



Samenvatting Hoorcolleges TOE

VOCUS heeft deze samenvatting te danken aan Karlijn Born.

Het gebruik van deze samenvatting is bedoeld als studeerhulp na het lezen van de verplichte literatuur. Gebruik van deze samenvatting is geheel voor eigen risico.

Soms wordt er verwezen naar bladzijden of tabellen in het originele boek.

Succes met studeren!



Inhoud

Samenvatting Hoorcolleges TOE	1
Hoorcollege 1 – Correlational Research	3
Hoorcollege 2 – Measurement and Surveys	8
Hoorcollege 3 – Measurement Quality	12
Hoorcollege 4 – (Multiple) regressie	17



Hoorcollege 1 – Correlational Research

Surveys

Kwalitatief

Onderzoek met interviews; je bent benieuwd naar de beweegredenen is handig omdat je kunt doorvragen.

Kwantitatief

Onderzoek met cijfers, door o.a. middel van vragenlijsten is anoniem (handig voor sociaal kwetsbare groepen als vluchtelingen) en kan op grotere schaal worden afgenomen.

Onderzoeksdoelen

Fundamenteel wetenschappelijk onderzoek

heeft als doel nieuwe kennis verzamelen en volgt hierbij de 'empirische cyclus'

Voorbeeld: taalontwikkeling bij kinderen of hoe gemeenschappen in Tanzania omgaan met de dood.

Toegepast wetenschappelijk onderzoek

praktijkgericht; een probleem in de praktijk moet worden opgelost.

Empirische Cyclus (fundamenteel onderzoek)

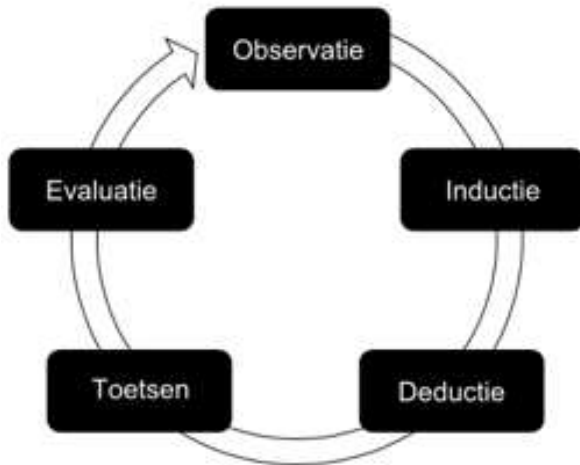
Observatie = startpunt onderzoek

Inductie = op basis van wat je hebt geobserveerd ga je een voorlopige conclusie trekken (theorievorming); nog erg breed

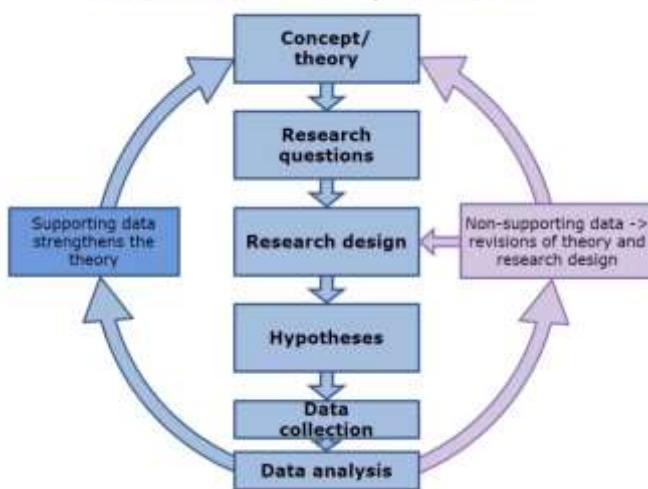
Deductie = vanuit de theorievorming ga je een voorspelling formuleren; je gaat dus een specifieke richting op

Toetsing = kijken of gegevens jouw voorspelling ondersteunen

Evaluatie = heb ik nu het antwoord op mijn onderzoeksvraag? Is het kennisprobleem opgelost? Waarschijnlijk heb je ondertussen andere vragen gekregen, waar je weer een nieuw onderzoek volgens de empirische cyclus mee start.



Remember the research cycle from KOM?



Survey data is *designed* data, wat betekent dat deze vragenlijst speciaal voor de steekproefpopulatie is ontworpen.

Correlational Data

Deze data is, net als een survey, speciaal voor de onderzoeksgroep ontworpen.

Correlatieel onderzoek heeft drie doelen:

1. Description

De realiteit in kaart brengen; hoe zit de populatie / verschijnsel in elkaar?

2. Causation

Causale verbanden blootleggen; wat is de oorzaak van het fenomeen?.

Causaliteit houdt in dat variabele X de directe oorzaak is van variabele Y. Er mag dus geen sprake mag zijn van een derde (mediërende) variabele die het verband overdraagt of een alternatieve verklaring.

3. Prediction

Generalisatie: resultaten van de steekproef zijn representatief voor de gehele populatie.

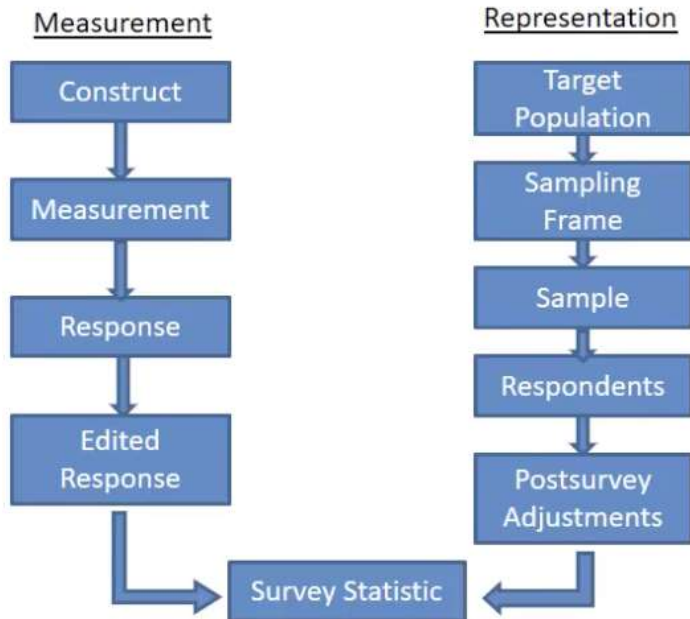
Survey Lifecycle

De *representation* gaat over je steekproefpopulatie; bij wie ga je de survey afnemen, wat is je doelgroep?

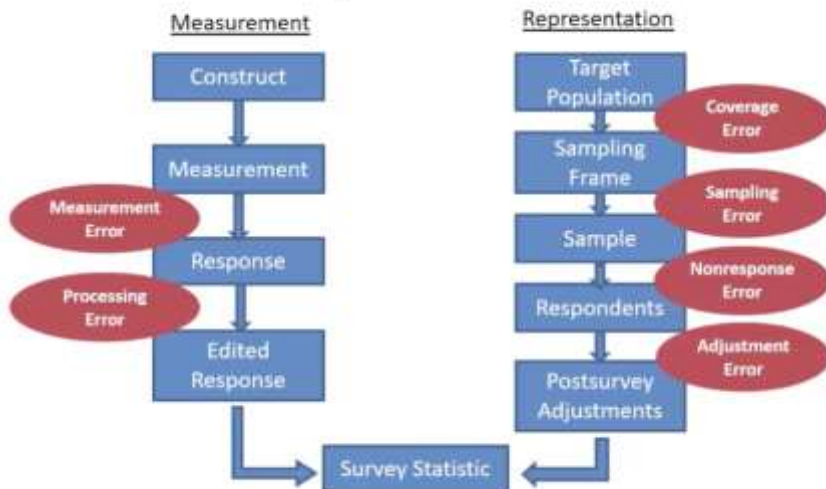
Bij de *measurement* kant ga je nadenken wat je in de survey gaat vragen en op welke manier.

Hier kunnen echter ook fouten (*errors*) optreden:

Survey Lifecycle



Total Survey Error Framework



Sampling error

A *sampling error* is a statistical error that occurs when an analyst does not select a sample that represents the entire population of data and the results found in the sample do not represent the results that would be obtained from the entire population.

De steekproef is dus niet representatief voor de gehele populatie, waardoor je de onderzoeksresultaten niet kunt terug generaliseren naar de gehele populatie.

Coverage error

Coverage error is a type of non-sampling error that occurs when there is not a one-to-one correspondence between the target population and the sampling frame from which a sample is drawn.

This can bias estimates calculated using survey data.

Oftewel: de frame- en target population komen niet met elkaar overeen.

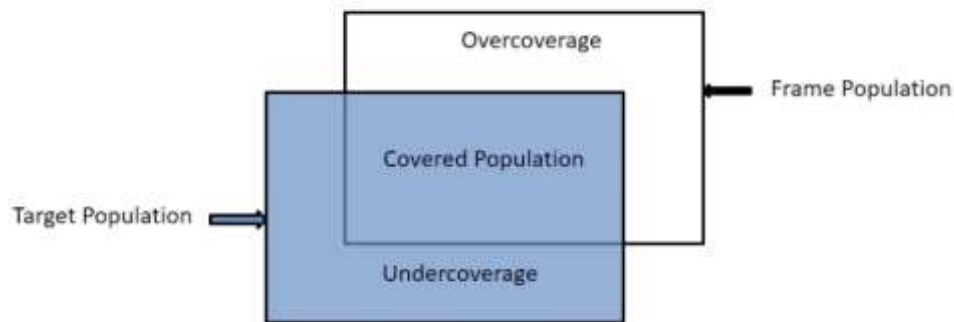
Example:

A researcher may wish to study the opinions of registered voters (target population) by calling residences listed in a telephone directory (sampling frame).

Undercoverage may occur if not all voters are listed in the phone directory.

Overcoverage could occur if some voters have more than one listed phone number.

Bias could also occur if some phone numbers listed in the directory do not belong to registered voters.



Non-response error

Dit houdt in dat iemand die geselecteerd is uit de frame populatie voor de target population (steekproef), niet reageert / niet de gewenste informatie geeft.

Frame population zijn de mensen die kans maken in de steekproef voor te komen. Zij maken dus deel uit van de populatie die je gaat onderzoeken. Stel dat jij het alcoholgebruik in de provincies Utrecht en Noord-Holland met elkaar gaat vergelijken, hebben de inwoners van deze regio's kans om aan jouw steekproef deel te nemen.

De **target population** is de doelgroep.

Het is een probleem als jouw *target population* verschilt van de *frame population*; dat houdt in dat de mensen die je wil onderzoeken, geen kans maken om in de steekproef voor te komen.

Adjustment error

Dit houdt in dat door coverage en non-response fouten, het onderzoeksresultaat gebaseerd op deze data niet overeenkomt met de gehele populatie.

Er zijn enkele manieren om deze fout te corrigeren:

- *Weighting*

Dit houdt in dat je sommige vragen uit de enquête meer 'gewicht' geeft dan anderen, waardoor ze zwaarder meetellen en alle groepen alsnog evenveel vertegenwoordigd worden; "*Weighting up the underrepresented*".

- *Imputation*

Dit is het geval als op sommige vragen geen antwoord gegeven wordt. Hier pas je *imputation* op toe, waardoor je kunt werken met het beantwoorde deel van de enquête en het geïmputeerde deel.

Survey methoden

Er zijn verschillende manieren waarop je een survey kunt afnemen:

- Face-to-face



- Schriftelijk (brief)
 - Telefonisch
 - Internet
 - Mixed
- o.a. deel schriftelijk (tijdens de werkgroep) en deel online (in vrije tijd) of het ontvangen van een schriftelijk verzoek om een digitale vragenlijst in te vullen.

Zowel bij de face-to-face als de telefonische methode is er een interviewer aanwezig. De invloed van de interviewer is echter hoger bij het face-to-face afnemen van een vragenlijst (door diens fysieke aanwezigheid) dan bij een telefonische vragenlijst.

Daarnaast heeft een face-to-face interview de minste privacy, wederom door de fysieke aanwezigheid van de interviewer. Dit in tegenstelling tot een schriftelijke of digitale vragenlijst, die je volledig alleen kunt invullen.

Survey modes in comparison

Attribute	Face-to-face	Mail	Telephone Interview	Web-based survey
Cost	high	low	moderate	low*
Response Rate	high	low	high	moderate
Researcher control over interview	high	low	moderate	moderate
Interviewer effects	high	low	moderate	low

Cross-

sectional survey

Deze survey wordt slechts één keer afgenomen.

Panel survey

Deze survey wordt gedurende een bepaalde tijd meerdere keren afgenomen, zoals een evaluatie/panelgesprek na elk blok. De vragen en onderwerpen blijven meestal hetzelfde, maar kunnen ook veranderen. Hiermee kun je niet alleen een mening, maar ook eventuele veranderingen en causaliteit in beeld brengen.

Een nadeel is wel, dat mocht er halverwege een participant afhaken, je deze niet meer kunt vervangen omdat je al middenin het onderzoeksproces zit (meerdere vragenlijsten hebt afgenomen).

Hoorcollege 2 – Measurement and Surveys

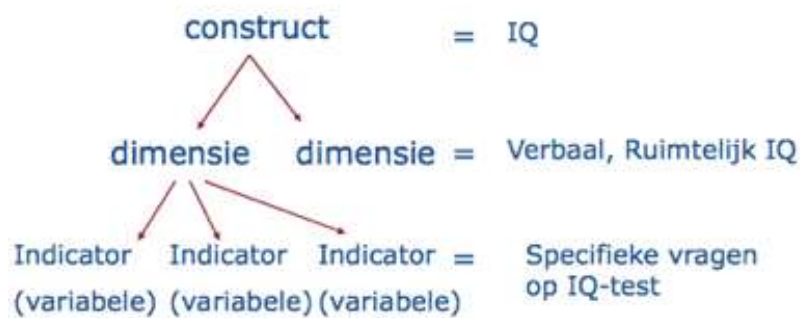
Measurement (meting)

Bij het begrip *measurement* hoort het volgende onderzoeksschema, wat als voorbeeld het onderwerp depressie meet:

- Stap 1: Onderzoeksvraag *Wat is de mate van depressie waarin iemand zit?*
- Stap 2: Conceptuele definitie *Depressie betekent ...*
- Stap 3: Operationele definitie *We meten het hoofdconstruct depressie als volgt*
- Stap 4: Deelconstructen opstellen *Depressie verdelen in verschillende deelconstructen zodat je het makkelijker kunt meten. Deze deelconstructen vormen samen dan het antwoord op de onderzoeksvraag.*
- Stap 5: Enquete afnemen
- Stap 6: Resultaten analyseren

Construct

Tijdens het meten begin je met een *construct*; een begrip waar je niet direct naar kan vragen. Daarom verdeel je dit begrip in observeerbare variabelen, zodat je het toch kan meten. Voorbeeld: IQ.



Betrouwbaarheid

De *betrouwbaarheid* van een onderzoek of meting houdt in dat het resultaat ervan niet varieert, hoe vaak je het ook uitvoert/herhaald.

