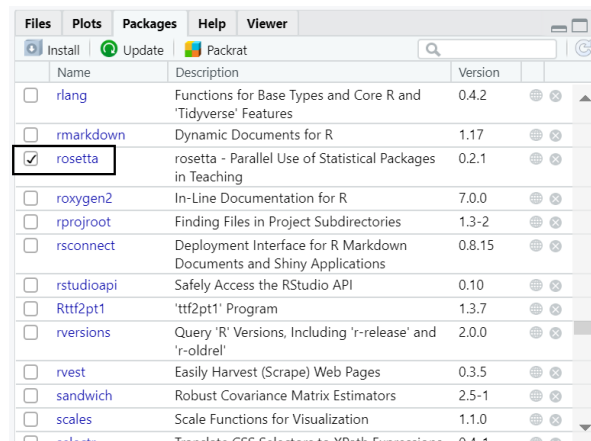


Figure 2.7: Een package laden met RStudio.

(of andere statistische pakketten) werkt, de meeste tijd niet wordt besteed aan het schrijven van code, maar aan het *debuggen* (i.e., fouten oplossen) van code, loont dit snel, ondanks de lichte inefficiëntie van het altijd moeten typen van de pakketnamen.

Een tweede reden om de pakketten expliciet te specificeren bij het aanroepen van functies, is dat het R-code draagbaarder maakt. Als u vertrouwt op geladen pakketten en u fragmenten van R-code naar een ander script kopieert, zal de code fouten genereren totdat u eraan denkt om die functies ook bovenaan uw code te laden. Identificeren welke pakketten u moet laden om een codefragment te kunnen uitvoeren, is onmogelijk, tenzij u toevallig uit uw hoofd weet welke functies uit welke pakketten komen.

Een derde nauw verwante reden is dat het expliciet specificeren van pakketten in functieaanroepen uw code gemakkelijker leesbaar maakt voor anderen. Ze hoeven niet op en neer te scrollen om te proberen te achterhalen welke functies u precies gebruikt.

Ten slotte is een vierde reden van toepassing op meer gevorderde gebruikers: als u R-pakketten ontwikkelt, is het ook een goede gewoonte om functies altijd aan te roepen met de operator `::`. Daarom is het het beste om dit vanaf het begin te automatiseren.

Als u bovenaan uw code wilt controleren of er een pakket bestaat, is er een functie genaamd `[checkPkgs ()]` (<https://r-packages.gitlab.io/ufs/reference/checkPkgs.html>) in het `ufs`-pakket.

2.2.3 Drie benaderingen in R

In R zijn er grofweg drie benaderingen. Veel analyses kunnen worden gedaan met base R zonder extra pakketten te installeren. Het `rosetta`-pakket vergezelt dit boek en heeft tot doel een output te leveren die vergelijkbaar is met die van jamovi en SPSS met eenvoudige commando's. Ten slotte is de `tidyverse` een populaire verzameling pakketten die consistent proberen samen te werken, maar een andere onderliggende logica implementeren die op R (en dus het `rosetta`-pakket) gebaseerd is.

2.3 SPSS

SPSS (<https://www.ibm.com/analytics/spss-statistics-software>) is populaire data-analysesoftware die eigendom is van IBM. SPSS biedt zowel een GUI (graphical user interface), dat wil zeggen het werken via de menustructuur, als een syntaxisoptie voor analyse, hoewel mensen SPSS voornamelijk gebruiken vanwege de eenvoud van de GUI. SPSS wordt geleverd met een vast aantal analytische opties in de GUI en deze kunnen niet door gebruikers worden gewijzigd. Deze tutorial gebruikt SPSS versie 26.

In dit boek beperken we ons tot de syntax van SPSS. Voor wie met de menustructuur wil werken, zijn er op internet veel tutorials te vinden over alle in dit boek gebruikte analyses.

2.3.1 De actieve dataset

In SPSS wordt elke dataset in een apart venster geopend. Er kan slechts één dataset tegelijk worden gebruikt. De gebruiker moet dus handmatig schakelen tussen datasets.

Bij gebruik van de *GUI* kan een dataset worden geactiveerd door op het bijbehorende venster te klikken. Daarom is het belangrijk om ervoor te zorgen dat u werkt in de dataset die u wilt gebruiken.

Gebruik het commando `DATASET ACTIVATE` gevolgd door de naam van de dataset om aan te geven welke dataset geactiveerd moet worden met *SPSS Syntax*. Om bijvoorbeeld een dataset met de naam `dat` te activeren, kan men gebruik maken van de syntax:

```
DATASET ACTIVATE dat.
```

Merk op dat wanneer u slechts één dataset heeft geopend, deze dataset altijd actief is en u deze dus niet hoeft te activeren.

Chapter 3

Datasets

De voorbeelden in dit boek maken gebruik van een aantal vrij beschikbare datasets. Deze datasets worden hier beschreven. In hoofdstuk 4 wordt uitgelegd hoe datasets geladen kunnen worden.

3.1 pp15

De dataset `pp15` is afkomstig van het “partypanel” studie uit 2015. Dit is een determinantenonderzoek opgezet om de determinanten van het gebruik van een hoge dosis MDMA (is het actieve ingrediënt in tabletten die als ecstasy worden verkocht) te meten. De meeste variabelen hebben betrekking op de subdeterminanten van de drie variabelen van de “Reasoned Action Approach”: houding, gepercipieerde normen en gepercipieerde gedragscontrole. Om deze data direct te downloaden, gebruik: <https://gitlab.com/r-packages/rosetta/raw/prod/data/pp15.csv?inline=false>.

De dataset `pp15` bevat de volgende variabelen, met labels:

column	variable
gender	Gender
age	Age (in years)
hasJob	Employment status
jobHours	Hours of employment
currentEducation	If in education, study type
currentEducation_cat	If finished education, study type
prevEducation	Categorized current education type
prevEducation_cat	Categorized previous education type
xtcUseDoseHigh	MDMA dose (in mg) that is considered a high dose
xtcUseDosePref	Preferred MDMA dose (in mg)
xtcUsePillHigh	Number of pills that is considered a high dose
xtcUsePillPref	Preferred number of pills
weight_other	Body weight (in kg)
highDose_intention	RAA intention: planning to
highDose_attitude	RAA intention: want to
highDose_perceivedNorm	RAA intention: expect to
highDose_pbc	Intention frequency: number of times
highDose_IntentionRAA_intention	Intention frequency: multiplier
highDose_IntentionRAA_want	Attitude (direct): bad-good
highDose_IntentionRAA_expectation	Attitude (direct): unpleasant-pleasant
highDose_IntentionFrq_freq.0	Attitude (direct): dumb-smart
highDose_IntentionFrq_freq.1	Attitude (direct): unhealthy-healthy
highDose_AttGeneral_good	Attitude (direct): boring-exciting
highDose_AttGeneral_prettig	Injunctive norm (direct): people important to me
highDose_AttGeneral_slim	Injunctive norm (direct): people whose opinion I
highDose_AttGeneral_gezond	Descriptive norm (direct): people I respect and a
highDose_AttGeneral_spannend	Descriptive norm (direct): people like me
highDose_NormGeneral_in1	Perceived behavioral control (direct): would not/
highDose_NormGeneral_in2	Perceived behavioral control (direct): easy-hard
highDose_NormGeneral_dn1	Perceived behavioral control (direct): no/complet
highDose_NormGeneral_dn2	Perceived behavioral control (direct): no/many e
highDose_PBCgeneral_ifwanted	Perceived behavioral control (direct): under other
highDose_PBCgeneral_easy	Perceived reasons for using a high dose of MDMA
highDose_PBCgeneral_control	Perceived reasons for *not* using a high dose of MDMA
highDose_PBCgeneral_externalFactors	Expectation (attitudinal belief): shorter/longer t
highDose_PBCgeneral_notOnlyMe	Expectation (attitudinal belief): less/more intens
highDose_OpenWhy	Expectation (attitudinal belief): less/more waste
highDose_OpenWhyNot	Expectation (attitudinal belief): less/more energy
highDose_AttBeliefs_long	Expectation (attitudinal belief): less/more eupho
highDose_AttBeliefs_intensity	Expectation (attitudinal belief): less/more self-in
highDose_AttBeliefs_intoxicated	Expectation (attitudinal belief): less/more conne
highDose_AttBeliefs_energy	Expectation (attitudinal belief): easier/harder to
highDose_AttBeliefs_euphoria	Expectation (attitudinal belief): feel less/more lil
highDose_AttBeliefs_insight	Expectation (attitudinal belief): forget problems
highDose_AttBeliefs_connection	Expectation (attitudinal belief): less/more social
highDose_AttBeliefs_contact	Expectation (attitudinal belief): can probe my b
highDose_AttBeliefs_sex	Expectation (attitudinal belief): music sounds wo
highDose_AttBeliefs_coping	Expectation (attitudinal belief): less/more halluc
highDose_AttBeliefs_isolated	Expectation (attitudinal belief): time passes slow
highDose_AttBeliefs_boundaries	Expectation (attitudinal belief): remember less/n
highDose_AttBeliefs_music	Expectation (attitudinal belief): less/more health
highDose_AttBeliefs_hallucinate	Expectation (attitudinal belief): experience worse
highDose_AttBeliefs_timeAwareness	Expectation (attitudinal belief): worry less/more

3.2 Politiek_Media

De dataset `Politiek_Media` is eigenlijk een kruistabel, en bevat de volgende variabelen, met labels:

column	variable
partij	Politieke partij (GL, PVV, VVD, CDA, D66, PvdA)
media	Mediagebruik kiezers(Twitter, Facebook, Krant, Televisie, Online)

3.3 statistiekangst

Deze data bevat een experimentele interventie: een cursus om statistiekangst te verminderen. Statistiekangst is zowel voor als na de interventie gemeten. Cursustype is gemanipuleerd en deelnemers zijn gerandomiseerd. De toegepaste cursus is ontwikkeld om een verbetering te zijn ten opzichte van de traditionele cursus. De dataset `statistiekangst` bevat de volgende variabelen, met labels:

column	variable
Respnum	Respondent nummer
Gender	Geslacht (1 = man, 2 = vrouw)
Age	Leeftijd (in jaren)
Education	Opleidingsniveau (1=laag,2=laag/middel, 3=middel,4=middel/hoog 5=hoog)
Cursus	1=Traditioneel, 2=Toegepast
Statken1	Statistiekkennis voor de cursus
Statken2	Statistiekkennis na de cursus
Angst1	Statistiekangst voor de cursus
Angst2	Statistiekangst na de cursus
dif_Stat	Toename statistiekkennis
dif_Angst	Toename statistiekangst

3.4 schoolsucces

De dataset `schoolsucces` bevat de volgende variabelen, met labels:

column	variable
geslacht	Geslacht (0 = jongen, 1 = meisje)
leeftijd	Leeftijd (in jaren) bij examen
uitslag	Uitslag examen (0 = geslaagd, 1 = gezakt)
cijfer	Gemiddeld cijfer behaald op eindtoets
IQ	Intelligentie score
inkomen	Inkomen ouders in drie categorieën

3.5 sportcasus

Bij een ICT-bedrijf met in totaal 108 medewerkers is men zich bewust dat men meer moet bewegen om gezond te blijven. Vrijwel alle medewerkers zitten namelijk de hele dag achter hun computer. Om gezondheidsklachten, en dus ook ziekteverzuim, te voorkomen heeft het bedrijf een sportzaal ter beschikking gesteld waarin de medewerkers van het bedrijf onder werktijd elke dag een half uur mogen sporten. Het bedrijf heeft echter de indruk dat de medewerkers weinig gebruikmaken van deze faciliteit en wil hen motiveren er wel gebruik van te maken. Hiervoor zetten ze een experiment op. Voorafgaand aan het experiment, wordt allereerst aan alle medewerkers gevraagd in te vullen hoeveel keer per week zij een half uur sporten in de sportzaal. Omdat de onderzoekers vermoeden dat leeftijd hier een belangrijke rol bij speelt, werd de medewerkers ook naar leeftijd gevraagd. Medewerkers zijn daarna ingedeeld in jonge medewerkers, dat wil zeggen: jonger dan 45 jaar, en oudere medewerkers, medewerkers vanaf 45 tot en met 65 jaar. Ook is gevraagd naar het aantal uren dat ze volgens contract per week werken.

Er is in samenspraak met een onderzoeker en de directie van het bedrijf besloten het personeel voorlichting te geven over de voordelen van bewegen en de effecten van gezondheid. Een groep van de werknemers krijgt elke keer als ze willen sporten een persoonlijk advies. Een andere groep van de medewerkers krijgt als ze de tijd reserveren om te gaan sporten een digitale voorlichtingsfolder over de voordelen van sporten als popup. Als laatste is er een groep die zonder enige extra voorlichting naar de sportschool gaat. Om de effecten van het type voorlichting te onderzoeken wordt eerst onderzocht hoe vaak de sportschool bezocht is. Dan wordt de persoonlijke voorlichting of de digitale folder aangeboden (afhankelijk van in welke groep iemand is ingedeeld). Direct na afloop van het experiment wordt nogmaals op dezelfde wijze onderzocht hoe vaak de sportzaal bezocht is. Na een maand wordt deze meting herhaald, hierbij wordt vastgesteld hoe vaak op dat moment de sportzaal wordt bezocht door de deelnemers aan het experiment.

De dataset `sportcasus` bevat de volgende variabelen, met labels:

column	variable
leeftijd	Leeftijd (1 = tot en met 45, 2 = ouder dan 45)
voorlichting	Type voorlichting (1 = controlegroep, 1 = digitale folder, 2=persoonlijk advies)
sportscore1	Aantal keren gesport eerste meting (0 tot en met 5)
sportscore2	Aantal keren gesport tweede meting (0 tot en met 5)
sportscore3	Aantal keren gesport derde meting (0 tot en met 5)
sportverschil	Verschil tussen meting 1 en 2 (positieve score geeft toename aan)
urenaanwezig	Aantal uren contractueel aanwezig

3.6 contraProductiefGedrag

In een bedrijf wordt onderzoek gedaan naar contraproductief gedrag (CPG) op de werkvloer (zoals: te laat komen, stelen, roddelen over collega's, lage kwaliteit van werk). Men vermoedt dat dit samenhangt met het vertrouwen in het management, met een cynische houding ten opzichte van de organisatie en van de stemming van een werknemer. Verder denkt men dat verschillende vormen van gepercipieerde rechtvaardigheid invloed kunnen hebben op dit CPG.

De dataset `contraProductiefGedrag` bevat de volgende variabelen, met labels:

column	variable
nummer	Respondentnummer
bedrijf	Organisatienummer
sekse	vrouw, man
leeftijd	Leeftijd respondent
opleiding	Opleiding respondent
vertrouwen	Vertrouwen in management
cynisme	Cynisme over organisatie
negatiefAffect	Negatief affect van respondent
contraProductiefGedrag	Contraproductief gedrag door respondent
procedureleRechtvaardigheid	Ervaren procedurele rechtvaardigheid
distributieveRechtvaardigheid	Ervaren distributieve rechtvaardigheid
interactioneleRechtvaardigheid	Ervaren interactionele rechtvaardigheid

3.7 wiskunde

In verschillende klassen zijn data verzameld om te kijken of de score op wiskundetoetsen kunnen worden voorspeld uit de intelligentie van de leerling en uit enkele karakteristieken van de klas en de leerkracht.

De dataset `wiskunde` bevat de volgende variabelen, met labels:

column	variable
klas	Klasnummer
wiskunde	Wiskundescore
intelligentie	Intelligentiescore
ervaring	Ervaring leerkracht
klasGrootte	Aantal leerlingen in klas
sekseLeerling	Sekse van de leerling
sekseLeerkracht	Sekse van de leerkracht

3.8 roken

In een intensieve longitudinale studie onder mensen die willen stoppen met roken, wordt gekeken of het roken van een sigaret samenhangt met de stress die men op een moment op de dag ervaart.

De dataset **roken** bevat de volgende variabelen, met labels:

column	variable
respondent	Respondentnummer
leeftijd	Leeftijd
meting	Volgnummer binnen persoon
zin	Zin in een sigaret
stress	Mate van ervaren stress
roken	Heeft u gerookt (ja/nee)
samen	Bent u in gezelschap (alleen/samen)

Chapter 4

Data laden

In dit hoofdstuk wordt uitgelegd hoe je data laadt om de diverse analyses. De meeste voorbeeld data staan in de OSF repository, die bereikbaar is via de link: https://osf.io/ra5b4/?view_only=ca72c7b65a0f457a88656f95e50ff3e5.

De data staan in SPSS format, dus met de extensie “.sav”. Deze data kunnen rechtstreeks worden ingelezen in jamovi en uiteraard ook in SPSS. Ook kunnen de data eenvoudig worden geïmporteerd in R, zoals hieronder wordt aangegeven.

4.1 jamovi

Open eerst het hoofdmenu van jamovi door op het “hamburgermenu” (de drie horizontale streepjes) in de linkerbovenhoek te klikken, net links van de tekst “Data”. Hierdoor wordt een menu geopend. Klik in dat menu op de tekst “Openen”. Hierdoor worden twee opties in dit menu uitgebreid: “This pc” en “Data Library”. Welke je kiest, hangt ervan af of je een bestand vanaf uw pc wil openen of dat je een dataset ergens anders wil openen.

Om een bestand op je pc te openen, klik je op “This pc” in het menu “Openen” van jamovi. Om een andere dataset te openen, klik je op “Data Library” in het menu “Openen” van jamovi. Standaard wordt de jamovi-databibliotheek al geleverd met een aantal datasets, maar geïnstalleerde modules kunnen meer datasets toevoegen.

4.2 R

Voor R installeer eerste (eenmalig) het package “haven” en gebruik de volgende opdracht om de dataset in te lezen in een dataframe met de zelfgekozen naam,

b.v. “dat”:

```
dat <- haven::read_sav("bestandsnaam.sav")
```

Als je RStudio gebruikt, is er een heel eenvoudige manier om gegevens te laden: je kunt het menu “Gegevensset importeren” gebruiken in het menu rechtsboven in RStudio. Van daaruit selecteer je het type gegevens dat je wilt importeren en volg je de aanwijs- en klikmenu’s. De dataset wordt in R geïmporteerd met de naam van het opgeslagen bestand. Dit kan relatief lang zijn en zo vervelend om te blijven typen, dus het is gebruikelijk om deze datasets te hernoemen naar een kortere naam, zoals “dat”. Om dit te doen, kun je de dataset kopiëren naar een andere naam met de pijl.

Om te controleren of de dataset (dataframe in R) correct is ingelezen, kun je de variabelen bekijken die erin zijn opgeslagen:

```
names(dat);
```

Je kunt de gegevens ook in spreadsheetformaat bekijken. Als je RStudio gebruikt, verschijnt het als een nieuw tabblad naast je R-script. Je kunt heen en weer schakelen tussen de twee.

```
View(dat);
```

4.3 SPSS

Om een SPSS dataset te openen in SPSS, kun je de volgende stappen doorlopen. Klik op “File”, “Open”, “Data” in het menu. Selecteer vervolgens het bestand dat je wil openen.

