**Eenheden, variabelen en waarden**

* Onderzoekseenheden: over wie/wat doe je straks uitspraken?
	+ Bijv. individuen, huishouden, dorpen, etc.
* Variabelen: eigenschap van de onderzoekseenheden.
	+ Bijv. sekse, leeftijd, inkomen, aantal inwoners dorp.
* Waarden: scores die de onderzoekseenheden kunnen halen op de variabelen. Deze haal je uit de interviews, enquête.
	+ Bijv. Sekse 1=man, 2=vrouw. Leeftijd: variabelen 0 - 100 jaar.

Eerste beslissingspunt

*Variabele* = kenmerk wat gemeten kan worden en wat kan. Aantal variabelen:

* *Univariate analyse:* beschrijft één variabele
* *Bivariate analyse:* onderzoek van twee variabelen
* *Trivariate analyse:* drie variabelen

Voorbeeld bivariaat: beschrijft een verband.

*Onderzoeksvraag:* Zijn er verschillen tussen ouderen en jongeren

in het belang dat zij hechten aan geloof?

*Onderzoekseenheid:* individuen

*Variabelen 1:* leeftijd

*Waarden:* jongeren/ouderen

*Variabelen 2:* belang dat zij hechten aan geloof

*Waarden:* …

Tweede beslissingspunt

- *Numerieke analyse:* tabel met getallen

- *Grafische analyse:* plaatje (om illustratie te geven)

Derde beslissingspunt

- Meetniveau van je variabele

Zegt iets over de eigenschappen van de waarden.

- Kies een analyse die past bij het meetniveau van jou variabele

**Meetniveaus**

4 meetniveaus (van laag naar hoog geordend):

1. Nominaal: categorieën alleen onderscheiden door naamgeving
	1. Bijv. geloof, burgerlijke staat. Categorieën die je niet kunt ordenen!
2. Ordinaal: categorieën + ordening van categorieën mogelijk
	1. Bijv. opleidingsniveau: LO/MBO/HBO/WO. De ene categorie is van hoger opleidingsniveau dan de andere > ordening van categorieën. MAAR: er zijn geen vaste verschillen tussen de categorieën.
3. Interval: categorieën + ordening + even grote verschillen
	1. Bijv. temperatuur in graden Celsius, geboortejaar. De verschillen tussen de categorieën zijn even groot. Nulpunt is willekeurig gekozen.
4. Ratio: categorieën + ordening + verschillen + absoluut nulpunt (dus verhoudingen kunnen geïnterpreteerd worden)
	1. Bijv. leeftijd

Nominaal & ordinaal = kwalitatief

Interval & ratio = kwantitatief

Dan zijn er ook dichotome variabelen. Die bestaan uit slechts twee antwoordcategorieën zoals geslacht. Dichotome variabelen krijgen een 0-1 codering, daarom gedragen zij zich als ratio meetniveau. Ook al is geslacht methodisch gezien nominaal.

***Grafische beschrijvingen van een univariate variabele:***

Plaatjes zijn vaak manipulatief, de waarde van de x-as en y-as kunnen het beeld vervormen.

***Kwalitatieve variabelen* (nominaal of ordinaal meetniveau)**

* Staafdiagram (bar-chart) 🡪 absolute frequenties
	+ *Losse staven* die even breed zijn
	+ Categorieën op x-as (niet te veel)
	+ Frequentie uitgedrukt in hoogte staaf y-as.
	+ Inzicht in onderlinge verhoudingen
	+ *Analyze > descriptive statistics > frequencies > charts > bar charts*
	+ *Of Graphs > legacy dialogs > bar…*
* Taartdiagram (pie-chart) 🡪 relatieve frequenties (percentages)
	+ Frequentie uitgedrukt in oppervlakte
	+ Inzicht in delen t.o.v. geheel
	+ *Graphs > legacy dialogs > Pie (Slices Respresent % of cases)*
	+ De percentages staan standaard uit in SPSS, opnemen door te dubbelklikken op taartdiagram dan elements (show data labels)