Zo is je geboorteplaats heel betrouwbaar, maar iemands mening over de EU niet.

Validiteit

De *validiteit* gaat na in hoeverre jouw metingen overeenkomen met de werkelijkheid; meet je wat je daadwerkelijk wil meten (ruwe score)?

Er zijn twee soorten validiteit:

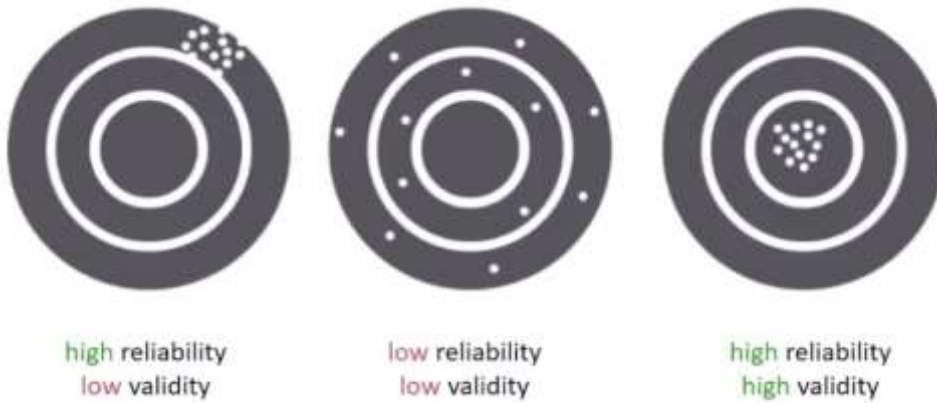
1. Interne validiteit

de mate waarin het redeneren en de handelingen binnen een onderzoek correct zijn uitgevoerd. Belangrijk hierbij is dat de resultaten van het onderzoek geldig zijn voor de onderzochte groep; de variabelen moeten een causale relatie hebben met elkaar.

2. Externe validiteit

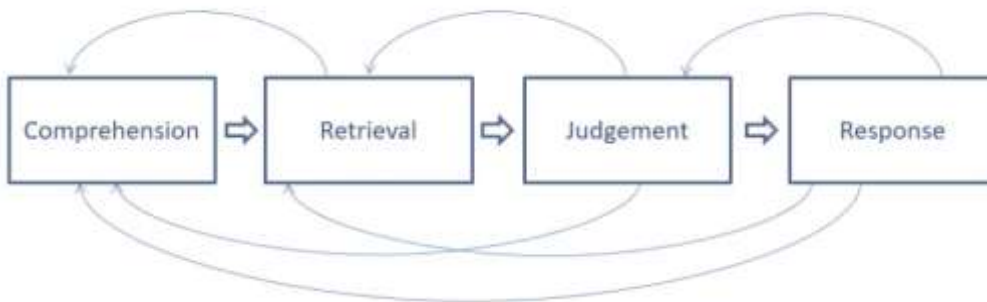
In hoeverre de resultaten van een onderzoek te generaliseren zijn. De resultaten gelden dan niet slechts voor de steekproefpopulatie, maar voor de gehele populatie.

V O C U S



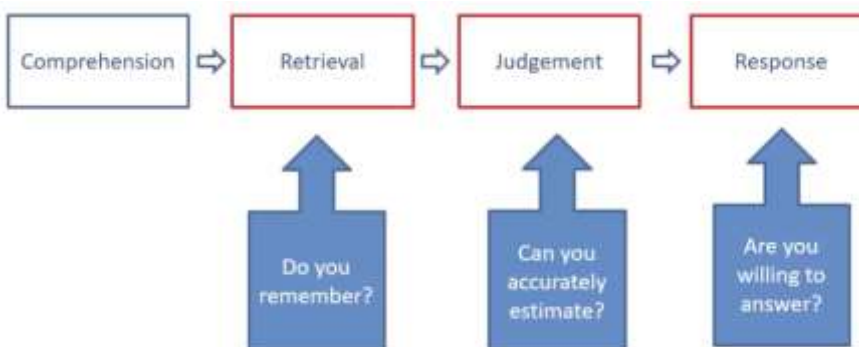
The response process

Voorbeeld onderzoeksvraag: Hoeveel glazen alcohol heb je afgelopen week gedronken?



Eerst moeten de participanten de vraag begrijpen (*comprehension*), waarna ze in hun geheugen op zoek gaan naar het antwoord (*retrieval*). Vervolgens moeten ze daar een schatting van maken; heb ik überhaupt gedronken en zo ja, hoeveel glazen waren dat (*judgement*)? Tot slot geven ze het antwoord (*response*).

Dit proces hoeft niet in chronologische volgorde doorlopen te worden; regelmatig worden er stappen overgeslagen of juist teruggegaan naar eerdere stappen.



Een probleem hierbij is *social desirability*; welk antwoord wil de interviewer graag horen en wordt geaccepteerd door de maatschappij? Dit is tevens een probleem voor de validiteit, omdat de meting niet



meer overeenkomt met de werkelijkheid.

Voorbeeld: zeggen dat je elke zondag naar de kerk gaat, omdat je in een christelijke gemeenschap leeft, terwijl je in realiteit slechts één keer per maand gaat.

Leading question

Deze vraag leidt de participant al naar een bepaald antwoord.

Response process

Dit gebeurt in de *retrieval* en *judgement* fase, omdat ze slechts selectieve informatie uit hun geheugen ophalen (*retrieval*) en hiermee naar één bepaalde schatting (*judgement*) geleid worden.

Validiteit of betrouwbaarheid?

Hierdoor treden er *validiteits*-problemen op, omdat de participant door de vraag al in de richting van één bepaald antwoord geduwd wordt, terwijl hijzelf een ander antwoord gegeven zou hebben. De meting komt dus niet volledig overeen met de werkelijkheid.

Double-barreled question

Er wordt binnen één vraag naar twee constructen gevraagd.

Voorbeeld: Ben je het eens met Barack Obama's miljoenen beslissing over de banking industry?

De participant wordt hier zowel om een mening over Barack Obama (*persoon*) gevraagd, als over een beslissing over het bankleven (*gedrag*).

Response process

De double-barreled question heeft betrekking op de *comprehension* en de *response* fase.

Participanten wordt gevraagd twee constructen te begrijpen (*comprehension*) en twee antwoorden tot één te formuleren (*response*).

Validiteit of betrouwbaarheid?

Dit levert problemen met de *validiteit* op, omdat je twee constructen meet in plaats van één. Een oplossing hiervoor is het opsplitsen in twee losse vragen.

Ordering effect

Hierbij hangt het antwoord dat gegeven wordt af van de volgorde waarin de vragen gesteld worden.

Response process

Dit vindt plaats in de *retrieval*, *judgement* en *response* fase; het wordt overgedragen van fase één, naar twee en tot slot drie.

Zo wordt er slechts bepaalde informatie opgehaald (*retrieval*), wat de schatting beïnvloedt (*judgement*) en hiermee ook het antwoord op de vraag (*response*).

Validiteit of betrouwbaarheid?

Het heeft invloed op de *validiteit*, omdat je participanten al aanzet tot het denken aan een bepaald antwoord in plaats van ze dit volledig zelf te laten formuleren. Hiermee komt de meting niet helemaal overeen met de werkelijkheid.



Telescoping effect

Participanten vallen terug op schatten, omdat ze het precieze antwoord (niet meer) weten.

Voorbeeld: Hoe vaak heb je de afgelopen twaalf maanden de bioscoop bezocht?

Response process

Dit gebeurt in de *retrieval* en *judgement* fase. De participant herinnert het exacte aantal bioscoopbezoeken niet meer (*retrieval*), wat hun schatting van het antwoord (*judgement*) zal beïnvloeden.

Validiteit of betrouwbaarheid?

Dit is zowel een probleem voor de *validiteit* als voor de *betrouwbaarheid*.

De participant schat namelijk het antwoord, dus de kans is groot dat indien hetzelfde onderzoek nogmaals onder dezelfde omstandigheden wordt afgenomen, de antwoorden toch van elkaar verschillen (*betrouwbaarheid*).

Daarnaast is de kans groot dat de participant enkele bioscoopbezoeken vergeet, waardoor de resultaten geen representatieve weergave van de werkelijkheid vormen (*validiteit*).

Sensitive question

Dit is een gevoelige vraag, die participanten vervelend vinden of zelfs liever niet beantwoorden.

Voorbeeld: Heb je veilige seks? / Gebruik je drugs? / Kijk je af tijdens tentamens?

Response process

Dit heeft invloed op de *response* fase. Zelfs als participanten het antwoord wel weten, willen ze het liever niet geven. Ook treed hier *social desirability* op, omdat het antwoord op de vraag het liefst 'ja' is.

Validiteit of betrouwbaarheid?

Dit is een probleem voor de *validiteit*, omdat mensen liegen of zelfs geen antwoord geven. Hierdoor wordt het construct dus niet volledig gemeten.

Satisficing

We nemen aan dat de participanten het gehele response process doorlopen, maar dat gebeurt niet altijd. Zo kan er **satisficing** optreden: participanten zoeken naar het eerste en beste antwoord dat hen tevreden stelt, in plaats van het hele *response process* te doorlopen.

Hier zijn vijf varianten:

1. *Social desirability*

Het gewenste antwoord wordt gegeven.

2. *Acquiescence*

Standaard met de stelling akkoord gaan, waarover deze ook gaat (geen eigen mening).

3. *Primary / recency effects*

Primary is dat je de eerste antwoordoptie kiest die gegeven wordt, terwijl bij *recency* juist één van de laatste antwoordopties gekozen wordt.

4. *Fence sitting*

Bij een antwoordschaal de middelste optie kiezen, in plaats van een van de extremen.

5. *Straightlining / non-differentiation*

Alle antwoordopties worden als hetzelfde beschouwd; er wordt niet meer tussen gedifferentieerd (niet gekeken naar de schaal).

Hoorcollege 3 – Measurement Quality

Schaalscore berekenen

Je hebt de *itemscore* en de *schaalscore*.

De *itemscore* is de behaalde score op slechts één vraag uit de vragenlijst, dus bij het eerste voorbeeld is de itemscore 2 (uit 4 antwoordopties).

De *schaalscore* is de totaalscore die behaald wordt op de gehele vragenlijst.

Dit kan op twee manieren:

1. Gemiddelden

Je telt alle individuele itemscores bij elkaar op en deelt deze door het aantal vragen.

Hier komt een getal uit: de *schaalscore*.

2. Gewogen gemiddelde (weighted mean)

Dit gebruik je als sommige vragen zwaarder meetellen dan anderen. Zo kun je bepaalde vragen keer twee doen en vervolgens alsnog delen door het totale aantal vragen. Op deze manier krijgen die vragen een zwaardere weging.

	not at all 1	a little 2	rather much 3	very much 4
It was difficult for me to express my feelings	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
I tried to avoid memories of events	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
I did not remember important details of past events	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was irritable	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
I had a physical reaction as a result of events that made me remember past events	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
I felt attached to others	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
I had frightening dreams	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
I had less interest in important events	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Diagram annotations: A blue arrow labeled 'Item score' points to the first row. A green arrow labeled 'Item score = 1' points to the second row. A green arrow labeled 'Item score = 2' points to the third row.

Hercoderen

In dit voorbeeld met PTSD geldt: hoe hoger de behaalde score, des te hoger de mate van PTSD.

In vragenlijsten moet je soms **hercoderen (recode)**, omdat bij niet alle vragen een hoge score veel PTSD inhoudt; alle vragen moeten positief of juist negatief geformuleerd zijn, zodat de scores bij elke vraag hetzelfde betekenen.

Voorbeeld:

1. I could easily concentrate (positief)
2. It was difficult for me to express my feelings (negatief)

Schaal:

1 = not at all

- 2 = a litte
- 3 = rather much
- 4 = very much

Een hoge score op vraag één betekent juist weinig PTSD, terwijl een hoge score op vraag twee veel PTSD betekent.

Dit kun je ook zien in de SPSS-output.

Alle gemiddelden zijn rond de 1,0, terwijl het gemiddelde van Q6 3,82 is. Deze vraag moet gehercodeerd worden.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Q1	95	1	3	1,29	,503
Q2	95	1	3	1,22	,465
Q3	95	1	3	1,13	,364
Q4	95	1	3	1,15	,412
Q5	95	1	2	1,08	,279
Q6	95	1	4	3,82	,565
Q7	95	1	3	1,19	,468
Q8	95	1	3	1,19	,490
Q9	95	1	3	1,09	,329
Q10	95	1	3	1,19	,420
Q11	95	1	3	1,14	,375
Q12	95	1	4	3,73	,659
Valid N (listwise)	95				

Het resultaat na het hercoderen in SPSS is dan als volgt:

Vervolgens gebruik je de gehercodeerde vragen om de schaalscore (gemiddelde) te berekenen:

Q6						Q6_R					
		Frequency	Percent	Valid Percent	Cumulative Percent			Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	2	2,0	2,1	2,1	Valid	1	84	84,8	88,4	88,4
	2	2	2,0	2,1	4,2		2	7	7,1	7,4	95,8
	3	7	7,1	7,4	11,6		3	2	2,0	2,1	97,9
	4	84	84,8	88,4	100,0		4	2	2,0	2,1	100,0
	Total	95	96,0	100,0			Total	95	96,0	100,0	
Missing	System	4	4,0			Missing	99	4	4,0		
Total		99	100,0			Total	99	100,0			

$$\text{Mean}(Q1, Q2, Q3, Q4, Q5, \text{Q6_R}, Q7, Q8, Q9, Q10, Q11, \text{Q12_R})$$

Correlatie

Je meet de sterkte van een relatie tussen twee continue variabelen X en Y met behulp van de correlatiecoëfficiënt (*r*).

Dit gebeurt aan de hand van *Pearson's r*, die een waarde kan aannemen tussen de -1 en 1.

Hierbij geldt: hoe dichterbij de -1 / 1, des te sterker is de relatie.

Deze plus of min geeft de richting aan. Indien het plus is, betekent dat als X in waarde toeneemt, neemt Y



ook in waarde toe.

Neemt X nou toe, maar Y af? Dan is er een negatieve relatie (min).

0	no correlation
.2	weak
.5	medium
.8	high
1/-1	perfect correlation

Een voorwaarde voor het gebruiken van de *Pearson's r*, is dat de twee variabelen van ratio/interval meetniveau zijn. Indien dit niet het geval is, gebruik je de *Spearman's r*.

Een andere voorwaarde is dat het verband tussen de twee variabele X en Y lineair is.

Hypotheses

De nul- en alternatieve hypothese bij een *correlatie* zijn altijd als volgt verdeeld:

H0 = geen verband / relatie tussen X en Y

H1 = wel verband / relatie tussen X en Y

Indien de p-waarde groter of gelijk is aan α , dan wijkt deze significant af. De nulhypothese (H0) wordt vervolgens verworpen en de alternatieve hypothese (H1) wordt aangenomen; er is sprake van een relatie tussen X en Y.

