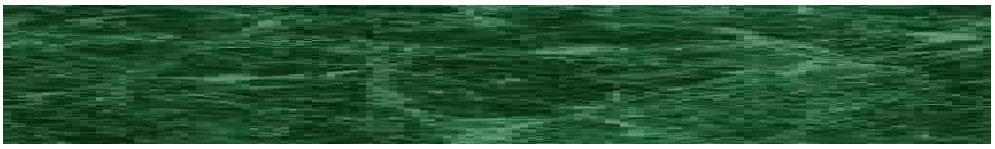


DOCENT : Valkeneers Guido
STUDIEPUNTEN : 4

A. Evaluatie types

- schriftelijk examen

B. Omschrijving

1e examenperiode (januari)

Er is een schriftelijk examen voorzien, waarin kennis en inzicht in aan bod komt. Tijdens dit geslotenboek examen mogen de studenten gebruik maken van een eenvoudig rekenapparaat en een formularium dat door de docent beschikbaar gesteld wordt. Tijdens dit examen komen zowel open vragen als multiple choice items aan bod. De MC items zullen voor 50% het resultaat bepalen. Bij deze MC items zal een giscorrectie toegepast worden.

De studenten zullen - desgevraagd - deelnemen als proefpersoon en/of als proefleider aan wetenschappelijk onderzoek.

2e examenperiode (juni)

3e examenperiode (augustus/september)

Er is een schriftelijk examen voorzien, waarin kennis en inzicht in aan bod komt. Tijdens dit geslotenboek examen mogen de studenten gebruik maken van een eenvoudig rekenapparaat en een formularium dat door de docent beschikbaar gesteld wordt. Tijdens dit examen komen zowel open vragen als multiple choice items aan bod. De MC items zullen voor 50% het resultaat bepalen. Bij deze MC items zal een giscorrectie toegepast worden.

Examencontract

De inhoud en evaluatie van het examencontract is gelijk aan het diplomacontract.

DEEL 1 : UNIVARIATE ANALYSE (1 VARIABELE)

HOOFDSTUK 1 : plaats van de statistiek in het onderzoek

Doelstellingen

- De student kent de diverse fasen van het onderzoek en begrijpt de rol van de statistiek in dat kader
- De student verstaat de besproken begrippen zoals variabele, afhankelijke en onafhankelijke variabele, operationalisering, populatie versus steekproef, beschrijvende versus inductieve statistiek, …
- De student kent de diverse vormen van steekproeftrekking en tevens de mogelijkheden en beperkingen hiervan
- De student kan een datafile aanmaken in SPSS

Statistiek speelt een belangrijke rol bij het verzamelen,samenvatten en interpreteren van gegevens in empirisch onderzoek. Dit geldt voor de ***beschrijvende statistiek,*** *waarin we proberen de verzamelde gegevens zo goed mogelijk weer te geven. Het geldt ook voor de* ***toetsende statistiek****, waarin we proberen uit een (kleine) steekproef conclusies te trekken over de (veel grotere) populatie waaruit deze steekproef afkomstig is.*

Inleiding

Twee typen statistiek

- beschrijvende statistiek: verzamelen, samenvatten van gegevens, analyseren resultaten

- inductieve statistiek: betekenis resultaten tot populatie, schattingsprobleem

Onderzoek

- Gaat over onderzoek dat op basis van *waarnemingen* probeert ware en *algemene uitspraken* te doen over de *werkelijkheid*.
- Een uitspraak is een bewering waarin een of meerdere objecten een eigenschap wordt toegeschreven.
- uitspraken maken die geldigheid hebben ‘boven’ het individu Bv; niet: Jan is slimmer dan Greet. Maar wel: Mannen zijn gemiddeld slimmer dan vrouwen.

Een onderzoek is wetenschappelijk verantwoord als :

- Objectief is
- Controleerbaar is
- Herhaalbaar is
- Systematiek heeft(volgt bepaalde lijn, een bepaald onderzoek)

- deterministische uitspraken Bv; wet van de zwaartekracht
  = met zekerheid iets zeggen over een individu

VERSUS

- probabilistische uitspraken Bv; frustratie bevordert agressie
  = niet met zekerheid iets zeggen over een individu

2 typen onderzoek

- Het experiment
  Doelbewust worden één of enkele variabelen gemanipuleerd en we onderzoeken de effecten hiervan op de afhankelijke variabele. bv. welk is de relatie tussen de attitude en het gedrag? Het experiment van de verloren brief (Nuttin en Beckers)

- Het veldonderzoek/ surveyonderzoek/ enquêteonderzoek
  Hier worden geen variabelen gemanipuleerd
  bv. hebben jongens meer aanleg voor de wiskunde dan meisjes? (Van Peet e.a.)

Variabele

= eigenschap/kenmerk van een onderzoekseenheid (persoon/huishouden…)
bv. Geslacht

- Kan diverse waarden (uitkomsten) aannemen. Mensen verschillen op het vlak van deze eigenschappen
  bv. man/vrouw

- Tegengestelde van een constante

2 typen variabele

- Onafhankelijke variabele (gebruikt om het verschil te verklaren)
  verschillen in deze variabelen worden gezien als oorzaak (?) van verschillen in de afhankelijke variabele

- Afhankelijke variabele = variabele ter studie(als we de onafhankelijke manipuleren , hoe verandert de afhankelijke dan?)
  verschillen in deze variabelen worden gezien als gevolg (?) van verschillen in de onafhankelijke variabele (= hetgeen je bestuurd)

- Er zijn diverse niveaus van meting (komt nog terug in samenvatting) (vb.: man / vrouw = niet te ordenen terwijl het IQ = wel te ordenen)

De vraagstelling

De vraagstelling vloeit voort uit de theorie ofwel uit een concreet probleem
Vraag naar secundaire gegevens: wat hebben andere onderzoekers reeds vastgesteld?
Vraag naar primaire gegevens via experiment of veldonderzoek

Relatie tussen de attitude en het gedrag

- Relatie tussen milieubesef (als attitude) en het milieuvriendelijk consumentengedrag (effectief toepassen van)?
  Bv. besef van de opwarming van de aarde en het gebruik van een terreinwagen.

- Nuttin stelt zich de vraag wat het verband is tussen de attitude en het gedrag door het experiment van de verloren brief

Wat is een attitude?
= "is een door ervaring verworven interne dispositie die belangrijke en duurzame richtinggevende determinant is van de evaluatieve aspecten van open gedrag tegenover een object dat zich leent tot gunstige of ongunstige waardering"
- cognitief aspect
- affectief aspect
- intenties
Hoe kan dit gemeten worden?
Wat is de relatie met concreet gedrag?

Experiment van Nuttin en Beckers

- vraagstelling: Welk is de relatie tussen de attitude en het gedrag van een persoon?
- experiment: de verloren brief

- attitude = een door ervaring verworven interne dispositie die belangrijke en duurzame richtinggevende determinant is van de evaluatieve aspecten van open gedrag tegenover een object dat zich leent tot gunstige of ongunstige waardering
- hoe kan dit gemeten worden?

- experiment van Nuttin en Beckers:
  - **attitude:** Vraag aan proefpersonen; wat zou u doen als u een brief zou vinden met daarop een adres? (interview, enquête) Bijkomende vraag; Zou het iets uitmaken als de postzegel los zit? Het adres in Wallonië ligt? Etc.
  - **gedrag:** Brieven worden effectief verloren gelegd, volgens factorieel design. Zullen de proefpersonen hetzelfde doen als in de interviews?
  - Factorieel design: 2x2x2x3 opzet:
    2 regio, 2 adres, 2 vindplaats, 3 postzegels (los, aantal)
  - Onafhankelijke variabelen: regio, adres, vindplaats, postzegels
  - Resultaten: Uit het experiment bleek dat de attitude die de proefpersonen stelde niet overeenkwam met het gedrag. Dit wil zeggen dat de proefpersonen de brieven niet zonder twijfel postte als het erop aankwam.
  - Besluit: Externe omstandigheden bepalen dit prosociaal gedrag. Attitudes spelen dus een beperkte rol in de verklaring van dit gedrag.

Voorbeelden van een veldonderzoek

- Vraagstelling
  Hebben jongens meer aanleg voor de wiskunde dan meisjes?
- Onderzoekshypothese
  Jongens hebben meer aanleg voor de wiskunde dan meisjes
- Het verzamelen van de gegevens behoort tot de beschrijvende statistiek. De analyse van de resultaten tot de inductieve statistiek.
  Wat betekenen deze resultaten voor de populatie? Hierbij werken we met twee tegengestelde veronderstellingen.

Onderzoeksfasen

Fase 1: vraagstelling, probleem, theorie
Fase 2: meetbaar maken (operationaliseren)
Fase 3: steekproefopzet
Fase 4: verrichten van metingen en/of verzamelen van gegevens
Fase 5: beschrijven en analyseren van gegevens
Fase 6: formuleren van statische conclusies
Fase 7: verband tussen resultaten en theorie

1.1. FASE 1 : vraagstelling , probleem , theorie

Onderzoekshypotheses

Onderzoekshypothese wordt gesteld in termen van meetbare kenmerken, d.i. **variabelen**.

Afhankelijke variabele: de verschillen in deze variabele dienen we te verklaren (bv. aanleg voor wiskunde)
Onafhankelijke variabele: deze verschillen kunnen een verklaring bieden. (bv. geslacht)

vraagstelling = Hebben verschillen in de onafhankelijke variabelen effect op de verschillen in de afhankelijke variabele?

- Onderzoeksvraag

  - onderzoek begint met een vraagstelling waarover we in het onderzoek uitsluitsel wil krijgen

  - vraagstelling vloeit uit theorie, deze moet houdbaar zijn dus ook empirisch aangetoond

  - type vraagstelling: Hoeveel %? (prevalentie), verschil?, samenhang?

  - onderzoeksvraag: Bv; Hebben jongens meer aanleg voor wiskunde dan meisjes?

- Onderzoekshypothese

  - als de theorie juist is kunnen we uit de onderzoeksvraag een uitspraak afleiden die een antwoord hierop is

- onderzoekshypothese: Bv; Jongens hebben meer aanleg voor wiskunde dan meisjes.

- Gebruik een vraagvorm
  bv. Niet: het eetgedrag van jongeren

- Specificatie van de begrippen
  bv. Niet: wat is het medicijngebruik in de psychiatrie?
  Beter: wat is het % van de in 2006 opgenomen psychiatrische patiënten in Vlaanderen die gedurende de eerste maand van de observatieperiode ten minste dagelijks een antidepressivum voorgeschreven kregen?

- Geen oordelende vragen
  bv. Niet: zijn er voldoende psychiaters in Vlaanderen?
  Bv. Niet: hoeveel % van de jongeren eet gezond?

- Een rijtje in plaats van een volzin
  Hoofdvraag en deelvragen
  Bv. Hebben kinderen van handarbeiders lagere schoolcijfers in het basisonderwijs dan deze van hoofdarbeiders?
  Krijgen deze kinderen als ze dezelfde cijfers behalen een ander advies?