Ook kun je een dataset openen via de syntax:

```
GET
```

```
FILE='C:\[de volledige locatie van het bestand]\\"bestandsnaam.sav"'.  
DATASET NAME DataSet1 WINDOW=FRONT.
```

Om datasets van andere bestandstypen te openen in SPSS (bijvoorbeeld .xls of .txt), selecteer “File”, “Open”, “Data” in het menu. Normaliter is bij “files of type” SPSS databestanden geselecteerd. Als je hierop klikt, dan kun je hier andere bestandstypen selecteren. Selecteer het bestand dat je wil openen. Per bestandstype zullen er andere stappen in het pop-up menu verschijnen. Doorloop deze stappen om de data te importeren.

Een voorbeeld van hoe een bestand uit excel kan worden ingeladen via de syntax:

```
GET DATA /TYPE=XLSX
  /FILE='C:\[de volledige locatie van het bestand]\\"bestandsnaam.xlsx"'
  /SHEET=name 'Sheet 1'
  /CELLRANGE=full
  /READNAMES=on
  /ASSUMEDSTRWIDTH=32767.
EXECUTE.
DATASET NAME DataSet1 WINDOW=FRONT.
```

Verschillende opties voor Excel zijn bijvoorbeeld het overnemen van de eerste rij van het bestand als de labels van de variabelen. Kijk dus altijd eerst goed in het originele bestand hoe de data eruit ziet.

Part II

Analyses vooraf

Chapter 5

Data screening

5.1 Intro

Voor je aan de slag gaat met ingewikkelde analyses, is het altijd goed om even door je dataset heen te lopen en alle variabelen te bekijken. Je kunt alvast exploratief naar de verdeling van de data kijken, en eventuele aanpassingen doen voor je met analyses verder gaat. Acties als hercoderen vallen hier ook onder, maar dat wordt in een apart hoofdstuk beschreven. Hier zal het gaan om het hernoemen van variabelen, het maken van simpele grafieken, beschrijvende statistieken, en frequentietabellen.

5.1.1 Voorbeeld dataset

Bij deze uitleg gaan we werken met de dataset `statistiekangst` die wordt besproken in hoofdstuk 3. Informatie over hoe je de datasets kunt inladen kun je vinden in hoofdstuk 4. In deze voorbeelden werken we met verschillende variabelen uit deze dataset.

5.2 jamovi

5.2.1 Variabelen bekijken

Zodra je in jamovi een dataset hebt geopend, verschijnt deze in je scherm. Het is handig om de variabelenamen goed te bekijken om zeker te weten dat de juiste dataset is ingeladen.

5.2.2 Variabelen hernoemen

Om een variabele te hernoemen kun je twee dingen doen. Je kunt met de rechtermuisknop klikken op de naam van de variabele. De variabelenamen staan boven de eerste rij in de dataset. Vervolgens kun je op **Setup** klikken. Hier kun je ook komen door naar het tabblad **Data** te gaan. Hier kun je ook op **Setup** klikken, en dan zal de variabele die je als laatst hebt aangeklikt verschijnen in het menu.

Het vakje helemaal bovenin met de dikgedrukte letters geeft de naam van de variabele weer. Als je hierop klikt kun je de naam wijzigen. Daaronder staat het label van de variabele, ook dit kun je wijzigen of toevoegen indien een variabele geen label heeft.

5.2.3 Categorische variabelen onderzoeken

Er zijn verschillende dingen die je kunt doen met een categorische variabele. In jamovi kun je veel van deze opties vinden op dezelfde plek. Om een frequentietabel te maken, kun je naar het tabblad **Analyses**. Hier klik je op **Exploration** en vervolgens **Descriptives**. Nu verschijnt er een menu in de linkerkant van het scherm. Je kunt de variabele waarvan je de tabel wilt maken verplaatsen naar het vak **variables** door op de pijl te klikken. Vervolgens kun je rechtsonder klikken op **Frequency tables**. Hier zie je ook symbooltjes bij staan. Dat houdt in dat jamovi alleen frequentietabellen maakt van variabelen die geclassificeerd zijn als categorisch.

In ditzelfde menu kun je ook aangeven dat je een staafdiagram van de data wilt. Hiervoor klap je het menu **Plots** uit en klik je op **Bar plot**. Nu verschijnt aan de rechterkant automatisch een staafdiagram. Als je dit doet voor de variabele **Education**, dan kun je zien dat de meeste mensen in de middelste categorie zitten.

De onderliggende syntax voor een frequentietabel met staafdiagram ziet er als volgt uit:

```
jmv::descriptives(  
  data = data,  
  vars = Education,  
  freq = TRUE,  
  bar = TRUE)
```

Het valt je misschien op dat je hierbij automatisch ook een tabel krijgt met beschrijvende statistieken die niet zo veel zegt bij categorische data. Om alleen een frequentietabel te maken, kun je ook gebruik maken van de **rosetta** module. Hier kun je klikken op **Frequencies** en weer de variabele toevoegen aan het vak. Dat ziet er in de syntax als volgt uit:


```
rosetta::freqjmv(  
  data = data,  
  vector = Education)
```

5.2.4 Continue variabelen onderzoeken

Bij continue variabelen heb je niet veel aan frequentietabellen, omdat er zo veel verschillende waarden zijn. Daarom zijn bepaalde beschrijvende maten een stuk handiger, zoals het gemiddelde, de mediaan of de standaarddeviatie. Daarnaast kun je ook continue variabelen grafisch weergeven om in één oogopslag een beeld te krijgen van de verdeling. Ga hiervoor naar het tabblad **Analyses** en klik op **Exploration**. Klik hier vervolgens op **Descriptives**. Zodra je een variabele in het vak voor **variables** stopt, verschijnt rechts een tabel met beschrijvende statistieken. Als je meer variabelen toevoegt verschijnen ze naast elkaar.

Ook kun je in ditzelfde menu verschillende dingen aanpassen of opvragen. In het kopje **Statistics** kun je kiezen welke maten je graag wilt zien. Zo kun je bijvoorbeeld de skewness van een variabele laten zien om een beeld te krijgen van de verdeling. In het kopje **Plots** kun je verschillende opties aanvinken. Enkele handige grafieken die je vaak nodig hebt zijn een histogram en een boxplot. Hiermee kun je zien of een variabele normaal verdeeld is en of er univariate uitbijters zijn.

Hieronder zie je de syntax voor het opvragen van de beschrijvende statistieken van de variabelen **Age** en **Statken1**, waarbij ook voor een histogram en een boxplot is gekozen:

```
jmv::descriptives(  
  data = data,  
  vars = vars(Age, Statken1),  
  hist = TRUE,  
  box = TRUE)
```

5.3 R

5.3.1 Variabelen bekijken

Het is handig om de variabelenamen goed te bekijken om zeker te weten dat de juiste dataset is ingeladen. In R kan je snel en makkelijk de variabelenamen bekijken door gebruik te maken van het commando **names**.

```
names(dat)
```

Om de gehele dataset te bekijken zodat je ook de gegevens kunt zien, kun je echter beter gebruik maken van het commando `View`. Hiermee opent R een spreadsheetachtig overzicht met je data en de bijbehorende variabelenamen. Als alternatief voor `View` kan je ook dubbelklikken op de naam van de dataset onder het tabblad “Environment” (standaard in het panel rechtsboven in R-studio).

```
View(dat)
```

5.3.2 Variabelen hernoemen

Soms is het nodig je variabelen te hernoemen. Probeer hiervoor zo simpel en systematisch mogelijk te werk te gaan. Het is niet gebruikelijk om spaties te gebruiken, in plaats daarvan kun je een laag streepje gebruiken (`_`) of door elk nieuw woord met een hoofdletter te laten beginnen. Dit heet `lowerCamelCase` als het eerste woord met een kleine letter begint, en `UpperCamelCase` als het eerste woord ook met een hoofdletter begint.

Het volgende commando hernoemt een variabele in R:

```
names(dat)[names(dat) == 'oudenaam'] <- 'NieuweNaam'
```

Hierbij wordt opnieuw het commando `names` gebruikt, maar omdat we vervolgens slechts één variabele specificeren, wordt alleen daarvan de naam veranderd.

5.3.3 Categorische variabelen onderzoeken

Om te bekijken hoe een categorische variabele verdeeld is, kun je frequentieverdelingen en staafdiagrammen opvragen. Dit kan met de volgende commando's.

Met het commando `freq` kun je voor één variabele een frequentieverdeling en een staafdiagram opvragen.

```
rosetta::freq(dat$Education, plot=TRUE)
```

Door `plot` te gebruiken geef je aan dat je een staafdiagram wil. Met het commando `frequencies` kun je van meerdere variabelen tegelijk een frequentietabel opvragen:

```
rosetta::frequencies(dat$Education, dat$Cursus, plot=TRUE)
```

5.3.4 Continue variabelen onderzoeken

Omdat met continue variabelen berekeningen uitgevoerd kunnen worden, zijn er meer mogelijkheden om de datareeksen van continue variabelen te verkennen. De betreffende beschrijvingsmaten, histogrammen, Q-Q-plots en boxplots vraag je met de volgende commando's op.

Met het commando `descr` vraag je de beschrijvingsmaten op (descriptives).

```
descr(dat$Age)
```

Vervolgens kun je een histogram maken met `hist`.

```
hist(dat$Age)
```

In plaats van een histogram kun je ook een boxplot laten maken, die we hier met een iets formelere wijze aanroepen:

```
rosetta::ggBoxplot(data = statistiekangst, y = Age)
```

Alternatieve commando's om beschrijvingsmaten op te vragen zijn `summary` en `describe` (dit commando zit in de `psych` package, deze moet je wel eerst laden). Met `describe` kun je ook de hele dataset in één keer laten beschrijven, of een specifieke selectie maken:

```
describe(dat[,c('Age', 'Statken1', 'Statken2')])
```

5.4 SPSS

In SPSS kun je de data gelijk zien bij het laden. Je kunt hierbij kiezen tussen een data view waarbij je alle losse cases ziet, en een variable view waarbij je de verschillende variabelen ziet en hun bijbehorende gegevens. Het is overigens heel belangrijk om vanaf het begin de gewoonte aan te leren nooit iets te veranderen in deze Data View of Variable View. Alle wijzigingen dienen traceerbaar te zijn en dat kan alleen door ze via de Syntax uit te voeren.

5.4.1 Variabelen bekijken

Zoals hierboven al kort beschreven hoef je in SPSS weinig moeite te verrichten om de variabelen te bekijken. In de data view zie je een spreadsheetachtige overzicht met daarin alle data. Elke kolom stelt een variabele voor, en elke rij een case. In de variable view zie je de cases niet, daar is een lijst met alle variabelen. Hier staat bijvoorbeeld ook bij of de variabele nominaal is, hoe missings gecodeerd zijn, en welk label de variabele heeft.

5.4.2 Variabelen hernoemen

Soms is het nodig je variabelen te hernoemen. Probeer hiervoor zo simpel en systematisch mogelijk te werk te gaan. Het is niet gebruikelijk om spaties te gebruiken, in plaats daarvan kun je een laag streepje gebruiken (`_`) of door elk nieuw woord met een hoofdletter te laten beginnen. Dit heet `lowerCamelCase` als het eerste woord met een kleine letter begint, en `UpperCamelCase` als het eerste woord ook met een hoofdletter begint.

In de syntax in SPSS kun je dat als volgt doen:

```
RENAME VARIABLES oudenaam = NieuweNaam.
```

5.4.3 Categorische variabelen onderzoeken

Om te bekijken hoe een categorische variabele verdeeld is, kun je frequentieverdelingen en staafdiagrammen opvragen. Dit kan met het volgende commando.

```
FREQ VARIABLES=Education Cursus
  /BARCHART FREQ.
```

Het laatste argument, `/BARCHART FREQ`, geeft aan dat ook de staafdiagram weergegeven moet worden; dit kun je eventueel weglaten. In plaats van `FREQ` kun je overigens ook `FREQUENCIES` gebruiken.

Het opvragen van een frequentieverdelingen kun je ook via het menu doen door te gaan naar `Analyze -> Descriptive statistics -> Frequencies`. Vervolgens selecteer je de variabelen die je wilt weergeven en geef je aan welke grafiek je wilt zien.

5.4.4 Continue variabelen onderzoeken

Met continue variabelen kun je vaak een stuk meer kanten op. Dit houdt in dat er vaak ook meerdere manieren te zijn om tot hetzelfde resultaat te komen. Wil je inzoomen op één variabele en hier veel informatie over, dan kun je dat het beste doen via `Analyze -> Descriptive statistics -> Explore`. Hier kun je variabelen selecteren, en vervolgens kiezen uit allerlei verschillende statistieken en plots. Om beschrijvingsmaten, een histogram en een boxplot op te vragen, kun je het volgende commando gebruiken.

```
DATASET ACTIVATE dat.
EXAMINE VARIABLES=Age
  /PLOT BOXPLOT HISTOGRAM.
```

De resultaten van dit EXAMINE commando kun je ook apart opvragen voor de verschillende niveaus van een categorische variabele, oftewel per subgroep. Dit kan door een BY-variabele te specificeren.

```
DATASET ACTIVATE dat.  
EXAMINE VARIABLES=Age BY Education  
/PLOT=BOXPLOT.
```

Wil je alleen een histogram van de data, dan kan dat makkelijk op twee manieren. Je kunt opnieuw gebruik maken van het `FREQ` commando, of je kunt het via Graphs doen. Als je dit via het menu wilt doen, ga je naar Graphs -> Legacy dialogs -> Histogram. Als je `FREQ` gebruikt, kun je ervoor zorgen dat de frequentietabel niet wordt weergegeven door ruimte te besparen. Dit doe je als volgt:

```
FREQUENCIES VARIABLES=Age  
/FORMAT=NOTABLE  
/HISTOGRAM NORMAL.
```

```
GRAPH  
/HISTOGRAM=Age.
```

Wil je veel variabelen tegelijk bekijken of makkelijk vergelijken, dan kun je ook gebruik maken van Descriptives in plaats van Explore. Ook dit vind je bij Analyze -> Descriptive Statistics.

```
DESCRIPTIVES VARIABLES=Statken1 Statken2 Angst1 Angst2  
/STATISTICS=MEAN STDDEV MIN MAX.
```


Chapter 6

Ontbrekende waarden

6.1 Intro

Bij het verzamelen van data, zullen er vaak ontbrekende data (missings) zijn. Hier kan op heel veel manieren mee worden omgegaan. Hier worden enkele basale aanpakken getoond. Allereerst de listwise en pairwise verwijderingsmethoden en daarna imputatiemethoden.

6.1.1 Voorbeeld data

Voor dit voorbeeld zullen we de `contraProductiefGedrag` dataset gebruiken. Deze dataset staat beschreven in het hoofdstuk Datasets. We zullen ons voornamelijk focussen op het verbanden tussen `leeftijd`, `vertrouwen` en `cynisme`.

6.2 jamovi

Jamovi gebruikt als standaard de pairwise verwijdering bij correlaties. Dit is te zien aan de `n`, die bij elk paar anders kan zijn als er missing waarden zijn.

```
jmv::corrMatrix(  
  data = data,  
  vars = vars(leeftijd, vertrouwen, cynisme),  
  n = TRUE,  
  plots = TRUE)
```

Om een analyse op complete observaties (listwise verwijderen) te doen, kan het symbool “Filters” in het tabblad “data” worden gebruikt. Definieer hierin de missings in de drie variabelen:

```
= leeftijd != NA and vertrouwen != NA and cynisme != NA
```

Hierdoor worden alle regels waarin een van deze drie variabelen een missing heeft verwijderd, dat wil zeggen niet meegenomen in de analyse. Dit is te zien aan de lichtgrijze arcering van deze regels.

6.2.1 Imputeren

6.3 R

Voor de berekenen van correlaties kan je het subcommando `use` gebruiken in de `cor` functie om aan te geven hoe er met misisngsmoet worden omgegaan. Pairwise verwijdering gaat via `use= "pairwise.complete.obs"` en listwise via `use = "complete.obs"`.

```
cor(contraProductiefGedrag[,c("leeftijd","cynisme","vertrouwen")], use = "complete.obs")
```

6.3.1 Imputeren

In R bestaat het package “mice”

6.4 SPSS

Standaard zal SPSS de complete case analyse toepassen (listwise verwijderen). Met het subcommando `missing` kan je aangeven hoe je wil dat SPSS met de ontbrekende waarden omgaat. Bijvoorbeeld door daar “pairwise” in te vullen.

```
correlations leeftijd cynisme vertrouwen  
/missing pairwise.
```

6.4.1 Imputeren

Chapter 7

Variabelen samenvoegen

7.1 Intro

Een veel voorkomende taak is het samenvoegen van meerdere variabelen (kolommen in een dataset) tot één nieuwe variabele (kolom). U wilt bijvoorbeeld de gemiddelde score op de items van een vragenlijst berekenen.

7.1.1 Voorbeeld data

In dit voorbeeld wordt de Rosetta voorbeelddataset “pp15” gebruikt (zie hoofdstuk @ref(datasets) voor informatie over de datasets en hoofdstuk @ref(loading-data) voor uitleg over het laden van datasets).

Van deze dataset gebruikt dit voorbeeld variabelen “.

We zullen deze aggregeren in de variabele `highDose_attitude` (merk op dat deze variabele al in de dataset bestaat, en dat de bestaande variabele ook het gemiddelde is van die vijf variabelen).

7.2 jamovi

Klik op het tabblad “Data” op de knop “Compute” Typ de nieuwe variabelenaam in het tekstveld bovenaan, met het label “COMPUTED VARIABLE”. Klik vervolgens op de functiekноп, gemarkeerd met f_x , selecteer de MEAN-functie in het vak “Functies” en dubbelklik op alle variabelen waarvan u het gemiddelde wilt in het vak “Variabelen” (rechterkolom), terwijl u een komma typt tussen elke variabele naam. Als alternatief kunt u de functienaam en de lijst met variabelen rechtstreeks typen zonder het dialoogvenster functie (f_x) te gebruiken.

Als u ontbrekende waarden wilt toestaan, kunt u het `ignore_missing = 1` argument specificeren. In dat geval typ je:

```
MEAN(highDose_AttGeneral_good, highDose_AttGeneral_prettig,  
      highDose_AttGeneral_slim, highDose_AttGeneral_gezond,  
      highDose_AttGeneral_spannend, ignore_missing=1)
```

Het is nog niet mogelijk om aan te geven hoeveel geldige waarden vereist zijn; ofwel zijn er helemaal geen missings toegestaan, ofwel wordt een willekeurig aantal ontbrekende waarden geaccepteerd.