Betrouwbaarheid meten

De betrouwbaarheid van een vragenlijst wordt bepaald met de *Cronbach's Alpha*.

Dit meet de *internal consistency*, oftewel: de mate waarin de verschillende items die hetzelfde construct meten met elkaar samenhangen (correlatie).

"It measures whether several items that propose to measure the same general construct produce similar scores".

Indien de vragen niet allemaal in dezelfde richting (positief of negatief) gesteld zijn, moet je *hercoderen*.

De Cronbach's Alpha ziet er als volgt uit in SPSS:

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.745	.767	12

- $\alpha < 0.6$ poor
- $0.6 < \alpha < 0.7$ medium
- $0.7 < \alpha < 0.8$ good
- $\alpha > 0.8$ great



Je kunt de betrouwbaarheid van een vragenlijst vergroten door vragen te verwijderen.

Onder *Cronbach's Alpha if Item Deleted* kun je zien wat de betrouwbaarheid wordt als deze vraag uit de vragenlijst gehaald wordt.

De rood omcirkelde vragen vergroten de betrouwbaarheid (>0,745).

Bij *Corrected Item-Total Correlation* wordt er gekeken naar de correlatie van dat specifieke item met de gehele vragenlijst. Als deze waarde <0,30 is, betekent dit dat er vrijwel geen sprake is van een correlatie en kan de vraag verantwoord uit de vragenlijst verwijderd worden.

Voordat je een vraag definitief mag verwijderen, check je dus eerst ook de *Corrected Item-Total Correlation*!

Alpha = .745

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Q1	12,83	6,248	,540	,469	,707
Q2	12,91	7,172	,188	,240	,752
Q3	13,00	6,809	,481	,582	,720
Q4	12,98	6,276	,684	,571	,694
Q5	13,04	6,977	,543	,532	,720
Q6_R	12,95	7,114	,141	,236	,765
Q7	12,94	6,209	,612	,811	,698
Q8	12,94	6,209	,576	,807	,702
Q9	13,03	7,095	,373	,489	,731
Q10	12,94	7,570	,047	,055	,764
Q11	12,99	7,117	,299	,526	,737
Q12_R	12,85	6,255	,354	,568	,739

Regressie

Dit is een uitbreiding van de correlatie, omdat ook hier gekeken wordt naar een relatie / verband tussen een afhankelijke (Y)- en een onafhankelijke variabele (X).

Het verschil met een *correlatie* is als volgt:

Correlatie = sterkte relatie *Pearson's r*
 Regressie = soort relatie *Lineair of niet?*

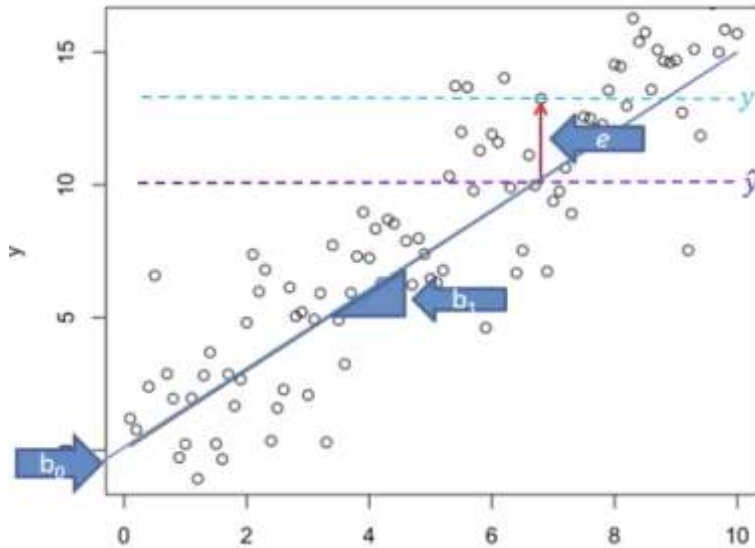
Een correlatie heeft twee *continue* variabelen, die elke mogelijke waarde kunnen aannemen.

Bij een regressie daarentegen, is de Y een continue variabele, maar kan de X ook een *categorische variabele* zijn. Dit betekent dat de variabele geen meetbare waarde kan aannemen, zoals kleuren, voornamen, geslacht, onderwijsniveau en geboortemaanden.

Bij een regressie wordt ook gekeken naar de afstand tussen de *predicted value* (\hat{Y}) en de *observed value* (Y).

In de grafiek heeft $X = 7$ een verwachte waarde van $\hat{Y} = 10$, terwijl de werkelijk geobserveerde waarde $Y = 14$.

Er zit dus een verschil van 4 tussen het residu (Y) en de regressielijn (\hat{Y}). Dit verschil noem je de *error* (e). Deze *error* moet zo klein mogelijk zijn, dus het verschil tussen de Y en \hat{Y} moet zo klein mogelijk zijn.



De regressievergelijking die hoort bij de *observed value* is als volgt:

$$Y = b_0 + b_1x + e$$

b_0 = het intercept; $x = 0$ (begin)

b_1 = slope; helling van de lijn kan positief/negatief zijn

e = error

De regressievergelijking die hoort bij de *predicted value* is als volgt:

		Coefficients ^a		
		Unstandardized Coefficients		Standardized Coefficients
Model		B	Std. Error	Beta
1	(Constant)	20,731	1,143	
	Weight	,665	,014	,860

a. Dependent Variable: What is your goal weight? One time asked ques

$$\text{Goal weight} = 20.717 + .665 \times \text{weight}$$

$$\hat{Y} = b_0 + b_1x$$

Oftewel: $Y - e$

Hoorcollege 4 – (Multi)pele regressie

Sum of Squares

Er zijn drie soorten *sum of squares* die van toepassing zijn bij een regressievergelijking.

De eerste is voor het gehele model, **SSmodel**:

\hat{Y} = de *predicted value*
 \bar{Y} = de *mean value* van alle data

Deze vergelijking gebruik je per individueel residu. Je kijkt naar het verschil tussen de gemiddelde waarde van alle residuen samen en de voorspelde waarde van één residu.

Hoe doe je dat?

Je zoekt de *predicted value* voor dat ene residu en haalt hier vervolgens het gemiddelde van alle residuen samen vanaf. Tot slot doe je de uitkomst in het kwadraat.

Je telt al deze uitkomsten van alle residuen bij elkaar op. Zo krijg je uiteindelijk de som van alle residuen, die samen het gemiddelde (*SSmodel*) vormen.

$$SS_M = \sum (\hat{y} - \bar{y})^2$$

De tweede is voor één enkel residu, **SSresidu**:

Y = de *geobserveerde waarde*
 \hat{Y} = de *voorspelde waarde*

$$SS_R = \sum (y - \hat{y})^2$$

Voor sommige residuen zal deze uitkomst nul zijn, omdat ze precies op de lijn (het gemiddelde) liggen. Er geldt: hoe verder van het gemiddelde (lijn) af, des te hoger de waarde.

De derde is het totaalplaatje, **SStotal**:

Y = de *geobserveerde waarde*
 \bar{Y} = de *mean value* van alle data

$$SS_T = \sum (y - \bar{y})^2$$

Voor al deze waarden kijk je in SPSS bij ANOVA.

Variantie (R^2)

Dit is de totale spreiding van alle residuen t.o.v. de regressielijn noem je de *variantie* (R^2).

De variantie zegt iets over hoe goed het model bij de geobserveerde data past. Liggen de residuen dichtbij



of juist ver van de regressielijn?

Er geldt: $0 \leq R^2 \leq 1$.

Hoe dichter bij 1, des te groter is de variantie en des te kleiner de afstand tussen de residuen en de regressielijn.

Hoe groter de variantie, des te beter past het model bij de geobserveerde data en des te kleiner is de e (*error, voorspellingsfout*).

De formule voor de variantie (R^2) is als volgt:

$$R^2 = \frac{SS_M}{SS_T} \quad SS_T = SS_M + SS_R$$

Met:

Voor de R^2 kijk je in SPSS bij de *Model Summary*.

Multipelere Regressie

Bij een *multipelere regressie* zijn er meerdere X 'en (*predictoren*), wat betekent dat er meerdere relaties kunnen zijn. Deze X 'en kunnen bovendien onderling samenhangen.

Voorbeeld:

Je doet onderzoek naar risicofactoren (X) op het ontstaan van een onderwijsachterstand (Y).

Hierbij kunnen X meerdere factoren zijn; opvoeding, leefomstandigheden, geslacht etc.

Zo zijn er meerdere relaties mogelijk tussen de X en Y .

Predictoren

Uiteindelijk wil je met zo min mogelijk risicofactoren (X 'en) de schoolprestaties van een kind (Y) kunnen verklaren.

Je gaat dus kijken welke risicofactor (*predictor*) de grootste invloed heeft; de sterkste relatie met Y .

Vaak hebben alle predictoren niet dezelfde waarde:

X1 = Man / Vrouw (0-1)

X2 = Leefomstandigheden (0-3)

X3 = Opvoeding (0-7)

Om alle predictoren met elkaar te vergelijken, moeten ze gelijk aan elkaar zijn en gemeten worden op dezelfde schaal. Dit doe je door de *standaardiseren*.

In de SPSS output vind je vervolgens het antwoord onder de *Standardized Coefficients Beta*. De predictor



met de hoogste score, heeft de sterkste relatie (grootste invloed) met de Y.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	69,223	,268		257,929	,000	68,696	69,750
	centered_weight	,564	,013	,730	42,706	,000	,538	,590
	male	6,748	,394	,294	17,140	,000	5,975	7,521
	centered_age	,057	,012	,075	4,833	,000	,034	,080

a. Dependent Variable: What is your goal weight? One time asked question at weighing scales registration

Hoorcollege 5 – Kwalitatief onderzoek

Kwalitatieve interviews

Je bent benieuwd naar de beweegredenen, gedachten, ideeën, meningen en gevoelens van een participant over een bepaald onderwerp of verschijnsel. Aan de hand van interviews kun je goed op antwoorden doorvragen.

Het gaat hierbij niet alleen om het stellen van de juiste vragen (verbaal gedrag) om zoveel mogelijk informatie uit de participant te krijgen, maar ook om non-verbaal gedrag; de interviewer moet een motiverende houding hebben.

Wat maakt een interview goed?

1. Reciprocal interaction

Een interview is niet slechts vragen stellen en hier een antwoord op krijgen, maar ook een kwestie van anticiperen. Zo beïnvloeden antwoorden soms de volgende vraag die je gaat stellen; je vraagt bijvoorbeeld door of moet bepaalde dingen in jouw volgende vraag iets aanpassen.

2. Responsiveness

Je moet ervoor zorgen dat de participant antwoorden geeft, ook al is deze brak, het niet met je eens, spreekt hij de taal niet etc. Je moet je dus aanpassen aan de situatie.

3. Mutual trust

Niet alleen tijdens de dataverzameling (interviews), maar ook tijdens het verwerken van deze data dien je er vertrouwelijk mee om te gaan en garandeer je anonimiteit.

Wat doet een goede interviewer?

1. Een rapport opstellen

2. Actief luisteren

De antwoorden van de participant niet alleen aanhoren, maar ook daadwerkelijk begrijpen.

3. Relevante vervolgvragen stellen

Hiervoor heb je een 'topic guide', zodat je je tijdens het interview op de participant kunt focussen en niet actief hoeft na te denken over wat je straks gaat vragen.

Interview vormen

De meest voorkomende vijf interview varianten zijn:

1. Face-to-face interview

Hier kun je naast verbale communicatie ook gedrag observeren (non verbale communicatie).

2. Telephone

Alleen verbale communicatie.

3. Online (o.a. Skype)

Zowel verbale- als nonverbale communicatie.

4. Go-along interview

Je gaat mee in de dagelijkse gang van zaken van de participant, dus komt bijvoorbeeld langs op diens werk en loopt een rondje in de pauze waar je vragen stelt. Dit wordt gedaan bij participanten die heel druk zijn, zoals CEO's, politici etc.



5. Etnografisch interview

Je volgt de participant niet een aantal uur, maar maanden of jaren. Zo verzamel je je data in episodes.

Het interview-proces

Stage 1: Arrival and introduction

- Establishing rapport
- Hosting the interaction

Stage 2: Introducing the research

- Seeking informed consent
- Scope of the interview

Stage 3: Beginning the interview

- Contextual background information

Stage 4: During the interview

- Breadth and depth of coverage

Stage 5: Ending the interview

- Advance notice
- Positive note

Stage 6: After the interview

- How the information will be treated and used
- Stay and help change the mode
- Listen for doorstep data

Focus Groups

Wat is een Focus groep?

Een groep met meerdere participanten, onder leiding van een interviewer, die gezamenlijk praten over het onderzoeksonderwerp.

Zo kun je ook informatie halen uit de interactie tussen verschillende participanten, die er tijdens een interview (één-op-één) niet is. Het gaat dus niet alleen om de antwoorden, maar ook om het gedrag.

“It is a carefully planned series of discussions designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment”.

Je vraagt dus meerdere participanten tegelijkertijd iets over een bepaald onderwerp, waarna zij mogen antwoorden op zowel jou (interviewer) als anderen participanten. Participanten krijgen hier de mogelijk



om te luisteren naar, reflecteren op en heroverwegen van hun eigen standpunt. Halverwege dit gesprek kan de data dus veranderen, omdat iemand van mening verandert door de perspectieven van andere participanten.

Let op: het is géén groepsinterview, want de participanten gaan ook met elkaar in gesprek. Er is hier dus sprake van interactie tussen participanten, wat niet het geval is tijdens een interview. Jij bent daar de enige die vragen stelt en er is alleen interactie tussen de interviewer en participant.

Moderator

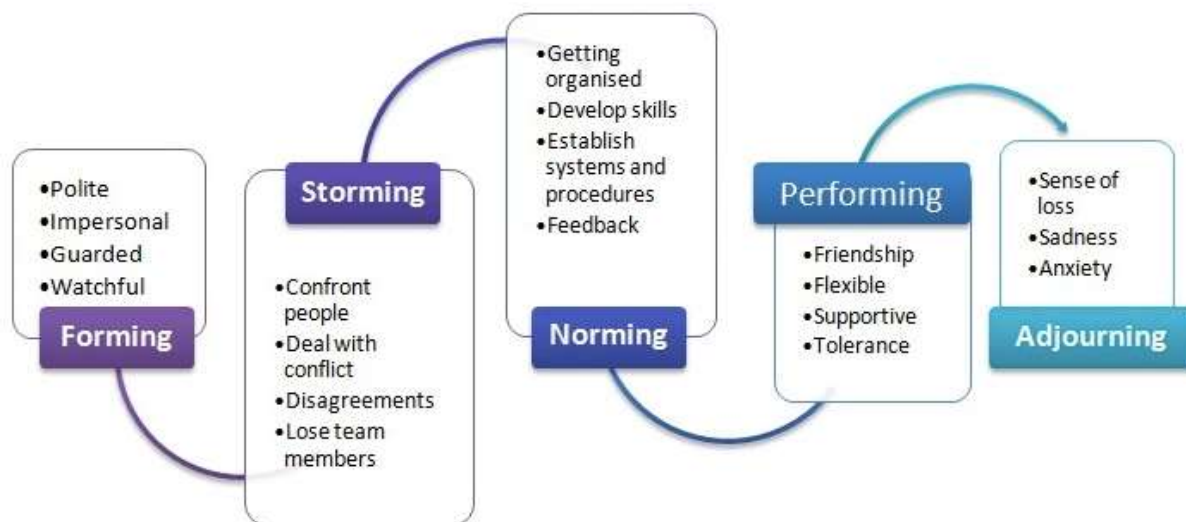
De onderzoeker treedt in een Focus-groep op als moderator. Hij zorgt ervoor dat de gesprekken soepel verlopen, dat iedereen aan het woord komt en dat er niet van het onderwerp afgedwaald wordt.