  Drie typen van vragen:

  - voorkomen van iets. Bv. hoeveel % van de Vlamingen is depressief?
  - verschillen tussen groepen. Bv. zijn vrouwen meer depressief dan mannen?
  - samenhang. Bv. bestaat er een samenhang tussen de leefsituatie en al dan niet depressief zijn?

1.2. FASE 2 : operationaliseren (meetbaar maken)

Operationaliseren van een begrip tot meetbare variabele = hoe kunnen we dit begrip concreet meten?

Voorbeelden van operationalisering

gemakkelijk	**moeilijker**
geslacht	intelligentie
leeftijd	aanleg voor wiskunde (veel geleerd?, hulp?)
diploma	arbeidstevredenheid
	moeilijk door externe omstandigheden

- het eerste begrip is bijvoorbeeld sekse, dit is makkelijk vast te stellen door bv. J / M op de test te laten zetten. Maar dan moet er nog de keuze gemaakt worden van leeftijdsgroep.
- het tweede begrip is bijvoorbeeld wiskunde aanleg. Vragen zoals 'welke wiskunde?' en 'wat is aanleg?' moeten beantwoordt worden.
- de onderzoekshypothese is dus feitelijk: 'Jongens van 12 à 13 jaar uit de brugklas behalen een hogere gemiddelde score op de wiskundeproefwerken dan meisjes van 12 à 13 jaar uit de brugklas.'

- nog enkele problemen zoals over de proefwerken, deze weergeven niet enkel de aanleg maar ook de ijver. Het proefwerkcijfer is dus niet erg valide voor het begrip 'wiskunde-aanleg'

1.3. FASE 3 : steekproefopzet

Het is duidelijk dat we voor een proef niet zomaar alle personen die ervoor in aanmerking komen kunnen testen, hiervoor is deze populatie veel te groot. Daarom nemen we een steekproef.

- Populatie: alle individuen waar we een uitspraak over willen doen en die bijgevolg in aanmerking komen voor het onderzoek. *Alle individuen testen?*
- Steekproef: selectie van individuen uit de populatie. *Hoe selecteren?*
  - **Aselecte steekproef (at random):** elk individu van de populatie heeft evenveel kans om in de steekproef terecht te komen. *Vereisten?*

    - vereist lijst van de deelnemers in de populatie = steekproefkader, elke persoon krijgt nummer

    - o.g.v. gegevens van dergelijke steekproef kan ik iets vertellen over de populatie.

    Schatting van de parameters (gemiddelde pop.) van de populatie, o.g.v. resultaten van de Steekproef

    - voordelen:
      - generalisering van de populatie is mogelijk
      - veel statistische technieken zijn mogelijk
      - keuze van aantal proefpersonen

    - nadelen:
      - bijkomende kosten en tijdinvestering (verspreide personen)
      - bij kleine steekproeven kan representativiteit een probleem geven

    - soorten steekproeven:
      - volledig aselect: pc kiest aantal nummer van personen
      - systematisch aselect: kies willekeurig 1e persoon, volgende persoon heeft telkens een nummer dat x aantal hoger ligt
      - gestratificeerd: populatie verdeel in deelpopulaties (strata), uit elke strata trekken we aselecte steekproef. (proportioneel/disproportioneel)
        Bv; ASO TSO BSO KSO
      - Clustersteekproef: populatie verdeel in subgroepen. Een aantal subgroepen geselecteerd en bevraagd.
  - **Niet aselecte steekproef:** elk individu heeft niet evenveel kans om in de steekproef terecht te komen. *Gevolgen?*

    - voordelen:
      - Geschikt voor verkennend onderzoek
      - Geschikt om meetinstrumenten te testen
      - Goedkoop en snel
      - Veel gebruikt in de psychologie

    - nadelen
      - Resultaten kunnen niet veralgemeend worden naar de populatie

- soorten steekproeven:
    - Convenience sampling: kies individuen die 'voor het grijpen liggen'
    - Judgement sampling: kies individuen die als bevoorrechte getuigen kunnen fungeren Bv; groep van zware gebruikers als je onderzoek gaat over heroïnegebruikers
    - Snowball sampling: de eerste persoon levert volgende persoon op, etc
    - Quota sampling: populatie in strata verdelen en uit deze een aantal proefpersonen willekeurig kiezen.
    - Random walk: de onderzoeker selecteert proefpersonen volgens een vooraf bepaald wandeltraject. Gebruikt wanneer een steekproefkader niet voor handen is.

- 1.4. FASE 4 : verrichten van metingen en / of verzamelen van gegevens

  Methoden voor verrichten van metingen:

  - Interview
  - Vragenlijst
  - Test/toets
  - Archiefonderzoek
  - Observatie
  - Experiment

→ gegevens worden systematisch weergegeven in een **datamatrix** in **SPSS**

1.5. fase 5 : beschrijven en analyseren van gegevens

Datamatrix is onoverzichtelijk, overzichtelijk maken door :

- Frequentieverdeling
- Grafische voorstelling van de resultaten

Bepaling van centrale tendens en variabiliteit

1.6. fase 6 : formuleren van statistische conclusies

- Deductief redeneren: vanuit een logische redenering, kunnen we conclusies opbouwen

  Bv; Alle mensen zijn sterfelijk, Jan is een mens, dus Jan is sterfelijk

- Inductieve/toetsende/generaliserende/inferentiële redenering: vanuit de empirische observaties trachten we een wet te formuleren

- Hume ziet een probleem: we kunnen niet alle mensen onderzoeken om te besluiten dat mensen sterfelijk zijn
- Statistische/inductieve redenering:

  - 2 tegengestelde beweringen (mensen onsterfelijk=nulhypothese, mensen sterfelijk=alternatieve hypothese)

  - Indien men uit empirie gegevens vind die de eerste stelling onderuit halen is de 2e stelling van toepassing

VOORBEELDEN VAN EEN ONDERZOEK (zie PowerPoint (onduidelijk))
Terug naar het voorbeeld.
- nulhypothese: jongens en meisjes hebben dezelfde aanleg voor wiskunde
- alternatieve hypothese: jongens en meisjes hebben niet dezelfde aanleg voor de wiskunde

Onderzoek 'Busters versus baby boomers. Eenheid en verscheidenheid'.

- Zijn er verschillen inzake levensstijl en koopgedrag tussen beide generaties?
Omschrijving levensstijl: hoe kijken consumenten naar het leven? Waarden, interesses, opvattingen…
Koopgedrag verwijst naar de wijze waarop consumenten hun aankoopgedrag verrichten.
Omschrijving busters (30plussers) en baby boomers (40 en 50plussers)… de operationalisering dus.

Hypothesen op grond van eerder onderzoek

Hypothesen m.b.t. levenstijl
Senioren hechten meer dan jongeren belang aan hun gezondheid (Joosten, 2000)
Ouderen begeven zich minder dan jongeren op het internet (Jacobs, 2003)
Busters zijn meer materialistisch ingesteld dan de baby boomers (De Pelsmacker et al., 2006)
Lachance (2003) stelde vast dat jongeren in sterke mate modebewust zijn, ouderen dus wellicht minder

Hypothesen m.b.t. koopgedrag
Henry (2002) heeft aangetoond dat jongeren bij het consumptiegedrag eerder belang hechten aan sociaal-expressieve aspecten van het product
Ouderen zijn meer geneigd om prijs en kwaliteit als verwisselbare begrippen te zien (Janssens en De Pelsmacker, 2002)
Valkeneers (2006) toonde aan dat twintigers dertigers en veertigers meer dan andere leeftijdsgroepen producten en diensten impulsief aankopen
Henry (2002) toonde eveneens aan dat ouderen meer dan busters belang hechten aan functionele aspect in hun consumptiegedrag

Romberts en Manolis (2000) toonden aan dat baby boomers minder compulsief koopgedrag vertoonden dan de busters.
Inzake attitude t.o.v. merken verwachten we dat de busters meer merkgevoelig zullen zijn dan de bayboomers (Lachance, 2003)

Methode
Steekproeftrekking: aangezien er geen steekproefkader voorradig is, gebruiken we een gemakkelijkheidssteekproef. Studenten BaTP zoeken elk drie proefpersonen in hun omgeving (n=402)

- Methode: beschrijving van de steekproef (I)

Geslacht

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	man	142	35,3	35,3	35,3
	vrouw	260	64,7	64,7	100,0
	Total	402	100,0	100,0	

nieuwleeftijd

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	buster	117	29,1	29,8	29,8
	boomer	276	68,7	70,2	100,0
	Total	393	97,8	100,0	
Missing	System	9	2,2		
Total		402	100,0		

- Methode: beschrijving van de steekproef (II)

Geslacht * nieuwleeftijd Crosstabulation

Count

		nieuwleeftijd		
		buster	boomer	Total
Geslacht	man	38	99	137
	vrouw	79	177	256
Total		117	276	393

Methode
Vragenlijst van het likertschaal type: geef aan in hoeverre u akkoord kunt gaan met volgende uitspraken

1. helemaal niet akkoord
2. niet akkoord

3. eerder niet akkoord
4. neutraal
5. eerder akkoord
6. akkoord
7. helemaal akkoord

22 items over levensstijl en 51 items over koopgedrag (overgenomen van andere auteurs)
5 onafhankelijke variabelen: leeftijd, geslacht, voornaamste verantwoordelijke voor de dagelijkse aankopen, woonplaats, hoogste diploma

Resultaten
Samenstelling van de schalen
- hercodering (omdraaien) van enkele items, zodanig dat alle items van één schaal in dezelfde richting wijzen
Via SPSS: transform, recode, into a different variable en draai de scores om.
- onderzoek de interne homogeniteit van de schalen. In SPSS via analyse, scale, reliabilty analysis…
Kies voor de Cronbach Alfa, en geef de items in van deze categorie
Welke items vormen samen de meest optimale schaal?

OEFENINGEN

opgave : 1.1.

leerlingen die in groepjes werken behalen hogere scores op de toets dan leerlingen die plenair les krijgen.
* afhankelijke variabele : de score
* onafhankelijke variabele : lesvorm

- plenair = 1 waarde, in groep = 1 waarde

Kan het dat een afhankelijke variabele een onafhankelijke variabele is? Ja dat kan

opgave 1.4.

ouders met een hoger opleidingsniveau hebben minder kinderen dan ouders met een lager opleidingsniveau
* afhankelijke variabele : hoeveel kinderen
* onafhankelijke variabele : het hoogste opleidingsniveau

HOOFDSTUK 2 : operationaliseren en meten & spss transform (zoals testen uit joepie-boekje, alles optellen, uitkomst,..)