***Kwantitatieve variabelen* (interval of ratio meetniveau)**

* Histogram
	+ Waarden indelen in klassen/intervallen (scores worden gegroepeerd, van 0-5 5-10 10-15 15-20 > klassebreedte is dan 5)
	+ *Staven staan tegen elkaar aan*
	+ Frequentie uitgedrukt in hoogte staaf
	+ Inzicht in verdeling van de variabele
	+ *Analyze > descriptive statistics > frequencies > charts > histogram*
	+ *Of Graphs > legacy dialogs > histogram*
* Stamdiagram (stem-and-leaf plot)
	+ Is vooral geschikt om de verdeling van een variabele te beschrijven als er niet te veel waarnemingen zijn. De histogram wordt echter vaker gebruikt omdat deze geschikter is bij grote aantallen.
	+ Analyze > descriptive statistics > explore > plots > stem-and-leaf
* Boxplot (*bij ordinaal, interval of ratiomeetniveau*)
	+ Onderste snor = minimum, bovenste snor = maximum 🡪 bereik
	+ Q1&Q3 (IKA)
	+ Q2/mediaan = de streep in het midden van de doos.
	+ Opsporen uitschieters
	+ *Analyze > descriptive Statistics > explore > display both*

***Numerieke weergave univariate variabelen:***

**Frequentietabel:** een veel gebruikte manier om de gegevens van een variabele, ongeacht van welk meetniveau, weer te geven is een frequentie tabel. Absolute frequentie, relatieve frequentie en cumulatieve relatieve frequenties worden aangegeven. Missende waarden worden niet meegenomen bij valid percent. De categorieën moeten niet te groot zijn want bij meer dan 10 categorieën verliest een frequentie tabel zijn overzichtelijkheid.

* *Analyze > descriptive statistics > frequencies*

**Frequentie verdeling**

Maten voor het *centrum* van een verdeling: waar ligt het ‘midden’?

* *Modus:* bij een nominaal meetniveau 🡪 welke score komt het meeste voor
* *Mediaan:* vereist minimaal een ordinaal meetniveau 🡪 getallen geordend en vervolgens het punt waar de helft van de waarnemingen is bereikt. Dus de middelste score. Robuuste maat, niet gevoelig voor uitschieters.
* *Gemiddelde:* vereist minimaal interval niveau 🡪 evenwichtspunt van de verdeling.
	+ Kan een niet-bestaande waarde zijn (1,3 kind).
	+ Is gevoelig voor uitschieters (extreme scores trekken het gemiddelde naar zich toe).
	+ Mediaan en gemiddelde vergelijken om te zien of er sprake is van een linksscheve (gemiddelde lager dan mediaan) of een rechtsscheve (gemiddelde hoger dan mediaan) verdeling.

Deze maten kunnen allemaal met de **Frequentietabel** bepaald worden.

*Analyze > descriptives statistics > frequencies > statistics > mode/median/mean.*

Analyse > descriptives statistics > descriptives > options > mean

Pagina 50 van boek heeft handig samenvattingsschema

Maten voor spreiding van een verdeling: liggen waarnemingen ver of dichtbij het midden?

* *Bereik* kun je gebruiken vanaf ordinaal meetniveau → verschil tussen maximale en minimale score. Zeer gevoelig voor extreme scores én zegt niets over de vorm.
	+ *Analyze > descriptive statistics > frequencies > statistics > range*
* *Interkwartielafstand (IKA)* vanaf ordinaal meetniveau. Bereik van de middelste 50% van alle waarnemingen. Groep verdelen in 4 delen van allemaal 25%. De uitschieters zitten altijd in het eerste en/of laatste kwartiel. Dus kijkt men naar de middelste twee kwartielen, robuuste maat. IKA=Q3-Q1, zelf berekenen. Zegt meer over de vorm.
	+ *Analyze > descriptives statistics > frequencies > statistics > quartiles > Q3-Q1*
* *Variantie (S^2)* is een spreidingsmaat die iets zegt over de groep. Gebaseerd op verschilscore/deviatiescore: hoeveel wijkt iemands score af t.o.v. gemiddelde. Negatieve scores zijn lager dan het gemiddelde van de groep. Positieve scores zijn hoger dan het gemiddelde. Verschilscores kwadrateren, optellen en dan het gemiddelde nemen. Een groep die ver uit elkaar ligt (grote spreiding heeft) heeft grote deviatiescores.