7.3 R

7.3.1 R: base R

```
dat$highdose_attitude <-  
  rowMeans(  
    dat[,  
      c(  
        'highDose_AttGeneral_good',  
        'highDose_AttGeneral_prettig',  
        'highDose_AttGeneral_slim',  
        'highDose_AttGeneral_gezond',  
        'highDose_AttGeneral_spannend'  
      )  
    ]  
  );
```

7.3.2 R: Rosetta

```
dat$highdose_attitude <-  
  rosetta::means(  
    data = dat,  
    'highDose_AttGeneral_good',  
    'highDose_AttGeneral_prettig',  
    'highDose_AttGeneral_slim',
```

```
'highDose_AttGeneral_gezond',
'highDose_AttGeneral_spannend'
);
```

Om aan te geven dat een bepaald aantal waarden geldig moet zijn (d.w.z. “niet ontbreekt”), kan het argument “requiredValidValues” worden doorgegeven. Om bijvoorbeeld vier geldige waarden te vereisen (in plaats van slechts één geldige waarde, de standaardwaarde), gebruikt u:

```
dat$highdose_attitude <-
  rosetta::means(
    data = dat,
    'highDose_AttGeneral_good',
    'highDose_AttGeneral_prettig',
    'highDose_AttGeneral_slim',
    'highDose_AttGeneral_gezond',
    'highDose_AttGeneral_spannend',
    requiredValidValues = 4
  );
```

7.4 SPSS

```
COMPUTE highdose_attitude =
  MEAN(
    highDose_AttGeneral_good,
    highDose_AttGeneral_prettig,
    highDose_AttGeneral_slim,
    highDose_AttGeneral_gezond,
    highDose_AttGeneral_spannend
  ).
```

Om aan te geven dat een bepaald aantal waarden geldig moet zijn (d.w.z. “niet-ontbrekend”), kan het commando “MEAN” worden toegevoegd met een punt en het aantal vereiste geldige waarden. Gebruik bijvoorbeeld om vier geldige waarden te vereisen:

```
COMPUTE highdose_attitude =
  MEAN.4(
    highDose_AttGeneral_good,
    highDose_AttGeneral_prettig,
```

```
highDose_AttGeneral_slim,  
highDose_AttGeneral_gezond,  
highDose_AttGeneral_spannend  
).
```

Chapter 8

Variabelen Standandaardiseren

8.1 Intro

Een variabele standaardiseren betekent het gemiddelde aftrekken van elke score in de gegevensreeks en de resulterende getallen delen door de standaarddeviatie van de variabele. Het resultaat is een variabele met een gemiddelde van 0 en een standaarddeviatie van 1. Centreren van een variabele betekent dat alleen het gemiddelde (of een andere waarde) van de de scores wordt afgetrokken.

8.1.1 Voorbeeld data

Dit voorbeeld gebruikt de variabele `xtcUsePillHigh` uit de voorbeelddataset “pp15” (zie Hoofdstuk 3 en Hoofdstuk 4).

8.2 jamovi

Ga naar de **Data** tabblad en klik daar op **Compute**. Hiermee kan er een nieuwe variabele worden berekend, die u een zelf gekozen naam kunt geven en waaraan u een beschrijving kunt toevoegen. Klik vervolgens op het functie-symbool f_x en scroll in de linkerkolom naar de functie **Z** onder het kopje “Statistical”. Dubbelklik hierop, waarna er $= Z()$ verschijnt in het venster. Scroll nu in de rechterkolom naar de variabele waarvan u de **Z**-score wil berekenen (bv `xtcUsePillHigh`) en dubbelklik daarop. In het venster staat nu $= Z(xtcUsePillHigh)$. Als u nu weer naar bijvoorbeeld het analyses tabblad gaat, dan is de nieuwe gestandaardiseerde variabele verschenen in het databestand. In deze instructievideo wordt de procedure ook uitgelegd.

Wanneer een variabele alleen gecentreerd moet worden, bijvoorbeeld rondom om het gemiddelde, dan moet ook weer **Compute** worden gebruikt. Geef weer een naam (b.v. `xtcUsePillHigh_centered`) en dubbelklik op het functie-symbool f_x . Selecteer in de rechterkolom de naam van de variabele die gecentreerd moet worden. Deze verschijnt dan in het venster. Type nu zelf een minteken achter deze naam en klik dan weer op het functie-symbool f_x . Selecteer nu in de linkerkolom de functie **VMEAN** (die het gemiddelde van een variabele berekent), dubbelklik en kies vervolgens in de rechterkolom weer de naam van de variabele. In het venster staat nu: `= xtcUsePillHigh-VMEAN(xtcUsePillHigh)`. Gaat u nu weer terug naar de analyses dan is de nieuwe gecentreerde variabele in het databestand verschenen.

8.3 R

Deze code slaat de gestandaardiseerde waarden op in een variabele genaamd `xtcUsePillHigh_standardized`:

```
dat$xtcUsePillHigh_standardized <-
  scale(dat$xtcUsePillHigh);
```

In R is het ook gemakkelijk om een variabele rond zijn gemiddelde te centreren (d.w.z. laat de deling door de standaarddeviatie uit de standaardisatieprocedure weg). De volgende opdracht slaat de gecentreerde waarden op in een variabele met de naam `xtcUsePillHigh_centered`:

```
dat$xtcUsePillHigh_centered <-
  scale(dat$xtcUsePillHigh, scale = FALSE);
```

8.4 SPSS

Dit commando vraagt om variabelebeschrijvingen, maar het `/SAVE` subcommando slaat ook de gestandaardiseerde waarden op. Deze krijgen dan de originele variabelenaam, voorafgegaan door 'Z', dus in dit geval 'ZxtcUsePillHigh', die automatisch wordt toegevoegd aan het databestand:

```
DESCRIPTIVES VARIABLES = xtcUsePillHigh
  /SAVE.
```

Chapter 9

Hercoderen van variabelen

9.1 Intro

Een veel voorkomende manier van het hercoderen van items is omschalen (inversion). Bijvoorbeeld, als antwoorden zijn gescoord op een schaal van 1-5, dan zal na hercodering, 1 een 5 worden, 2 een 4, 4 een 2, en 5 een 1.

Dit kan op verschillende manieren:

1. Neem de maximum waarde +1 en trek daar alle waarden van af (dus om scores op een 1-5 schaal om te schalen, trekt u alle waardes af van 6 om tot de hergecodeerde waarden te komen)
2. Voor elke waarde apart specificeren wat de nieuwe waarde moet zijn.

De tweede methode kun je gebruiken voor alle varianten van hercoderen, bijvoorbeeld als u groepen in categorieën wilt indelen of scores handmatig wilt wijzigen.

In dit hoofdstuk zullen beide methoden aan bod komen. Houd bij het maken van de nieuwe namen voor de variabelen de richtlijnen in gedachte voor het maken van geschikte namen.

9.1.1 Voorbeeld data

Voor dit voorbeeld zal gebruik gemaakt worden van de Rosetta stats voorbeeld-dataset “pp15”.(zie Hoofdstuk 3 en Hoofdstuk 4. De variabele die zal worden hergecodeerd is `highDose_IntentionRAA_intention`. Deze variabele heeft 7 antwoordmogelijkheden.

9.2 jamovi

9.2.1 aftrekken

Ga naar de *Data* tabblad en klik daar op *Compute*. Hiermee kan er een nieuwe variabele worden berekend, die u een zelf gekozen naam kunt geven en waaraan u een beschrijving kunt toevoegen. In het tekstveld rechtsonder kunt u vervolgens aangeven hoe de nieuwe variabele gemaakt moet worden. In dit geval specificeert u het getal waar u alle waarden van af wil trekken (voor het voorbeeld, 8), een minteken, en de variabele die aangepast wordt (`highDose_IntentionRAA_intention`). Het gehele tekstvak zegt dan dus: `8 - highDose_IntentionRAA_intention`

9.2.2 handmatig specificeren

Ga naar de *Data* tabblad en klik daar op *Transform*. In het dropdown menu gelabeld “Source variable” kunt u de variabele selecteren die u wilt aanpassen. In het tekstveld bovenaan kunt u de nieuwe naam opgeven van de getransformeerde variabele. Vervolgens kunt u in het dropdown menu gelabeld “using transform” op “Create new transform...” klikken. Als je deze transformatie wilt toepassen op meerdere variabelen, dan kunt u deze transformatie een bepaalde naam geven in de box onder “TRANSFORM”. Klik dan op “Add recode condition” om een nieuwe transformatie toe te voegen. Voeg net zo veel condities toe als het aantal labels dat u wilt aanpassen (in ons voorbeeld zijn dat er 7). Specificeer vervolgens de link tussen `$source` en `use` en welke waarden daarbij horen. Voor dit voorbeeld krijg je de volgende 7 condities:

```
if $source == 1 use 7
if $source == 2 use 6
if $source == 3 use 5
if $source == 4 use 4
if $source == 5 use 3
if $source == 6 use 2
if $source == 7 use 1
```

In de laatste rij staat “`else use`”, dit kunt u gebruiken om een waarde te geven aan alle overige observaties die niet binnen één van deze categorieën vallen. Zo zou missing data bijvoorbeeld de waarde 99 kunnen krijgen.

9.3 R

9.3.1 aftrekken

Om alle waarden van de maximum waarde +1 (in dit geval, 8) af te trekken, kun je het volgende commando gebruiken (let erop dat voor dit voorbeeld de dataset onder de naam `dat` moet zijn geïmporteerd):

```
dat$highDose_IntentionRAA_intention_recoded <-  
  8 - dat$highDose_IntentionRAA_intention;
```

9.3.2 handmatig specificeren

Om handmatig data te hercoderen kun je het volgende commando gebruiken (hiervoor moet de rosetta package zijn geïnstalleerd en de data onder de naam `dat` zijn opgeslagen):

```
dat$highDose_IntentionRAA_intention_recoded <-  
  rosetta::recode(  
    dat$highDose_IntentionRAA_intention,  
    "1=7; 2=6; 3=5; 4=4; 5=3; 6=2; 7=1"  
  );
```

9.4 SPSS

Zorg ervoor dat de juiste dataset is geactiveerd.

9.4.1 aftrekken

Om alle waarden van de maximum waarde +1 af te trekken (in dit geval, 8), kun je het volgende in de syntax typen:

```
COMPUTE highDose_IntentionRAA_intention_recoded =  
  8 - highDose_IntentionRAA_intention.
```

9.4.2 handmatig specificeren

Om handmatig alle waarden aan te passen, kun je het volgende commando gebruiken:

```
RECODE highDose_IntentionRAA_intention  
  (1=7) (2=6) (3=5) (4=4) (5=3) (6=2) (7=1)  
  INTO highDose_IntentionRAA_intention_recoded.
```

Chapter 10

Coefficient Alpha

10.1 Intro

Coefficient Alpha, beter bekend als Cronbach's Alpha, is een maat voor de interne consistentie, en kan als er aan bepaalde assumpties wordt voldaan de betrouwbaarheid van een set items bepalen. Een andere maat waarvoor minder strenge eisen gelden is McDonalds Omega, die in jamovi en R ook worden getoond wanneer je Alpha opvraagt.

10.1.1 Voorbeeld data

Dit voorbeeld gebruikt één van de datasets uit de Rosettat Stats package, namelijk "pp15". (Zie Hoofdstuk 3 voor informatie over de datasets en 4 over een uitleg hoe je datasets moet importeren). Uit deze dataset gaan we de Coefficient Alpha uitrekenen voor een groep items over hetzelfde onderwerp. Hiervoor gebruiken we `highDose_AttGeneral_good`, `highDose_AttGeneral_prettig`, `highDose_AttGeneral_slim`, `highDose_AttGeneral_gezond`, `highDose_AttGeneral_spannend`

10.2 jamovi

In het tabblad "Analyses", klik op de knop "Factor" en uit het menu kies dan "Reliability Analysis". Je kunt nu alle variabelen die je wilt gebruiken voor de analyse verplaatsen naar het vak dat "Items" heet. Coefficient Alpha (wordt nog Cronbach's Alpha genoemd door jamovi) staat automatisch aangevinkt, waardoor je meteen aan de rechterkant de uitkomsten zult zien. Links onderin kun je nog verschillende andere opties aanvinken. Zo kun je bij "Item statistics" bijvoorbeeld de waarde van Coefficient Alpha opvragen bij het verwijderen van

één item, onder Cronbach's Alpha (if item deleted). Ook kun je via “Reverse Scaled items” aangeven of sommige items in je lijst andersom gecodeerd zijn, zonder dat je ze handmatig hoeft te hercoderen. Als laatste kun je jamovi ook nog vragen om direct een somscore of gemiddelde score van de items op te slaan.

In syntax ziet het verkrijgen van Coefficient Alpha er als volgt uit in jamovi:

```
jmv::reliability(
  data = data,
  vars = vars(highDose_AttGeneral_good, highDose_AttGeneral_prettig, highDose_AttGen
```

10.3 R

In R kunnen we met het rosetta package ook de waarde van coefficient alpha uitrekenen door de functie `reliability` te gebruiken.

```
rosetta::reliability(
  data = dat,
  items = c(
    "highDose_AttGeneral_good",
    "highDose_AttGeneral_prettig",
    "highDose_AttGeneral_slim",
    "highDose_AttGeneral_gezond",
    "highDose_AttGeneral_spannend"
  )
);
```

10.4 SPSS

In SPSS kun je Coefficient Alpha uitrekenen door naar Reliability Analysis te gaan in het tabblad Analyse. In de Syntax gebruik je dan het volgende commando:

```
RELIABILITY
/VARIABLES =
  highDose_AttGeneral_good
  highDose_AttGeneral_prettig
  highDose_AttGeneral_slim
  highDose_AttGeneral_gezond
  highDose_AttGeneral_spannend
/MODEL =
  ALPHA.
```

Je kan ook hier weer extra informatie opvragen, zoals beschrijvende statistieken, de inter-item correlaties en andere statistieken over de schaal. Daarvoor kun je extra regels toevoegen aan de syntax, zoals hier:

```
RELIABILITY
/VARIABLES =
    highDose_AttGeneral_good
    highDose_AttGeneral_prettig
    highDose_AttGeneral_slim
    highDose_AttGeneral_gezond
    highDose_AttGeneral_spannend
/MODEL = ALPHA
/STATISTICS = DESCRIPTIVE SCALE CORR
/SUMMARY = TOTAL.
```


Part III

Analyses met twee variabelen

Chapter 11

Kruistabellen

11.1 Intro

In een kruistabel worden de frequenties van twee variabelen gezamenlijk getoond. Deze variabelen zijn categorisch, met een nominaal of ordinaal meetniveau. De tabel toont de frequentieverdeling van de ene variabele voor elke categorie van de andere variabele en vice versa. Om de samenhang tussen de twee categorische variabelen te quantificeren kun je **phi** statistiek of **Cramer's V** gebruiken, en om te toetsen of het verband statistisch significant is kun je de chi-kwadraattoets gebruiken.

11.1.1 Voorbeeld dataset

Voor dit voorbeeld zullen we de politieke partij en media dataset gebruiken. Deze dataset bevat informatie over de voorkeuren van kiezers van bepaalde politieke partijen voor een bepaald medium.

11.2 jamovi

Klik in het tabblad “Analyses” op “Frequencies”. Vervolgens staat onder het kopje **Contingency tables** de optie “Independent Samples - Chi-square test of association”. Hier kun je de twee variabelen specificeren die je in de rijen en kolommen van de kruistabel wilt hebben. Nu verschijnt er direct een tabel met onderaan de chi-kwadraat toets. Iedere regel telt nu één keer mee, maar als dat niet is wat je wil, sleep dan een variabele die de frequentie aangeeft naar **Counts (optional)**, omdat deze variabele aangeeft hoe vaak de combinatie voorkomt. Indien je extra informatie wilt opvragen, kun je dat doen bij **Statistics, Cells**

of **Plots**. Bij **Statistics** kun je bijvoorbeeld ook phi of Cramer's V aanklikken, bij **Cells** kun je percentages opvragen of de verwachte waarden per cel en bij **Plots** kun je een staafdiagram maken om de data te visualiseren.

Hieronder zie je de achterliggende R syntax van de kruistabel zonder extra opgevraagde informatie.

```
jmv::contTables(
  formula = ~ partij:media,
  data = data)
```

11.3 R

Met het rosetta package kun je gebruik maken van `crossTab` om een kruistabel te maken en de bijbehorende chi-kwadraat toets te zien en de waarde van Cramer's V.

```
rosetta::crossTab(Politiek_Media$partij,Politiek_Media$media)
```

11.4 SPSS

In SPSS kun je kruistabellen maken via "Descriptive Statistics" en "Crosstabs". Ook hier kun je allerlei extra opties aanvinken via de menu's aan de zijkant. Automatisch krijg je alleen de kruistabel zelf. Onder "Statistics" kun je de chi-kwadraat toets en de effectmaten phi en Cramer's V aanvinken. In de Syntax ziet dat er dan als volgt uit.

```
CROSSTABS
  /TABLES=partij BY media
  /FORMAT=AVALUE TABLES
  /STATISTICS=CHISQ PHI
  /CELLS=COUNT
  /COUNT=ROUND CELL.
```

Als je alleen de kruistabel wilt, dan haal je simpelweg de regel met `/STATISTICS` uit je syntax. Wil je bijvoorbeeld de verwachte waarden toevoegen, dan kun je bij `/CELLS` achter `COUNT` ook nog `EXPECTED` schrijven.

Chapter 12

T-test

12.1 Intro

Om het verschil tussen de twee onafhankelijke groepen te toetsen wordt de t-test gebruikt (independent samples t-test). Om een verschil in gemiddelden tussen twee variabelen te testen wordt de gepaarde t-test gebruikt (paired samples t-test).

12.1.1 Voorbeeld data

Voor dit voorbeeld zullen we de `statistiekangst` dataset gebruiken. Deze dataset staat beschreven in het hoofdstuk Datasets.

12.2 jamovi

Voor het vergelijken van twee groepen kies in het tabblad “Analyses” de “T-tests”. Kies de **Independent Samples T-Test**. Sleep de afhankelijke variabele, in dit geval de toename (of afname) van de statistiekangst (`dif_Angst`) naar het blok **Dependent Variables** en de variabele `Cursus` naar de **Grouping Variable**. Vervolgens kun je bij “Additional Statistics” ook “Effect Size” aanklikken, zie het hoofdstuk over Cohen’s d (13) voor meer informatie. De test die standaard is aangevinkt in jamovi is Student’s t . Deze test gaat ervan uit dat de varianties van de groepen aan elkaar gelijk zijn. Om een analyse te doen waarbij je deze assumptie niet maakt, kun je Welch’s t aanvinken. Je kunt ze ook allebei aanvinken, zo kun je zien wat het verschil is.

De onderliggende R-code voor een Student’s t -test ziet er zo uit:

```
jmv::ttestIS(
  formula = dif_Angst ~ Cursus,
  data = data,
  vars = dif_Angst,
  effectSize = TRUE)
```

Het toevoegen van Cohen's *d* aan de berekening van een t-test is vrij gemakkelijk. Je hoeft alleen een extra regel toe te voegen met `effectSize = TRUE`.