Waarom focusgroepen?

Je organiseert een focusgroep om diagnoses te stellen van bepaalde problemen, nieuwe ideeën te stimuleren en programma's te evalueren. Je organiseert deze focusgroep niet om tot overeenstemming met elkaar te komen, maar juist om verschillende meningen en stakeholders te verzamelen.

Je kunt ook een focusgroep organiseren voor een toelichting op de kwantitatieve data (cijfers) die je hebt verzameld.

Fases van een Focus-groep



Storming stage:

De participanten zijn nog bezig hun gelijk te halen en staan nog niet open voor andere perspectieven. In deze fase stel je dus niet jouw belangrijkste vragen, omdat ze nog niet in staat zijn hun daadwerkelijke, goed overwogen antwoord te geven.

Groep compositie en grootte

Bij het samenstellen van de Focus-groep moet je de volgende punten overwegen:

1. Heterogeniteit / homogeniteit

Kies je voor een heterogene groep met participanten van verschillende bedrijven / uit verschillende klassen of juist een homogene groep met participanten werkend bij hetzelfde bedrijf / uit dezelfde klas.



2. Vreemden, kennissen of al bestaande groepen

Soms praten participanten makkelijker als ze de andere participanten niet kennen. Anderen praten juist makkelijker als ze de andere participanten wel kennen of zelfs vrienden met hen zijn.

3. Groepsgrootte

Tussen zes-twaalf is ideaal.

Soorten focusgroepen

Er zijn verschillende soorten focusgroepen:

1. Two-way focusgroep

Er zijn twee aparte focusgroepen met twee aparte moderatoren. Groep 1 begint met discussiëren, terwijl groep 2 toekijkt. Hierna wisselen de rollen om. Het idee is dat door te luisteren naar de discussie van de andere groep, je betere discussies en conclusies krijgt.

2. Dual Moderator

Er is één focusgroep met twee moderatoren; een gespreksleider en een observator.

3. Dueling moderator

Er is één focusgroep met twee moderatoren. Deze moderatoren hebben allebei een andere mening over het onderwerp en voeren hier voor de focusgroep een discussie over. Zo krijgen de participanten het onderwerp vanuit verschillende perspectieven te zien.

4. Respondant Moderator

Een participant neemt de rol van de moderator tijdelijk over, om diepere / meer diverse antwoorden te krijgen. Wellicht praten participanten makkelijker tegen iemand die ze kennen dan een onbekende moderator.

5. Online focusgroep (o.a. Skype & WhatsApp)

- Chatroom

Allemaal op dezelfde plek en tegelijkertijd.

- Bulletin Board

Je maakt een document aan, waar gedurende een bepaalde tijd participanten comments in kunnen plaatsen.

Hulpmiddelen

Ook tijdens een focusgroep kun je als moderator gebruik maken van verschillende hulpmiddelen:

1. Topic guide

Een schema met onderwerpen die je afgaat, een soort rode draad.

2. Memos / fieldnotes

Aantekeningen die je tijdens de focusgroep maakt, o.a. steekwoorden, context, observaties etc.

Hoorcollege 6 – Kwalitatief onderzoek

Observaties

Er zijn drie soorten observaties:

1. Participant - Non-participant
2. Overt - Covert
3. Systematic - Non-systematic

Data Analyse

We kijken alleen naar de *analytic codes*; analytisch. Beschrijft aan de hand van literatuur en de verzamelde data.

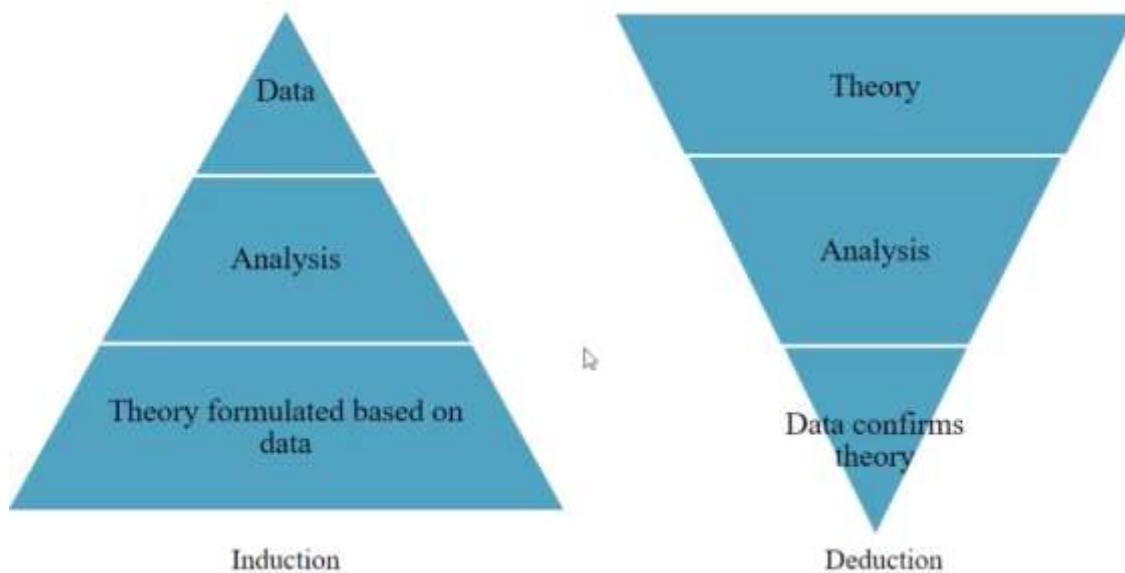
Inductie en deductie

Deductie: Er is een bestaande theorie, die je gaat toetsen om te kijken of deze nog steeds relevant / waar is. Het is dus een soort theorie-testen (*theory testing*).

= theorie testen

Inductie: Je begint met het verzamelen van de data, analyseert deze en gaat aan de hand hiervan een bepaalde theorie (conclusie) formuleren.

= theorie bouwen



Grounded Theory: Inductie

Sensitizing concepts

Je begint een inductieve analyse met het opstellen van **sensitizing concepts**. Deze worden gemaakt voordat de dataverzameling- en analyse plaatsvindt, dus je kunt ze vergelijken met de fundering van een huis; het begin van jouw onderzoek, een framework om mee te starten.

Het zijn nog redelijk vage, brede concepten. Ze geven suggesties voor jouw onderzoek en welke richting



dit op kan gaan, maar zijn nog niet bepalend of specifiek (zoals definitieve concepten). Er kunnen tijdens het onderzoek dus nog aanpassingen of veranderingen optreden.

“It gives the user general sense of reference and guidance, no constraining (tunnelvisie)”

Voorbeeld: Je onderzoekt hoe het schoolklimaat leraren beïnvloedt in hun leren. Schoolklimaat is hier het brede sensitizing concept. Je kunt dit concept meetbaar maken door het op te delen in kleinere, specifieke a-priori codes, zoals de relatie tussen de leraar en collega's, de relatie tussen de leraar en de leerlingen etc. Al deze codes samen meten het sensitizing concept 'schoolklimaat'.

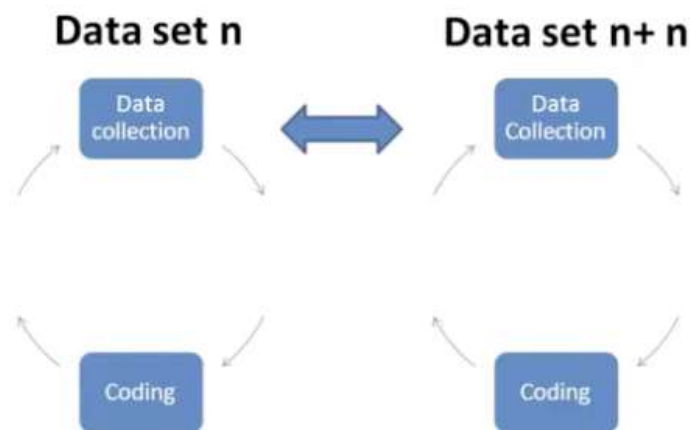
A-priori codes

Deze codes worden ontwikkeld voordat de dataverzameling- en analyse plaatsvindt. De codes worden samengesteld aan de hand van literatuur en eerdere onderzoeken.

Constant comparison

Je analyseert een interview stelt hierbij codes op. Dit doe je ook met de volgende interviews. Vervolgens leg je alle hieruit voortgekomen codes naast elkaar en zorgt dat deze overeenkomen, zodat je hieruit een gezamenlijke theorie kunt ontwikkelen.

Gedurende het analyseren van de data ben je dus constant aan het vergelijken; hoe verschilt hetgeen wat deze respondent zegt t.o.v. de vorige? Hoe is de vorige code anders dan deze code?

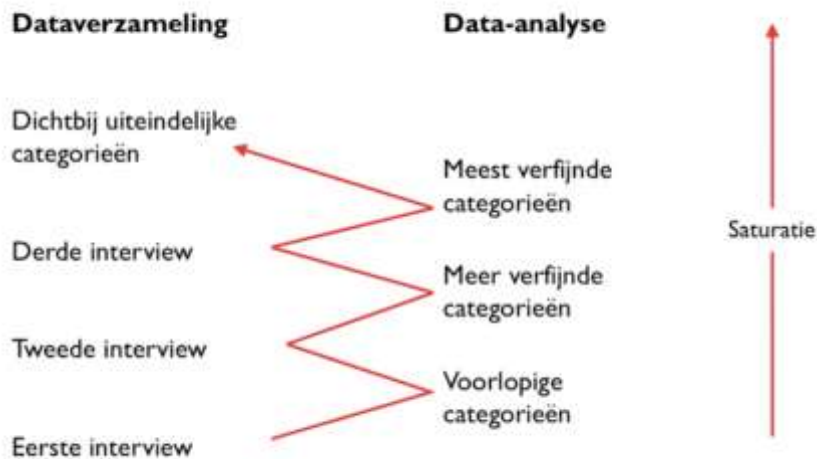


Grounded Theory: Het inductive coding process

De **Grounded Theory** is een systematische manier van kwalitatieve gegevens analyseren, waardoor je al tijdens het onderzoek doen een bepaalde theorie kunt vormgeven / ontwikkelen (= inductie). Je verwerkt

V O C U S

de data op zo'n robuuste manier, waardoor je op basis hiervan al een betrouwbare theorie kunt



ontwikkelen.

Deze analysetechniek bestaat uit twee fases:

Fase 1: Uiteenrafelen (*segmenting*)

Je deelt de data (interview) op in kleine stukjes en gaat deze op twee manieren coderen:

1. Open coderen

Dit doe je als eerst. Hierbij geef je de data bepaalde labels / hoofdthema's (codes), waardoor er een codeboom ontstaat.

Het einde van deze codeerfase noem je **saturatie**; de codeboom is toereikend voor de data en er hoeven geen nieuwe codes meer toegevoegd te worden.

Er zijn twee soorten codes:

- *in vivo-codes* gebaseerd op de data
- *constructed codes* gebaseerd op de literatuur (concepten, begrippen etc.)

- I: Welke problemen loopt u of uw leerlingen tijdens de begeleiding het meest tegenaan?
- D: Voor mij is dat tijd. Als je al geen tijd hebt. Als ik tijd zou hebben, zou ik kunnen zeggen, dat en dat zijn de problemen, maar ik heb al geen tijd dus.
- I: Maar als u tijd zou hebben, wat zijn dan de problemen?
- Ja dat weet ik niet, dat is allemaal speculatief, want ik heb te weinig tijd. Voor zover ik dan kan spreken van problemen binnen die beperkte tijd die ik heb, zijn, het moeilijke is dat de leerlingen niet doorhebben hoe moeilijk het is om een simpel ding te onderzoeken. Ze denken zoiezo als ze beginnen, "hier is geen informatie over", terwijl er al enorm veel informatie over te vinden is. Ze beginnen dan met een heel breed onderwerp, daar gaan ze dan een beetje dingen over vertellen, maar het is... Je wilt eigenlijk dat ze in de loop van het onderzoek een soort wetenschappelijke nieuwsgierigheid ontwikkelen.

Tijdgebrek
begeleiding

Probl. uitvoeren
onderzoek

Probl. formuleren
onderzoeks-
vraag

Laag tempo
Concentratie
Gebrek aan onderzoeksvaardigheden
Gebrek aan sociale vaardigheden
Samenwerkingsproblemen
Dominantie
Meeliften
Te moeilijk
Rumoer
Onrust
Tijd
Weinig zicht op proces

2. Axiaal coderen

Dit doe je na het open coderen. De codes die je net gevormd hebt, ga je met elkaar vergelijken en vervolgens onderbrengen in overkoepelende thema's / categorieën.

Je clustert ze dus als het ware en brengt eventuele verbanden in kaart. Ook krijgen de concepten die je onderzoekt hierdoor meer inhoud (conceptualisering).

Hierbij maak je de codeboom kleiner en gestructureerder, door het toevoegen van hoofd- en subcodes (clusters).

Het einde van deze codeerfase is ook **saturnatie**: de (nieuwe) codeboom is toereikend voor de data en er hoeven geen nieuwe codes meer toegevoegd te worden.