Doelstellingen hoofdstuk II

- De student begrijpt de begrippen variabele, operationalisering,…;
- De student kent en begrijpt de betekenis van de vier soorten meetniveaus;
- De student begrijpt de consequenties van deze vier typen meetniveaus.
- De student kan via SPSS variabelen bewerken via Transform-compute,recode en count

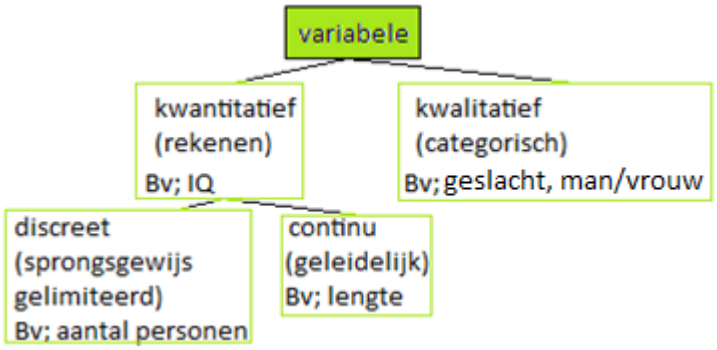
Operationaliseren: de vooropgestelde hypothesen bevatten begrippen, hoe gaan we deze meten?
Bv; Hoe meten we de attitude t.o.v. het milieu?

2.1. variabelen

- Variabelen: zijn kenmerken van de proefpersonen die diverse waarden aannemen

  Bv; geslacht

- Scores: waarden: zijn de individuele uitslagen op een variabele
  Bv; bij een vrouw: vrouw, bij een man: man



Kwalitatief : heeft geen volgorde

*Discreet : geen komma getallen, beperkt aantal waarden
*continue : wel komma getallen, onbeperkte waarden
-> discreet of continu heeft geen invloed op de analyse maar wel op de grafiek

Voorbeelden :

- Kwalitatief Bv;
  - Wel of niet in mileubewust
  - Militaire rang (soldaat, sergeant)
- Kwantitatief en discreet Bv;
  - Aantal inschrijvingen voor een cursus
  - Aantal kinderen in het huishouden

- Kwantitatief en continu Bv;
    - Lichaamslengte
    - Tijd nodig om het examen op te lossen

2.2. Meetniveaus (belangrijk)

er zijn 4 meetniveaus :

2.2.1. nominaal
2.2.2. ordinaal
2.2.3. interval
2.2.4. ratio

NOIR

2.2.1. nominaal meetniveau

Alle categorische variabelen zijn nominaal
bv. variabele geslacht kent twee waarden: jongen en meisje.
Kunnen getallen aan gekoppeld worden, maar deze hebben geen numerieke betekenis
- waarden kennen geen rangorde
- er is geen meeteenheid
- er is geen nulpunt

bv. geslacht is een nominale variabele

- Speciaal geval van nominale schaal:
  dichotomie: dwz er zijn slechts twee niveaus mogelijk voor deze variabele
  bv. Volgt u logopedie? Ja of neen?

- De onafhankelijke variabelen zijn vaak nominaal van niveau.

Voorbeelden:
Geslacht
    man
    vrouw
Burgerlijke stand
    alleenstaand
    gehuwd
    gescheiden
    weduwe/weduwnaar
-> niet goed : stel samenwonend?
Nationaliteit
    Belg
    Nederlander
    Duitser
    …….

2.2.2. ordinaal meetniveau

Categorieën hebben een bepaalde volgorde
Er kunnen getallen aan gekoppeld worden, maar het verschil tussen 2 opeenvolgende waarde heeft geen betekenis, enkel de volgorde.
- er is wel een rangorde
- er is geen meeteenheid (is de afstand tussen zeer slecht –en- slecht even groot als tusse slecht –en – neutraal?)
- er is geen nulpunt

bv. hoogst behaalde diploma is ordinale variabele
geen/lager secundair/hoger secundair/hoger onderwijs/unief

Voorbeeld
Welk is uw evaluatie van deze cursus?
zeer slecht
slecht
neutraal
goed
zeer goed
Welk is uw hoogst verworven diploma?
lager onderwijs
lager secundair
hoger secundair
bachelor
master

2.2.3. intervalmeetniveau

- Rangorde is belangrijk
- Verschil tussen twee opeenvolgende waarden is gelijk; er is een meeteenheid
- Geen absoluut nulpunt.
bv. IQ meting
volgorde is belangrijk; en er is een meeteenheid, maar IQ 120 is niet dubbel zo slim als IQ 60

Voorbeelden
Temperatuur in graden Celsius
het verschil tussen 6° en 30° is 24°
maar 30° is niet het vijfvoud van 6°
De jaartelling
IQ coëfficiënt

2.2.4. ratio meetniveau

- Rangorde is belangrijk
- Er is een meeteenheid
- Er is een absoluut nulpunt

bv. Lichaamslengte, gewicht, tijd om de test op te lossen, enz..
Bv. Iemand van 1,8 meter is dubbel zo groot als iemand van 0,9 meter

2.3. betekenis van dit meetniveau zie ook SPSS

- Een variabele kan van aard veranderen al naargelang de vraagstelling, bv. de leeftijd
  Hoe oud bent u? …… Dit is ratio niveau
  In welk jaar bent u geboren?.... Dit is interval
  Hoe oud bent u, maak uw keuze:
  0 20 à 30 jaar
  0 31 à 40 jaar
  0 41 à 50 jaar
  0 ouder dan 51….. Dit is ordinaal niveau

- Naarmate het meetniveau hoger is kunnen we meer bewerkingen uitvoeren

- Zorg voor variabelen – indien mogelijk - met een zo hoog mogelijk meetniveau, bv. leeftijd.

- Diverse variabelen met intervalniveau kunnen opgeteld worden… Zorg ervoor dat de afhankelijke variabele zo veel mogelijk op interval niveau gemeten wordt. Dit heeft eveneens consequenties voor de analyse van de gegevens

- De vraag naar houdingen, overtuigingen, gevoelens, etc… kan het best gebeuren aan de hand van een likertschaal
  1. helemaal akkoord
  2. akkoord
  3. neutraal
  4. niet akkoord
  5. helemaal niet akkoord

  (-> door deze cijfers is de afstand tussen helemaal akkoord – en – akkoord gelijk aan de afstand tussen akkoord – en – neutraal) anders zou het ordinaal zijn)

- De vraag naar feitelijkheden, bv. Bent u een man/vrouw? Woont u in Antwerpen? kan uiteraard niet met een likertschaal bevraagd worden

2.4. betrouwbaarheid en validiteit

- Betrouwbaarheid: meet de test iets?
  Heeft te maken met de stabiliteit van de meting
  - hertesting

- halvering
- parallel vorm
- interne homogeniteit (Chronbach Alfa)

- Validiteit: meet de test wat hij behoort te meten?
  Welk is de relatie van de testuitslag met een andere meting van dit begrip (bv. relatie intelligentie en schooluitslag)

Betrouwbaarheid is een voorwaarde voor validiteit

-> zie SPSS voor SPSS transform

Oefeningen

1)
Welke meetschaal past hier het best?
1. meetresultaat X_1 is hoger dan X_2
2. X_1 is 3 keer zo groot als X_2
3. X_1 is twee punten hoger als X_2
4. Het beroep X_1 valt in een andere categorie als X_2

1 = ordinaal
2 = scale – ratio
3 = interval
4 = nominaal

2)
Om te zeggen dat prestatie A een bepaald percentage beter is dan prestatie B moet men meten op een …
1. nominale schaal
2. ordinale schaal
3. interval schaal
4. ratio schaal

4 ratio schaal

3)
Bepaal het meetniveau van de volgende variabelen
- de bloedgroep van een mens (O, A, B en AB) nominaal
- het hoogste diploma van een persoon ordinaal
- het aantal verkeersongevallen per jaar op een bepaald kruispunt ratio - scale
- de uitslag van een loopwedstrijd (in volgorde? In tijd?) ordinaal / in tijd : scale
- het rugnummer van een wielrenner in de Ronde van Frankrijk nominaal
- indeling van renners (zelfde ronde) in wel of niet dopinggebruiker nominaal

HOOFDSTUK 3 : frequentieverdelingen

Doelstellingen hoofdstuk III

- De student verstaat een aantal begrippen, zoals frequentieverdeling, enz…
- De student kent de diverse vormen van grafische voorstelling van gegevens;
- De student weet hoe het schaalniveau een impact heeft op de wijze van grafische voorstelling;
- De student kan percentielen/decielen/kwartielen bepalen uit een frequentietabel;
- De student kan via SPSS een frequentietabel maken en een eenvoudige grafische voorstelling van de gegevens.

3.1. frequentietabel

- Datamatrix: bevat de resultaten vh onderzoek (onoverzichtelijk)
- Frequentieverdeling: geeft een beter overzicht

- Één-dimensionale tabel: 1 variabele (geslacht) :

Geslacht

		Frequency	Percent
Valid	man	142	35,3
	vrouw	260	64,7
	Total	402	100,0

Eendimensionale tabel
= Frequentie is het aantal keren dat een bepaalde waarde voorkomt (1 variabele en hoeveel deze voorkomt)
- absolute frequentie
- relatieve frequentie (%)
- absolute cumulatieve frequentie
- cumulatieve proportie (%)
(beide laatste enkel voor interval waarden)

- Meer-dimensionale tabel/kruistabel: meer variabelen (geslacht, nieuw: leeftijd)

Geslacht * nieuwleeftijd Crosstabulation

Count

		nieuwleeftijd		
		buster	boomer	Total
Geslacht	man	38	99	137
	vrouw	79	177	256
Total		117	276	393

frequentietabellen en grafieken

- Elke tabel/grafiek krijgt een volgnummer en een titel waaruit blijkt wat de inhoud is van de tabel/grafiek
- Vermeld steeds de herkomst van de gegevens (n=402)
- Tabellen kun je overnemen via copy-paste special-picture. Je kunt ook een nieuwe tabel maken in Word

3.2. grafieken

3.2.1.Nominale waarden

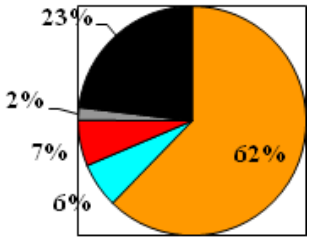
3.2.2.Ordinale waarden

3.2.3.Interval/ratio niveau

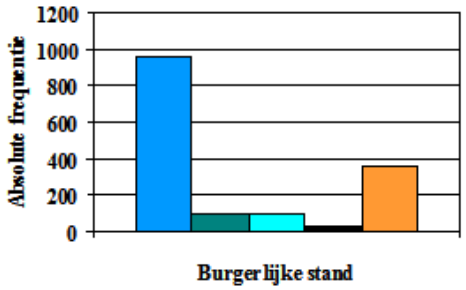
3.2.1.Nominale waarden

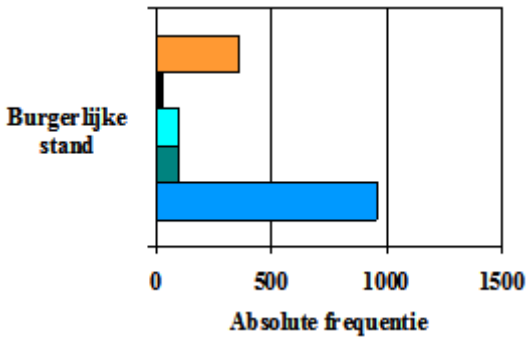
APS-SURVEY 2004: Burgerlijke stand		
Burgerlijke stand	*Absolute Frequentie*	*Relatieve Frequentie*
Gehuwd	957	62,2%
Weduwe/weduwnaar	98	6,4%
Wettelijk gescheiden	100	6,5%
Feitelijk gescheiden	28	1,8%
Ongehuwd	355	23,1%
TOTAAL	1538	100,0%