Voor het vergelijken van twee variabelen kies je in het tabblad "Analyses" de "T-tests". Kies vervolgens de **Paired Samples T-Test**. Sleep de twee afhankelijke variabelen, die je wil vergelijken, in dit geval de statistiekangst voor en na de interventie (`Angst1` en `Angst2`) naar het blok **Paired Variables**. Je kan ook de **Effects size** optie met het betrouwbaarheidsinterval aanvinken.

De onderliggende R-code ziet er zo uit:

```
jmv::ttestPS(
  data = data,
  pairs = list(
    list(
      i1="Angst1",
      i2="Angst2")),
  effectSize = TRUE,
  ciES = TRUE)
```

12.3 R

In R zijn er diverse packages waarmee je een t-test voor onafhankelijke groepen uit kunt rekenen. Om het verschil in de toename of afname in statistiekangst te toetsen tussen beide cursussen in basis R:

```
t.test(statistiekangst$dif_Angst ~ statistiekangst$Cursus)
```

Hierbij wordt er standaard vanuit gegaan dat de varianties ongelijk zijn.

In het `rosetta` package kan de volgende functie worden gebruikt:

```
rosetta::meanDiff(statistiekangst$dif_Angst ~ statistiekangst$Cursus, var.equal = "no")
```

Hierbij is nu aangegeven met de optie `var.equal = "no"` dat de varianties in beide groepen niet gelijk is. Dit is beter, maar als je zeker weet dat de varianties wel gelijk zijn, dan kun je deze parameter wijzigen naar `"yes"`.

De gepaarde t-test kan in basis R simpelweg met dezelfde functie worden uitgevoerd als de onafhankelijke t-test, met de extra parameter `paired = TRUE`:

```
t.test(statistiekangst$Angst1, statistiekangst$Angst2, paired = TRUE)
```

12.4 SPSS

De syntax in SPSS voor de onafhankelijke t-test is als volgt.

T-TEST

```
/VARIABLES = dif_Angst  
/GROUPS = Cursus(1,2).
```

SPSS geeft je bij een t-test niet de mogelijkheid om te kiezen of de varianties gelijk of ongelijk zijn. In plaats daarvan krijg je altijd allebei de opties in één tabel. Je kunt in de output zien dat de eerste regel ervan uitgaat dat de varianties gelijk zijn, en de tweede regel ongelijk.

Voor de gepaarde t-test kent SPSS de volgende syntax.

```
T-TEST PAIRS = Angst1 WITH Angst2.
```

12.5 APA style report

Voor het rapporteren van een t-test zijn standaarden opgesteld. Hier worden de APA-7 regels toegepast. Bij het uitvoeren van een t-test is het gebruikelijk om ook de effectmaat Cohen's d te rapporteren. Meer informatie hierover vind je in het hoofdstuk Cohen's d (13). Enkele voorbeelden van het rapporteren van de resultaten staan hieronder:

Ouderen hadden significant meer last van eenzaamheid dan jongeren, $t(32) = 2.94$, $p = .006$, $d = 0.76$, 95% CI [0.46, 1.06].

De 25 participanten in de experimentele conditie ($M = 480$, $SD = 34.5$) hadden in vergelijking met de 28 participanten in de controle-groep ($M = 425$, $SD = 31$) een significant betere score op de test, $t(51) = 2.1$, $p = .04$, $d = 0.34$, 95% CI [.05, .63].

Chapter 13

Cohen's d

13.1 Intro

Cohen's d is het verschil tussen de twee groepen, maar dan gecorrigeerd voor de standaarddeviatie. De standaarddeviatie is een maat voor de meetschaal voor de betreffende variabele. Door het verschil tussen de gemiddelden hierdoor te delen krijgen we een schaalonafhankelijke maat voor het verschil tussen de groepen. Cohen's d wordt berekend door het verschil te nemen tussen twee gemiddelden of tussen een gemiddelde en een populatiewaarde (μ) en dat te delen door een standaarddeviatie. Bij het gebruik van één gemiddelde wordt de standaarddeviatie van deze sample gebruikt, als het om een verschil tussen twee gemiddelden gaat wordt een gepoolde standaarddeviatie van beide groepen gebruikt.

13.1.1 Voorbeeld data

Voor dit voorbeeld zullen we de `schoolsucces` dataset gebruiken. Deze dataset staat beschreven in het hoofdstuk Datasets. We zullen voor de Cohen's D nu voornamelijk focussen op de variabele IQ.

13.2 jamovi

In jamovi is Cohen's d geen losse berekening, maar zit dit standaard ingebouwd in de uitwerking van de t-toetsen. Die kun je vinden in het tabblad "Analyses" onder "T-tests". We gaan hier eerst kijken naar Cohen's d voor een one-sample t-toets. Zo kunnen we kijken of de gemiddelde IQ score in onze steekproef verschilt van het standaard gemiddelde (100). Doe dit door IQ in het vak rechts

te slepen zodat de analyse wordt uitgevoerd. Vervolgens kun je bij “Additional Statistics” ook “Effect Size” aanklikken. Nu zie je direct rechts in de uitwerking de waarde van Cohen's d verschijnen.

```
jmv::ttestOneS(
  data = data,
  vars = IQ,
  testValue = 100,
  effectSize = TRUE)
```

Cohen's d kun je ook berekenen voor de paired samples t-toets en de independent samples t-toets. Zo kunnen we bijvoorbeeld kijken hoe groot het verschil in IQ is tussen jongens en meisjes. Klik hiervoor in de categorie “T-tests” op “independent samples t-test”. Nu verschijnt er een tweede vak, waar je geslacht heen kunt slepen, genaamd “Grouping variable”.

```
jmv::ttestIS(
  formula = IQ ~ geslacht,
  data = data,
  vars = IQ,
  effectSize = TRUE)
```

Zoals je misschien al kunt zien is het toevoegen van Cohen's d aan de berekening van een t-toets vrij gemakkelijk. Je hoeft alleen een extra regel toe te voegen met `effectSize = TRUE`.

13.3 R

In R zijn er diverse packages waarmee je Cohen's d uit kunt rekenen. Elke package heeft zijn eigen extra mogelijkheden die handig kunnen zijn voor jouw specifieke doeleinden. Voor het vergelijken van twee gemiddelden is er binnen het rosetta package de mogelijkheid om de functie `meanDiff` te gebruiken, hiermee wordt ook automatisch Cohen's d berekend. Wil je bijvoorbeeld IQ vergelijken tussen jongens en meisjes, dan kan dat als volgt:

```
rosetta::meanDiff(schoolsucces$IQ ~ schoolsucces$geslacht)
```

Dit werkt echter alleen voor het vergelijken van twee groepen. Wil je een enkele score vergelijken met een vaste waarde (μ), dan kun je daar bijvoorbeeld het package `rstatix` voor gebruiken.

De formule om de IQ score uit de sample te vergelijken met een vaste waarde van 100 is als volgt:


```
cohens_d(data=schoolsucces, IQ ~ 1, mu=100)
```

Met deze package kun je ook direct meerdere groepen vergelijken, zonder dat je steeds per paar een t-toets hoeft uit te voeren. In het geval dat je een categorische variabele hebt met meer dan twee opties, dan kun je met het argument `comparisons` aangeven welke groepen je met elkaar wilt vergelijken, bijvoorbeeld `comparisons = list(c("A", "B"), c("B", "C"))` als je groep A met groep B wilt vergelijken en groep B met groep C.

13.4 SPSS

Het is in SPSS niet mogelijk om Cohen's d uit te rekenen.

Chapter 14

Correlaties

14.1 Intro

De correlatiecoëfficiënt of simpelweg correlatie of Pearson's r drukt uit hoe sterk twee continue variabelen (dus minimaal intervalniveau) met elkaar samenhangen. Correlaties worden voor van alles gebruikt en zijn vaak een eerste stap richting meer complexere analyses. Naast de Pearson r bestaan er ook andere vormen van correlaties die vaak op dezelfde manier met slechts kleine aanpassingen in de uitvoering ook berekend kunnen worden, zoals de Spearman's ρ .

14.1.1 Voorbeeld data

Om correlaties te illustreren zullen we de Palmer Pinguïns dataset gebruiken. Deze dataset bevat informatie over drie verschillende pinguïnsoorten. Er is meer informatie beschikbaar op <https://allisonhorst.github.io/palmerpenguins> om de data te downloaden of te bekijken.

14.2 jamovi

Om de samenhang tussen flipperlengte en gewicht te visualiseren, maken we eerst een scatterplot. Klik in het tabblad "Analyse" op "Exploration" en vervolgens op "Scatterplot". Vul de juiste variabelen in bij de X-as en de Y-as. In de syntax ziet het er dan als volgt uit.

```
scatr::scat(  
  data = data,  
  x = flipper_length_mm,  
  y = body_mass_g)
```

Om vervolgens de correlatie uit te rekenen, klik op “Regression” en dan “Correlation matrix”. Kies hier alle variabelen waarvan je de samenhang wilt berekenen. In de verschillende opties onderaan kun je kiezen welke coëfficiënt berekend moet worden, en of je extra informatie wilt zoals de betrouwbaarheidsintervallen.

```
jmv::corrMatrix(
  data = data,
  vars = vars(flipper_length_mm, body_mass_g))
```

14.3 R

Om de samenhang tussen flipperlengte en gewicht te visualiseren, kunnen we eerst een scatterplot maken. Let erop dat de pinguïndataset opgeslagen is onder de naam `data`.

```
rosetta::ggScatterPlot(data$flipper_length_mm,data$body_mass_g)
```

Vervolgens zijn er in R verschillende opties voor het berekenen van de correlatiecoëfficiënt. Voor het verkrijgen van de coëfficiënt zelf, kun je `cor()` gebruiken.

```
cor(data$flipper_length_mm,data$body_mass_g, use="complete.obs", method=c("pearson"))
```

De functie `cor` gaat ervan uit dat er geen missing data in de dataset is. Indien dat wel het geval is zul je met de optie `use` aan moeten geven of je pairwise of listwise deletion wilt gebruiken. Bij `method` kun je kiezen uit Pearson, Spearman of Kendall.

Voor het toetsen van één correlatiecoëfficiënt kun je ook base R gebruiken, maar wil je statistische inferenties maken over meerdere coëfficiënten tegelijk en bijvoorbeeld p-waarden of betrouwbaarheidsintervallen verkrijgen, dan heb je andere packages nodig, zoals `Hmisc`.

```
cor.test(data$flipper_length_mm,data$body_mass_g, use="complete.obs", method=c("pearson"))

library(Hmisc)
rcorr(data$flipper_length_mm,data$body_mass_g, type="pearson")
```

14.4 SPSS

Om de samenhang tussen flipperlengte en gewicht te visualiseren, maken we eerst een scatterplot met het volgende commando.

GRAPH

```
/SCATTERPLOT(BIVAR)=flipper_length_mm body_mass_g  
/MISSING=LISTWISE.
```

Vervolgens kunnen we via “Correlate” en “Bivariate” ook een correlatiematrix maken. In de syntax gebruik je daarvoor het volgende commando.

CORRELATIONS

```
/VARIABLES=flipper_length_mm body_mass_g  
/PRINT=TWOTAIL NOSIG  
/MISSING=LISTWISE.
```


Chapter 15

ANOVA

15.1 Intro

Een ANOVA, ook wel genaamd variantieanalyse, is een manier om twee of meer groepen met elkaar te vergelijken. De ANOVA ligt in het verlengde van de t-test voor onafhankelijke steekproeven, maar in plaats van dat slechts twee gemiddelden worden getoetst, kun je er met een anova meer toetsen. Bij een ANOVA is de uitkomstvariabele een continue variabele, en de predictorvariabele is een categorische variabele, ook wel factor genoemd. Als je slechts één factor gebruikt, dan noemen we dat een oneway ANOVA. Verder bestaan er veel uitbreidingen op de ANOVA, zoals de factoriële ANOVA en de ANOVA voor herhaalde metingen, en nonparametrische alternatieven voor de ANOVA. Hier zullen we kijken naar de one-way ANOVA en de factoriële ANOVA. Bij een factoriële ANOVA heb je twee of meer hoofdeffecten, maar ook interactieeffecten. Deze effecten kijken of de uitkomst verschilt voor een bepaalde combinatie van factoren.

15.1.1 Voorbeeld dataset

In dit voorbeeld gaan we werken met de dataset `wiskunde`. Informatie over deze dataset is te vinden in hoofdstuk 3 en informatie over hoe je data kunt laden staat beschreven in hoofdstuk 4. voor dit voorbeeld gaan we een oneway ANOVA uitvoeren om te kijken of de wiskundescores verschillen tussen de 10 klassen. Daarnaast zullen we een factoriël model maken waarbij we kijken of de sekse van de leerling een effect heeft op de wiskundescore en of er mogelijk een interactieeffect is tussen de klas en de sekse.

15.2 jamovi

In het tabblad “Analyses” kun je klikken op “ANOVA”. Klik in het menu dat verschijnt vervolgens op “One-Way ANOVA”. Selecteer de variabele die je wilt vergelijken, in dit geval `wiskunde` en verplaats deze naar het vak “Dependent Variables”. Selecteer vervolgens de factor die je wilt gebruiken en verplaats deze naar het vak “Grouping Variable”. Nu verschijnt automatisch de F-test aan de rechterkant van het scherm. Automatisch staat de Welch’s ANOVA aangevinkt, die er vanuit gaat dat de varianties in de verschillende groepen niet aan elkaar gelijk zijn. Als je de assumpties hebt gecheckt en de varianties zijn wel gelijk, dan kun je dit aanpassen naar Fisher’s ANOVA. De assumpties kun je in jamovi direct opvragen door onder het kopje “Assumptions” de verschillende plots aan te klikken.

Daarnaast kun je ook andere informatie toevoegen, zoals de gemiddelden per groep, of post-hoc toetsen. Deze post-hoc toetsen zijn onafhankelijke t-toetsen die per paar kijken of die twee groepen van elkaar verschillen, aangezien de F-test alleen aangeeft dat er ergens een verschil is tussen alle groepen.

In de syntax ziet een One-Way ANOVA er als volgt uit:

```
jmv::anovaOneW(
  formula = wiskunde ~ klas,
  data = data,
  phFlag = TRUE,
  duplicate = 4)
```

Nu gaan we dit model uitbreiden door sekse toe te voegen. Open opnieuw het menu voor “ANOVA” maar klik nu op “ANOVA” in plaats van op “One-Way ANOVA”. Voeg opnieuw de afhankelijke variabele toe aan het vakje “Dependent Variable”. Voeg vervolgens de factors toe aan het vak “Fixed Factors”. Nu verschijnt automatisch een tabel met de twee hoofdeffecten en het interactieeffect. Indien gewenst kun je linksonderin ook de test voor het gehele model toevoegen.

Ook kun je allerlei extra informatie toevoegen, zoals effectmaten, assumptiechecks, of post-hoc toetsen.

De factoriële ANOVA met toevoeging van sekse ziet er als volgt uit in de syntax:

```
jmv::ANOVA(
  formula = wiskunde ~ klas + sekseLeerling + klas:sekseLeerling,
  data = data,
  modelTest = TRUE)
```


15.3 R

In R kun je met de rosetta package een ANOVA uitvoeren met de functie `fanova`. Met deze functie kun je zowel One-Way ANOVAs als factoriële ANOVAs uitvoeren afhankelijk van hoe je het model opbouwt. Daarnaast kun je voor One-Way Anovas ook `oneway` gebruiken. Voor ons voorbeeld kun je de volgende code gebruiken:

```
rosetta::oneway(y=dat$wiskunde, x=dat$klas)

rosetta::fanova(dat=dat, y='wiskunde', between = 'klas')
```

Om dit model uit te breiden kunnen we sekse toevoegen aan de formule. Dit levert de volgende code op:

```
rosetta::fanova(dat=dat, y='wiskunde', between='klas', 'sekseLeerling')
```

Let op: soms kan de functie `fanova` een foutmelding geven, met name als je een bestand hebt geïmporteerd vanuit SPSS met de `haven` package. Om zeker te weten dat de analyse werkt, kun je voor de zekerheid even je databestand (nogmaals) converteren naar een dataframe met `dat <- as.data.frame(dat)`.

15.4 SPSS

In SPSS zijn er twee manieren om een ANOVA uit te voeren. Een One-Way ANOVA kun je uitvoeren via het tabblad “Analyze”, “Compare Means” en dan “One-Way ANOVA”. Voor factoriële ANOVAs moet je echter naar “Analyze”, “General Linear Model”, en dan “Univariate”. Je kunt hiermee ook One-Way ANOVAs uitvoeren, en hoewel deze manier iets ingewikkelder is, heb je wel meer opties voor extra statistieken. Hieronder zie je eerst de syntax via de One-Way ANOVA manier en daaronder hoe je hetzelfde kunt doen via General Linear Model.

```
ONEWAY wiskunde BY klas
/MISSING ANALYSIS.
```

```
UNIANOVA wiskunde BY klas
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/CRITERIA=ALPHA(0.05)
/DESIGN=klas.
```

Om voort te bouwen op het model via General Linear Model kunnen we hier ook direct de tweede factor `sekseLeerling` aan toevoegen. Dit levert de volgende syntax op:

```
UNIANOVA wiskunde BY klas sekseLeerling
/METHOD=SSTYPE(3)
/INTERCEPT=INCLUDE
/CRITERIA=ALPHA(0.05)
/DESIGN=klas sekseLeerling klas*sekseLeerling.
```

15.5 APA style report

Voor het rapporteren van een ANOVA zijn verschillende regels opgesteld. Hier worden de APA-7 regels toegepast. Hieronder enkele voorbeelden voor het rapporteren van een ANOVA:

We vonden een statistisch significant effect van leeftijdsgroep op gebruik van sociale media, $F(3, 117) = 3.19$, $p = .026$, $\eta^2 = .21$.

Een eenweg ANOVA liet zien dat er een significant verschil was tussen heteroseksuele mensen ($M = 41.2$, $SD = 5.7$), biseksuele mensen ($M = 33.8$, $SD = 4.9$), en homoseksuele mensen ($M = 30.1$, $SD = 8.2$) wat betreft gevoelens van discriminatie, $F(2, 133) = 38.89$, $p < .001$, $\eta^2 = .34$.

Chapter 16

Lineaire regressieanalyse

16.1 Intro

Univariate regressieanalyse (met één predictor) kan handig zijn om een schatting te krijgen van een verandering in de afhankelijke variabele als functie van de predictorvariabele. In dit hoofdstuk richten we onze aandacht op predictorvariabelen die continu of dichotoom zijn.

16.1.1 Voorbeeld data

In dit voorbeeld wordt de `schoolsucces` dataset gebruikt. De afhankelijke variabele is schoolsucces gemeten met een schoolcijfer. Deze variabele gaan we proberen te voorspellen aan de hand van IQ.