→ variantie = S^2 = gemiddelde gekwadrateerde afstand tot het gemiddelde

* + *Analyze > descriptive > frequencies > statistics > variance.*
* *Standaardafwijking (S)* = gemiddelde afstand tot het gemiddelde, spreidingsmaat. Men verkiest de standaardafwijking boven de variantie omdat de meeteenheid dan gelijk is aan de variabele zelf (i.p.v. in het kwadraat). De standaardafwijking kan allen voor interval en ratio

Variantie en standaardafwijking vanaf interval meetniveau. Werken met gemiddelde dus gevoelig voor uitschieters.

* *Analyze > descriptives statistics > frequencies > statistics > std. deviation*
* *Analyze > descriptives statstics > descriptives > options > std. deviation*

*De doosdiagram (boxplot), die bij de grafische weergave hoort, bevat de maten bereik, mediaan, kwartielen en interkwartielafstand. Deze diagram wordt gebruikt voor het opsporen van bijzonder hoge en lage scores en voor bivariate vergelijkingen van verdelingen. Hoge scores kunnen ontstaan door foutieve invoer, door scores als weet niet (999) niet te verwijderen of doordat er gewoon geldige metingen zijn van analyse-eenheden die nu eenmaal een extreme hoge score hebben. Q1-3 IKA en Q3+3 IKA zijn de grenzen en buiten deze grenzen zijn het mogelijk uitschieters die de uitkomsten te sterk beïnvloeden. Er is echter niet te zien hoeveel respondenten achter de extreme scores, aangegeven met een \*, schuilen. Dit is te vinden in een frequentie tabel. Er moet opgepast worden dat de scores 999 ‘weet niet’ als missings worden opgegeven.*

**Relatieve scores**

Absolute score *in vergelijking met* de omgeving, relatieve positie verschilt.

Relatieve plaats in de verdeling. Waar staat jou score in de verdeling?

Scoor je onder of boven het gemiddelde?

Ook kijken naar de spreiding. Is het een zeldzame score of niet?

Percentiel score: Percentage respondenten dat dezelfde of een lagere waarde scoort op de variabele.

Af te lezen bij cumulatieve percentages (naar beneden afronden!).

Alleen bruikbaar bij ordinaal, interval en ratio.

* *Analyze > descriptives statistics > frequencies*

Z-score

* Bij interval en ratio meetniveau.
* Scoort een onderzoekseenheid boven (+) of onder (-) het gemiddelde? Zegt iets over relatieve positie in de groep 🡪 *individuele score.*
* Relatieve score Z = (absolute score - gemiddelde) / S
* *Analyze > descriptive Statistics > descriptives > variabele selecteren > vink ‘save standardized values as variables’ aan.* Er komt een nieuwe kolom bij waar alle Z-scores in staan.
	+ Uitkomst is bijvoorbeeld 3 🡪 iemand scoort dan 3 standaarddeviaties boven het gemiddelde.
	+ Kijk naar empirische regel of regel van Chebyshev om te interpreteren OF overschrijdingskans (p) berekenen via Syntax.
	+ Standaarddeviatie berekening in schrift

*Empirische regel:* bij symmetrische, heuvelvormige verdeling. Ongeveer 68% tussen -1S en 1S. Ongeveer 95% tussen -2S en 2S en ongeveer 99,7% tussen -3S en 3S.

*Regel van Chebyshev:* bij onregelmatige verdelingen. Minimaal 75% tussen -2S en 2S. Minimaal 88,9% tussen 3S en -3S.

*Overschrijdingskans (p) berekenen via Syntax.*

Z-waarde invullen bij vraagteken. Run > All. In output kun je de overschrijdingskans interpreteren.

Bij de syntax wordt berekend hoeveel % van de mensen extremer scoort dan jij.

File: open syntaxfile. Bij compute z waarde =? De z score invoeren. En dan Run > all.

**Grafische beschrijving van bivariate variabelen:**

X = onafhankelijk → op horizontale as

Y = afhankelijk → op verticale as

Grafische analyses = wat is de *aard* van het verband

**Boxplot** (Y minimaal ordinaal meetniveau, X niet te veel categorieën)

Kijk naar hoe de mediaan verschuift.

* *Graphs > legacy dialogs > boxplot*

**Spreidingsdiagram/scatterplot** (interval & ratio)

Alle scores worden met een puntje afgebeeld in een puntenwolk. Hoe platter de scatterplot hoe sterker de relatie. Hoe boller de scatterplot hoe zwakker de relatie. Zoek een trendlijn in de puntenwolk. Deze kan positief of negatief zijn. Positief: hoe hoger leeftijd (X) des te hoger gewicht (Y). Negatief: hoe hoger leeftijd (X) des te lager gewicht (Y). Horizontale lijn: er is geen verband/relatie.