TAARTDIAGRAM :



STAAFDIAGRAM:

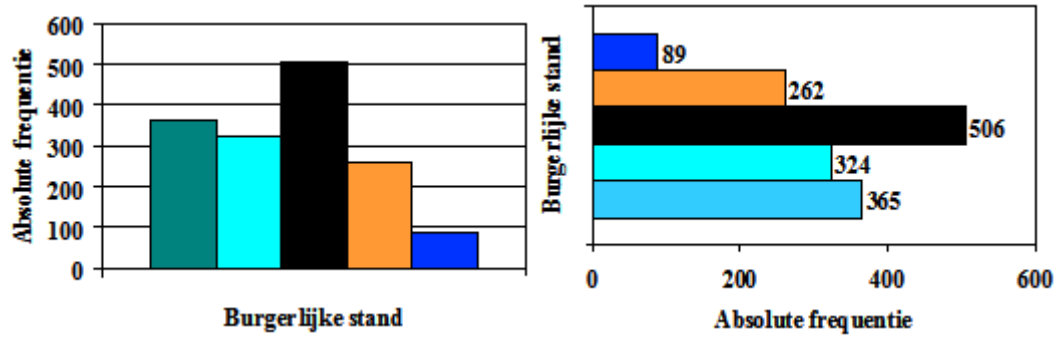




3.2.2.Ordinale waarden

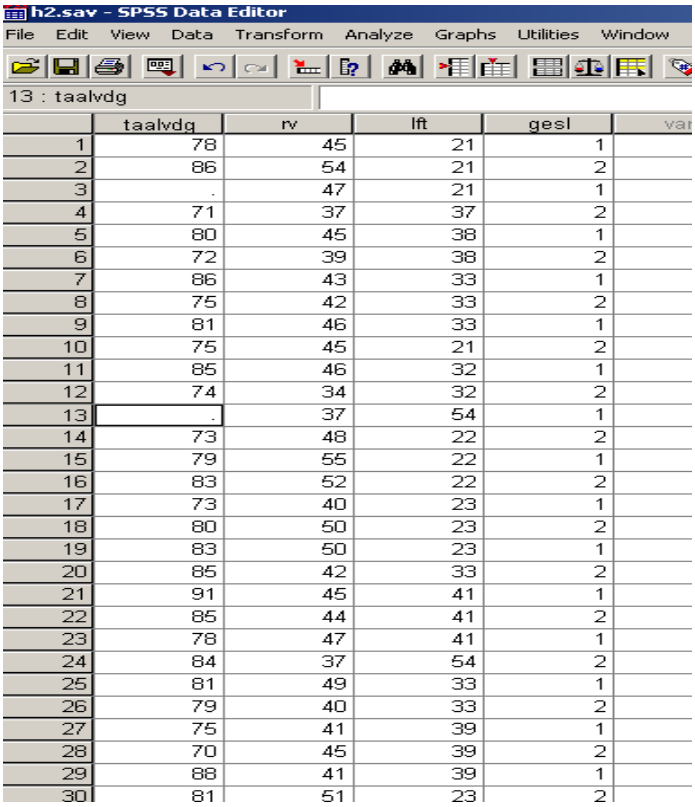
APS-SURVEY 2004: Hoogste diploma		
Diploma	*Absolute Frequentie*	*Relatieve Frequentie*
Geen/LO	365	23,6%
Lager secundair	324	21,0%
Hoger secundair	506	32,7%
Niet universitair HO	262	16,9%
Universitair HO	89	5,8%
TOTAAL	1546	100,0%

STAAFDIAGRAM:



3.2.3.Interval/ratio niveau

Datamatrix,Taalvaardigheid,Rekenvaardigheid,Leeftijd en Geslacht


h2.sav - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window

13 : taalvdg

	taalvdg	rv	lft	gesl
1	78	45	21	1
2	86	54	21	2
3	.	47	21	1
4	71	37	37	2
5	80	45	38	1
6	72	39	38	2
7	86	43	33	1
8	75	42	33	2
9	81	46	33	1
10	75	45	21	2
11	85	46	32	1
12	74	34	32	2
13	.	37	54	1
14	73	48	22	2
15	79	55	22	1
16	83	52	22	2
17	73	40	23	1
18	80	50	23	2
19	83	50	23	1
20	85	42	33	2
21	91	45	41	1
22	85	44	41	2
23	78	47	41	1
24	84	37	54	2
25	81	49	33	1
26	79	40	33	2
27	75	41	39	1
28	70	45	39	2
29	88	41	39	1
30	81	51	23	2

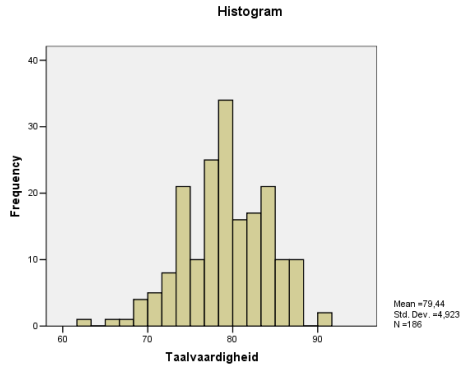
Frequentietabel

SPSS biedt een
overzicht van de resultaten,
middels een frequentietabel

Taalvaardigheid

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	62	1	,5	,5	,5
	66	1	,5	,5	1,1
	68	1	,5	,5	1,6
	70	4	2,1	2,2	3,8
	71	5	2,7	2,7	6,5
	72	1	,5	,5	7,0
	73	7	3,7	3,8	10,8
	74	8	4,3	4,3	15,1
	75	13	6,9	7,0	22,0
	76	10	5,3	5,4	27,4
	77	9	4,8	4,8	32,3
	78	16	8,5	8,6	40,9
	79	19	10,1	10,2	51,1
	80	15	8,0	8,1	59,1
	81	16	8,5	8,6	67,7
	82	7	3,7	3,8	71,5
	83	10	5,3	5,4	76,9
	84	9	4,8	4,8	81,7
	85	12	6,4	6,5	88,2
	86	10	5,3	5,4	93,5
	87	6	3,2	3,2	96,8
	88	4	2,1	2,2	98,9
	91	2	1,1	1,1	100,0
	Total	186	98,9	100,0	
Missing	System	2	1,1		
Total		188	100,0		

HISTOGRAM:

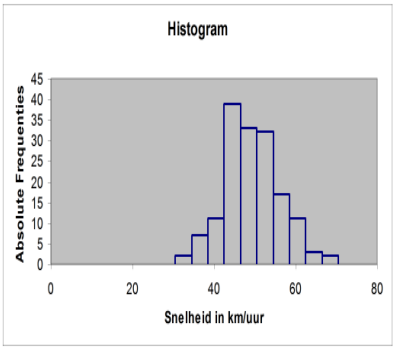


Gegroepeerde frequentietabel

- Enkel om de gegevens overzichtelijk voor te stellen; informatie gaat verloren = functie
- Voor de komst van SPSS gebruikelijke wijze van voorstelling;
  Hoeveel klassen? Turven van aantallen, enz… verwijzen we naar het pre-SPSS tijdperk.
- ***Geen verdere analyse van de gegevens aan de hand van dergelijke tabel.***

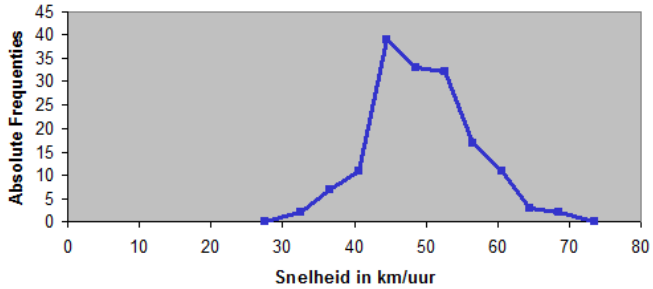
Klasse	onder grens	boven grens	centrum	absolute frequentie	relatieve frequentie	gecumuleerde absolute frequentie	gecumuleerde relatieve frequentie
31-34	30,5	34,5	32,5	2	1,3%	2	1,3%
35-38	34,5	38,5	36,5	7	4,5%	9	5,7%
39-42	38,5	42,5	40,5	11	7,0%	20	12,7%
43-46	42,5	46,5	44,5	39	24,8%	59	37,6%
47-50	46,5	50,5	48,5	33	21,0%	92	58,6%
51-54	50,5	54,5	52,5	32	20,4%	124	79,0%
55-58	54,5	58,5	56,5	17	10,8%	141	89,8%
59-62	58,5	62,5	60,5	11	7,0%	152	96,8%
63-66	62,5	66,5	64,5	3	1,9%	155	98,7%
67-70	66,5	70,5	68,5	2	1,3%	157	100,0%
TOTAAL				157	100,0%		

Klasse	onder grens	boven grens	centrum	absolute frequentie	relatieve frequentie	gecumuleerde absolute frequentie	gecumuleerde relatieve frequentie
31-34	30,5	34,5	32,5	2	1,3%	2	1,3%
35-38	34,5	38,5	36,5	7	4,5%	9	5,7%
39-42	38,5	42,5	40,5	11	7,0%	20	12,7%
43-46	42,5	46,5	44,5	39	24,8%	59	37,6%
47-50	46,5	50,5	48,5	33	21,0%	92	58,6%
51-54	50,5	54,5	52,5	32	20,4%	124	79,0%
55-58	54,5	58,5	56,5	17	10,8%	141	89,8%
59-62	58,5	62,5	60,5	11	7,0%	152	96,8%
63-66	62,5	66,5	64,5	3	1,9%	155	98,7%
67-70	66,5	70,5	68,5	2	1,3%	157	100,0%
TOTAAL				157	100,0%		