16.2 jamovi

Klik in het tabblad “Analyse” op de knop “Regressie” en selecteer “Linear regression” in het menu dat verschijnt. Selecteer in het vak aan de linkerkant de afhankelijke variabele (`cijfer`) en verplaats deze naar het vak met het label “Afhankelijke variabele” met behulp van de knop met het naar rechts wijzende pijltje. Selecteer in het vak aan de linkerkant de numerieke voorspellers (hier: `IQ`) en verplaats ze naar het vak met het label “Covariaten” met behulp van de knop met het naar rechts wijzende pijltje. U kunt naar beneden scrollen om aanvullende analyses te specificeren, bijvoorbeeld om meer details over de coëfficiënten op te vragen door de sectie “Modelcoëfficiënten” te openen. Als u bijvoorbeeld het betrouwbaarheidsinterval voor de coëfficiënt en de

gestandaardiseerde (geschaalde) coëfficiënten wilt opvragen, vinkt u de bijbehorende selectievakjes aan.

Hieronder kunt u zien hoe een regressie er volgens de jamovi syntax uitziet.

```
jmv::linReg(  
  data = data,  
  dep = cijfer,  
  covs = IQ,  
  blocks = list(  
    list(  
      "IQ")),  
  refLevels = list())
```

16.3 R

In base R, de `lm` (lineair model) functie kan worden gecombineerd met de `summary` functie om de meest belangrijke resultaten te tonen. R beschouwt variabelen die factoren zijn automatisch als categorisch.

In het `rosetta` package, de `regr` functie maakt gebruik van R's `lm` functie, maar geeft de resultaten wat anders weer.

```
rosetta::regr(cijfer ~ IQ,  
             data=schoolsucces)
```

16.4 SPSS

In SPSS, wordt het `REGRESSION` commando gebruikt.

```
REGRESSION  
  /STATISTICS COEFF OUTS CI(95) R ANOVA  
  /DEPENDENT cijfer  
  /METHOD=ENTER IQ.
```

16.5 APA style report

Voor het rapporteren van een regressie zijn standaarden opgesteld. Hier worden de APA-7 regels toegepast. Enkele voorbeelden van het rapporteren van de resultaten staan hieronder:

Regressieanalyse is gebruikt om te testen of IQ-scores de cijfers van schoolkinderen konden voorspellen. De resultaten van de regressie wijzen aan dat IQ-scores 20% van de variantie in cijfers verklaarde, $R^2 = 0.2$, $F(1,298) = 76.71$, $p < .001$, 95% CI [.09, 0.29].

We hebben geprobeerd uitgaansgedrag te voorspellen aan de hand van een viertal variabelen. Het gehele model verklaarde 43.2% van de variantie, $R^2 = 0.432$, $F(3,165) = 36.4$, $p < .001$, 95% CI [.31, .55]. Extraversie was een significante voorspeller van uitgaansgedrag, $b = 2.76$, $p < .001$, evenals leeftijd, $b = -0.58$, $p = .021$.

Chapter 17

Non-parametrische toetsen

17.1 Intro

In sommige gevallen voldoe je niet aan de assumpties om de testen uit te voeren die je graag had willen doen. Zo geldt bijvoorbeeld voor de t-toetsen dat de variabele(n) normaal verdeeld moeten zijn. Indien dit niet het geval is, kun je alternatieven gebruiken. Eén van die alternatieven is het gebruik van non-parametrische toetsen. Voor een aantal veelvoorkomende statistische testen zijn nonparametrische alternatieven gemaakt. Zo heb je voor de gepaarde t-toets als alternatief de Wilcoxon's signed rank test, voor de onafhankelijke t-toets de Wilcoxon's rank sum test (ook wel Mann-Whitney U test genoemd), en voor de one-way ANOVA de Kruskal-Wallis test.

17.1.1 Voorbeeld dataset

Voor deze dataset zullen we gebruik maken van de dataset `statistiekangst`. Informatie over deze dataset is te vinden in 3 en informatie over hoe je data kunt laden staat beschreven in 4. Voor de Wilcoxon's signed rank test zullen we kijken of er een verschil is tussen de statistiekangst voor en na het volgen van een cursus. Voor de Wilcoxon's rank sum test zullen we kijken of er een verschil is in de toename in statistiekangst tussen de twee cursussen, en voor de Kruskal-Wallis test zullen we kijken of er een verschil is in initiële statistiekangst tussen mensen van een verschillend opleidingsniveau.

17.2 jamovi

17.2.1 Wilcoxon's signed rank

In het tabblad “Analyses”, klik op “T-tests”. Kies vervolgens voor “Paired samples T-test”. Linksonder is een kopje genaamd “Tests”. Hier staat automatisch `student's` aangevinkt. Dit is de normale t-test. Deselecteer deze en klik in plaats daarvan op `Wilcoxon rank`. Verplaats nu de score van het eerste meetmoment naar het vakje “Paired Variables” en vervolgens de score van het tweede meetmoment. Als het goed is staan deze scores nu naast elkaar in het vak en verschijnt rechts de uitkomst van de analyse.

De onderliggende syntax voor deze analyse is als volgt:

```
jmv::ttestPS(  
  data = data,  
  pairs = list(  
    list(  
      i1="Angst1",  
      i2="Angst2")),  
  students = FALSE,  
  wilcoxon = TRUE)
```

17.2.2 Mann-Whitney U

In het tabblad “Analyses”, klik op “T-tests”. Kies vervolgens voor “Independent samples T-test”. Linksonder is een kopje genaamd “Tests”. Hier staat automatisch `student's` aangevinkt. Dit is de normale t-test. Deselecteer deze en klik in plaats daarvan op `Mann-Whitney U`. Verplaats nu de afhankelijke variabele naar het vak “Dependent Variables” en verplaats de dichotome variabele naar “Grouping Variables”. Nu verschijnt aan de rechterhelft automatisch de uitkomst van de test.

De onderliggende syntax voor deze analyse is als volgt:

```
jmv::ttestIS(  
  formula = dif_Angst ~ Cursus,  
  data = data,  
  vars = dif_Angst,  
  students = FALSE,  
  mann = TRUE)
```


17.2.3 Kruskal-wallis

In het tabblad “Analyses”, klik op “ANOVA”. Vervolgens kies je uit het menu onderaan bij non-parametric voor “One-Way ANOVA Kruskal-Wallis”. Verplaats de afhankelijke variabele naar het vak “Dependent variables” en verplaats de factor naar “Grouping Variable”.

De onderliggende syntax voor deze analyse is als volgt:

```
jmv::anovaNP(
  formula = Angst1 ~ Education,
  data = data)
```

17.3 R

In R kun je zowel de Wilcoxon’s signed rank als de Wilcoxon’s rank sum test uitvoeren met de functie `wilcox.test`. De Kruskal-wallis toets kunnen we uitvoeren met `kruskal.test`. Om meer te leren over extra argumenten die je kunt gebruiken om je test uit te breiden, bijvoorbeeld het instellen van een eenzijdige hypothese, kun je altijd kijken in de help functie door `?wilcox.test` in te typen.

17.3.1 Wilcoxon’s signed rank

Met onderstaande code kun je de Wilcoxon’s signed rank test uitvoeren in R:

```
wilcox.test(mydata$Angst1, mydata$Angst2, paired= TRUE)
```

17.3.2 Wilcoxon’s rank sum

Met de onderstaande code kun je de Wilcoxon’s rank sum test uitvoeren in R:

```
wilcox.test(dif_Angst ~ Cursus, data=mydata)
```

17.3.3 Kruskal-Wallis

Met de Kruskal-Wallis test kunnen we meerdere groepen vergelijken. Hiervoor gebruiken we de volgende code:

```
kruskal.test(Angst1 ~ Education, data=mydata)
```

17.4 SPSS

In SPSS zijn er twee manieren om tot een nonparametrische test te komen. Onder “Analyze”, in het tabblad “Nonparametric tests” zijn drie categorieën te vinden: Related samples, Independent Samples en One Sample. Als je hierop klikt kom je in een vrij ingewikkeld keuzemenu waarbij je zelf testen kunt selecteren of SPSS laat bepalen welke test geschikt is voor uw data. Daaronder vindt je echter ook de optie *Legacy Dialogs*, en dan kun je direct kiezen om wat voor test het gaat.

17.4.1 Wilcoxon’s signed rank

Aangezien de Wilcoxon’s signed rank test een test van gepaarde data is, klik bij *Legacy Dialogs* vervolgens op “2 Related Samples”. Hier kun je vervolgens de paren aangeven en welke test je wilt doen, waarbij Wilcoxon automatisch is aangevinkt. Je kunt ook de volgende syntax gebruiken:

```
NPARTESTS
  /WILCOXON=Angst1 WITH Angst2 (PAIRED)
  /MISSING ANALYSIS.
```

17.4.2 Mann-Whitney U

De Wilcoxon’s rank sum of Mann-Whitney U test is een test voor onafhankelijke groepen. Kies bij *Legacy Dialogs* dan ook voor “2 Independent Samples”. Hier kun je vervolgens aangeven wat de afhankelijke variabele is en welke groepen er zijn. Dit levert de volgende syntax op:

```
NPARTESTS
  /M-W= dif_Angst BY Cursus
  /MISSING ANALYSIS.
```

17.4.3 Kruskal-Wallis

De Kruskal-Wallis vergelijkt meerdere groepen op een afhankelijke variabele. Hiervoor kun je bij *Legacy Dialogs* klikken op “K Independent Samples”. Selecteer de variabele die je wilt vergelijken en geef de range van de groepen aan. Dit levert de volgende syntax op:

```
NPARTESTS
  /K-W= Angst1 BY Education(1 5)
  /MISSING ANALYSIS.
```

Part IV

Analyses met meer variabelen

Chapter 18

Exploratorieve Factoranalyse

18.1 Intro

Exploratieve Factoranalyse (EFA) is een manier om te zoeken naar onderliggende factoren om een set variabelen te verklaren. Deze methode wordt vaak toegepast om schalen te construeren uit een grote set items. Deze set items wordt teruggebracht tot een kleiner aantal factoren.

18.1.1 Voorbeeld data

Dit voorbeeld gebruikt de Rosetta Stats voorbeeld dataset “pp15” (Zie Hoofdstuk 3 voor informatie over de dataset en hoofdstuk 4 voor een uitleg hoe datasets in te laden).

Uit deze dataset gebruiken we de volgende variabelen: `highDose_AttBeliefs_long`, `highDose_AttBeliefs_intensity`, `highDose_AttBeliefs_intoxicated`, `highDose_AttBeliefs_energy`, `highDose_AttBeliefs_euphoria`, `highDose_AttBeliefs_insight`, `highDose_AttBeliefs_connection`, `highDose_AttBeliefs_contact` & `highDose_AttBeliefs_sex`.

18.2 jamovi

In het tabblad “Analyses”, klik op de knop “Factor” en uit het menu dat verschijnt, kies “Exploratory Factor Analysis”. Uit de variabelen in je dataset, selecteer de items die je wilt includeren en verplaats ze naar het vak genaamd

“Variables” aan de rechterkant. Nu verschijnt automatisch de factoranalyse in het vak met de resultaten met de standaardinstellingen. Onderaan de analyse kun je allerlei verschillende instellingen specificeren en aanpassen. Zo kun je bijvoorbeeld kiezen dat je alleen factoren wilt behouden met een eigenwaarde groter dan 1, dat je de correlatie tussen de factoren wilt zien, de scree plot, en nog meer extra's.

Uiteindelijk ziet de onderliggende code er dan als volgt uit:

```
jmv::efa(  
  data = data,  
  vars = vars(highDose_AttBeliefs_long, highDose_AttBeliefs_intensity, highDose_AttB  
  nFactorMethod = "eigen",  
  screePlot = TRUE,  
  factorCor = TRUE,  
  factorScoresOV = list(  
    synced=list())
```

18.3 R

In R kun je de `rosetta` package gebruiken om de EFA uit te voeren met het volgende commando:

```
rosetta::factorAnalysis(  
  data = dat,  
  items = c(  
    'highDose_AttBeliefs_long',  
    'highDose_AttBeliefs_intensity',  
    'highDose_AttBeliefs_intoxicated',  
    'highDose_AttBeliefs_energy',  
    'highDose_AttBeliefs_euphoria',  
    'highDose_AttBeliefs_insight',  
    'highDose_AttBeliefs_connection',  
    'highDose_AttBeliefs_contact',  
    'highDose_AttBeliefs_sex'  
  ),  
  nfactors = "eigen"  
);
```

Let erop dat deze functie je verplicht om direct aan te geven hoeveel factoren je wilt overhouden aan het eind, met het `nfactors` argument. In het bovenstaande voorbeeld is dit gespecificeerd met “eigen”, wat wil zeggen dat het

Kaiser criterion gebruikt wordt. Dit criterium geeft je de mogelijkheid om de minimale eigenwaarde te bepalen (staat standaard ingesteld op 1)

Om overige informatie aan te vragen, zoals een samenvatting, de correlaties tussen de factoren, een scree plot, of residuen, kun je extra opties specificeren in het commando. Ook kun je bijvoorbeeld item labels toevoegen aan je set met variabelen.

```
rosetta::factorAnalysis(  
  data = dat,  
  items = c(  
    'highDose_AttBeliefs_long',  
    'highDose_AttBeliefs_intensity',  
    'highDose_AttBeliefs_intoxicated',  
    'highDose_AttBeliefs_energy',  
    'highDose_AttBeliefs_euphoria',  
    'highDose_AttBeliefs_insight',  
    'highDose_AttBeliefs_connection',  
    'highDose_AttBeliefs_contact',  
    'highDose_AttBeliefs_sex'  
  ),  
  itemLabels = c(  
    'Expectation that a high dose results in a longer trip',  
    'Expectation that a high dose results in a more intense trip',  
    'Expectation that a high dose makes you more intoxicated',  
    'Expectation that a high dose provides more energy',  
    'Expectation that a high dose produces more euphoria',  
    'Expectation that a high dose yields more insight',  
    'Expectation that a high dose strengthens your connection with others',  
    'Expectation that a high dose facilitates making contact with others',  
    'Expectation that a high dose improves sex'  
  ),  
  nfactors = "eigen",  
  summary = TRUE,  
  correlations = TRUE,  
  scree = TRUE,  
  residuals = TRUE  
);
```

18.3.1 SPSS

In SPSS kun je het FACTOR commando gebruiken. Belangrijke argumenten om te gebruiken zijn /VARIABLES om de items te specificeren, /CRITERIA om te

bepalen hoeveel factoren er moeten overblijven (of hoe dat beslist moet worden: gebruik MINEIGEN(1) om alleen factoren met een eigenwaarde hoger dan 1, of bijvoorbeeld FACTORS(2) om 2 factoren te extraheren), /EXTRACTION om de manier waarop de factoren worden berekend te kiezen (zoals OLS voor ordinary least squares regressie, of ML voor de maximum likelihood methode), en /ROTATION om de rotatie vast te stellen (bijv. NOROTATE als je geen rotatie wilt, of VARIMAX als je orthogonale rotatie wilt).

```

FACTOR
/VARIABLES
  highDose_AttBeliefs_long
  highDose_AttBeliefs_intensity
  highDose_AttBeliefs_intoxicated
  highDose_AttBeliefs_energy
  highDose_AttBeliefs_euphoria
  highDose_AttBeliefs_insight
  highDose_AttBeliefs_connection
  highDose_AttBeliefs_contact
  highDose_AttBeliefs_sex
/CRITERIA =
  MINEIGEN(1)
/EXTRACTION =
  ULS
/ROTATION =
  OBLIMIN
.

```

Voor specifieke uitkomsten kun je /PRINT en /PLOT gebruiken om bijvoorbeeld een scree plot te zien:

```

FACTOR
/VARIABLES
  highDose_AttBeliefs_long
  highDose_AttBeliefs_intensity
  highDose_AttBeliefs_intoxicated
  highDose_AttBeliefs_energy
  highDose_AttBeliefs_euphoria
  highDose_AttBeliefs_insight
  highDose_AttBeliefs_connection
  highDose_AttBeliefs_contact
  highDose_AttBeliefs_sex
/CRITERIA =
  FACTORS(1)
/PRINT =
  INITIAL

```



```
EXTRACTION
UNIVARIATE
CORRELATION
REPR
/PLOT =
  EIGEN
/EXTRACTION =
  ULS
/ROTATION =
  OBLIMIN
```

Meer informatie is te vinden in de SPSS manual:

The SPSS FACTOR manual section

Chapter 19

Factoriele ANOVA

19.1 Intro

Factoriele ANOVA is een uitbreiding van de gewone oneway ANOVA naar meerdere factoren. Er zijn hier dus twee of meer nominale variabelen die ieder twee of meer categorieën, de experimentele condities, bevatten. We kunnen het effect van elk van de variabelen bekijken, en ook van het interactie effect tussen beide variabelen.

19.1.1 Voorbeeld dataset

In dit voorbeeld gaan we werken met de dataset `sportcasus`. Informatie over deze dataset is te vinden in hoofdstuk 3 en informatie over hoe je data kunt laden staat beschreven in hoofdstuk 4. Hier willen we testen of de sportscore door de interventie (type informatie) is beïnvloed en of dit effect eventueel afhangt van de leeftijdsklasse.

19.2 jamovi

In het tabblad “Analyses” kun je klikken op “ANOVA”. Klik in het menu dat verschijnt vervolgens op de tweede optie evneens “ANOVA” genaamd. Selecteer de variabele die je wilt vergelijken, in dit geval `sportscore2` en verplaats deze naar het vak “Dependent Variables”. Selecteer vervolgens de factoren die je wilt gebruiken en verplaats deze naar het vak “Fixed Factors”, in dit voorbeeld `voorlichting` en `leeftijd`. Nu verschijnt automatisch de F-test aan de rechterkant van het scherm. Het model dat je toets bevat nu automatisch alle termen, dus bij twee variabelen wordt ook de interactieterm tussen deze

twee bijgevoegd. Als je dat niet wil, kan je het model menu uitklappen en daarin de interactieterm verwijderen, door deze term te selecteren en dan op het pijltje naar links te klikken. Je kan effect sizes aanvinken die dan meteen worden getoond in de tabel. De checks van de assumpties kun je in jamovi direct opvragen door onder het kopje “Assumptions” de verschillende plots aan te klikken.

Daarnaast kun je ook andere informatie toevoegen, zoals de gemiddelden per groep, of post-hoc toetsen. Ook is het mogelijk om contrasten te specificeren per variabele.