* *Graphs > legacy dialogs > scattered*

**Lijndiagram** (interval & ratio)

Geeft de gemiddelde score per variabele van de x-as weer, en trekt een lijn tussen deze punten. In een lijn diagram zie je wel fluctuaties. Je ziet duidelijker een trendlijn dan in een scatterplot. Positieve of negatieve trendlijn.

* *Graphs > legacy dialogs > line > vink ‘Other statistic’ aan > onafhankelijk X bij Category Axis*

**Numerieke beschrijving bivariate variabelen:**

Numerieke analyses = wat is de *sterkte* van het verband

**Samenhang tussen kwalitatieve variabelen**

Maak gebruik van een kruistabel. Onafhankelijk X in de kolommen, afhankelijk Y in de rijen.

Percenteer in de richting van de onafhankelijke variabelen (X), ook wel verticaal percenteren genoemd. De totalen van de kolommen moeten op 100% uitkomen.

* *Analyze > descriptive statistics > crosstabs > Cells > percentages Column*

Als het gaat om de relatie tussen variabelen met relatief weinig categorieën (liefst minder dan 10), dan is een kruistabel handig om te gebruiken. Tenzij de marginalen gelijk zijn kunnen absolute verschillen niet gebruikt worden om verbanden vast te stellen, je moet dan percenteren.

**Samenhangsmaten *kwalitatieve* variabelen.**

Nominaal

* d% = het verschilpercentage. Verschil tussen twee cellen. Weinig gebruikt omdat het slechts bruikbaar is bij twee kolommen/rijen en omdat het vaak niet gelijk is bij horizontaal percenteren en verticaal percenteren.
* (Chi-kwadraat &) De Chi-kwadraat vergelijkt de geobserveerde celfrequentie (observed) met de verwachte celfrequentie (expected) indien er geen samenhang is. Hoe groter de samenhang, des te groter is het Chi-kwadraat. Alle verschillen worden gekwadrateerd, gedeeld door de verwachte score en opgeteld.
	+ Het is een slechte maat voor samenhang omdat een bepaalde waarde in een grote steekproef een kleine samenhang lijkt te hebben, en in een kleine steekproef een grote samenhang lijkt te hebben. Hangt af van grootte steekproef en tabel.
	+ *Analyze > descriptive statistics > crosstabs > statistics > Chi-square*
* Cramer’s V: minstens 1 van de variabelen moet nominaal zijn. Vergelijkt de gevonden Chi-kwadraat met de maximaal te behalen waarde in de tabel. Ligt altijd tussen 0 (geen samenhang) en 1 (perfecte samenhang). Getal 🡪 sterkte
	+ *Analyze > descriptive statistics > crosstabs > statistics > Phi and Cramer’s V*

>0 tot 0,1 = zeer zwak

0,1 tot 0,25 = zwak

0,25 tot 0,35 = middelmatig

0,35 tot 0,45 = sterk

0,45 en hoger = zeer sterk

Ordinaal

* Kendalls Tau rangcorrelatie (Maatschappij wetenschappen)
	+ T=1 of T=-1 perfecte samenhang, T=0 geen samenhang.
	+ Tau B: bij vierkante kruistabellen
	+ Tau C: bij niet-vierkante kruistabellen
* Spearmans rangcorrelatie (Rs) (vink Correlations aan) (Psychologie)
	+ Gebaseerd op rangorde, kijkt naar rangnummers.
	+ Rs = 1 perfect positief verband, Rs = -1 perfect negatief verband, Rs = 0 geen verband.
	+ Spearman heeft als voordeel dat het wat berekening betref gelijkenis vertoont met de correlatiecoëfficiënt (Pearson).

*🡪 Analyze > correlate > bivariate > aankruisen*

**Wanneer Spearman & wanneer Pearson**

Als Spearman groter is dan Pearson is een verband NIET lineair. Als Pearson groter of gelijk is dan Spearman dan is het wel lineair. Geldt voor de getalswaarde, +/- maakt niet uit.