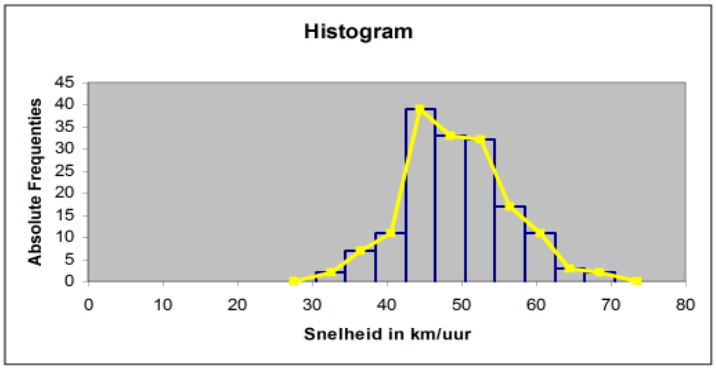


Klasse	onder grens	boven grens	centrum	absolute frequentie	relatieve frequentie	gecumuleerde absolute frequentie	gecumuleerde relatieve frequentie
31-34	30,5	34,5	32,5	2	1,3%	2	1,3%
35-38	34,5	38,5	36,5	7	4,5%	9	5,7%
39-42	38,5	42,5	40,5	11	7,0%	20	12,7%
43-46	42,5	46,5	44,5	39	24,8%	59	37,6%
47-50	46,5	50,5	48,5	33	21,0%	92	58,6%
51-54	50,5	54,5	52,5	32	20,4%	124	79,0%
55-58	54,5	58,5	56,5	17	10,8%	141	89,8%
59-62	58,5	62,5	60,5	11	7,0%	152	96,8%
63-66	62,5	66,5	64,5	3	1,9%	155	98,7%
67-70	66,5	70,5	68,5	2	1,3%	157	100,0%
TOTAAL				157	100,0%		

Polygoon

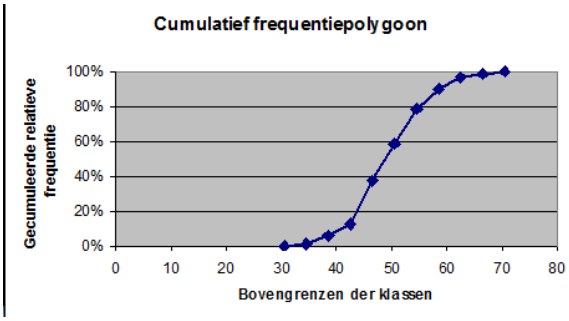


Polygoon (gecombineerd met histogram)



Cumulatief Frequentiepolygoon

Klasse	onder grens	boven grens	centrum	absolute frequentie	relatieve frequentie	gecumuleerde absolute frequentie	gecumuleerde relatieve frequentie
31-34	30,5	34,5	32,5	2	1,3%	2	1,3%
35-38	34,5	38,5	36,5	7	4,5%	9	5,7%
39-42	38,5	42,5	40,5	11	7,0%	20	12,7%
43-46	42,5	46,5	44,5	39	24,8%	59	37,6%
47-50	46,5	50,5	48,5	33	21,0%	92	58,6%
51-54	50,5	54,5	52,5	32	20,4%	124	79,0%
55-58	54,5	58,5	56,5	17	10,8%	141	89,8%
59-62	58,5	62,5	60,5	11	7,0%	152	96,8%
63-66	62,5	66,5	64,5	3	1,9%	155	98,7%
67-70	66,5	70,5	68,5	2	1,3%	157	100,0%
TOTAAL				157	100,0%		



Frequentietabel en histogram met SPSS ZIE SPSS

3.3. positie van een score in een verdeling van uitslagen

Het percentiel P van een ruwe score is het percentage metingen dat kleiner is (of gelijk aan) dan deze ruwe score.

Dus hoeveel procent van de observaties ligt beneden deze score?

Voorbeeld: op een taaltest behaalde Jan een score van 112/120.

Is dat een goede score?

Kijk hiervoor naar het percentiel.

Als 20% van de leerlingen een betere score behaalde, zeggen we dat de uitslag 112 het 80^{ste} percentiel is, ofwel $P_{80}=112$

Als 70% van de leerlingen een betere score behaalde, zeggen we dat deze uitslag het 30^{ste} percentiel is, ofwel $P_{30}=112$

Varianten van de percentielscores

- Decielscore. We verdelen de uitslagen in 10 delen, zodanig dat in elk onderdeel 10% van de observaties zich situeren;
  dus D_1 = Pc10, D_2 = Pc20, enz…

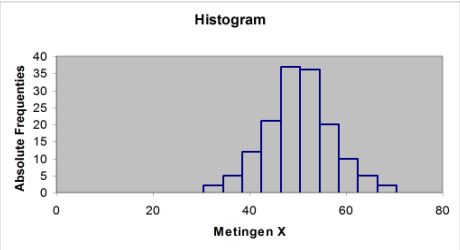
- Kwartielscores. We verdelen de uitslagen in vier onderdelen, die elk 25% van de observaties bevatten.
  dus Q_1 = Pc25; Q_2 = Pc 50 en Q_3 = Pc75

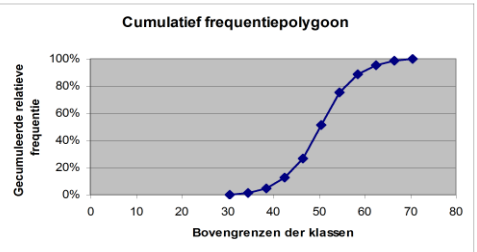
3.4. frequentieverdeling verder ontleed

Verschillen bestaan er op vlak van

- Centrale tendens: centrum 'perfect' of niet
- Mate van variabiliteit
- Symmetrie / scheefheid
- Scherpe top of afgevlakte top

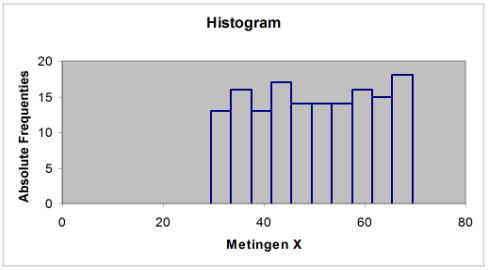
- **Voorbeeld 1: normaalachtige verdeling**
  - Centrum is 'perfect' in het midden van de verdeling
  - 'kleine' spreiding rond de centrale waarden
  - Symmetrie
  - Mooie welving, ééntoppig (unimodaal)
  - Bv. IQ-uitslagen

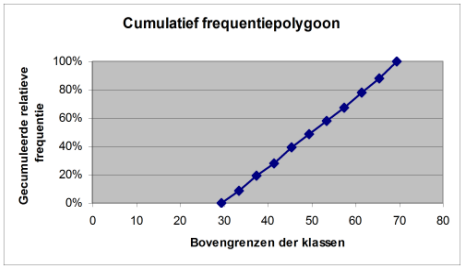




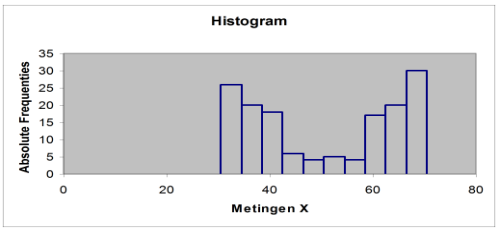
- **Voorbeeld 2: Uniformachtige verdeling**

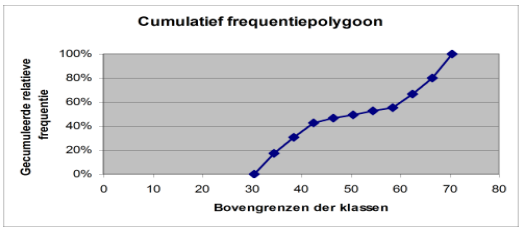
- Relatief grote spreiding van uitslagen rond de centrale waarden
- Symmetrische verdeling
- Geen duidelijke top
- Bv. Leeftijd van groep volwassenen (30-50j)



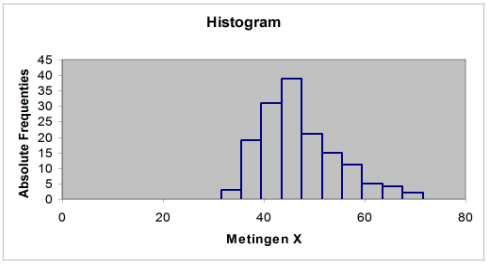


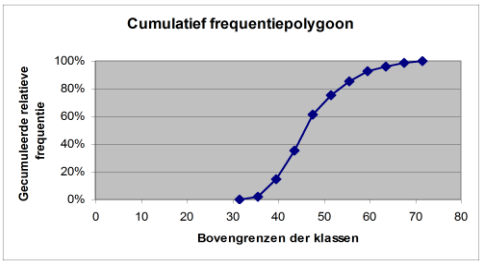
- **Voorbeeld 3: U-achtige verdeling**
  - Centrum is 'perfect' in het midden
  - Grote spreiding rond de centrale waarden
  - Symmetrie
  - Centraal een 'ingeslagen' top, tweetoppig (bimodaal)
  - Bv; leeftijd in pretpark bezocht door kinderen met grootouders



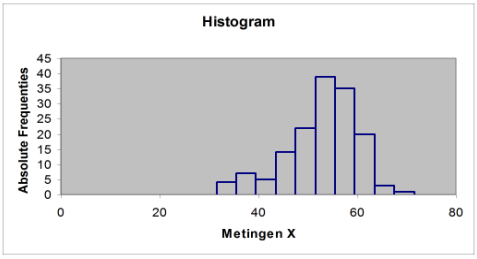


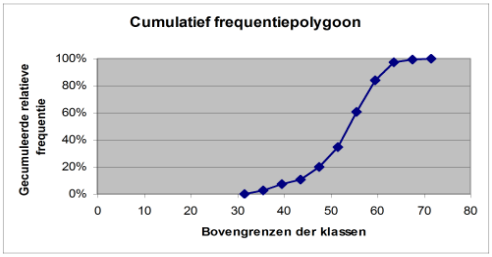
- **Voorbeeld 4: Rechtsscheve verdeling**
  - Centrum schuift op naar links, staart naar rechts
  - Kleine spreiding rond de centrale waarden
  - Asymmetrie (scheef)
  - Mooie welving, eentoppig (unimodaal)
  - Bv; verdeling uitslagen van een heel moeilijke test (vloereffect)





- **Voorbeeld 5: Linksscheve verdeling**
    - Centrum schuift op naar rechts, staart naar links
    - Kleine spreiding rond de centrale waarden
    - Asymmetrie (scheef)
    - Mooie welving, eentoppig (unimodaal)
    - Bv; testuitslag van een heel gemakkelijke test (plafondeffect)





AANMAKEN VAN TABEL IN SPSS -> ZIE SPSS

HOOFDSTUK 4 : centrummaten

Doelstellingen hoofdstuk IV

- De student kent de diverse begrippen over de centrummaten;
- De student kent de impact van de aard van de schaal op de bepaling van de centrale tendens;
- De student kan – handmatig - de centrale tendens berekenen voor een verdeling van uitslagen;
- Via SPSS kan de student de centrale tendens van een reeks gegevens bepalen.