In de syntax ziet een Factoriele ANOVA er als volgt uit:

```
jmv::ANOVA(
  formula = sportscore2 ~ voorlichting + leeftijd + voorlichting:leeftijd,
  data = data,
  effectSize = c("eta", "omega"))
```

Je kunt het probleem ook benaderen als een lineair model met twee interacterende nominale predictoren. Het grote voordeel hiervan is dat je ook schattingen krijgt voor iedere aparte categorie van het effect op de afhankelijke variabele. Voor deze analyse ga je in het “Analyses” tabblad naar “Regression” en “Linear regression”. De dependent variabele is wederom `sportscore2` en bij “Factors” zet je de twee nominale predictoren. Vouw nu het menu “model Builder” open en selecteer beide predictoren. Vouw het onderste pijltje open en selecteer “interaction”. Hiermee wordt de interactieterm toegevoegd aan het model. Het menu kent nog meer opties. Zo kan de F tests worden opgevraagd, de betrouwbaarheidsintervallen van de geschatte parameters en de geschatte gemiddelden.

19.3 R

In R kun je met de `rosetta` package een ANOVA uitvoeren met de functie `fanova`. Met deze functie kun je zowel One-Way ANOVAs als factoriële ANOVAs uitvoeren afhankelijk van hoe je het model opbouwt. Voor ons voorbeeld kun je de volgende code gebruiken:

```
rosetta::fanova(dat=sportcasus,
  y='sportscore2',
  between=c('voorlichting', 'leeftijd'))
```

Let op: soms kan de functie `fanova` een foutmelding geven, met name als je een bestand hebt geïmporteerd vanuit SPSS met de `haven` package. Om zeker te weten dat de analyse werkt, kun je voor de zekerheid even je databestand (nogmaals) converteren naar een dataframe met `dat <- as.data.frame(dat)`.

19.4 SPSS

In SPSS moet je naar “Analyze”, “General Linear Model”, en dan “Univariate”. Hieronder zie je hoe je de analyse kunt doen via General Linear Model.

```
UNIANOVA sportscore2 BY voorlichting leeftijd
  /CONTRAST(voorlichting)=Helmert
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /PLOT=PROFILE(voorlichting*leeftijd)
  /EMMEANS=TABLES(OVERALL)
  /EMMEANS=TABLES(voorlichting) COMPARE ADJ(BONFERRONI)
  /EMMEANS=TABLES(leeftijd) COMPARE ADJ(BONFERRONI)
  /EMMEANS=TABLES(voorlichting*leeftijd)
  /PRINT=ETASQ HOMOGENEITY DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /DESIGN=voorlichting leeftijd voorlichting*leeftijd.
```

Deze analyse specificeert het Helmert contrast voor de vier condities van voorlichting. Tevens wordt er een plot opgevraagd van de scores van de twee leeftijdsgroepen met voorlichting, en tenslotte worden alle verwachte gemiddelden opgevraagd.

19.5 APA style report [AANPASSEN]

Voor het rapporteren van een ANOVA zijn verschillende regels opgesteld. Hier worden de APA-7 regels toegepast. Hieronder enkele voorbeelden voor het rapporteren van een ANOVA:

We vonden een statistisch significant effect van leeftijdsgroep op gebruik van sociale media, $F(3, 117) = 3.19$, $p = .026$, $\eta^2 = .21$.

Een eenweg ANOVA liet zien dat er een significant verschil was tussen heteroseksuele mensen ($M = 41.2$, $SD = 5.7$), biseksuele mensen ($M = 33.8$, $SD = 4.9$), en homoseksuele mensen ($M = 30.1$, $SD = 8.2$) wat betreft gevoelens van discriminatie, $F(2, 133) = 38.89$, $p < .001$, $\eta^2 = .34$.

Chapter 20

Covariantie analyse

20.1 Intro

Covariantie analyse is een uitbreiding van een gewone enkelvoudige (oneway) of factoriele anova naar een model waarin ook numerieke predictoren aanwezig zijn. Die numerieke variabelen worden vaak covariaten genoemd. Er zijn hier dus een of meer nominale variabelen (factoren) en een of meer numerieke variabelen (covariaten). We kunnen het effect van de experimentele condities bekijken, gecorrigeerd voor de covariaten.

20.1.1 Voorbeeld dataset

In dit voorbeeld gaan we werken met de dataset `sportcasus`. Informatie over deze dataset is te vinden in hoofdstuk 3 en informatie over hoe je data kunt laden staat beschreven in hoofdstuk 4. Hier willen we testen of de sportscore op tijdstip 2 door de interventie (type informatie) is beïnvloed, waarbij dit effect is gecorrigeerd door de sportscore op tijdstip 1.

20.2 jamovi

In het tabblad “Analyses” kun je klikken op “ANOVA”. Klik in het menu dat verschijnt vervolgens op de tweede optie eveneens “ANCOVA” genaamd. Selecteer de variabele die je wilt vergelijken, in dit geval `sportscore2` en verplaats deze naar het vak “Dependent Variables”. Selecteer vervolgens de factoren die je wilt gebruiken en verplaats deze naar het vak “Fixed Factors”, in dit voorbeeld `voorlichting`. Sleep vervolgens de variabele `sportscore1` naar het vak “Covariates”. Nu verschijnt automatisch de tabel met de F-test aan de rechterkant

van het scherm. Het model dat je toets bevat nu automatisch alle termen, dus bij twee variabelen wordt ook de interactieterm tussen deze twee bijgevoegd. Als je dat niet wil, kan je het model menu uitklappen en daarin de interactieterm verwijderen, door deze term te selecteren en dan op het pijltje naar links te klikken. Je kan effect sizes aanvinken die dan meteen worden getoond in de tabel. De checks van de assumpties kun je in jamovi direct opvragen door onder het kopje “Assumptions” de verschillende plots aan te klikken.

Daarnaast kun je ook andere informatie toevoegen, zoals de gemiddelden per groep, of post-hoc toetsen. Ook is het mogelijk om contrasten te specificeren per variabele.

In de syntax ziet een covariantie analyse er als volgt uit:

```
jmv::ancova(
  formula = sportscore2 ~ sportscore1 + voorlichting,
  data = data)
```

Je kunt het probleem ook benaderen als een lineair model met een nominale en numerieke predictor. Het grote voordeel hiervan is dat je ook schattingen krijgt voor iedere aparte categorie van het effect op de afhankelijke variabele. Voor deze analyse ga je in het “Analyses” tabblad naar “Regression” en “Linear regression”. De dependent variabele is wederom `sportscore2` en bij “Factors” zet je de nominale predictor `voorlichting` en bij “Covariates `sportscore1`”. Vouw nu het menu “model Builder” open en selecteer beide predictoren. Vouw het onderste pijltje open en selecteer “interaction”. Hiermee wordt de interactieterm toegevoegd aan het model, die bij een covariantie analyse niet significant moet zijn. Het menu kent nog meer opties. Zo kan de F tests worden opgevraagd, de betrouwbaarheidsintervallen van de geschatte parameters en de geschatte gemiddelden.

20.3 R

In R kun je met de `rosetta` package een ANOVA uitvoeren met de functie `fanova`. Met deze functie kun je zowel One-Way ANOVAs als factoriële ANOVAs uitvoeren, en ook covariantie analyses, afhankelijk van hoe je het model opbouwt. Voor ons voorbeeld kun je de volgende code gebruiken:

```
rosetta::fanova(dat=sportcasus,
  y="sportscore2",
  between="voorlichting",
  covar = "sportscore1")
```

Let op: soms kan de functie `fanova` een foutmelding geven, met name als je een bestand hebt geïmporteerd vanuit SPSS met de `haven` package. Om zeker

te weten dat de analyse werkt, kun je voor de zekerheid even je databestand (nogmaals) converteren naar een dataframe met `dat <- as.data.frame(dat)`.

20.4 SPSS

In SPSS moet je naar “Analyze”, “General Linear Model”, en dan “Univariate”. Hieronder zie je hoe je de analyse kunt doen via General Linear Model.

```
UNIANOVA sportscore2 BY voorlichting WITH sportscore1
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /EMMEANS=TABLES(OVERALL) WITH(sportscore1=MEAN)
  /PRINT=HOMOGENEITY DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /DESIGN=sportscore1 voorlichting.
```

Hier worden alle verwachte gemiddelden opgevraagd.

20.5 APA style report [AANPASSEN]

Voor het rapporteren van een ANOVA zijn verschillende regels opgesteld. Hier worden de APA-7 regels toegepast. Hieronder enkele voorbeelden voor het rapporteren van een ANOVA:

We vonden een statistisch significant effect van leeftijdsgroep op gebruik van sociale media, $F(3, 117) = 3.19$, $p = .026$, $\eta^2 = .21$.

Een eenweg ANOVA liet zien dat er een significant verschil was tussen heteroseksuele mensen ($M = 41.2$, $SD = 5.7$), biseksuele mensen ($M = 33.8$, $SD = 4.9$), en homoseksuele mensen ($M = 30.1$, $SD = 8.2$) wat betreft gevoelens van discriminatie, $F(2, 133) = 38.89$, $p < .001$, $\eta^2 = .34$.

Chapter 21

Herhaalde metingen analyse

21.1 Intro

Herhaalde metingen analyse, meestal aangeduid met de Engelse term Repeated Measures analysis, is een uitbreiding van enkelvoudige (oneway) anova, factoriele anova of ancova, naar een model waarin de afhankelijke variabele meerdere keren (herhaald) is gemeten. Er zijn hier dus een of meer nominale variabelen (factoren) en nul of meer numerieke variabelen (covariaten) en twee of meer versies van de afhankelijke variabele. We kunnen met deze analyse het effect van de experimentele condities bekijken op de verandering van de afhankelijke variabelen, eventueel gecorrigeerd voor covariaten.

21.1.1 Voorbeeld dataset

In dit voorbeeld gebruiken we de dataset `sportcasus`. Informatie over deze dataset is te vinden in hoofdstuk 3 en informatie over hoe je data kunt laden staat beschreven in hoofdstuk 4. Hier willen we testen of de sportscore op de drie tijdstippen beïnvloed wordt door de interventie (type informatie). Dus is er een verschil tussen de drie tijdstippen en hangt dit verschil af van de interventie.

21.2 jamovi

In het tabblad “Analyses” kun je klikken op “Repeated Measures Anova”. Klik in het bovenste vak “repeated Measures Factor” op “RM Factor 1” en hernoem dat naar bijvoorbeeld ‘sportscores’ (je kan het ook gewoon zo laten staan). In het tweede vak staan standaard twee levels, die kan je een andere naam geven, bijvoorbeeld tijdstip 1, tijdstip 2, en maak dan ook een tijdstip 3 aan. Vervolgens

sleep je de drie sportscore variabelen naar hun levels. Dus `sportscore1` naar level 1 (of tijdstip 1 als je die hebt hernoemd) etc. Selecteer vervolgens de factoren die je wilt gebruiken en verplaats deze naar het vak “Between Subject Factors”, in dit voorbeeld `voorlichting`. Eventuele covariaten gaan naar het vak “Covariates”. Nu verschijnt automatisch de tabel met de F-test aan de rechterkant van het scherm. Het model dat je toets bevat nu automatisch alle termen, dus bij twee variabelen wordt ook de interactieterm tussen deze twee bijgevoegd. Als je dat niet wil, kan je het model menu uitklappen en daarin de interactieterm verwijderen, door deze term te selecteren en dan op het pijltje naar links te klikken. Je kan effect sizes aanvinken die dan meteen worden getoond in de tabel. De checks van de assumpties kun je in jamovi direct opvragen door onder het kopje “Assumptions” de verschillende plots aan te klikken.

Daarnaast kun je ook andere informatie toevoegen, zoals de gemiddelden per groep, of post-hoc toetsen. Ook is het mogelijk om contrasten te specificeren per variabele.

In de syntax ziet een covariantie analyse door jamovi er als volgt uit:

```

jmv::anovaRM(
  data = data,
  rm = list(
    list(
      label="RM Factor 1",
      levels=c("Level 1", "Level 2", "tijdstip 3"))),
  rmCells = list(
    list(
      measure="sportscore1",
      cell="Level 1"),
    list(
      measure="sportscore2",
      cell="Level 2"),
    list(
      measure="sportscore3",
      cell="tijdstip 3")),
  bs = voorlichting,
  rmTerms = ~ `RM Factor 1`,
  bsTerms = ~ voorlichting)

```

21.3 R

In R kun je met de rosetta package een ANOVA uitvoeren met de functie `fanova`. Met deze functie kun je zowel One-Way ANOVAs als factoriële ANOVAs uitvoeren, en ook covariantie analyses, maar ook herhaalde metingen anova, afhanke-

lijk van hoe je het model opbouwt. Voor ons voorbeeld kun je de volgende code gebruiken:

```
rosetta::fanova(dat=sportcasus,
               y=c("sportscore1", "sportscore2", "sportscore3"),
               between="voorlichting")
```

Let op: soms kan de functie `fanova` een foutmelding geven, met name als je een bestand hebt geïmporteerd vanuit SPSS met de `haven` package. Om zeker te weten dat de analyse werkt, kun je voor de zekerheid even je databestand (nogmaals) converteren naar een dataframe met `dat <- as.data.frame(dat)`.

21.4 SPSS

In SPSS moet je naar “Analyze”, “General Linear Model”. Hieronder zie je hoe je de analyse kunt doen via General Linear Model.

```
GLM sportscore1 sportscore2 sportscore3
  /WSFACTOR=Sportscore 3 Repeated
  /METHOD=SSTYPE(3)
  /PLOT=PROFILE(Sportscore)
  /PRINT=DESCRIPTIVE
  /CRITERIA=ALPHA(.05)
  /WSDESIGN=Sportscore.
```

21.5 APA style report [AANPASSEN]

Voor het rapporteren van een ANOVA zijn verschillende regels opgesteld. Hier worden de APA-7 regels toegepast. Hieronder enkele voorbeelden voor het rapporteren van een ANOVA:

We vonden een statistisch significant effect van leeftijdsgroep op gebruik van sociale media, $F(3, 117) = 3.19$, $p = .026$, $\eta^2 = .21$.

Een eenweg ANOVA liet zien dat er een significant verschil was tussen heteroseksuele mensen ($M = 41.2$, $SD = 5.7$), biseksuele mensen ($M = 33.8$, $SD = 4.9$), en homoseksuele mensen ($M = 30.1$, $SD = 8.2$) wat betreft gevoelens van discriminatie, $F(2, 133) = 38.89$, $p < .001$, $\eta^2 = .34$.

Chapter 22

Multivariate Regressie

22.1 Intro

Multivariate regressieanalyse kan nuttig zijn om een model te verkrijgen om de afhankelijke variabele te voorspellen als een functie van twee of meer voorspellende variabelen en om te schatten welk deel van de variantie van die afhankelijke variabele kan worden begrepen met behulp van de voorspellende variabelen. De voorspellende variabelen kunnen numeriek zijn of categorisch, of te wel, een interval of een nominaal niveau hebben.

22.1.1 Voorbeeld data

In dit voorbeeld wordt de `schoolsucces` dataset gebruikt. De afhankelijke variabele is `schoolsucces` gemeten met een schoolcijfer. Deze variabele wordt onder andere voorspeld door intelligentie (IQ) en het inkomen van de ouders (drie categorieën: hoog, middel, laag).

22.2 jamovi

Klik in het tabblad “Analyse” op de knop “Regressie” en selecteer “Linear regression” in het menu dat verschijnt. Selecteer in het vak aan de linkerkant de afhankelijke variabele (`cijfer`) en verplaats deze naar het vak met het label “Afhankelijke variabele” met behulp van de knop met het naar rechts wijzende pijltje. Selecteer in het vak aan de linkerkant de numerieke voorspellers (hier: IQ) en verplaats ze naar het vak met het label “Covariaten” met behulp van de knop met het naar rechts wijzende pijltje. Categorische voorspellers (hier: `inkomen`) worden verplaatst naar het vak met het label “Factoren”. Je kunt naar

beneden scrollen om aanvullende analyses te specificeren, bijvoorbeeld om meer details over de coëfficiënten op te vragen door de sectie “Modelcoëfficiënten” te openen. Als je bijvoorbeeld het betrouwbaarheidsinterval voor de coëfficiënt en de gestandaardiseerde (geschaalde) coëfficiënten wilt opvragen, vink je de bijbehorende selectievakjes aan.

Hieronder kunt u zien hoe een regressie er volgens de jamovi syntax uitziet.

```
jmv::linReg(  
  data = data,  
  dep = cijfer,  
  covs = IQ,  
  factors = inkomen,  
  blocks = list(  
    list(  
      "IQ",  
      "inkomen")),  
  refLevels = list(  
    list(  
      var="inkomen",  
      ref="laag")))
```

Een regressie waarbij aanvullende analyses zijn gespecificeerd kan er zo uitzien. Hier is bijvoorbeeld de F test voor het gehele model aangevraagd, en het betrouwbaarheidsinterval voor de coëfficiënten. Je zult zien dat het eerste stuk van de syntax hetzelfde is als van het simpele model. Zo kun je dus gemakkelijk dezelfde code kopiëren en uitbreiden om de gegevens te verkrijgen die je wilt.

```
jmv::linReg(  
  data = data,  
  dep = cijfer,  
  covs = IQ,  
  factors = inkomen,  
  blocks = list(  
    list(  
      "IQ",  
      "inkomen")),  
  refLevels = list(  
    list(  
      var="inkomen",  
      ref="laag")),  
  modelTest = TRUE,  
  ci = TRUE)
```


22.3 R

In base R, de `lm` (linear model) functie kan worden gecombineerd met de `summary` functie om de meest belangrijke resultaten te tonen. R beschouwt variabelen die factoren zijn automatisch als categorisch.

In het `rosetta` package, de `regr` functie maakt gebruik van R's `lm` functie, maar geeft de resultaten wat anders weer.

```
rosetta::regr(cijfer ~ IQ + inkomen,  
             data=schoolsucces)
```

Net als `lm` in basis R, is het commando hetzelfde voor een continue voorspeller als voor een categorische voorspeller. Extra output kan worden opgevraagd met bijvoorbeeld het argument `collineariteit = TRUE`. Via de notatie “:” tussen twee variabelen, wordt de interactie tussen deze twee meegenomen.

```
rosetta::regr(cijfer ~ IQ + inkomen + leeftijd + IQ:leeftijd,  
             data=schoolsucces,  
             collinearity=TRUE)
```

22.4 SPSS

In SPSS, wordt het `REGRESSION` commando gebruikt.

```
REGRESSION  
  /STATISTICS COEFF OUTS CI(95) R ANOVA  
  /DEPENDENT cijfer  
  /METHOD=ENTER IQ inkomen.
```

22.5 APA style report

Voor het rapporteren van een regressie zijn standaarden opgesteld. Hier worden de APA-7 regels toegepast. Bij complexe analyses wordt aangeraden om gebruik te maken van tabellen om het overzicht te bewaren, zodat je in de tekst minder statistische resultaten hoeft te verwerken. Enkele voorbeelden van het rapporteren van de resultaten staan hieronder:

We hebben geprobeerd uitgaansgedrag te voorspellen aan de hand van een viertal variabelen. Het gehele model verklaarde 43.2% van de variantie, $R^2 = 0.432$, $F(3,165) = 36.4$, $p < .001$, 95% CI [.31, .55]. Extraversie was een significante voorspeller van uitgaansgedrag, $b = 2.76$, $p < .001$, evenals leeftijd, $b = -0.58$, $p = .021$.