**Samenhang tussen kwantitatieve variabelen.**

De *sterkte* numeriek uitdrukken met Pearson en de *aard* aflezen met regressieanalyse.

Pearson correlatiecoëfficiënt (r)

* Mate voor sterkte van *lineaire* samenhang. Pearson geeft de verandering van Y aan wanneer X met 1 toeneemt. Perfecte samenhang r = -1 of r = 1. Bij 0 is er geen lineaire samenhang (misschien Spearman gebruiken).
* Eis is dat het verband lineair is, controleer dit grafisch met lijndiagram en numeriek door R met Rs te vergelijken.
* Nadeel van Pearson is dat het gevoelig is voor uitschieters.

*🡪 Analyze > correlate > bivariate > aankruisen.*

Regressieanalyse

OLS-regression lijn voorspelling. Als punten van de scatterplot boven of onder de lijn liggen zijn er voorspellingsfouten gemaakt (e). Er wordt een lijn gezocht met zo min mogelijk voorspellingsfouten.

Wiskundige beschrijving van de lijn gewicht = a + b\*lengte (y = ax + b)

* b-coëfficiënt is eigenlijk de richtingscoëfficiënt van de lijn: als lengte met 1 cm toeneemt, zal gewicht met b kilo veranderen.
* Intercept/constante a definieert snijpunt van de lijn met de verticale as: voorspeld gewicht bij lengte 0. Met de b en a kun je de formule opstellen en de aard beschrijven.
	+ *Analyze > regression > lineair > OK > a aflezen bij (Constant) b aflezen daaronder*

*Kwaliteit van de voorspelling:*

* Proportie verklaarde variantie = R square (r\*r=r^2). Hoeveel procent van de verschillen in Y (BMI) wordt veroorzaakt door verschillen in X (opleiding). Naarmate R square hoger is hoe beter de kwaliteit.
	+ *Analyze > regression > lineair > OK > aflezen*
* Genuanceerder als je meerdere predictoren toevoegt aan de regressie. Bijvoorbeeld: gewicht = a + b1\*lengte + b2\*opleiding (opleiding in aantal jaren, kwantitatief, ratio).
* De belangrijkste predictor lees je af bij Standardized Coefficients, de hoogste getalwaarde (los van + of -) is de belangrijkste predictor. (Kan je niet gebruiken als je meerdere dummy’s nodig hebt om 1 variabele te interpreteren!)
	+ *Analyze > regressoin > lineair > OK > aflezen*

*Kwalitatieve predictoren in regressieanalyse*

* Kwalitatieve variabelen kunnen niet direct in de regressieanalyse. Er moeten dan dummies gemaakt worden. Dummyvariabelen zijn dichotoom, dus kwantitatief.
* Als je kwalitatieve variabelen drie categorieën heeft maakt je twee dummyvariabelen, als je kwalitatieve variabelen vier categorieën heeft maak je drie dummyvariabelen etc.
* De groep die een 1 scoort op de dummy wordt vergeleken met de referentiegroep (die alleen nullen scoort). De referentiecategorie wordt NIET in de analyse gegooid. Dit is een multiple regressieanalyse.

*🡪 Analyze → Regression → lineair*

*Voorbeeld:*

Opleidingsniveau (kwalitatief, ordinaal) niet rechtstreeks in regressie vergelijking.

Referentie groep Dj zijn de jongeren. Dm zijn middelbare leeftijd en Do zijn ouderen.

Geloof = a + b1\*Dm + b2\*Do + b3\*leeftijd

Geloof = 0,415+0,090\*Dm+0,507\*Do-0,059\*leeftijd

Middelbare leeftijd scoort 0,09 hoger dan jongeren. Ouderen scoren 0,507 hoger dan jongeren.

Als de opleiding gelijkt blijft en leeftijd gaat 1 omhoog scoort men 0,059 minder op geloof.

Voer de predictor en twee dummy’s in (niet de referentiedummy)!!