4.1. de modus

= Is de waarde met de hoogste frequentie

Bijvoorbeeld scores op een Likertschaal (1-5)

Ik vind de opwarming van de aarde een groot probleem (helemaal akk….. helemaal niet akk)

score	frequentie
1 helemaal akk	13
2 akkoord	12
3 weet niet	3
5 helemaal niet akk	1

- Welk is de modus? Score 1 'helemaal akkoord

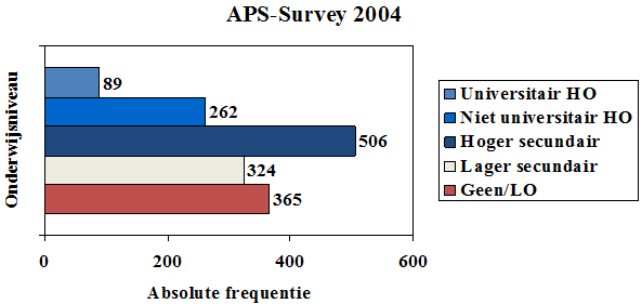
modus kan : nominaal en ordinaal voorkomen

Voorbeeld van nominale gegevens :

APS-SURVEY 2004: Burgerlijke stand		
Burgerlijke stand	*Absolute Frequentie*	*Relatieve Frequentie*
Gehuwd	957	62,2%
Weduwe/weduwnaar	98	6,4%
Wettelijk gescheiden	100	6,5%
Feitelijk gescheiden	28	1,8%
Ongehuwd	355	23,1%
TOTAAL	1538	100,0%

De modus hier is : gehuwd

Voorbeeld van ordinaal meetniveau :



de modus hier is hoger secundair onderwijs

De modus :

- Zal vooral gebruikt worden voor nominale waarden; maar kan in principe altijd bepaald worden. Is meteen duidelijk in de frequentietabel
- Meer dan één modus is mogelijk, bij een bimodale verdeling zijn er twee modi. (vb.: in een pretpark ondervraagt men de klanten , dit zijn bejaarde (modus 1) en kinderen (modus2)
- Gebruikt weinig informatie uit de gegevens

4.2. de mediaan

- De mediaan is de middelste waarde wanneer de observaties in volgorde van laag naar hoog zijn gezet. (niet mogelijk voor nominale waarden)
- Bij een oneven aantal observaties precies de middelste, en bij een even aantal observaties het midden tussen de twee middelste scores;
- Komt dus overeen met percentiel 50.
    - Welk is de mediaanwaarde van
      2, 4, 6, 8, 10?
      De mediaanwaarde is 6, als middelste waarde
    - Welk is de mediaanwaarde van
      2, 4, 6, 7, 8, 10?
      De mediaan is 6,5 zijnde het midden tussen 6 en7.
    - Welk is de impact van een wijzing van de laatste observatie 10 in 20?
      Verandert hierdoor de mediaan?
      ipv 2,4,6,7,8,**10** – 2,4,6,7,8,**20**
      neen hierdoor verandert de mediaan niet , want een mediaan bevat weinig informatie over de uitslagen

APS-SURVEY 2004: Hoogste diploma		
Diploma	*Absolute Frequentie*	*Relatieve Frequentie*
Geen/LO	365	23,6%
Lager secundair	324	21,0%
Hoger secundair	506	32,7%
Niet universitair HO	262	16,9%
Universitair HO	89	5,8%
TOTAAL	1546	100,0%

Voorbeeld van een nominaal meetniveau.

De mediaan is : hoger secundair onderwijs

50% is het midden, dat valt daar

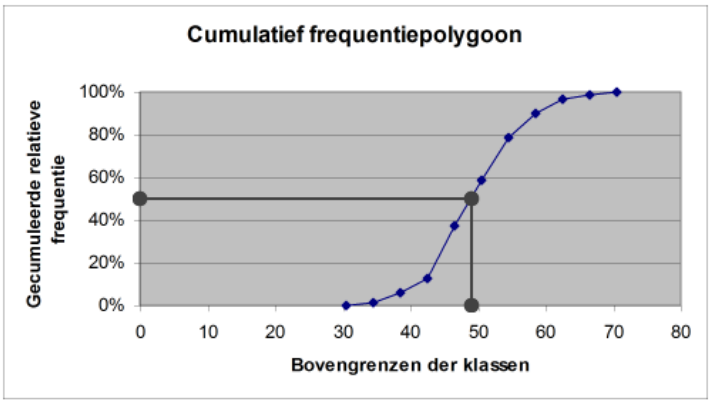
Voorbeeld van ordinaal meetniveau

de mediaan is hier de grens tussen slecht en neutraal (35, waar ligt dat)

- bepaal een mediaan uit een tabel :

$$\text{Mediaan} = \frac{14 + 15}{2} = 14{,}5$$

- de mediaan kan grafisch afgeleid worden uit de cumulatieve frequentiepolygoon (50%)



De mediaan :

- Kan niet gebruikt worden bij nominale waarden;
- Is niet afhankelijk van extreem hoge of lage uitslagen.
  Gebruikt weinig info uit de gegevens;
- Kan gezien worden in vergelijking met het rekenkundig gemiddelde;
- Is gemakkelijk te begrijpen/uit te leggen/grafisch voor te stellen.

-> hoe maak ik een boxplot met spss zie spss (2de methode om boxplot te maken)

4.3. het gemiddelde

Het gemiddelde is de som van alle scores gedeeld door het aantal scores.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Is enkel mogelijk voor interval en ratio meetniveaus, bv. IQ, schooluitslagen, testuitslagen, leeftijd,…

Voorstelling van het gemiddelde :

- in de steekproef : $\bar{X}$
- in de populatie : μ

Een voorbeeld

Score	Frequentie
4	9
6	15
8	21

gemiddelde: (9*4 + 15*6 + 21*8)/45 = 6,53

4.3.1. het gemiddelde bij een samengestelde steekproef

Veronderstel je beschikt over twee steekproeven n_1 en n_2 met een respectievelijk gemiddelde X_1 en X_2, welk is dan het zgn. gewogen gemiddelde?

$$\bar{X} = \frac{n_1 \cdot \bar{X}_1 + n_2 \cdot \bar{X}_2}{n_1 + n_2}$$ gemiddelde optellen

Een voorbeeld

- Tien jongens kijken gemiddeld 3 uur per dag tv en vijf meisjes kijken gemiddeld 2 uur per dag tv. Wat is dan het gemiddelde van de gezamenlijke proefgroep?
- Oplossing
  de jongens kijken 30 uur tv
  de meisjes kijken 10 uur tv
  totaal: 40 uur;
  dit is gemiddeld 40/15 = 2,67 (=gewogen gemiddelde)

$$\bar{X} = \frac{n_1 \cdot \bar{X}_1 + n_2 \cdot \bar{X}_2}{n_1 + n_2} = \frac{10 \cdot 3 + 5 \cdot 2}{10 + 5} = \frac{40}{15} = 2,67$$

Een analoge eigenschap voor de mediaan bestaat niet. Om de mediaan van de samengestelde steekproef te kennen, moet je alle metingen kennen

4.3.2. het getrimde gemiddelde

Het rekenkundig gemiddelde van het deel van de waarnemingsgetallen dat overblijft na weglating van de P% kleinste en P% grootste.

Bv; gemiddelde: 18,2

1
3
~~6~~
7
8
9
~~10~~
~~14~~
15
16
17
19
21
23
25
28
30
33
39
40

getrimde gemiddelde: 17,94

1
3
6
7
8
9
10
14
15
16
17
19
21
23
25
28
30
33
39
40

- Eigenschappen van het rekenkundig gemiddelde
  - Som van de afwijkingen van de waarnemingsgetallen tot het rekenkundig gemiddelde is gelijk aan 0

X_i	$X_i - \bar{X}$
18	18-14=4
13	13-14=-1
17	17-14=3
16	16-14=2
10	10-14=-4
09	9-14=-5
15	15-14=1
	SOM=0

  - Bij een lineaire transformatie van de scores, wordt het rekenkundig gemiddelde op dezelfde wijze getransformeerd; d.w.z. als je alle waarnemingsgetallen met b vermenigvuldigt en daar een constante a bijtelt, dan wordt het rekenkundig gemiddelde op dezelfde manier getransformeerd

    Bv; je meet de volgende temperaturen met de schaal van Celsius:

    18°C 13°C 17°C 16°C 10°C 09°C 15°C: gemiddelde 14°C

    Via eenvoudige transformatie kan je de waarden omzetten naar de schaal van Fahrenheit:

    F= 32 + 1,8 . X°C

    64,4F 55,4F 62,6F 60,8F 50F 48,2F 59F: gemiddelde 57,2F

  - Gevoelig voor extreme waarden
  - Steeds berekenen bij interval en ratio waarden

4.4. gebruik van centrummaten

- Modus: bij **nominale**, ordinale, interval en ratio waarden
- Mediaan: bij ordinale, interval en ratio waarden
- Gemiddelde: bij interval en ratio waarden
- Gemiddelde versus mediaan?
    - Gemiddelde gebruikt meer informatie dan de mediaan; de mediaan gebruikt enkel de rangorde van de getallen, dus bij interval waarden….
    - Invloed van 'uitbijters'/''outliers'? Uitbijters hebben geen invloed op de mediaan, wel op het gemiddelde.
    - Bij de mogelijkheid van extreme waarden kan getrimde gemiddelde een oplossing bieden.
    - Getrimde gemiddelden worden berekend zonder rekening te houden met bv. de 5% hoogste en 5% laagste waarden.
    - Gemiddelde versus mediaan:
      Het gemiddelde varieert minder van steekproef tot steekproef t.o.v. de mediaan. Dus het gemiddelde wordt meer gebruikt in de toetsende statistiek om het centrum van de populatie te schatten.
    - Gemiddelde is algebraïsch aardiger. We kunnen gegevens van subgroepen samenvoegen om gewogen gemiddelde te berekenen, … dit kan niet bij een mediaan.
    - Het gemiddelde verdient de voorkeur bij interval/ratio schalen.
    - Onderlinge positie van gemiddelde en mediaan zegt iets over de mate van scheefheid van de verdeling.

4.5. vergelijking van centrummaten

Eigenschappen	**Modus**		**Mediaan**		**Gemiddelde**
meetniveau	nominaal of hoger		ordinaal of hoger		interval of hoger
interpretatie	typische waarde		middelste waarde		gemiddelde waarde
bij symmetrische verdeling	modus	=	mediaan	=	gemiddelde
bij positief scheve verdeling (staart rechts)	modus	<	mediaan	<	gemiddelde
bij negatief scheve verdeling (staart links)	modus	>	mediaan	>	gemiddelde
gevoelig voor extreme waarden (uitbijters)	nee		nee		ja
gebruik info uit de gegevens	weinig		tamelijk veel		het meest
gevoelig voor klassenindeling	groot		minder		minst

verschil van steekproef tot steekproef	grootst		minder		minst
algebraïsch hanteerbaar	nee		nee		ja

Besluit:

1. De vorm van de verdeling heeft invloed op de onderlinge positie van de centrummaten.
2. Indien mogelijk maak gebruik van het rekenkundig gemiddelde als maat van centrale tendens.