Multipel regressie is gebruikt om te onderzoeken of zelfvertrouwen en sociale steun de scores op mentaal welzijn van de participanten kon voorspellen. De resultaten van de regressie lieten zien dat het model 41.8% van de variantie verklaarde, en dat het model een significante voorspeller was van mentaal welzijn, $F(2,26) = 9.34$, $p = .001$. Sociale steun droeg significant bij aan het model, $b = 0.668$, $p < .001$, maar zelfvertrouwen niet, $b = 0.295$, $p = .071$.

Chapter 23

Logistische Regressie

23.1 Intro

Logistische regressieanalyse kan handig zijn om een model te maken dat een dichotome uitkomstvariabele heeft (ook wel binaire variabele genoemd). Deze dichotome uitkomst kan worden voorspeld aan de hand van één of meer predictorvariabelen.

23.1.1 Voorbeeld data

In dit voorbeeld wordt de `schoolsucces` dataset gebruikt. De afhankelijke variabele is de dichotome variabele `uitslag`, met de waarden “gezakt” of “geslaagd”. Deze variabele wordt onder andere voorspeld door intelligentie (IQ) en het inkomen van de ouders (drie categorieën).

23.2 jamovi

Klik op het tabblad “Analyse” op de knop “Regressie” en selecteer in het gedeelte “Logistische regressie” in het menu dat verschijnt “2 resultaten (binomiaal)”. Selecteer in het vak aan de linkerkant de afhankelijke variabele en verplaats deze naar het vak met het label “Afhankelijke variabele” met behulp van de knop met het naar rechts wijzende pijltje. Selecteer in het vak aan de linkerkant alle numerieke voorspellers en verplaats ze naar het vak met het label “Covariaten” met behulp van de knop met het naar rechts wijzende pijltje. Als er categorische voorspellers zijn, verplaats ze dan naar het vak met het label “Factoren”. De resultaten worden direct getoond in het rechter “Resultaten”

paneel. Je kunt naar beneden scrollen om aanvullende analyses op te geven, bijvoorbeeld om meer details over de regressiecoëfficiënten op te vragen kun je de sectie “Modelcoëfficiënten” openen. Als je bijvoorbeeld de odds-ratio’s en hun betrouwbaarheidsintervallen wilt ordenen, vink je de bijbehorende selectievakjes aan.

Hieronder zie je hoe deze resultaten tot stand zijn gekomen in de syntax.

```
jmv::logRegBin(
  data = data,
  dep = uitslag,
  covs = IQ,
  factors = inkomen,
  blocks = list(
    list(
      "IQ",
      "inkomen")),
  refLevels = list(
    list(
      var="uitslag",
      ref="1"),
    list(
      var="inkomen",
      ref="laag")))
```

Voor het toevoegen van aanvullende analyses kun je simpelweg extra regels toevoegen aan de syntax. Zie hieronder een voorbeeld van dezelfde logistische regressie, maar dan met odds-ratio’s, betrouwbaarheidsintervallen, en een classificatietabel.

```
jmv::logRegBin(
  data = data,
  dep = uitslag,
  covs = IQ,
  factors = inkomen,
  blocks = list(
    list(
      "IQ",
      "inkomen")),
  refLevels = list(
    list(
      var="uitslag",
      ref="1"),
    list(
      var="inkomen",
      ref="laag")),
```

```

ci = TRUE,
OR = TRUE,
class = TRUE,
duplicate = 10)

```

23.3 R

In basis R, de `glm` (generalized linear model) functie kan worden aangeroepen met het argument `family = binomial(link = "logit")` om een logistische regressie uit te voeren. De `summary` functie toont de resultaten.

```

result <-
  glm(uitslag ~ IQ + inkomen,
      data = schoolsucces,
      family = binomial(link = "logit"))

summary(result);

```

In het `rosetta` package, de `logRegr` functie gebruikt base R's `glm` functie om de output te presenteren gelijkend op SPSS. Met meerdere predictoren, gebruik `collinearity=TRUE`, om collinearity diagnostics te bekijken. Als er maar één voorspeller is, kun je een visualisatie van de ruwe data, de (binned) gemiddelden en de logistische curve opvragen door gebruik te maken van `plot=TRUE`.

```

rosetta::logRegr(uitslag ~ IQ + inkomen,
  data = schoolsucces,
  plot = TRUE)

```

23.4 SPSS

In SPSS, wordt het `LOGISTIC REGRESSION` commando gebruikt. Omdat inkomen een categorische variabele is, moet dat apart worden aangegeven. Dit hoeft dus niet in een analyse met alleen maar intervalvariabelen.

```

LOGISTIC REGRESSION VARIABLES uitslag
  /METHOD=ENTER IQ inkomen
  /CONTRAST (inkomen)=Indicator
  /CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).

```

Ook hier kunnen aan de syntax extra commando's worden toegevoegd, zoals betrouwbaarheidsintervallen voor de odds ratio's en classificatieplots.

```
LOGISTIC REGRESSION VARIABLES uitslag  
/METHOD=ENTER IQ inkomen  
/CONTRAST (inkomen)=Indicator  
/CLASSPLOT  
/PRINT=CI(95)  
/CRITERIA=PIN(.05) POUT(.10) ITERATE(20) CUT(.5).
```

Chapter 24

Moderatieanalyse

24.1 Intro

Als er een verband bestaat tussen twee variabelen, dan kan het zo zijn dat dit effect alleen aanwezig is bij een bepaalde groep mensen of dat het verband sterker is bij de ene groep in vergelijking tot een andere groep. De variabele die het verschil tussen de groepen aangeeft (bijvoorbeeld opleiding of sekse) wordt dan een moderator genoemd. Een moderator is dus een variabele die het verband tussen twee variabelen beïnvloedt. Dit betekent dat voor mensen die hoog scoren op de moderator het effect van de predictor op de afhankelijke variabele anders is dan voor mensen die laag scoren op de moderator. Een moderator kan een groepsindeling representeren, zoals bij opleiding of sekse. De moderator is dan een categorische of dichotome (twee categorieën) variabele, maar een moderator kan ook een intervalvariabele zijn, zoals een eigenschap die iemand in meer of mindere mate bezit (bijvoorbeeld veerkracht). Bij een intervalvariabele zijn er geen natuurlijke groepen, maar we kunnen nog steeds mensen onderscheiden die relatief hoog scoren op zo'n variabele en mensen die er relatief laag op scoren.

24.1.1 Voorbeeld dataset

Voor het eerste voorbeeld wordt gebruik gemaakt van de dataset *Statistiekangst*. Hiermee gaan we kijken of iemands toename in statistiekkennis wordt beïnvloed door de cursus die is gevolgd, en of dit wordt gemodereerd door iemands geslacht. Voor de laatste twee voorbeelden wordt gebruik gemaakt van de dataset *roken*. Hier gaan we eerst uitzoeken of de zin in een sigaret beïnvloedt wordt door de mate van stress en of dit gemodereerd wordt door het gezelschap waarin de persoon zich bevindt (samen/alleen). Daarna gaan we uitzoeken of de leeftijd een moderatie-effect heeft op de mate van stress. Informatie over

deze datasets is te vinden in 3 en informatie over hoe je data kunt laden staat beschreven in 4.

24.2 jamovi

24.2.1 Moderatie met categorische variabelen

Voor een moderatie met categorische variabelen, waarbij de uitkomst wel een intervalvariabele is, kunnen we gebruik maken van de variantieanalyse. Voor meer informatie over het uitvoeren hiervan, kun je eerst het hoofdstuk 15 lezen.

Voor het uitvoeren van een moderatieanalyse hoef je hier weinig op uit te breiden. We nemen het voorbeeld van een factoriële anova. Je begint met het toevoegen van de variabele **Cursus** en vervolgens klik je ook op de variabele **Gender**. Jamovi voegt dan behalve de twee hoofdeffecten ook automatisch het interactie-effect toe. Verder kan het helpen om een grafiek te maken om het moderatie-effect te visualiseren. Dit kun je doen door onderaan bij de opties op **Estimated Marginal Means** te klikken, en daar dan vervolgens beide variabelen in de box gelabeld “Marginal Means, Term 1” in te plaatsen.

Volgens de syntax ziet het er als volgt uit om een moderatieanalyse met grafiek uit te voeren:

```
jmv::ANOVA(  
  formula = dif_Stat ~ Cursus + Gender + Cursus:Gender,  
  data = data,  
  emMeans = ~ Cursus:Gender)
```

24.2.2 Moderatie met predictor op intervalniveau

Om een moderatie uit te voeren waarbij één van de twee predictoren op intervalniveau is en de ander is dichotoom, dan moet je eerst een regressieanalyse uitvoeren. Instructies hierover kun je vinden onder 16 en 22.

Voor het uitvoeren van een moderatieanalyse moeten we eerst een regressie doen met de predictorvariabele. Deze moeten we centreren, omdat we anders straks het interactie-effect niet kunnen uitrekenen. Een variabele centreren doe je als volgt:

```
c_stress = stress - VMEAN(stress)
```

In het menu doe je dit door in het tabblad Data op **Compute** te klikken. Hier kun je dan vervolgens de nieuwe naam (c_stress) en de bijbehorende formule (stress - VMEAN(stress)) invullen.

Vervolgens kunnen we een regressie uitvoeren. Hierbij voegen we de variabele `zin` in bij `Dependent variable`, en de predictoren bij `Covariates`. Nu berekent jamovi echter niet uit zichzelf het interactie-effect, dus dat moet je handmatig toevoegen. Dit doe je door naar `Model Builder` te gaan en daar zowel `c_stress` als `samen` te selecteren, en dan op het onderste pijltje te klikken. Hier kun je nu kiezen voor “Interaction”. De uitgevoerde regressie ziet er in de onderliggende syntax als volgt uit:

```
jmv::linReg(
  data = data,
  dep = zin,
  covs = vars(c_stress, samen),
  blocks = list(
    list(
      "c_stress",
      "samen",
      c("c_stress", "samen"))),
  refLevels = list(),
  modelTest = TRUE)
```

Omdat een dichotome variabele eigenlijk een speciale vorm is van een intervalvariabele, kun je deze soort variabele makkelijk meenemen in een analyse. Je kunt echter ook direct aangeven dat het een factor is in je model in plaats van een covariaat. Hiervoor moet binnen jamovi de variabele wel als nominaal geassocieerd zijn, anders snapt het programma de berekening niet. Deze manier maakt het wel makkelijker om het interactie-effect te visualiseren. Je kunt als je de regressie hebt uitgevoerd zoals hierboven namelijk onderaan naar “Estimated Marginal Means”. Hier kun je weer beide variabelen tegelijk selecteren en verplaatsen naar het vak rechts. Als het goed is, staat “marginal plots” automatisch aangevinkt, waardoor je nu een grafiek krijgt met stress op de x-as, zin in roken op de y-as, en de twee categorieën van samen als twee afzonderlijke lijnen. De syntax verandert hierdoor een beetje, zie hieronder:

```
jmv::linReg(
  data = data,
  dep = zin,
  covs = c_stress,
  factors = samen,
  blocks = list(
    list(
      "c_stress",
      "samen",
      c("c_stress", "samen"))),
  refLevels = list(
    list(
```

```

      var="samen",
      ref="nee, alleen")),
emMeans = ~ c_stress:samen)

```

24.2.3 Moderatie met intervalvariabelen

Een analyse waarbij het moderatie-effect bestaat uit twee intervalvariabelen kan simpelweg met een regressieanalyse worden uitgevoerd. De procedure hierbij is hetzelfde als wat hierboven beschreven staat in het stuk over een predictor op intervalniveau, alleen zijn nu allebei de variabelen op intervalniveau. Wel moet je steeds handmatig toevoegen welke interacties je wilt laten toevoegen via Model Builder, dus in dit geval willen we dat doen voor de combinatie van leeftijd en stress. Ook hier is het wel weer van belang om beide variabelen eerst te centreren of standaardiseren, anders kun je de waarden niet goed interpreteren. Mocht je ook deze interactie willen visualiseren, dan kan dat via Estimated Marginal Means. Het verschil wat je nu zult zien met een dichotome variabele, is dat er nu drie lijnen zullen zijn. De eerste lijn is voor de gemiddelde waarde van de moderator (0 als je hebt gestandaardiseerd), en de twee lijnen daaromheen zijn voor 1 standaarddeviatie boven en 1 standaarddeviatie onder dit gemiddelde. Zo kun je een beetje een beeld krijgen wat het effect is van de predictor voor mensen die hoog of laag scoren op de moderator. Voor ons voorbeeld ziet de syntax er daarvoor als volgt uit:

```

jmv::linReg(
  data = data,
  dep = zin,
  covs = vars(c_stress, c_leeftijd),
  blocks = list(
    list(
      "c_stress",
      "c_leeftijd",
      c("c_stress", "c_leeftijd"))),
  refLevels = list(),
  emMeans = ~ c_stress:c_leeftijd)

```

24.3 R

24.3.1 Moderatie met categorische variabelen

Voor een moderatie met categorische variabelen, waarbij de uitkomst wel een intervalvariabele is, kunnen we gebruik maken van de variantieanalyse. Voor meer informatie over het uitvoeren hiervan, kun je eerst het hoofdstuk 15 lezen.

Voor het uitvoeren van een moderatieanalyse hoef je hier weinig op uit te breiden. We nemen het voorbeeld van een factoriële anova. In R wordt hiermee automatisch ook het interactie-effect toegevoegd, dus je hoeft daar niets in te veranderen. Wel kun je deze moderatie laten visualiseren door `plot = TRUE` toe te voegen aan de code.

```
fanova(data=dat, y='dif_Stat', between = c('Cursus', 'Gender'), plot=TRUE)
```

In R wordt een interactie weergegeven met de dubbele punt, omdat de asterisk die we vaak zien al gebruikt wordt voor de significantie van de p-waarden. In dit geval zie je eerst de hoofdeffecten voor Cursus en Gender, en zie je daaronder Cursus:Gender staan, dit is dan het interactie-effect.

24.3.2 Moderatie met predictor op intervalniveau

Om een moderatie uit te voeren waarbij één van de twee predictoren op intervalniveau is en de ander is dichotoom, dan moet je eerst een regressieanalyse uitvoeren. Instructies hierover kun je vinden onder 16 en 22.

Hierbij geldt wel dat een interactie moeilijk te interpreteren is als de variabelen niet gecentreerd of gestandaardiseerd zijn. In R kun je als volgt een variabele centreren:

```
c_stress <- scale(roken$stress, scale = FALSE)
roken <- cbind(roken, c_stress)
```

De laatste regel is om de nieuwe variabele die we gemaakt hebben weer toe te voegen aan de originele dataset. In R komen variabelen namelijk niet automatisch in een dataset, maar allemaal los als objecten.

Zoals in het hoofdstuk over meerdere regressie al is uitgelegd, kun je vrij gemakkelijk een interactie toevoegen. In ons voorbeeld zou je dan het volgende commando krijgen:

```
rosetta::regr(zin ~ c_stress + samen + c_stress:samen,
             data = roken)
```

Via deze functie is het echter niet mogelijk om makkelijk een visualisatie te krijgen van het interactie-effect. Om dit wel te kunnen doen, kunnen we een extra package installeren genaamd `sjPlot`. Hiermee kun je vervolgens in twee stappen een plot genereren. Eerst moeten we via `lm` opnieuw de regressie uitvoeren en opslaan. Dit model kunnen we vervolgens in de functie voor de grafiek stoppen:

```
fit <- lm(zin ~ c_stress * samen, data= roken)
plot_model(fit, type="int")
```

24.3.3 Moderatie met intervalvariabelen

Een moderatieanalyse met twee intervalvariabelen is in R eigenlijk exact hetzelfde als wat hierboven al beschreven is. Let er hierbij wel op dat het makkelijker is als beide variabelen gecentreerd of gestandaardiseerd zijn, omdat anders de interpretatie van de verschillende coëfficiënten erg ingewikkeld wordt.

In ons huidige voorbeeld zou dat tot de volgende code leiden:

```
rosetta::regr(zin ~ c_stress + c_leeftijd + c_stress:c_leeftijd,  
             data = roken)
```

Om dit te visualiseren, kunnen we ook weer gebruik maken van de package `sjPlot`. Omdat de moderator nu niet dichotoom is, zijn er heel veel verschillende waarden mogelijk. Dit maakt het iets lastiger om te visualiseren, waardoor we extra na moeten denken. Binnen de functie `plot_model` kun je de waarden van de moderator specificeren via `mdrt.values`. Hier kun je bijvoorbeeld kiezen uit de minimale waarde en de maximale waarde van de moderator. Een veel gebruikte manier is om het gemiddelde te nemen, en 1 standaarddeviatie boven en onder het gemiddelde. Zo krijg je een beeld van het effect van de predictor voor mensen die hoog en laag op de moderator scoren. Dit is de methode die is gekozen voor de onderstaande code.

```
fit2 <- lm(zin ~ c_stress * c_leeftijd, data=roken)  
plot_model(fit2, type = "int", mdrt.values = "meansd")
```

24.4 SPSS

24.4.1 Moderatie met categorische variabelen

Voor een moderatie met categorische variabelen, waarbij de uitkomst wel een intervalvariabele is, kunnen we gebruik maken van de variantieanalyse. Voor meer informatie over het uitvoeren hiervan, kun je eerst het hoofdstuk 15 lezen.

In principe verandert er weinig ten opzichte van het uitvoeren van een tweeweg anova, aangezien SPSS automatisch een “Full factorial model” uitrekent, wat inhoudt dat zowel hoofdeffecten als interactie-effecten worden meegenomen.

Wel kun je er nog voor kiezen om dit interactie-effect te visualiseren. Als je in het menu bent van “Univariate”, kun je aan de zijkant op “Plots” klikken. Hier kun je aangeven welke variabele je op de x-as wilt hebben, en voor welke variabele je verschillende lijnen wilt. Meestal is het gebruikelijk dat de predictor op de x-as komt en de moderator als lijnen wordt weergegeven.

Je kunt dit ook via de syntax doen, dan voeg je een regel toe voor `PLOT`. De gehele analyse ziet er dan zo uit:

```

UNIANOVA dif_Stat BY Cursus Gender
  /METHOD=SSTYPE(3)
  /INTERCEPT=INCLUDE
  /PLOT=PROFILE(Cursus*Gender) TYPE=LINE ERRORBAR=NO MEANREFERENCE=NO YAXIS=AUTO
  /CRITERIA=ALPHA(.05)
  /DESIGN=Cursus Gender Cursus*Gender.

```

24.4.2 Moderatie met predictor op intervalniveau

Om een moderatie uit te voeren waarbij één van de twee predictoren op intervalniveau is en de ander is dichotoom, dan moet je eerst een regressieanalyse uitvoeren. Instructies hierover kun je vinden onder 16 en 22.