Problemen II.	Cognitief	Laag tempo
		Concentratie
		Gebrek aan onderzoeksvaardigheden
		Te moeilijk
		Gebrek aan sociale vaardigheden
Problemen I.	Organisatorisch	Rumoer
		Onrust
		Tijd
Problemen I.	Begeleiding	Weinig zicht op proces

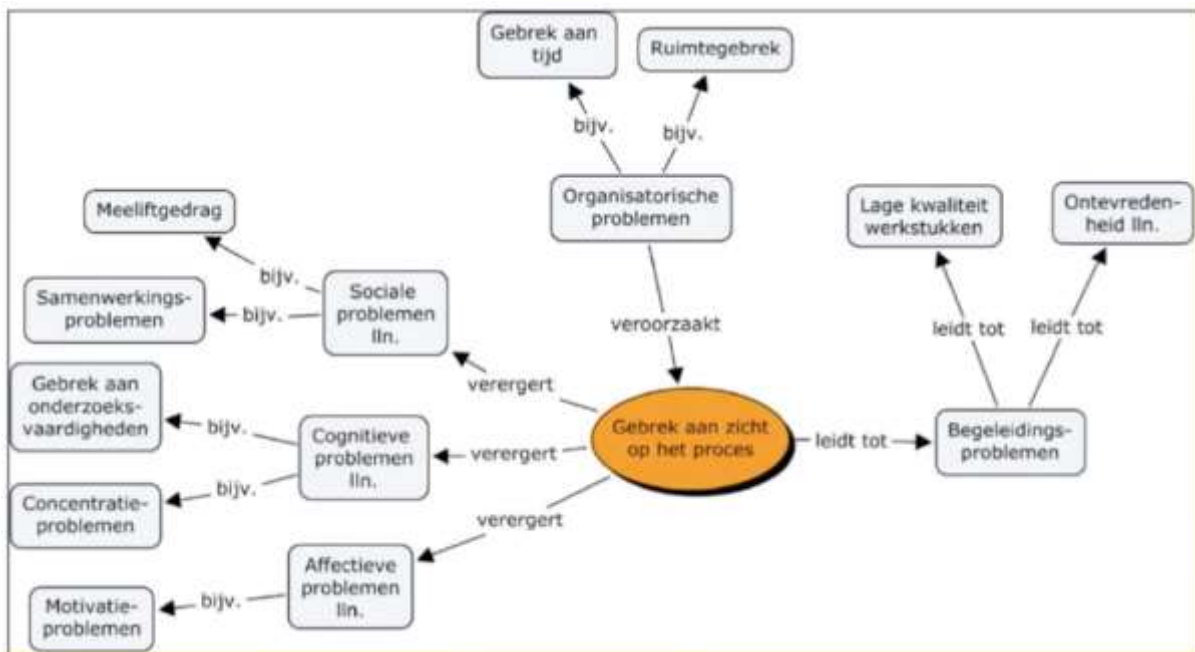
Fase 2: Structureren (*reassembling*)

Je gaat het interview dat je hebt opgedeeld in kleine, uiteengegafte stukjes weer samenvoegen, waardoor er een grounded theory ontstaat.

Hierbij hoort een derde vorm van coderen:

3. *Selectief coderen*

Je bepaalt de belangrijkste categorie (kerncategorie) in de data; welke codes leveren de theorie waarnaar jij op zoek bent en helpen bij het beantwoorden van de onderzoeksvraag? Op deze manier construeer je dus een grounded theory.



Kwalitatieve observaties

Je maakt gebruik van **observaties** bij jouw dataverzameling om complexe processen of interacties (met o.a. mensen, de omgeving etc.) in kaart te brengen. Ook kun je goed kijken naar onbewust, instinctief gedrag en onuitgesproken normen en waarden. Je gebruikt dus observaties om te begrijpen wat er gebeurt, niet waarom iets gebeurt (daar heb je interviews voor nodig). Dit noem je **familiarisation**.

Verzamel je data uitsluitend aan de hand van observaties, noem je dit een *core method*. Maak je naast observaties ook gebruik van o.a. interviews (= triangulatie) om meer begrip te creëren, noem je dit een *supplementary method*. Zo kan iemand een interview afnemen en is er ook een observant aanwezig die op non-verbale gedragingen let.

Het is belangrijk dat een onderzoeker niet **biased** (bevooroordeeld) is, omdat hij dan kan focussen op bepaalde dingen die hij verwacht te zien, en hiermee andere belangrijke punten mist. Een oplossing hiervoor is het maken van memo's / fieldnotes en die gebruiken tijdens het analyseren van de data.

Er zijn twee soorten observaties:

1. **Systematische observatie**
 - Situatie is van tevoren bepaald
 - Meer dan één observator



- *inter-observer agreement*: observators moeten allemaal dezelfde situatie zien
- *intra-observer agreement*: observators moeten gedrag hetzelfde beoordelen, ongeacht wanneer of hoe de observaties zijn gedaan
- Onderzoeksvraag en meetinstrumenten zijn van tevoren bepaald
- Codes zijn van tevoren bepaald
- De duur is van tevoren bepaald
- Deductief
- Observator is *complete observator*

2. **Non-systematische observatie**

- Situatie kan tijdens het onderzoek nog veranderen
- Slechts één observator
- Onderzoeksvraag kan nog veranderen a.d.h.v. wat je observeert
- Codes zijn niet van tevoren bepaald, worden a.d.h.v. de data gemaakt
- De duur is flexibel, hangt ervan af hoe lang de observatie doorgaat
- Inductief
- Observator is *complete participant*

Het observatie-proces



Stap 1: Een (sociale) situatie kiezen.

Stap 2: Data verzamelen.

Welk fenomeen ga je wanneer en bij wie? Ook kijk je naar bijzaken waar je rekening mee moet houden (= side arrangements).

Rol

Als observant moet je kiezen welke rol je aanneemt:

- Complete participant

Je bent als observator ook participant. Je neemt deel aan de actie die plaatsvindt, gaat volledig in je rol op en doet de observatie er slechts naast. Op deze manier kun je ook achter de gedachtegang en beweegredenen van de participanten komen.

Voorbeeld: je wordt als student gevraagd de leraar te observeren tijdens een hoorcollege. Je bent gewoon student (= participant), volgt het hoorcollege (= neemt deel aan de actie) en het observeren van de docent is slechts iets wat je ernaast doet.



= overt en active

- Complete observer

De observator neemt geen deel aan de actie en maakt zijn rol niet kenbaar. Deze manier is objectief.

Voorbeeld: Je gaat in Hoog Catherijne zitten en observeert de voorbijgangers (= neemt een deel aan de actie).

= covert en passive

Recording observaties

Je kunt observaties ook opnemen, maar dan moeten ze tenminste de volgende punten bevatten:

- **Primary observation**
Datum, tijd, locatie, aanwezig, situatie, wat jij geobserveerd hebt (alles objectief)
- **Secondary observation**
Zaken die anderen geobserveerd hebben; hebben ze iets nieuws gezien wat jij over het hoofd hebt gezien?
- **Experiential data**
Als observant ben jij het meetinstrument, dus moet je reflecteren op jouw state of mind en emotionele toestand die eventueel het onderzoeksproces of de interpretatie van de data kunnen beïnvloeden.
- **Circumstantial- en background data**
Achtergrondinformatie over jouw participant / bedrijf etc.

Internet observatie

Een observatie via het internet heeft een aantal voordelen, namelijk dat het geen geografische restricties heeft en alle discussies zichtbaar zijn die in het echte leven wellicht 'hidden' waren. Hoe gaat dat in z'n werk en waar moet je op letten:

- Websites selecteren (data- en informatierijk)
- Hier toegang tot krijgen
- Selectie van alle data maken (steekproef)
- Beslissen over de rol van de observator; doet hij mee aan de online discussies of kijkt hij slechts toe?
- Wat ga je precies observeren?
- Hoe ga je de data bewaren? Foto's maken? Downloaden?

Hoorcollege 7 – Kwalitatief onderzoek

Archival data

Dit is al bestaande data, zoals archiefstukken, nieuwsberichten, dagboeken etc.

Er komen hier dus geen observaties, interviews of focusgroepen aan te pas.

Je kunt archival data gebruiken voor **triangulatie**, waarbij je gebruik maakt van meerdere manieren om data te verzamelen. Zo kun je namen of bedrijven die jouw participant tijdens het interview noemt opzoeken, checken of beweringen die de participant over o.a. een bepaald programma maakt ook echt waar zijn om zo jouw interviews te ondersteunen of te vergelijken.

Er zijn twee soorten archival data:

1. Manifest data

Punten die de observator ziet en opvallen worden beschreven, maar verder (nog) niet geïnterpreteerd.

Voorbeeld: Hoe vaak studenten elkaar bij naam aanspreken of hoe vaak er gerookt wordt tijdens een tv-show.

2. Latent data

Je gaat de data ook daadwerkelijk interpreteren.

Voorbeeld: Het vermogen van studenten om namen te onthouden of hoe worden mensen die roken op tv in beeld gebracht (cool, neutraal, vies etc.).

Tekstuele archival data

Dit is het gebruiken en analyseren van transcripten van interviews en focusgroepen gebruiken, o.a. door:

- **Documenten verzamelen**

- Is zonder tussenkomst van de observator, dus geen observer bias
- Voorbeelden: beleidsdocumenten, jaarverslagen, websites, wetgeving, nieuwsartikelen, brieven, aantekeningen, transcripten.
- Worden gepubliceerd door instanties / bedrijven zelf, door derden over de instanties / bedrijven (o.a. de media) of persoonlijk, onofficieel en informeel
- Toegankelijkheid: documenten kunnen publiekelijk zijn, met toestemming toegankelijk (instantie) of privé (persoon)

Visueel archival data

Dit is het gebruiken van afbeeldingen en foto's.

- **Afbeeldingen en foto's**

- Kan gebruikt worden om dingen te herinneren (meer informatie) of participanten laten vertellen a.d.h.v. een foto (vertellen makkelijker en meer).
- Voorbeelden: advertenties, websites, kranten, magazines, foto's uit privé-collecties (fotoalbums), tekeningen.
- Worden gemaakt door de participanten zelf (personal), door de onderzoeker of universiteit zelf (professioneel) of door bedrijven (commercieel).
- Toegankelijkheid: rekening houden met copyrights; is het voor persoonlijk / professioneel gebruik of commerciële doeleinden?



Kwalitatieve content analyse: Deductieve data analyse

Bij deductief onderzoek hoort **theory testing**, waarbij al eerder geformuleerde theorieën systematisch worden getest en gekeken wordt of ze nog relevant / waar zijn. Je kunt ze eventueel ook uitbreiden. Soms zitten er ook **negative cases** tussen; hier blijkt de theorie / het resultaat dat jij vindt niet overeen te komen met eerdere soortgelijke theorieën en resultaten. Dit zijn outliers en kunnen er zowel negatief als positief bovenuit schieten.

Categorization Matrix

Dit is vergelijkbaar met de sensitizing concepts van inductief analyseren. Ook hier ga je voorafgaand aan het onderzoek op zoek naar enkele hoofdconcepten- en variabelen die de basis vormen voor het verdere onderzoek. In tegenstelling tot sensitizing concepts zijn deze minder flexibel en zullen ze minder snel veranderen tijdens het onderzoek (kunt niks verwijderen, alleen dingen toevoegen). Deze zet je vervolgens in een **categorization matrix**.

Een ander verschil met sensitizing concept is dat de concepten in de categorization matrix gebaseerd op eerder onderzoek, literatuur, bestaande modellen etc.

An example from a categorization matrix

	Dependence	Worries	Sadness	Guilt
What kind of mental well-being threats does diabetes have for adolescents?				

Stappenplan deductief analyseren

Stap 1: Categorization matrix maken

Stap 2: Data verzamelen

Stap 3: Coderen

- gebruik de concepten uit de categorization matrix
- ook data die hier niet bij past labelen ('un-fitting')
- kijk hoe goed de data bij de matrix past
- kijk naar de achterliggende redenen (latent)

Stap 4: Is jouw verwachting waar? Geldt de bestaande theorie ook onder andere omstandigheden? Indien dit niet zo is, waarom niet, wat zijn alternatieve verklaringen?

Inductief vs. Deductief

Inductive	Deductive
<ul style="list-style-type: none"> • Bottoms-up • Uses sensitizing concepts (for giving the researcher a general sense of reference and guidance) • Example of coding process – open -> axial -> selective • Flexibility regarding new (unexpected) data or codes – High • Often theory building 	<ul style="list-style-type: none"> • Top-down • Uses categorization matrices (for retesting key concepts and variables) • Example of coding process - apply the matrix -> label 'un-fitting' data -> compare data-matrix fit -> explore underlying (latent) content • Flexibility regarding new (unexpected) data or codes – Low to moderate • Often theory testing

Hoorcollege 8 – Kwalitatief onderzoek

Rigor

Met behulp van Rigor kun je de betrouwbaarheid en validiteit van kwalitatief onderzoek testen. De checklist hiervoor bestaat uit vier onderdelen:

1. Reliability

Er is geen random error (E), waardoor elke keer dat de studie exact hetzelfde wordt uitgevoerd er ook exact dezelfde resultaten uitkomen (= betrouwbaarheid). De error is namelijk ook elke keer precies hetzelfde.

Om reliability te garanderen, gebruik je een database en study protocol, waarin alles precies is vastgelegd en iedere andere onderzoeker kan nagaan- en doen wat jij hebt gedaan.

2. Internal validity

De juiste conclusies trekken a.d.h.v. de data, waarbij je wil dat de relatie tussen de twee variabelen causaal is (direct oorzaak-gevolg).

Om internal validity te garanderen, gebruik je een research framework (sensitizing concepts / categorization matrix), om een basis voor het onderzoek vast te stellen. Ook gebruik je triangulatie, zodat je een fenomeen vanuit meerdere invalshoeken bekijkt en zekerder kunt zijn van je resultaat.

3. Construct validity

De onderzoeksmethode die je gebruikt is de best passende methode bij jouw onderzoek. Het gehele construct moet hierbij gemeten worden, waardoor de uitspraak die jij doet overeenkomt met de realiteit.

Om construct validity te garanderen, gebruik je triangulatie, zodat je de best passende methode vindt en legt uit hoe en waarom je op deze manier de data analyseert. Ook gebruik je feedback van de participanten / peers / collega's, kijk je onder welke omstandigheden jouw onderzoek is afgenomen en vergelijkt deze met de werkelijkheid (= generaliseren).

4. Generalizability (= external validity)

Empirisch: In hoeverre jouw resultaat generaliseerbaar is naar de gehele populatie (= external validity).

Theoretisch: in hoeverre de data past bij de geformuleerde theorie.

Om generalizability te garanderen, gebruik je een *rationale* (= reden / uitleg) voor het selecteren van de participanten en geef je het fenomeen dat je onderzoekt context om te kijken of de data er echt goed bij past.

Ethiek

Er zijn vijf perspectieven wat betreft ethiek:

1. Universalism approach

Ethische regels mogen nooit gebroken worden.

2. Utilitarian approach

Je bekijkt een actie vanuit diens consequenties en uitkomsten en weegt de voor- en nadelen



tegen elkaar op. Op deze manier probeer je het volgende te bereiken:

“It strives to achieve the greatest good for the greatest number while creating the least amount of harm or preventing the greatest amount of suffering”.

3. **Virtue ethics**

Een persoon kiest op basis van zijn eigen normen en waarden zelf wat hij / zij doet of wil, hier heeft de onderzoeker geen grip op.

4. **Relational ethics**

De relatie tussen de onderzoeker en de participant; als deze elkaar al kennen zullen ze (etisch) anders met elkaar omgaan in vergelijking met als ze elkaar nog nooit gezien hebben.

5. **Situational / principled relativism**

Case-by-case approach, oftewel: “considering each case / participant individually rather than considering several cases / participants together as a whole.”

Elke stap, participant en beslissing binnen een onderzoek apart en gelijkwaardig bekijken.