SPSS en de centrummaten zie SPSS

BESLUIT centrummaten :
Bij centrummaten gaat het om het centrum van de geobserveerde scores. We onderscheiden modus, mediaan en gemiddelde. De mediaan is niet gevoelig voor extreme waarden, het gemiddelde wel. Het gemiddelde varieert minder (in vergelijking met de mediaan) wanneer men uit een populatie meerdere steekproeven trekt.

OPGAVE

Er zijn 4 deelgroepen bestaande uit 15, 20, 10 en 18 personen. Het gemiddeld gewicht per deelgroep bedraagt respectievelijk 162, 148, 153 en 140 pond. Wat is het gemiddeld gewicht van alle personen (63 in totaal) ?

$$\frac{15\cdot 162+20\cdot 148+10\cdot 153+18\cdot 140}{15+20+10+18}=149,84$$

OPGAVE 2

bij het invoeren van een reeks waarden wordt het getal 26 als 62 ingetypt, na correctie blijkt dat 26 de grootste waarde is , welk van onderstaande maten blijft

1) het gemiddelde
2) de mediaan

--> de mediaan

OPGAVE 3

Welke centrummaat wordt gebruikt bij een populatie?

… het gemiddelde
… de mediaan
… de modus

--> Het gemiddelde

HOOFDSTUK 5 : spreidingsmaten en spss descriptives

Doelstellingen hoofdstuk 5

- De student kent de diverse maten van variabiliteit en de voor- en nadelen van elk van deze maten;
- U kunt uit een eenvoudig bestand deze maten handmatig en via SPSS berekenen. Bovendien kunt u deze resultaten interpreteren;
- De student begrijpt het effect van lineaire transformaties op het rekenkundig gemiddelde en de spreidingsmaten;
- U kunt omgaan met Z-waarden en op deze wijze uiteenlopende scores met mekaar vergelijken.

Spreidingsmaten :

- Naast de centrale tendens vormt de *mate van verscheidenheid* van de uitslagen een belangrijk gegeven
- Spreiding over de schaal klein: scores allemaal dicht bij centrum
- Spreiding over de schaal groot: scores allemaal ver uit elkaar en ver van het centrum
- Enkel bepaald bij interval en ratio schalen (bij ordinale maten ook interkwartielbereik)

5.1. gemiddelde afwijkingsscores (= deviatie score)

In hoeverre wijkt elke individuele uitslag af van het rekenkundig gemiddelde?

$$X_i - \bar{X}$$

Eigenschap:
De som van de afwijkingscores bedraagt 0; gemiddelde afwijkingscore dus ook

$$\sum_{i=1}^{k}\left(X_i - \bar{X}\right) = 0$$

Dus deze maat kunnen we niet gebruiken als maat van variabiliteit

voorbeeld:

Reeks 1	Reeks 2	$X_i - \bar{X}_1$	$X_i - \bar{X}_2$
9	2	-4	-11
11	5	-2	-8
11	7	-2	-6
12	9	-1	-4
12	10	-1	-3
13	15	0	2
14	16	1	3
15	20	2	7
16	22	3	9
17	24	4	11

$$\sum_{i=1}^{k}\left(X_i - \bar{X}\right) = 0$$

$$\bar{X}_1 = 13$$

$$\bar{X}_2 = 13$$

Reeks 1 : gemiddelde = 13 , -> 9-13 =-4 , 13-13 = 0, 17- 13 = 4 ,…

- Om de gemiddelde afwijkingsscore te bepalen moeten we de afwijkingsscores optellen, maar als we dat doen zien we echter dat we 0 uitkomen, dit zegt niets over de mate vd spreiding dus moeten we een andere oplossing vinden

5.2. variatiebreedte , bereik, range

De range of het bereik is de hoogste minus de laagste waarde.

Probleem?
Wordt enkel beïnvloed door twee uitslagen; we houden geen rekening met frequenties
Zinvol?
Niet zinvol, geen goede maat want het gaat hier voornamelijk over het verschil tussen hoogste en laagste scores dus dat word dan ook bepaald door deze scores.

- Voorbeeld 1: ordinaal niveau, bereik: zeer slecht tot goed (zeer goed heeft geen scores)

Oordeel	**Absolute frequentie**
Zeer slecht	15
Slecht	20
Neutraal	18
Goed	10
Zeer goed	00
TOTAAL	**63**

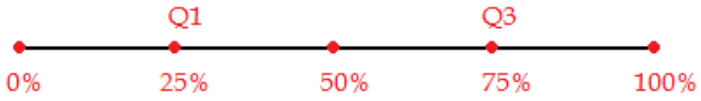
- Voorbeeld 2: interval/ratio niveau, bereik: 17 – 9 = 8 en 24 – 2 = 22

Reeks 1	Reeks 2
9	2
11	5
11	7
12	9
12	10
13	15
14	16
15	20
16	22
17	24

5.3. interkwartielafstand

- Het gebied op de X-as waartussen de middelste helft (50%) van alle waarnemingen valt, is de interkwartielafstand (Q).
  Welk is het verschil tussen Q_3 en Q_1?
  Biedt een goed inzicht in de variabiliteit van de uitslagen.
- Speelt geen rol in de inductieve statistiek, enkel in beschrijvende.

$$Q = P_{75} - P_{25} = Q_3 - Q_1$$



- Voorbeeld 1: ordinaal niveau

APS-SURVEY 2004: Hoogste diploma

Diploma	*Absolute Frequentie*	*Relatieve Frequentie*	*Cum. Rel. Freq*
Geen/LO	365	23,6%	23,6%
Lager secundair	324	21,0%	44,6%
Hoger secundair	506	32,7%	77,3%
Niet universitair HO	262	16,9%	94,2%
Universitair HO	89	5,8%	100,0%
TOTAAL	1546	100,0%	

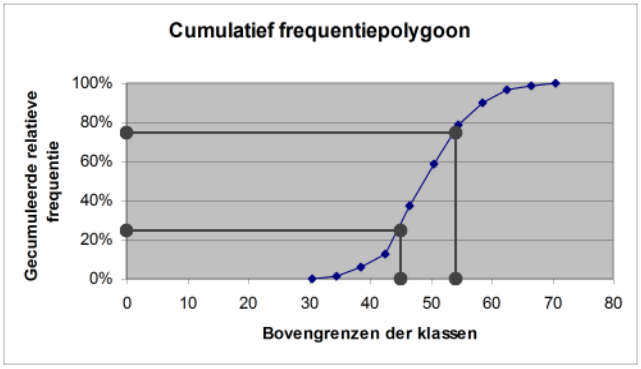
P75: hoger secundair

P25: lager secundair

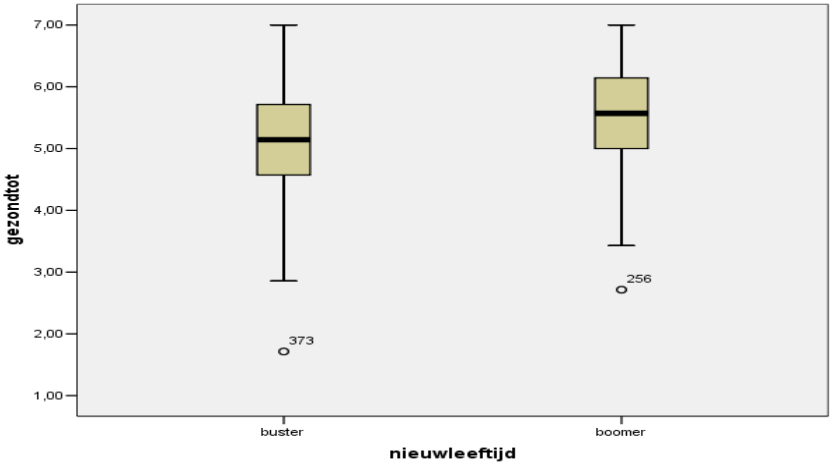
Q: HSO – LSO

- Grafische afleiding interkwartielafstand

Grafische afleiding interkwartielafstand

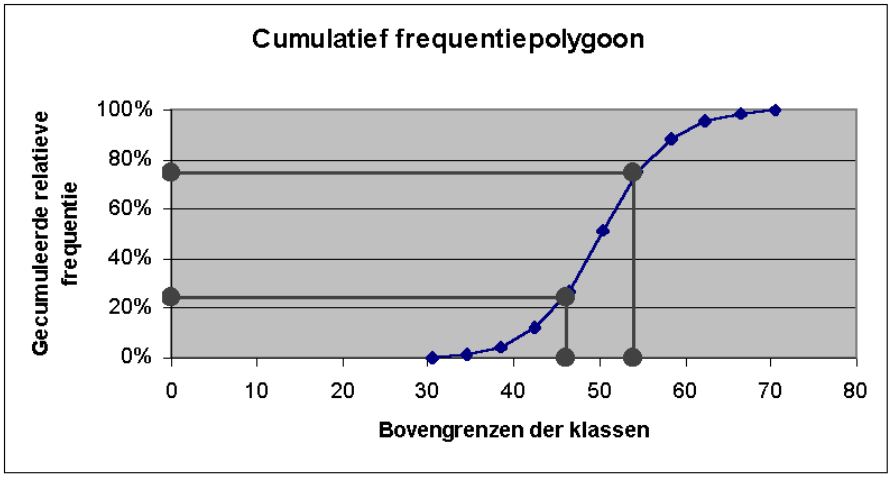


- Via de doos van de boxplot. Boxplot biedt grafische voorstelling van alle observaties. De doos van deze boxplots bevat de 50% middelste observaties en geeft derhalve een beeld van het interkwartielbereik

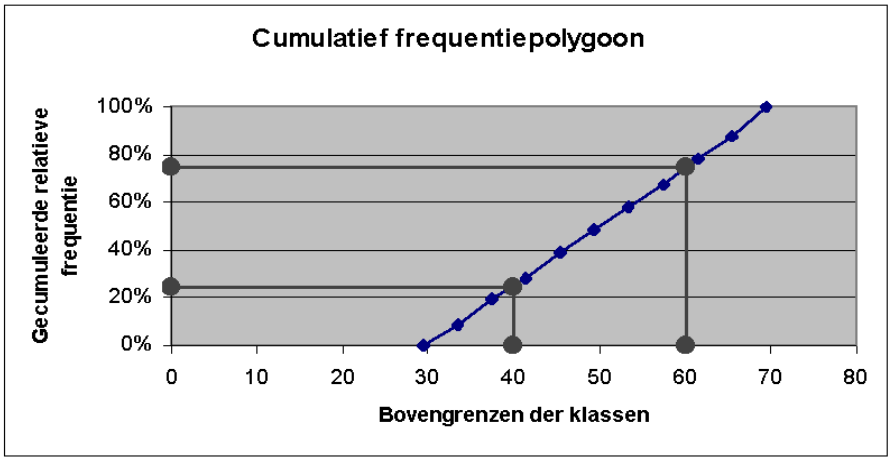


Interkwartiel afstand :

- Bij normaal verdeling Me= 50 Q=8



- Bij uniforme verdeling Me= 50 Q=20



5.4. gemiddelde absolute afwijking

- Bij het berekenen van een deviatiescore/afwijkingsscore zochten we een betere oplossing

  = we maken een rekenkundig gemiddelde van de ABSOLUTE (zonder tekens) waarden van de afwijking van elke score t.o.v. het gemiddelde

Bv;
X – gemiddelde = (absolute) afwijkingsscore
4 - 20 = 16
12 – 20 = 8
20 – 20 = 0
28 – 20 = 8
36 – 20 = 16
Gemiddelde: 9,6 = gemiddelde absolute afwijking

- Geeft een goed idee van de variabiliteit van de scores
- Maat wordt in de praktijk weinig gebruikt

(minnen vallen dus weg)

5.5. de variantie (s^2) !!