**Odds ratio (kan bij elk meetniveau)**

Als *relatieve samenhangsmaat* kan men de Odds ratio gebruiken.

|  |  |  |
| --- | --- | --- |
|  | Allochtoon | Autochtoon |
| Laag inkomen | 75% | 50% |
| Hoog inkomen | 25% | 50% |
|  | 100% | 100% |

Odds = 75/25 = 3 en 50/50 = 1

Odd ratio = 3/1

🡪 Inkomen is bij allochtonen 3x zo scheef als bij autochtonen

* Bij een odds ratio van 1 zijn beide kans verhoudingen gelijk aan elkaar en is er naar verhouding geen verschil tussen de categorieën van x. (Niet scheef)
* Er is geen maximale waarde bij de Odds ratio en je kan niks zeggen over de sterkte, er zijn geen vuistregels bij de Odds ratio.
* Vink ‘Risk’ aan in SPSS:
	+ *Analyze > descriptive statistics > crosstabs > statistics > Risk (ook bij Cells percenteren naar de kolom - onafhankelijk X)*

**Schaalscores maken**

Je kunt sommige variabelen samenvoegen tot een nieuwe variabele. Hierbij worden de scores opgeteld omdat zij hetzelfde meten (pagina 45 basiscursus)

**Stappenplan:**

1. Zijn de items dezelfde kant op gericht.

* Bij scores 1, 2, 3, 5 en 6 hoe hoger je op de items scoort hoe hoger je bent ingebed. Maar bij score 4 hoe hoger je scoort hoe lager je bent ingebed. Item vier is dus niet dezelfde kant op gericht en moet je daarom spiegelen. Wat een 1 was wordt een 5 wat een 2 was wordt een 4, 3 blijft 3, 4 wordt 2, 5 wordt 1.
* *Transform > Recode into different variables > name: (vul iets in) > change > old an new values > old value: 1 new value: 5 > old value: 2 new value: 4 > old value: 3 new value: 3 etc.*
* OF Transform > compute variabele > (name) > 6 - item4 (dan moeten ze zich wel naast elkaar bevinden in het databestand). Bij driepuntschaal 4 - oorspronkelijk variabele. Bij vierpuntschaal 5 - oorspronkelijke variabele
* Pas als je hebt gespiegeld kun je een schaalscore maken met MEAN.

2. Mag je alle items meenemen? 🡪 correlatiematrix

* Items die zwak correleren met de rest haal je eruit, vuistregels op pagina 83 in het boek.
	+ Ordinale variabelen Spearman (komt bijna nooit voor)
	+ Kwantitatieve variabelen Pearson Correlation
	+ Schaal van 1 tot 5 of meer wordt altijd beschouwd als interval meetniveau, dus met Pearson Correlatie.
* *Analyze > correlate > bivariate > zet alle enquetevragen erin > OK*
* Als een item negatief correleert had je misschien moeten spiegelen, had je dat al gedaan, dan moet je het item eruit halen.

3. Schaalscore berekenen

Voorbeeld:

* + Geloof = Mean.3(v1 to v5) De 3 na Mean. geeft aan hoeveel scores er minimaal ingevuld moeten zijn. Bij 5 moet dus alles ingevuld zijn, bij 3 maar drie van de vijf vragen, bij 2 maar twee van de vijf vragen en bij 1 maar één van de vijf vragen.
	+ Geloof = Mean.3(v1,v2,v3,v4,v5)
	+ *Tranform 🡪 compute variable… 🡪 Target Variable: vul nieuwe naam in 🡪 selecteer MEAN en breng met pijltje naar boven 🡪 type in: MEAN.3(v1,v2,v3,v4,v5) óf MEAN.3(v1 to v5)*
	+ Kijk naar nieuwe kolom in de dataset. Voer correlaties uit om te kijken of de schaal zich gedraagt zoals het in theorie werd verwacht.

***SPSS***

**Het maken en bewerken van grafieken:**

*Graphs > legacy dialogs > …..*

*Of*

*Analyze > descriptives statistics > frequencies > charts > …..*

*(P. 62 bcs)*

Category axis = x

Variable = y

**Missings eruit halen**

*Variable view > missings > missings invullen*

**Kruistabel:**

*Analyze > descriptive statistics > crosstabs > cells - (P 70 bcs)*

Per kolom percenteren!

*Analyze > descriptive statistics > crosstabs > ­­­statistics*

Samenhangsmaten.