Onderzoeksethiek

Bij kwalitatief wetenschappelijk onderzoek zijn er een aantal ethische principes:

1. **Informed Consent**

De participant moet volledig op de hoogte zijn van wat deelname aan het onderzoek precies inhoudt, wat de mogelijke nadelen zijn en dat je je op elk moment weer kunt terugtrekken.

2. **Privacy**

Deelname is volledig anoniem (anonimiteit), wat inhoudt dat je geen gegevens van de participanten verzameld die hun identiteit kunnen onthullen voor buitenstaanders (o.a. naam, geboortedatum).

Ook moeten de onderzoekers vertrouwelijk omgaan met de gegevens (vertrouwelijkheid) en moet de rapportage van de gegevens op zo'n manier gedaan worden, dat mensen die het lezen er niet achter kunnen komen om welke deelnemers het gaat.

3. **Schade**

Je mag participanten geen kwaad doen, oftewel fysieke of psychische schade aanrichten bij de participanten.

Een voorbeeld van ethische schending is het Milgram onderzoek, waar de participanten gevraagd werden hun 'leerling' (acteur) bij elk fout antwoord een steeds hogere stroomschok te geven. Dit kan psychische schade aanrichten bij de participanten, die gedwongen worden de 'leerling' steeds meer pijn te doen (emotionele stress). Ook moesten ze blijven zitten, ook al vonden ze het erg vervelend en wilden ze stoppen met het onderzoek. Tot slot waren de participanten niet volledig op de hoogte van het onderzoek (informed consent), daar ze niet wisten dat de leerling gespeeld werd door een acteur.

Reflexivity

Je kijkt hierbij terug op het onderzoek dat jij hebt uitgevoerd; alsof je in de spiegel kijkt en nagaat wat je kan veranderen om het beter te maken. Er zijn twee soorten:

1. **Personal reflexivity**

Je reflecteert op jouw betrokkenheid als onderzoeker, jouw gedrag en gedachten en hoe dit eventueel het onderzoek / resultaat beïnvloed kan hebben (= zelfreflectie).



2. **Epistemological reflexivity**

Je reflecteert op de gebruikte onderzoeksmethoden en geformuleerde theorie.

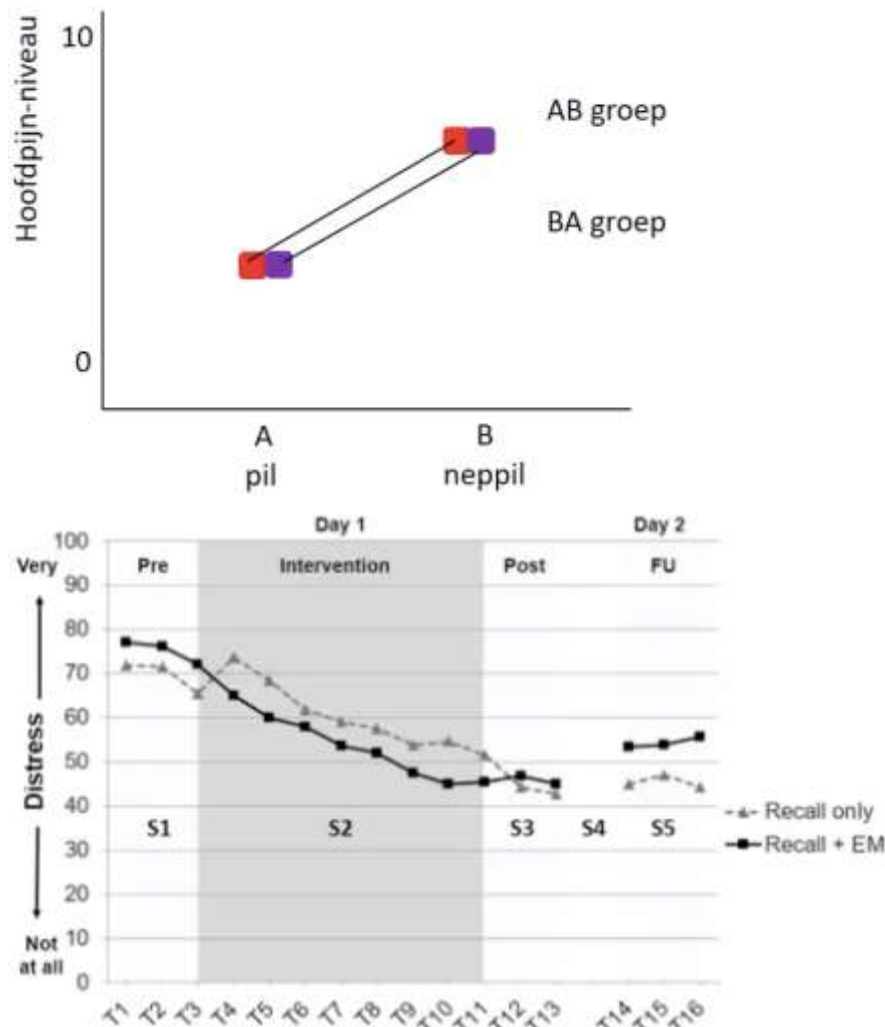
Je garandeert reflexivity als volgt:

- Assumpties checken
- Zaken als normen, waarden, sekse, economische status etc. van zowel jouzelf als de participant mogen geen invloed hebben op het resultaat
- Jouw emoties en relatie met de participant mogen geen invloed hebben op het resultaat
- De fysieke omgeving en logistiek mogen geen invloed hebben op het resultaat

De theorieën en onderzoeksmethoden die je gebruikt mogen ook geen invloed hebben op de resultaten

Hoorcollege 9 – Experimenteel onderzoek

Within-subject: Er is slechts één onderzoeksgroep (dus geen controle/experimenteel), waarbij elke participant aan dezelfde factoren wordt blootgesteld. Je kijkt vervolgens welke effecten over tijd hebben plaatsgevonden.



Voorbeeld: Je onderzoekt in hoeverre de participanten beïnvloed worden door de media in hun politieke voorkeur. In de voormeting (T1) laat je de participanten hun politieke voorkeur uitspreken, waarna ze zes maanden veelvuldig worden blootgesteld aan politiek-gerelateerde berichten in de media.

Na deze zes maanden wordt een nameting (T2) gedaan, waar de participanten nog eens gevraagd hun politieke voorkeur uit te spreken. Hieruit blijkt dat ze naar rechts zijn opgeschoven. Tijdens de periode die er tussen T1 en T2 zit is de politieke voorkeur dus veranderd.

Er zijn drie factoren die ervoor kunnen zorgen dat de resultaten binnen één groep variëren: een meetfout, individuele verschillen en externe verstoringen.

Between-subject: Hierbij kijk je naar de invloed van verschillende factoren op twee of meer groepen (controle/experimenteel), vergelijkt ze en bepaalt vervolgens welke factor het meest gewenste effect heeft.



Voorbeeld: Je wil weten of depressies verholpen kunnen worden door psychiatrische hulp in plaats van door medicijnen. Een groep patiënten wordt verdeeld in een controle- en experimentele groep. De controlegroep krijgt de medicijnen, terwijl de experimentele groep wordt behandeld met psychiatrische hulp. Vervolgens wordt na zes maanden gekeken welke van de twee het meeste effect heeft gehad op het verhelpen van depressies.

Between factor: Dit is de factor (onafhankelijke variabele / manipulated variabele) die de participanten in verschillende groepen verdeeld, o.a. leeftijd, sekse, tijd of een experimentele- en controlegroep.
= categorische variabelen

Wat is experimenteel onderzoek?

Experimenteel onderzoek is onderzoek dat bestaat uit twee soorten variabelen:

1. **Manipulated variabele**

De gecontroleerde variabele. De onderzoekers bepalen zelf tot welke groep (controle/experimenteel) de participanten behoren.
= onafhankelijke variabele

2. **Measured variabele**

De variabele die gemeten wordt, zoals het vertonen van bepaald gedrag of denkpatronen.
= afhankelijke variabele

Voorbeeld: Onderzoekers zijn benieuwd of het maken van aantekeningen via de laptop of met de hand invloed heeft op de leerprestaties. Ze verdelen de klas in tweeën, waarvan de ene helft aantekeningen maakt met de hand en de andere helft op de laptop. Vervolgens nemen ze een toets af en vergelijken de leerprestaties met elkaar.
Hier is de manier waarop de participanten aantekeningen maken bepaald door de onderzoekers, oftewel de *manipulated variabele*. Vervolgens worden de leerprestaties gemeten en vergeleken, wat dit de *measured variabele* maakt.

Causaliteit

Bij experimenteel onderzoek ga je op zoek naar een **causaal verband** tussen twee variabelen, oftewel: de directe oorzaak voor een gevolg. **Causaliteit** houdt in dat de onafhankelijke variabele X de directe oorzaak is voor de afhankelijke variabele Y.

Deze twee moeten na elkaar plaatsvinden, dus niet tegelijkertijd. Er mag verder geen sprake mag zijn van een derde (mediërende) variabele die het verband overdraagt, een toevallig schijnverband of een alternatieve verklaring.

Voorbeeld: Je onderzoekt welke soort therapie het beste werkt bij het behandelen van PTSD. Tetris blijkt hierbij goed te werken; mensen die voor en na een traumatische gebeurtenis Tetris speelden, hadden een snellere daling in de mate van PTSD over tijd dan mensen die natuurlijk herstelden (natuurlijk herstel = **naturation threat**).

Later bleek echter dat dit komt doordat Tetris bepaalde synapsen in je hersenen blokkeert, waardoor traumatische gebeurtenissen minder goed onthouden worden. Tetris is dus niet de enige factor die dit kan oproepen, maar ook o.a. Marihuana roken of Snake spelen.

Bij het voorbeeld hierboven heb je dus geen causaal verband; het maakt niet uit of je Tetris speelt, Snake speelt of Marihuana rookt, het gaat erom dat die synapsen uitgeschakeld worden, waardoor je de traumatische ervaring minder goed onthoudt en er sneller van hersteld.

Kortom:

Directe oorzaak: blokkeren synapsen.

Mediërende variabele: Tetris spelen, Snake spelen, Marihuana roken etc.

Direct gevolg: snellere afname PTSD

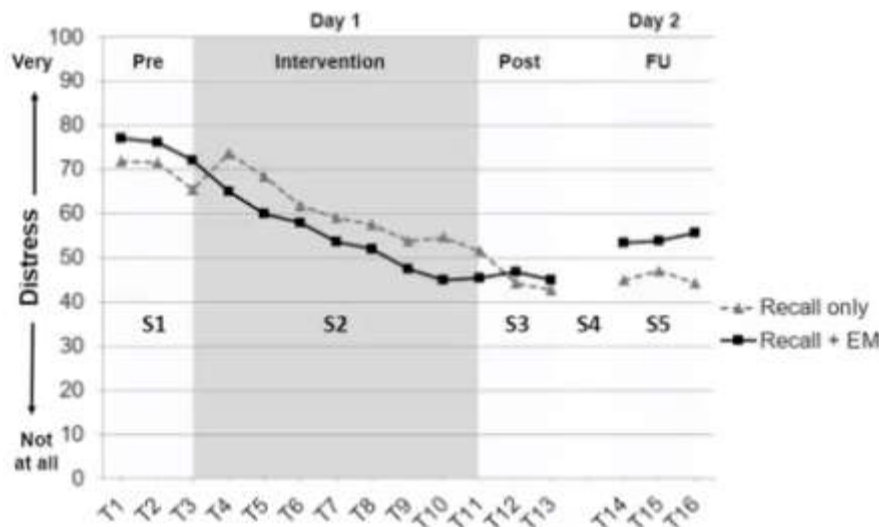
Randomisatie

Randomisatie is enorm belangrijk bij het vaststellen van een causaal verband; je wil zeker weten dat de verdeling van de groepen niet zorgt voor een alternatieve verklaring, bijv. omdat de experimentele groep toevallig mensen heeft met een mindere traumatische ervaring (mindere mate van PTSD) dan de controlegroep, waardoor zij sws een snellere afname in PTSD hebben (dan maakt het spelen van Tetris dus niet uit). Of dat de ene groep uit alleen maar meisjes bestaat en de andere groep uit alleen maar jongens, waardoor uiteindelijk sekse een snellere afname van PTSD bepaald in plaats van het spelen van Tetris.

In de afbeelding hieronder zie je bijv. dat de twee groepen op een ander niveau beginnen, waardoor je kunt concluderen dat ze niet gelijk aan elkaar zijn en je ze niet goed kunt vergelijken. Hier is de randomisatie dus mislukt.

Threats

Er zijn verschillende soorten threats die op kunnen treden bij het doen van onderzoek en van invloed zijn op de interne validiteit (= het trekken van de juiste conclusies adhv de data).



1. **Maturation threat**

Een verandering in gedrag of denken die over tijd spontaan optreedt en niet veroorzaakt wordt door een bepaalde factor of interventie. Denk hierbij aan een kind dat steeds beter leert lopen en praten of een plant die steeds groter wordt.

Voorbeeld: Een school gaat op kamp en een docent merkt dat de vijftien jongens van huisje vier erg druk zijn. Ze besluit hen twee dagen geen suiker meer te geven. Na deze twee dagen zijn ze inderdaad rustiger geworden, maar komt dit echt door de vermindering van suikerinname? Een alternatieve verklaring is dat de jongens in een nieuwe omgeving terecht kwamen, zonder ouderlijk toezicht. Ze werden hierdoor erg druk, maar na een aantal dagen waren ze steeds meer gewend aan de nieuwe situatie, waardoor deze drukte en excitement vanzelf afnam (= maturation threat).



Dit kan worden opgelost door een controle- en experimentele groep te gebruiken, waarbij de experimentele groep geen suiker meer krijgt en de controlegroep nog wel en deze later te vergelijken.

2. **History threat**

Tegelijkertijd met het experiment, beïnvloed een eerdere (historical) of externe factor het grootste deel van de participanten op dezelfde manier. Hierdoor is het niet meer duidelijk of de verandering die optreedt is veroorzaakt door het experiment of de eerdere/externe factor.

Voorbeeld: Een universiteit start een 'Go-Green' campagne waarbij ze elektriciteitsgebruik onder studenten willen verminderen. Voorafgaand aan het onderzoek wordt het verbruikte aantal kilowatt opgemeten, waarna de campagne van start gaat. Na twee maanden wordt wederom het aantal verbruikte kilowatt berekend, waarna blijkt dat dit is afgenomen. Komt dit echt door de 'Go-Green' campagne? Een alternatieve verklaring is dat het ondertussen kouder weer werd en er minder airco is gebruikt (= history threat).

3. **Regression threat**

Dit houdt in dat tijdens de voormeting blijkt dat het gemiddelde van de participanten zodanig hoog is, dat deze gedurende het onderzoek alleen maar kan verminderen (ongeacht de factor die deze afname zou moeten veroorzaken). Dit noem je ook wel het **plafondeffect**.