Voor het uitvoeren van een moderatieanalyse waarbij er intervalvariabelen bij betrokken zijn, is het van belang om deze variabelen te centreren. Als je dit niet doet, dan wordt het bij ingewikkelde analyses bijna onmogelijk om de uitkomst te interpreteren.

In SPSS kun je een variabele als volgt centreren:

```

AGGREGATE
  /OUTFILE=* MODE=ADDVARIABLES
  /BREAK=
  /stress_mean=MEAN(stress).

COMPUTE c_stress=stress - stress_mean.
EXECUTE.

```

Via het menu doe je dit door op Data te klikken, dan op Aggregate. Verplaats hier je gekozen variabele naar het rechterscherf in het vak “Aggregated variables - Summaries of variable(s)”. Klik op OK, nu is er een nieuwe variabele voor het gemiddelde berekend. Vervolgens kun je via Transform en Compute Variable de nieuwe variabele maken door die variabele voor het gemiddelde af te trekken van de originele variabele.

In SPSS kun je helaas niet zomaar een interactie-effect toevoegen in de regressieanalyse zelf. Dit kan wel met de *Process* macro die ook wordt gebruikt in het hoofdstuk 25, maar het uitleggen van deze macro is voor een simpele moderatie niet nodig. Je moet dus handmatig de twee predictoren met elkaar vermenigvuldigen, en dit als variabele toevoegen in de analyse. Deze interactieterm maken doe je ook weer via Transform en Compute Variabele.

```

COMPUTE inter_stress_samen=c_stress * samen.
EXECUTE.

```

Vervolgens voer je een regressieanalyse uit, waarbij je zowel de twee variabelen los invoert, als de nieuwe variabele die we net gemaakt hebben.

```
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT zin
  /METHOD=ENTER c_stress samen inter_stress_samen.
```

In SPSS kun je via Graphs allerlei grafieken maken. Om de moderatie te visualiseren kiezen we voor een scatterplot. Klik hiervoor op Graphs, Legacy Dialogs, en dan Scatter/Dot. Kies voor “Simple Scatter”. In het menu selecteer je de afhankelijke variabele voor op de y-as, de variabele die interval is voor op de x-as, en de dichotome variabele verplaats je naar “Set markers by”.

```
GRAPH
  /SCATTERPLOT(BIVAR)=c_stress WITH zin BY samen
  /MISSING=LISTWISE.
```

SPSS laat hierbij nu niet automatisch lijnen zien, alleen de kleuren van de observaties verschillen. Als je de grafiek wilt bewerken kun je erop dubbelklikken. Je komt dan in een menu waarin je allerlei dingen kunt aanpassen, zoals de kleuren, de labels en de legenda. Om lijnen weer te geven per groep, klik je op “Add Fit line at Subgroups”. Dit is helaas iets wat je niet via de syntax kunt doen.

24.4.3 Moderatie met intervalvariabelen

Een moderatie met intervalvariabelen werkt eigenlijk hetzelfde als het voorbeeld hierboven. Ook hier zul je de variabelen moeten centreren of standaardiseren, en we moeten weer handmatig een interactieterm maken door de variabelen met elkaar te vermenigvuldigen. Vervolgens kun je alle drie de termen in je regressieanalyse stoppen. Dit ziet er in de syntax uiteindelijk als volgt uit:

```
/MISSING LISTWISE
/STATISTICS COEFF OUTS R ANOVA
/CRITERIA=PIN(.05) POUT(.10)
/NOORIGIN
/DEPENDENT zin
/METHOD=ENTER c_stress c_leeftijd inter_stress_leeftijd.
```

Deze moderatie is helaas niet makkelijk te visualiseren in SPSS. Het is handmatig wel mogelijk, maar het is niet aan te raden. Het is beter om hiervoor één van de andere programma's te gebruiken. Online zijn ook nog tutorials te vinden hoe je dit in Excel zou kunnen doen, mocht je geen jamovi of R hebben.

Chapter 25

Moderatie-Mediatie Modellen

25.1 Intro

Mediatie verwijst naar een causaal model waarin de relatie tussen de voorspeller en de afhankelijke variabele wordt gemedieerd door een of meer andere variabelen (“de mediators”). Bedenk dat mediation per definitie sterke aannames met zich meebrengt met betrekking tot causaliteit, waarvoor longitudinale ontwerpen en bijna altijd experimentele onafhankelijke manipulatie van de voorspeller (en indien mogelijk mediator) nodig zijn. Moderatie- en mediatiemodellen kunnen ook worden gecombineerd. Het causale pad tussen de voorspeller en de mediator(en) kan worden gemodereerd door een moderator. Evenzo kan het pad tussen de mediators en de afhankelijke variabele worden gemodereerd door een andere moderator. Een moderator versterkt of verzwakt de relatie tussen twee variabelen. We duiden dit type modellen aan met de Engelstalige term “moderated mediation”. Het pad van voorspeller naar bemiddelaar kan worden gemodereerd door een moderator die wordt aangeduid als “w”, terwijl het pad van bemiddelaar naar afhankelijke variabele kan worden gemodereerd door een moderator die wordt aangeduid als “v”.

25.1.1 Voorbeeld data

We gebruiken voor deze analyse de dataset `ContraProductiefGedrag` die verder beschreven staat in hoofdstuk 3. In de voorbeelden is de voorspeller (de eerste variabele in de causale keten) `procedureleRechtvaardigheid`, de mediators zijn `vertrouwen`, `cynisme`, (de tussenvariabelen in de causale keten), de afhankelijke variabele heet `contraProductiefGedrag` (de laatste variabele

in de causale keten), en covariaten zijn `distributieveRechtvaardigheid`, `cynisme`. Ten slotte worden de moderator `sekse` gebruikt.

25.2 jamovi

Download eerst de `jamm` module. Ga naar het Medmod menu en selecteer ‘GLM mediation model’. Continue predictoren kun je in het vak genaamd ‘Covariates’ plaatsen, in dit voorbeeld de variabele `procedureleRechtvaardigheid`. Dit wordt de predictor variabele, de eerste in de causale keten, en daarom bij voorkeur een experimenteel gemanipuleerde variabele. Mediatoren worden in het vak ‘Mediators’ geplaatst, in dit voorbeeld de variabelen `vertrouwen` en `cynisme`. Het eerste voorbeeld is een mediatiemodel met twee mediators zonder moderatoren.

In jamovi-code ziet dit model er als volgt uit:

```
jamm::jammGLM(
  formula = list( vertrouwen ~ procedureleRechtvaardigheid,
                 cynisme ~ procedureleRechtvaardigheid,
                 contraProductiefGedrag ~ vertrouwen + cynisme + procedureleRechtv
  data = data)
```

In voorbeeld twee is een moderator, `leeftijd`, toegevoegd op het pad van predictor naar mediator, en er is maar één mediator gebruikt: `vertrouwen`. Het tweede voorbeeld heeft extra uitbreidingen nodig in de analyse. Voeg naast de predictor (`procedureleRechtvaardigheid`) de moderator (`leeftijd`) toe in het vak ‘Covariates’, net als in voorbeeld 1. Klik vervolgens op de sectie ‘Moderators’, en verplaats de moderator `leeftijd` naar het ‘Mediated Effects’ vak. Ga vervolgens naar de sectie ‘Full Model’ en verwijder daar de interactietermen uit het model: ‘`vertrouwen * leeftijd`’, ‘`procedureleRechtvaardigheid * leeftijd`’.

In jamovi R-code:

```
jamm::jammGLM(
  formula = list( vertrouwen ~ procedureleRechtvaardigheid + leeftijd + procedureleR
                 contraProductiefGedrag ~ vertrouwen + procedureleRechtvaardigheid
  data = data,
  moderatorsTerms = list(
    list("leeftijd")))
```

Bij het toevoegen van variabelen als mediators, covariaten of moderators, creëert Jamovi uit zichzelf een volledig model met alle termen erin, waaronder interacties. Door bepaalde termen uit dit model te halen, kunnen allerlei verschillende modellen worden gespecificeerd. In de output kun je in het pad diagram direct

zien of het model dat gespecificeerd is overeenkomt met het bedachte model. Zo kun je modellen maken met meerdere niveaus m.b.t. mediatie en moderatie, zoals een indirect pad via twee verschillende mediators, of een model waarbij alle stappen een moderator hebben.

25.3 R

In R zit een functie genaamd `gemm()`, die onderdeel is van de package `rosetta`. Een volledige tutorial kan hier gevonden worden: https://www.academia.edu/40039588/Tutorial_of_moderated_mediation_with_SEM_the_gemm_function of via DOI:10.13140/RG.2.2.22175.10404.

Voorbeeld 1 is een mediatie model zonder moderators en met twee mediators; dit model wordt als volgt gerund. De data zijn opgeslagen in het R object (data frame) `ContraProductiefGedrag`.

```
gemm(dat = ContraProductiefGedrag,
      xvar = "procedureleRechtvaardigheid",
      mvars = c("vertrouwen", "cynisme"),
      yvar = "contraProductiefGedrag");
```

In voorbeeld 2 is een moderator toegevoegd op het pad x-m en is er slechts één moderator gebruikt. Dit wordt als volgt gerund:

```
gemm(dat = ContraProductiefGedrag,
      xvar = "procedureleRechtvaardigheid",
      mvars = "vertrouwen",
      yvar = "contraProductiefGedrag",
      xmmod = "leeftijd")
```

In voorbeeld 3 is een covariaat (`distributieveRechtvaardigheid`) toegevoegd. Het is mogelijk om te kiezen op welke variabelen (mediators of de afhankelijke variabele) de covariaten een effect zouden hebben. In dit voorbeeld heeft de covariaat een verwacht effect op de afhankelijke variabele, omdat het achter de optie `cyvars = staat`. Daarnaast is in dit voorbeeld het m-y (mediator-afhankelijke variabele) pad gemodereerd. Voorbeeld 3 wordt als volgt gerund.

```
gemm(dat = ContraProductiefGedrag,
      xvar = "procedureleRechtvaardigheid",
```

```
mvars = "vertrouwen",
yvar = "contraProductiefGedrag",
mymod = "sekse",
cyvars = c("distributieveRechtvaardigheid" )
```

In voorbeeld 4 zijn zowel het pad x-m als het pad m-y uitgebreid met een moderator. Ook zijn er twee mediators. De covariaat heeft een verwacht effect op de afhankelijke variabele. Voorbeeld 4 wordt als volgt gerund.

```
gemm(dat = ContraProductiefGedrag,
      xvar = "procedureleRechtvaardigheid",
      mvars = c("vertrouwen", "cynisme"),
      yvar = "contraProductiefGedrag",
      xmmod = "sekse",
      mymod = "leeftijd",
      cyvars = "distributieveRechtvaardigheid" )
```

De output van de `gemm` functie kunnen worden opgeslagen in een R object om meer resultaten te bekijken. De `print` en `plot` functies zijn ontwikkeld om relevante output van deze analyse te tonen. Plots zijn alleen relevant als er moderatie is, omdat ze de slopes of de `index van moderated mediation` laten zien.

```
output <- gemm(dat = ContraProductiefGedrag,
               xvar = "procedureleRechtvaardigheid",
               mvars = c("vertrouwen", "cynisme"),
               yvar = "contraProductiefGedrag",
               xmmod = "leeftijd");

print(output)
plotIMM(output);
plotSS(output);
```

25.4 SPSS

In SPSS kan de PROCESS macro worden gebruikt om (moderated) mediation modellen te analyseren. Versie 3 van deze macro kan worden gedownload van <http://www.afhayes.com>. De SPSS code hieronder is gebaseerd op versie 2. Verschillende mediators kunnen worden toegevoegd, evenals verschillende

covariaten. Wel moet u per soort model de juist code opgeven. Voor meer informatie zie Hayes (2018). De PROCESS macro gebruikt verschillende modelnummers om deze modellen te onderscheiden.

Voorbeeld 1 heeft twee mediators en geen moderators of covariaten.

```
PROCESS vars = contraProductiefGedrag, procedureleRechtvaardigheid, vertrouwen, cynisme
  / y = contraProductiefGedrag
  / x = procedureleRechtvaardigheid
  / m = vertrouwen
      cynisme
  / model = 4.
```

Voorbeeld 2 heeft één mediator, één mediator op het pad x-m en geen covariaten.

```
PROCESS vars = contraProductiefGedrag, procedureleRechtvaardigheid mediatorVariable, moderatorVariable
  / y = contraProductiefGedrag
  / x = procedureleRechtvaardigheid
  / m = vertrouwen
  / w = leeftijd
  / model = 7.
```

Voorbeeld 3 heeft één mediator, één moderator op het pad van m-y, en een covariaat.

```
PROCESS vars = contraProductiefGedrag, procedureleRechtvaardigheid vertrouwen, moderatorVariable,
              distributieveRechtvaardigheid cynisme
  / y = contraProductiefGedrag
  / x = procedureleRechtvaardigheid
  / m = vertrouwen
  / v = leeftijd
  / c = distributieveRechtvaardigheid
  / model = 14.
```

Voorbeeld 4 heeft twee mediators, twee moderators (één voor x-m en één voor m-y) en een covariaat.

```
PROCESS vars = contraProductiefGedrag, procedureleRechtvaardigheid vertrouwen, moderatorVariable,
              distributieveRechtvaardigheid cynisme
  / y = contraProductiefGedrag
  / x = procedureleRechtvaardigheid
  / m = vertrouwen cynisme
  / w = sekse
  / v = leeftijd
  / c = distributieveRechtvaardigheid
  / model = 21.
```


Chapter 26

Multilevel analyse

26.1 Intro

Multilevel-analyse (MLA) wordt gebruikt voor de analyse van hiërarchische gegevens. Gegevens op het laagste niveau zijn geclusterd in hogere niveaus. Een belangrijke toepassing voor MLA is bij gegevens verkregen via de zogenaamde Experience Sampling Methode (ESM). Dit soort gegevens worden intensieve longitudinale gegevens genoemd. De subjecten (personen) zijn bijvoorbeeld gemeten op meerdere tijdstippen per dag. De tijdstippen zijn dan geclusterd binnen de subjecten.

26.1.1 Voorbeeld data

In deze voorbeelden is allereerst de dataset `wiskunde.sav` gebruikt, waarin het wiskundecijfer wordt voorspeld uit onder andere de intelligentiescores van de leerlingen die gegroepeerd zijn in klassen. De predictor heet `intelligentie`, de clustervariabele heet `klas`, de afhankelijke variabele heet `wiskunde`. Als voorbeeld van longitudinale data wordt de data set `roken.sav` gebruikt, waarin de zin in een sigaret en of men heeft gerookt in de afgelopen meetperiode wordt voorspeld uit onder andere de stress die men ervaart. In de data staat een index `meting`, die de herhaalde metingen aangeeft voor iedere persoon.

26.2 jamovi

MLA kan in jamovi worden uitgevoerd met de module “GAMLj: general analysis for linear models for jamovi”. Na het laden van deze module moet het menu “Lineaire modellen” worden geselecteerd en vervolgens de optie MIXED Model. In

het invoervenster worden de variabelen in de juiste velden geplaatst: de afhankelijke variabele (wiskunde) onder *Afhankelijke variabele*, onder *Factoren* de nominale predictoren (indien aanwezig) en onder *Covariaten* de numerieke variabelen (indien aanwezig), hier **intelligentie**. Onder *Clustervariabelen* kunnen één of meerdere variabelen (klas) worden geselecteerd die de hiërarchische cluster- of groepsstructuur definiëren. Selecteer vervolgens de REML-schattingsprocedure en vraag de 95%-betrouwbaarheidsintervallen aan. Onder het kopje *Fixed Effects* (vouw dit menu eerst uit met de pijl) worden de variabelen voor de fixed effects geselecteerd. Hier is het ook mogelijk om interactie-effecten te definiëren door twee (of meer) variabelen tegelijk te selecteren en naar het rechtervenster te sturen. Het vaste intercept kan worden aangevinkt of uitgevinkt.

Onder het kopje *Random Effects* zijn de variabelen voor de random effecten geselecteerd, en als er meer dan één zijn, kunnen de correlaties tussen deze effecten ook worden gespecificeerd door het vakje ‘gecorrleerd’ aan te vinken.

Voorbeeld 2 betreft een longitudinaal voorbeeld. Hier hebben we een extra variabele, **meting**, die een index is van tijd, bijvoorbeeld de dagen of andere meetmomenten. Deze variabele kan worden toegevoegd aan de fixed effecten. Verder kan er een lagged variabele, **zinLagged1**, opgenomen worden om de autocorrelatie te modelleren. Deze variabele moet eerste zelf worden geconstrueerd en kan dan ook worden toegevoegd aan de box met fixed effecten.

26.3 R

In R zijn er verschillende mogelijke pakketten voor analyse van multilevel data. Hier wordt de functie `lmer` gebruikt, die in het pakket `lme4` zit. De gegevens voor het eerste voorbeeld staan in de dataset `wiskunde`. Zorg ervoor eerst het `lme4`-pakket installeren en laad het in je R-sessie met `library(lme4)`.

```
model <- lmer(wiskunde ~ intelligentie +
             (1 + intelligentie | klas, data = wiskunde)
```

Voorbeeld 2, met longitudinal data, wordt aangeroepen met de volgende code. Hiervoor is eerst de lagged-1 variabele geconstrueerd.

```
modell1 <- lmer(zin ~ stress + meting + zinLag1 +
              (1 + stress | respondent, data = roken)
```

De resultaten kunnen in een R-object worden gestopt, bijvoorbeeld in ‘modell1’, met de `summary()` functie kunnen dan de resultaten verder worden bekeken.


```
summary(model1);
```

Als je p-waarden voor het fixed effect wilt zien, moet je het pakket `lmerTest` laden. Met dit pakket levert de output van de aanroepen naar `lmer` automatisch p-waarden.

26.4 SPSS

In SPSS the MIXED procedure wordt gebruikt voor een multilevel analysis. Een eenvoudig model met één voorspeller en twee random effecten, voor respectievelijk het intercept en de predictor, kan worden uitgevoerd met de volgende syntax.

```
MIXED wiskunde WITH intelligentie  
  /PRINT= SOLUTION TESTCOV  
  /METHOD= REML  
  /FIXED= INTERCEPT intelligentie  
  /RANDOM= INTERCEPT intelligentie | SUBJECT(klas) COVTYPE(UN).
```

De syntax van voorbeeld 2 met de longitudinale data is als volgt.

```
MIXED zin WITH stress meting  
  /PRINT=SOLUTION TESTCOV  
  /FIXED=INTERCEPT stress | SSTYPE(3)  
  /RANDOM=INTERCEPT stress | SUBJECT(respondent) COVTYPE(VC)  
  /REPEATED=meting | SUBJECT(respondent) COVTYPE(AR1).
```


Chapter 27

Referenties

Bibliography

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: a regression-based approach*. Guilford Press, New York.