Odds ratio = Risk

**Correlatie:**

*Analyze > correlate > …..*

*Grafische weergave voor univariate variabelen.*

|  |  |  |  |
| --- | --- | --- | --- |
| **Meetniveau** | **Grafische beschrijving** | **Numerieke beschrijving: CENTRUM** | **Numerieke beschrijving:****SPREIDING** |
| **Nominaal** | Staafdiagram, taartdiagram (als totaal optelt tot 100%) | Modus | Frequentietabel (aantal categorieën klein) |
| **Ordinaal** | Staafdiagram, doosdiagram | Modus\*Mediaan | Frequentietabel (aantal categorieën klein), bereik, IKA |
| **Interval/ ratio** | Doosdiagram, histogram, stamdiagram (bij weinig waarnemingen | Modus\*Mediaan\*Gemiddelde  | Frequentietabel\* (aantal categorieën klein), Bereik\*IKA\*variantie, standaardafwijking,  |

\* Deze beschrijving pas gebruiken als andere maten geen goed beeld geven, bijvoorbeeld als gevolg van extreme scheefheid of uitschieters.

|  |  |
| --- | --- |
| **Afhankelijke variabele (y)** | **Onafhankelijke variabele (x):** **Nominaal Ordinaal interval/ratio** |
| **Nominaal** | Geen | Geen | Geen |
| **Ordinaal** | Doosdiagram | Doosdiagram (bij weinig categorieën) |
| **Interval/ratio** | Spreidingsdiagram Lijndiagram |

*Grafische weergave voor bivariate variabelen.*

*Numerieke weergave voor Bivariate variabelen\**

|  |  |
| --- | --- |
|  | **Onafhankelijke variabele (X)** |
| **Afhankelijke variabele (Y)** | **Nominaal** | **Ordinaal** | **Interval** |
| **Nominaal** | Cramér's V(Odds Ratio) |
| **Ordinaal** | Cramér's V (Odds Ratio) | Kendall’s tau-b en cSpearmans rangcorrelatie (Rs) |
| **Interval** | Regressieanalyse (kwalitatieve predictoren als dummies! | Spearmans rangcorrelatie (Rs)Pearsons correlatie (r) Regressieanalyse |

\* in deze tabel moet je een dichotome variabelen als nominaal beschouwen, hier dus NIET als ratio.

**SPSS:**

**Hercoderen:**

Hercoderen van variabelen gebeurt om twee redenen:

* Men wil bepaalde waarden van een variabele wijzingen
* Men wil bepaalde waarden samenvoegen

Er zijn hierbij twee mogelijkheden, namelijk:

* Recode into same variables
* Recode into different variables

Recode into same variables: de variabele houdt dezelfde naam en labels en de nieuwe waarden worden in dezelfde kolom van het datavenster gezet. De consequentie is wel dat de originele gegevens verloren gaan. Er is dus geen controle en geen herstel mogelijk. Deze manier van hercoderen wordt daarom STERK AFGERADEN.

Recode into different variables: hierbij blijft de oorspronkelijke variabele bestaan wat ook vaak de bedoeling is. De waarden worden onder een nieuwe variabele naam toegevoegd aan het bestand in de eerste lege kolom van het datavenster.

(Edit > options > display names > ok)

*🡪 Transform > recode into different variables. (p. 43)*

**Nieuwe variabele maken:**

Je kunt sommige variabelen samenvoegen tot een nieuwe variabele. Hierbij worden de scores opgeteld omdat zij hetzelfde meten.

*Transform > compute variable (p. 45 bcs)*

**Selecteren van analyse-eenheden:**

Als je slechts een deel van de respondenten in het bestand wilt gebruiken kun je deze selecteren.

*Data > select cases > if condition is satisfied > if > invullen voorwaarde.*

*(p. 48 bcs)*

**Gemiddelde berekenen (+ minimum, maximum, std. Deviatie)**

*Analyze > decriptives statistics > descriptives*

**Het opsplitsen van bestanden:**

Als je voor meerdere groepen dezelfde statistische analyses afzonderlijk wilt uitvoeren.

*Data > split file > compare groups.*

**Analyse, descriptive statistics bevat:**

Frequencies:frequentieverdeling. Opsomming en aantallen en percentages. Kan worden weergegeven in de vorm van diagrammen.

Descriptives**:** statistische maten berekenen, zoals gemiddelde. Meer overzichtelijk en compact dan frequencies.

Explore:uitgebreide inspectie van variabelen. Kan ook grafisch weergegeven worden in stamdiagram en boxplot.

Cross tabs:kruistabellen met verbanden. Kan alleen bij **ordinaal en nominaal.**