Je hebt dus bijna het plafond bereikt van je schaal, waardoor het vrijwel niet anders kan dan dat de participanten de volgende meting(en) lager scoren; je kan maar één kant op, want de participant scoort in het begin al maximaal op de schaal.

4. **Attrition threat**

Wanneer twee participanten met de hoogste / laagste scores plotseling uit het onderzoek stappen, zal er een grote verschuiving in het gemiddelde plaatsvinden, ook al blijven de andere scores gelijk.

Indien de twee participanten die het onderzoek verlaten beide een gemiddelde score hadden, zal dit het gemiddelde vrijwel niet veranderen.

Voorbeeld: Groep 7 bestaat uit 15 kinderen en je berekent hun klassen- gemiddelde op de geschiedenisstoets. De toets van de vier leerlingen met het hoogste cijfer, wordt echter ongeldig verklaard. Hierdoor zal er een grote, negatieve verschuiving in het klassengemiddelde plaatsvinden.

5. **Testing threat**

Hierbij is het de vraag of een verandering optreedt naar aanleiding van het experiment of simpelweg door oefening.

Voorbeeld: Je neemt wekelijks een intelligentietest bij groep 7 af en de leerlingen scoren met de week hoger. Ze zijn echter niet slimmer geworden, maar herkennen en onthouden wat de vraag en het bijbehorende antwoord is. Om deze reden mag zo'n test slechts eens in de twee jaar worden afgenomen, zodat het een representatief beeld geeft van de intelligentie.



6. Instrumentation threat

Het meetinstrument verandert over tijd.

Voorbeeld: Idealiter stel je bij zo'n wekelijkse intelligentietest elke keer weer andere vragen, maar wel over hetzelfde onderwerp; je verandert het meetinstrument. De vraag is daarom of je al deze verschillende vragen toch met elkaar kunt vergelijken en ze alsnog samen intelligentie meten.

Voorbeeld: Je vraagt participanten zich vijf keer een traumatische PTSD-ervaring in te beelden en hier een cijfer wat betreft de levendigheid aan te geven. Hierna ga je via psychiatrische hulp proberen de levendigheid ervan te verminderen. Per keer laat je diegene echter een andere ervaring kiezen. Hierdoor wordt de vermindering van levendigheid niet bepaald aan de hand van de psychiatrische hulp, maar aan de hand van de ervaring die gekozen wordt; verandering van het meetinstrument.

Observer Bias

Als onderzoeker wil je heel graag dat jouw middel bij de participanten werkt. Stel dat je de therapeut bent die met PTSD-cliënten praat om de PTSD sneller af te laten nemen, maar tegelijkertijd ook de wetenschapper bent. Je weet dan wie er in de experimentele groep zitten en doet onbewust extra je best om de PTSD bij hen sneller af te laten nemen. Dit noem je **observer bias**.

Je kan dit voorkomen door **dubbelblind onderzoek** te doen. Niet alleen de participant weet niet in welke groep (controle/experimenteel) hij zit (= **blind onderzoek**), maar ook de wetenschapper die tegelijkertijd de therapeut is weet niet in welke groep zijn cliënt zit. Hiermee voorkom je dat je voor cliënten uit de experimentele groep beter je best doet om een positief effect te behalen.

Placebo-effect

Je denkt dat iets effect heeft, waardoor het ook daadwerkelijk gaat werken. Dit noem je het **placebo-effect**.

Bijvoorbeeld: Een patiënt krijgt een pil van de dokter, waarover de dokter zegt dat hij goed werkt. Jij gelooft de dokter, gebruikt de pil en gaat je daadwerkelijk beter voelen. Het blijkt echter dat deze pil een neppil is, zonder werkend stofje erin. Een bepaalde gedachte in plaats van het echte medicijn zorgt er dus voor dat jij je beter voelt.

Om dit placebo-effect onder controle te krijgen werk je dus zowel met een controle- als een experimentele groep. Zo kun je kijken of dat pilletje ook daadwerkelijk effect heeft, bovenop het placebo-effect (natuurlijke herstel = **naturation threat**).

Factorial Design: Hoofd- en interactie-effecten

Hoofdeffecten & interactie-effecten

Interactie-effect = grafieken lopen niet gelijk en zullen (ergens) kruisen.

AB-Design

Dit is een **counterbalanced design**; je gebruikt twee groepen die het tegenovergestelde doen. Dit is zowel een within subject-effect als een between-subject effect.

Within: omdat je de verandering van één groep over een bepaalde tijd bekijkt, in dit geval namelijk het effect van de pil; voelt iemand zich beter door de pil te gebruiken?

Between: effect van de volgorde waarin je de (nep)pil aanbiedt, dus je vergelijkt daarnaast twee groepen



met elkaar.

Het maakt in dit geval niet uit in welke volgorde je de pil krijgt, er is altijd een afname van hoofdpijn.

Quasi-experimenteel: je laat de onderzoeksgroep zichzelf al random verdelen, o.a. door te vragen wie er na een traumatische ervaring professionele hulp heeft gekregen en wie niet (groep professionele hulp en groep geen hulp). Je hebt als wetenschapper dus niet meer de volledige controle.

Puur experimenteel: je deelt de hele onderzoeksgroep zelf op een random manier in.

Hoorcollege 10 – Experimenteel onderzoek

P-waarde

De nul- en alternatieve hypothese zijn altijd als volgt verdeeld:

H_0 = geen verband / relatie tussen X en Y, oftewel: er is niets aan de hand

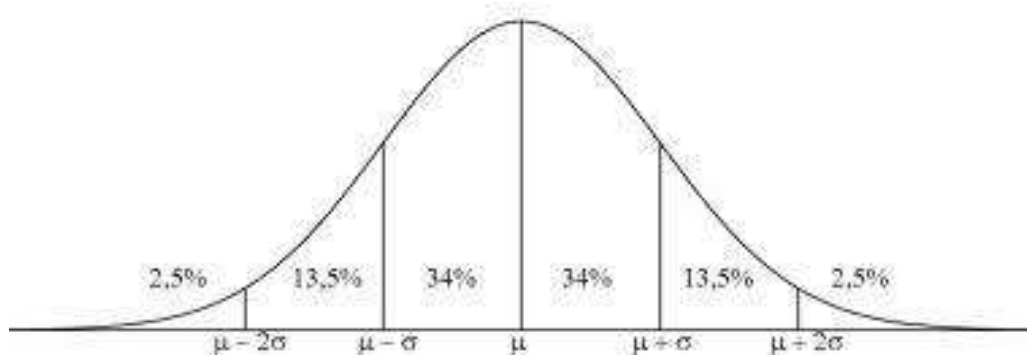
H_1 = wel verband / relatie tussen X en Y, oftewel: er is wel iets aan de hand

Indien de p-waarde kleiner of gelijk is aan *alpha* (α) (vaak $<.05$), dan wijkt deze significant af. De nulhypothese (H_0) wordt vervolgens verworpen en de alternatieve hypothese (H_1) wordt aangenomen; er is sprake van een relatie tussen X en Y.

Let op: Een p-waarde kan alleen de nulhypothese verwerpen, maar niet zeggen of de alternatieve hypothese voor waar aangenomen kan worden!

Een p-waarde geeft aan hoe extreem de data is. Het is een kans tussen de 0-1, hoe dichter bij nul des te extremer de data. Een wetenschapper moet zelf kiezen wanneer hij vindt dat de data extreem genoeg is om verworpen te worden (p-waarde).

Hoe extremer de data, des te verder zit hij in de staart van de normaalverdeling en des te kleiner is de p-waarde.



Voorbeeld: Je bent in Alaska, denkt een zwarte beer in de bosjes te zien en neemt een foto. Deze dieren leven daar eigenlijk niet, dus je gaat berekenen hoe groot de kans is dat dit daadwerkelijk een zwarte beer is.

De p-waarde is dan de kans hoe 'extreem' jouw foto is ten opzichte van foto's die genomen zijn in een land waar ook geen zwarte beren wonen.

Deze kans / p-waarde ligt in dit geval heel dicht bij nul ($p \leq .05$), oftewel: er zit daadwerkelijk een zwarte beer in de bosjes. De nulhypothese wijkt dan ook significant af en mag verworpen worden; er is wel degelijk iets aan de hand, zwarte beren leven namelijk wel in Alaska. De foto komt dus niet uit de nul-populatie, maar uit de alternatieve populatie.

Bij een **alternatieve hypothese** kan er nog van alles aan de hand zijn, je bent niet heel specifiek maar geeft alleen aan dat je verwacht dat iets afwijkt; je toetst tweezijdig.



Een **informatieve hypothese** daarentegen is eenzijdig in plaats van tweezijdig. Deze hypothese is specifieker.

Type-I fout = H_0 is waar, maar wordt ten onrechte verworpen. Dit is *alpha* (α) en staat vaak op .05. Je denkt dan dat je hele extreme data hebt die H_0 verwerpt, maar dat is eigenlijk niet zo.
Hoe kleiner de alpha, des te kleiner de kans op een Type-I fout.

Voorbeeld: Alpha laat dus extreme waarden zien op basis van toeval. Zo kun je in Nederland ook heel veel foto's maken tot er per ongeluk eentje tussen zit die gemaakt is in een dierentuin waar toevallig ook een zwarte beer woont. Deze ene foto is alleen bij lange na niet extreem genoeg om te kunnen zeggen dat er in Nederland zwarte beren in het wild leven, dus valt hij gewoon binnen de nul-populatie en wordt H_0 niet verworpen.

Type-II fout = H_0 is niet waar, maar wordt wel voor waar aangenomen.

Power

Dit is de kans dat de nulhypothese (H_0) terecht wordt verworpen en staat meestal op .80. Je concludeert dus terecht dat er iets aan de hand is.

Met een poweranalyse kun je van tevoren berekenen hoe groot jouw populatie (*sample size*) moet zijn om de nulhypothese met genoeg evidentie te mogen verwerpen, oftewel: hoeveel participanten heb je nodig om een power van .80. te krijgen?

Hierbij kijk je naar de grootte van de alpha en de grootte van het effect dat je verwacht. Hoe groter de alpha, des te minder mensen heb je nodig om een power van .80 te halen.

Voor de effectgrootte geldt hetzelfde: verwacht je een groot effect, dan betekent dit dat het waarschijnlijk duidelijk en bij veel mensen binnen de populatie aanwezig is, waardoor je niet zoveel participanten nodig hebt om dit te bewijzen.

Duidelijke uitleg alpha en p-waarde

De *alpha* is de grens die je als onderzoeker voor jezelf stelt: wanneer vind ik dat de waarden extreem genoeg zijn om te concluderen dat er toch iets aan de hand is (nulhypothese verwerpen)? Hoe kleiner de alpha, des te extremer deze waarden moeten zijn.

De p-waarde is uiteindelijk het effect dat je vindt. Dit vergelijk je met de alpha en is

$p \leq \alpha$, dan zijn de gevonden waarden extreem genoeg om te concluderen dat de nulhypothese verworpen kan worden.

Als blijkt dat $p > \alpha$, kun je concluderen dat de nulhypothese niet verworpen hoeft te worden. Het kan ook zo zijn, dat je sample size te klein is, waardoor geen goede conclusies kunt trekken over het wel of niet verwerpen van de nulhypothese. Als dat het geval is, kun je twee dingen doen:

- eenzijdig ipv tweezijdig toetsen (alpha wordt groter, kans op extreme waarden neemt toe)

- Bayesiaanse statistiek

Steekproefgroottes per groep voor een power van .80:

α	Effect Size Cohen's d		
	small = .20	medium = .50	large = .80
.01	586	95	38
.05	393	64	26
.10	310	50	20



Cohen's d (effect size)

= relevantie

De Cohen's d kan een waarde aannemen tussen de 0 en 1. Hoe dichterbij 1, des te groter het effect. Er geldt:

0.20 = klein effect

0.50 = medium effect

0.80 = groot effect

Variantie (R^2)

De totale spreiding van alle residuen t.o.v. de regressielijn noem je de *variantie* (R^2).

De variantie zegt iets over hoe goed het model bij de geobserveerde data past. Liggen de residuen dichtbij of juist ver van de regressielijn?

Er geldt: $0 \leq R^2 \leq 1$.

Hoe dichterbij 1, des te groter is de variantie en des te kleiner de afstand tussen de residuen en de regressielijn. Hoe groter de variantie, des te beter past het model bij de geobserveerde data en des te kleiner is de e (*error, voorspellingsfout*).

N^2 (eta squared)

Dit gebruik je als je >2 groepen in jouw onderzoek gebruikt en hun spreiding (variantie) wil bekijken. Hiervoor vergelijk je hun gemiddelden met elkaar.

Er geldt: $0 \leq N^2 \leq 1$

Hoe dichterbij 1, des te groter is de variantie en des te kleiner de afstand / spreiding tussen de twee gemiddelden.

Met N^2 kun je vervolgens iets zeggen over de effectsize (relevantie) van jouw onderzoek. Er geldt net als bij Cohen's d :

0.20 = klein effect

0.50 = medium effect

0.80 = groot effect

Een **post-hoc test** vergelijkt deze verschillende groepen vervolgens paarsgewijs. Per paar wordt er ook een p -waarde en effectgrootte berekend.

Elke p -waarde vergelijk je vervolgens weer met de α om te bepalen of hij significant afwijkt of niet. Hierdoor treedt er **alpha inflatie** op en neemt de kans op een Type-I fout toe (= **kanskapitalisatie**). Dit komt omdat je per paar toetst met elke keer weer 5% kans dat je H_0 ten onrechte verworpt.

Om dit te voorkomen, kun je deze 5% delen door het aantal toetsen dat je hebt uitgevoerd. Dit noem je de **bonferroni correctie**. Je krijgt dan: $p \leq \alpha$: aantal toetsen, waarna je alle p -waardes met de nieuwe α vergelijkt.

Hoorcollege 11 – Experimenteel onderzoek

Contrast Testing

= Tweezijdig toetsen

= Eenzijdig toetsen

Bayesiaanse statistiek

Dit bestaat uit de informatieve hypothese en het complement.

Bij H_0 is dus niet langer alles gelijk aan elkaar, maar werk je met een **informatieve hypothese** met ongelijkheids-restricties (**H_{i0}**). Dat betekent dat je eenzijdig toetst, een hele specifieke verwachting hebt en hierover een duidelijke voorspelling doet. Ook is de alternatieve hypothese (**H_c**) het **complement**, oftewel: precies het tegenovergestelde van de informatieve hypothese, $H_c = \text{niet } H_{i0}$.

H_0 = H_{i0} = informatieve hypothese

H_1 = H_c = complement

Bij de klassieke hypothese is het 'ik heb gelijk of ieder ander', terwijl het bij de Bayesiaanse hypothese is 'ik heb gelijk of jij hebt gelijk'.

Contrast Testing:	Null hypothesis Testing:
$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$	$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4$
vs.	vs.
$H_{i1} : \beta_1 > \beta_2 > \beta_3 > \beta_4$	$H_1 : \text{niet } H_0$

Bayes factor voor de klassieke hypothese

De Bayes factor is een ratio (verhouding) van steun tussen twee hypothesen.

Bayes factor = 1.0

beide hypothesen krijgen evenveel steun van de data

Bayes factor > 1.0

Nulhypothese (H_0) krijgt meer steun van de data

Bayes factor < 1.0

Alternatieve hypothese (H_1) krijgt meer steun van de data

Als de Bayes factor 10 is, krijg de nulhypothese tien keer meer steun van de data dan de alternatieve hypothese.

= BF_{0a}

Als de Bayes factor 0.64 is, betekent dit dat de alternatieve hypothese $1 / 0.64 = 1.50$ keer meer steun krijgt dan de nulhypothese.

= BF_{a0}

Posterior Model Kansen (PMK)

Hiermee bereken je de kans dat H_0 of H_1 waar is. Dit is dus het tegenovergestelde van de p-waarde,



hiermee wil juist aantonen dat H_0 niet waar is en verworpen kan worden.

Deze kans kan een waarde tussen de 0 en 1 aannemen; hoe dichterbij 1 des te groter is de kans dat deze hypothese waar is. Als de steun voor de ene hypothese toeneemt, neemt de steun voor de ander dus af. Je ziet in de tabel dan ook duidelijk dat hoe groter de Bayes factor, des te groter de kans dat de nulhypothese (PMK0) waar is en andersom.

Let op:

Dit verschilt dus van de p-waarde, omdat die vergelijkt hoe extreem jouw data is ten opzichte van de algehele nul-populatie.

Conditionele Type-I en Type-II fouten

Ook in de Bayesiaanse statistiek zijn er Type-I en Type-II fouten, alleen net iets anders:

- Conditionele Type-I fout

Dit geldt bij $PMK_0 = .80$, want als we in dit geval toch voor H_1 kiezen, is de kans dat we dit ten onrechte doen 80%.

- Conditionele Type-II fout

Dit geldt bij $PMK_1 = .20$, want als we in dit geval voor H_0 kiezen is de kans dat we dit ten onrechte doen slechts 20%.

Bayes factor voor de informatieve hypothese

Hierbij werk je met een *unconstrained* hypothese (*unc*). De Bayes factor bereken je als volgt:

C1 = Complexiteit

Je gaat na hoe complex de hypothese is en kijkt naar de hoeveelheid ongelijkheids-restricties. Er geldt: hoe meer ongelijkheids-restricties, des te specifieker is de hypothese en des te kleiner wordt C1.

F1 = Fit

Hoe goed passen de beperkingen (ongelijkheidsrestricties) die je hebt toegevoegd aan de hypothese bij de data? Hoe specifiek ben je geweest en past dat bij de data?

Wat doe je als je meerdere informatieve hypothesen hebt? Dit doe je wederom volgens de Post-Hoc tests, oftewel: je vergelijkt de informatieve hypothesen paarsgewijs. Per paar wordt er ook een p-waarde en effectgrootte berekend.

Belangrijk:

De Bayes factor kun je een beetje vergelijken met de p-waarde; die bepaalt welke hypothese voor waar aangenomen kan worden.

Turing test

Je weet niet meer of je nu met software praat, of met echte mensen.

Voorbeeld p-waarde vs Bayesiaanse statistiek:

Er werd via een Turing test onderzoek gedaan naar online speeddaten. Vrouwen speeddaten online met een computer of een echte persoon, waarna vervolgens gekeken werd of ze hier onderscheid in konden maken.



P-waarde:

Bij dit onderzoek gold dat $\alpha = .05$, met een gevonden $p = .68$. Hierbij werd geconcludeerd dat H_0 niet verworpen, maar dit zei niks over het feit of de vrouwen wel of geen onderscheid tussen de computer en de echte persoon konden maken.

Bayesiaanse statistiek:

Hier kwam een Bayes factor van 6.85 uit, wat betekent dat de nulhypothese (H_0) 6.85 keer meer steun kreeg van de data dan de alternatieve hypothese (H_1).

Via de Bayes factor kun je dus juist steun voor de nulhypothese berekenen, terwijl via de p-waarde alleen steun voor de alternatieve hypothese berekend kan worden door te kijken of H_0 verworpen kan worden.

Relevantie (= effectgrootte van p-waarde)

Wanneer is een Bayes factor nou relevant?

1 - 3	twijfelachtig
3 - 10	gemiddeld
10 - 30	sterk
30 - 100	heel sterk
> 100	extreem sterk

Hoorcollege 12 – Experimenteel onderzoek

Assumpties

1. Onafhankelijkheid

Alle participant zijn onafhankelijk van elkaar, ze mogen geen dingen met elkaar delen, zoals patiënten met dezelfde arts, inwoners van dezelfde stad, studenten met dezelfde studie etc. Deze participanten lijken namelijk meer op elkaar dan participanten die volledig willekeurig geselecteerd zijn.

Als je in dit geval op zoek bent naar een causaal verband, kan dit dus verklaard worden door de overeenkomsten tussen de participanten in plaats van door de werking van het experiment.

Genestheid betekent dat de participanten afhankelijk zijn van elkaar, bijvoorbeeld als ze allemaal in dezelfde klas zitten en dus hetzelfde onderwijs en dezelfde docent hebben. Als een participant de hele dag vervelend doet, heeft dit namelijk invloed op de rest van de participanten.

Genestheid schendt dus de aanname van onafhankelijkheid. Een oplossing voor deze **genestheid** is het uitvoeren van de **multilevel analyse**. Zo kun je wel één kind uit één klas van één school vragen, maar niet meerdere kinderen uit dezelfde klas, van dezelfde school. Je werkt dan dus met verschillende levels.

Voorbeeld: Je onderzoekt populariteit door extraversie binnen een klas. De participanten zijn dus niet onafhankelijk van elkaar. Door in SPSS een **random factor** toe te voegen, kun je het effect van dezelfde klas (= genestheid) eruit filteren. Het kan namelijk zo zijn dat een andere factor de relatie tussen populariteit en extraversie veroorzaakt. Zo'n random effect kan bijv. sekse, de plaats van de school (stad/platteland) of het aantal jaren ervaring van de docent zijn; hoe meer ervaring des te minder effect heeft extraversie op de populariteit van de participanten.

Let op: Als je toch onderzoek binnen één school of één klas wil doen, heb je hiervoor wel minimaal 20-25 klassen nodig.

2. Outliers

Een **outlier** is een datapunt dat verder weg ligt van de rest van de data.

Er zijn twee soorten:

1. Onmogelijke waarde

Voorbeeld: je neemt een vragenlijst af in een basisschoolklas, waar een participant van 46 jaar tussen zit. Dit is een onmogelijke waarde, omdat dit echt niet kan voorkomen. Waarschijnlijk heeft iemand de vragenlijst niet goed ingevuld of is er een invulfout in SPSS.

Een ander voorbeeld is een populariteitsscore van -3, dat is onmogelijk.

Hier **moet** je iets aan doen, door deze participant te corrigeren of te verwijderen.

2. Onwaarschijnlijke waarde

Voorbeeld: in een statistiek college neem je een vragenlijst af, waar wederom een participant van 46 jaar tussen zit. Dit is een onwaarschijnlijke waarde, omdat een ouder iemand in principe gewoon toegang heeft tot een studie en dit gewoon als derde carrière o.i.d. kan doen.

Een ander voorbeeld is dat een participant de laagste score heeft op depressie, maar de hoogste score op angst. Dit is onwaarschijnlijk, omdat deze twee factoren sterk samenhangen, maar niet onmogelijk.

Hier kun je drie dingen aan doen:

1. Negeren

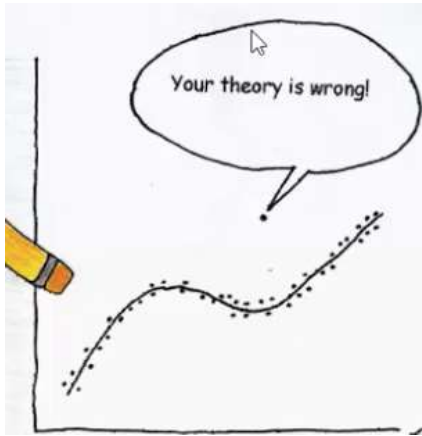
Dit doe je om te voorkomen dat de regressielijn er raar uit gaan zien.

2. *Verwijderen*

Je draait de analyse dan met en zonder outliers om te kijken of het impact heeft op de resultaten.

3. *Robust estimating*

Je laat de outliers erin, maar past de regressielijn slechts een beetje aan. Zo geef je participanten die verder weg zitten van de regressielijn minder gewicht.



3. **Normaliteit**

Er zijn twee soorten:

1. **Normaal verdeeld**

De variabele zelf waarin je geïnteresseerd bent moet in alle experimentele groepen **normaal verdeeld** zijn.

Om dit te controleren gebruik je een boxplot of histogram en kijkt of de spreiding van jouw data overeenkomt met de normale spreiding binnen deze grafieken. Rond het gemiddelde zullen dan de meeste observaties liggen, waarna waarden boven en onder het gemiddelde even waarschijnlijk zijn.

Hoe verder de waarden van het gemiddelde liggen, des te minder normaal (onwaarschijnlijk) is de variabele verdeeld.

2. **Residuen**

Een **residu** is de afstand van één enkele observatie / waarde ten opzichte van de regressielijn.

Als je alle residuen bij elkaar optelt (= **variantie**), moet daar nul uitkomen. Dit betekent dat alle residuen precies op de regressielijn liggen. De variantie kan een waarde tussen de 0-1 aannemen, waarbij geldt: hoe dichter bij 1, des te groter is de spreiding en afstand tot de regressielijn.

De residuen mogen wel wat van de regressielijn af liggen, maar moeten elkaar compenseren om bij elkaar opgeteld alsnog nul te zijn.

Indien jouw data niet normaal verdeeld is, zijn er een aantal oplossingen:

1. *Negeren*

Bij een grote steekproef maakt het niet uit dat je hier en daar normaliteit schendt.

2. *Robust estimating*

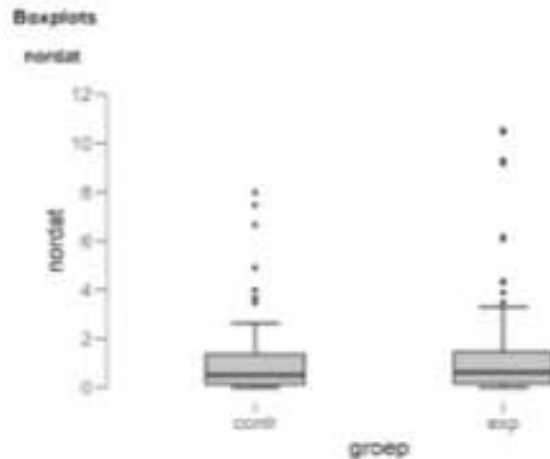
Je laat de outliers erin, maar past de regressielijn slechts een beetje aan. Zo geef je participanten die verder weg zitten van de regressielijn minder gewicht, zodat ze de regressielijn minder uit vorm trekken.

3. *Bootstrapping*

Hierbij ga je niet meer uit van de normaalverdeling (kijken naar grafieken), maar kijk je puur naar de getallen in jouw dataset.

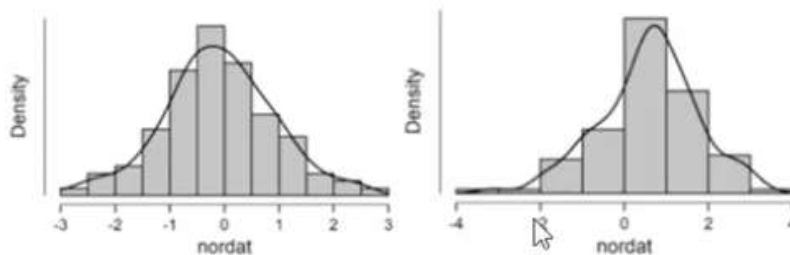
Uit deze dataset worden nieuwe, kleinere steekproeven getrokken. Op deze manier wordt jouw originele dataset dus de nieuwe populatie die je onderzoekt en gebruikt om een grotere populatie te krijgen zonder deze in fysieke aantallen aan te vullen (o.a. door vermenigvuldigen). Hoe groter de populatie, des te meer kans is er op een normale verdeling (outliers hebben dan immers minder invloed). Via bootstrapping maak je er dus alsnog een normaalverdeling van.

Voorwaarde: de participanten in jouw originele steekproef zijn heel representatief voor de



algehele populatie.

cont



4. Homoscedasticiteit

Hierbij ga je ervan uit dat de spreiding (variantie) binnen alle experimentele groepen ongeveer gelijk is.

- H0 = aanname van homoscedasticiteit wordt niet geschonden
 H1 = aanname van homoscedasticiteit wordt geschonden

Oftewel: de regressielijnen moeten parallel aan elkaar lopen, omdat dit betekent dat de spreiding binnen elke experimentele groep gelijk is. Lopen ze niet gelijk of kruisen ze, dan is de aanname van homoscedasticiteit geschonden.

Vuistregel:

Als de experimentele groepen kwa grootte niet meer dan een factor 4 van elkaar verschillen (dus 4x zo groot zijn), mogen de varianties een factor 10 van elkaar verschillen (dus 10x zo groot zijn), voordat deze aanname geschonden wordt.

Voorbeeld: de controlegroep bestaat uit 10 participanten en de experimentele groep uit 40 participanten (4x zo groot, factor 4 verschil). De variantie tussen deze twee groepen mag dan maximaal een factor 10 verschillen. Stel dat de variantie van de controlegroep 1 is, dan mag de



variantie van de experimentele groep maximaal 10 zijn om de aanname van homoscedasticiteit niet te schenden.

Is de variantie van de controlegroep nou 5 en de variantie van de experimentele groep 60 (12x zo groot, factor 12 verschil), dan is de aan aanname van homoscedasticiteit dus wel geschonden.

Replicatie en Sloppy Science

De **replicatie-crisis** houdt in dat indien je een onderzoek repliceert, je niet dezelfde significante resultaten vindt. Dit kan veroorzaakt worden door het feit dat wetenschappers outliers uit hun onderzoek verwijderen en dit niet vermelden, om alsnog significante resultaten te kunnen publiceren. Het is namelijk makkelijker en beter voor je reputatie om significante resultaten te publiceren dan niet-significante resultaten (= **publication bias**).

Indien dit gebeurt, noem je dat **sloppy science**; je gaat op zoek naar trucjes om alsnog een significant resultaat te vinden. Dit kan door o.a. de p-waarde net over de alpha heen te tillen, waardoor het alsnog significant is (bijv. eenzijdig ipv tweezijdig).

Om dit te voorkomen is het van belang als wetenschapper alles netjes en nauwkeurig te rapporteren, zodat indien het onderzoek opnieuw uitgevoerd wordt er hetzelfde resultaat uitkomt.

Een andere oplossing is **pre-registratie**, waar je van tevoren al vastlegt wat je o.a. met eventuele outliers gaat doen (je kunt ze niet meer zomaar verwijderen).