We gaan uit van het kwadraat van de afwijking t.o.v. het gemiddelde. Hiervan maken we het gemiddelde.
Aldus ontstaat de gemiddelde gekwadrateerde afwijkingsscore = variantie

VOORBEELD 1 :

Reeks 1	Reeks 2
9	2
11	5
11	7
12	9
12	10
13	15
14	16
15	20
16	22
17	24

Gemiddelde reeks 1 : 13 (alles optellen : 10 (aantal cijfers)
Gemiddelde reeks 2: 13 (“)

$X_i - \overline{X}_1$	$X_i - \overline{X}_2$
- 4	-11
-2	-8
-2	-6
-1	-4

-1	-3
0	2
1	3
2	7
3	9
4	11

Reeks 1 vakje 1 : 9 – 13 (= gemiddelde)= -4

reeks 2 vakje 1 : 2 -13 (= gemiddelde)= 11

$(X_i - \overline{X}_1)^2$	$(X_i - \overline{X}_2)^2$
16	121
4	64
4	36
1	16
1	9
0	4
1	9
4	49
9	81
16	121

de vorige tabel doen we nu tot de 2de waardoor het min teken ook zal wegvallen

hier het gemiddelde van nemen :

Reeks 1 : gemiddelde : (alles optellen : 10 (aantal cijfers)= 5,6

reeks 2 : gemiddelde : 51

(via spss zullen we dit doen) dus niet de handmatige formule kennen

5.6. de standaardafwijking (s)!!!

Standaardafwijking of standaarddeviatie is gelijk aan de wortel uit de variantie

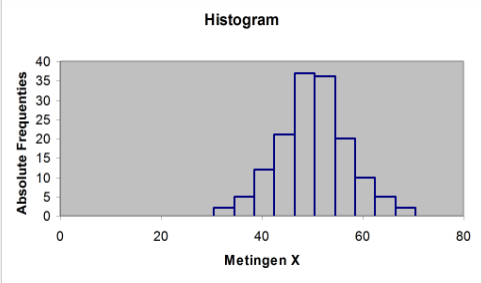
voorbeeldje :

	Reeks 1	Reeks 2	Reeks 3
	0	99	100
	100	100	100
	200	101	100
Gemiddelde $\bar{x}$	100	100	100
S (standaardafwijking)	81,65	0,82	0,00

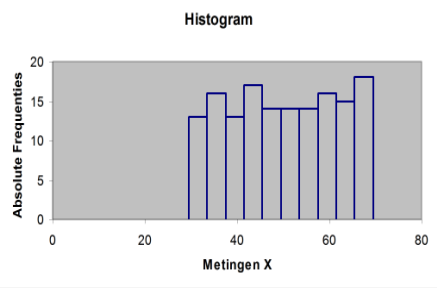
Standaardafwijking = 0 -> alle waarde zijn gelijk = reeks 3
standaardafwijking is tussen 0 en 1 -> kleine afwijking , kleine variatie= reeks 2
standaardafwijking is groter dan 1 -> grote variatie, grote afwijking = reeks 1

Normaalachtige verdeling
gemiddelde = 50,49

s = 6,84



Uniformachtige verdeling
gemiddelde = 50,25
s = 11,68



Opmerking :

- Standaardafwijking van de populatie (sigma: σ)
- Standaardafwijking van de steekproef (s)

Analoog voor de variantie; SPSS maakt altijd gebruik van N-1

Standaarddeviatie

- Standaarddeviatie en variantie zijn beiden steeds positief;
- Standaarddeviatie wordt steeds uitgedrukt in dezelfde eenheid als de scores; variantie als gekwadrateerde eenheid van de scores.
- Variantie wordt (samen met het rekenkundig gemiddelde) zeer veel gebruikt in de inductieve statistiek.

SPSS EN STANDAARD DEVIATIE -> zie SPSS

5.7. lineaire transformaties

Lineaire transformatie = alle waarden worden met bepaald getal gedeeld/vermenigvuldigd/opgeteld/afgetrokken

- Wat is het effect op het rekenkundig gemiddelde als elke waarde met b vermenigvuldigd wordt en a bij opgeteld?

$$\Rightarrow \bar{Y} = a + b \cdot \bar{X}$$

- Wat is het effect op de standaarddeviatie en de variantie van de nieuwe verdeling als elke waarde met b vermenigvuldigd wordt en er a bij opgeteld?

$$\Rightarrow s_Y = s_{a+bX} = |b| \cdot s_X$$

$$\Rightarrow s_Y^2 = s_{a+bX}^2 = b^2 \cdot s_X^2$$

DWZ : !!!!

- Als je alle waarden (X) met eender welke constante vergroot of verkleint, dan wordt het gemiddelde eveneens met die waarde vergroot of verkleind, maar dit heeft geen effect op de standaardafwijking, noch op de variantie.
- Als je alle waarden (X) met 3 vermenigvuldigt, dan wordt het gemiddelde 3 keer groter en de standaardafwijking ook 3 keer groter. De variantie wordt 9 keer groter.
- Als je alle waarden (X) met -3 vermenigvuldigt, dan wordt het gemiddelde met -3 keer vermenigvuldigd, maar de standaardafwijking wordt met 3 vermenigvuldigd. De variantie wordt 9 keer groter.

VOORBEELD:

X	Y = - 5 + 2X
9	13
11	17
11	17
12	19
12	19
13	21
14	23
15	25
16	27
17	29

Gemiddelde van de eerste tabel is

x = 13 (alles optellen : 10 (aantal cijfers))

Gemiddelde van de 2de tabel is

y = -5 + 2 . 13 = 21 (is het gemiddelde)

$(X_i - \overline{X})^2$	$(y_i - \overline{y})^2$
16	64
4	16
4	16
1	4
1	4
0	0
1	4
4	16
9	36
16	64

de eerste rij, het eerste vakje is

9 – 13 (het gemiddelde) = -4 en dit tot de 2de is 16

zo is dat met de hele tabel gedaan

als men hier het gemiddelde van neemt

= 5,6

bij de 2 de rij , het eerste vakje is

13 – 21 (gemiddelde van eerste tabel , 2de rij) = - 8,

-8 tot de 2de = 64

hier het gemiddelde van is : 22,4

Belangrijke transformaties !!!

Standaardscores of Z-scores geven aan hoeveel standaardafwijkingen een score van het gemiddelde ligt.

standaardscores:

Wat wordt het rekenkundig gemiddelde en variantie van deze Z-waarden?

- het gemiddelde zal altijd nul zijn
- de variantie en standaarddeviatie zullen altijd 1 zijn.

Deze omzetting noemen we standaardiseren.
Zinvolheid?

VOORBEELD standaardscores
Arie behaalt een 7 voor Frans, terwijl het gemiddelde 6 is en de standaarddeviatie 2
Arie behaalt voor Engels 6, terwijl het gemiddelde 5 was en de standaarddeviatie 1,5 bedroeg.
Welke prestatie is het beste?
$Z_F = (7 - 6)/2 = 0{,}50$

$Z_E = (6 – 5)/1{,}5 = 0{,}67$

Dus de prestatie voor Engels is beter dan deze voor Frans.

- Z-waarden zijn een dimensieloos getal, en kunnen zowel positief als negatief zijn.
- Een negatieve Z-waarde betekent dat deze uitslag zich links van het gemiddelde bevindt.
- Een positieve Z-waarde betekent dat deze uitslag zich rechts van het gemiddelde bevindt.

- De uitslagen kunnen omgezet worden in Z-waarden, maar ook omgekeerd, indien we het rekenkundig gemiddelde hebben en de standaarddeviatie van de oorspronkelijke gegevens kunnen we elke Z-waarde terug plaatsen in de oorspronkelijke verdeling.

VOORBEELD

Score voor taal	Score voor rekenen
$X_i = 84$	$Y_i = 90$
$\overline{X} = 76$	$\overline{Y} = 82$
$s_X = 10$	$s_Y = 16$

Welke van beide scores 84 of 90 is de beste prestatie voor deze persoon?

$$z_i = \frac{84-76}{10} = 0,8 \qquad z_i = \frac{90-82}{16} = 0,5$$

- Uit de Z-waarden kunnen de oorspronkelijke scores bepaald worden.

$$z_i = \frac{X_i - \overline{X}}{s_X} \qquad \Rightarrow X_i = \overline{X} + z_i \cdot s_X$$

$z_i = 0,8$

$\overline{X} = 76$

$s_X = 10$

$$\Rightarrow X_i = 76 + 0,8 \cdot 10 = 84$$

Score voor taal is 84.

- Standaardscores kunnen omgezet worden naar een verdeling met een bepaald gemiddelde, en dat via lineaire transformatie
  Omzetting naar T scores met gemiddelde van 50 en een standaarddeviatie van 10
  Of een omzetting naar C scores, met gemiddelde 5 en standaarddeviatie van 2.
  Hoe?

- Omzetting van een score voor taal via de Z-waarde in een T-score (met een gemiddelde van 50 en een s van 10)

- Omzetting in C scores (gemiddelde is 5 en s gelijk aan 2)?

-

- Let wel: het standaardiseren heeft geen effect op de vorm van de verdeling, m.a.w. scheef blijft scheef, symmetrisch blijft symmetrisch.
  Enkel het gemiddelde wordt nul en de standaarddeviatie wordt één.

  --> SPSS spss en standaardscores

5.8. slotopmerking

- De s kan, naast het rekenkundig gemiddelde gebruikt worden om groepen met elkaar te vergelijken
- De grootte van de s hangt ook af van het gemiddelde. Oplossing: de variatiecoëfficiënt $V = \frac{s_X}{\overline{X}}$

Door deze coëfficiënt wordt de s gecorrigeerd voor het gemiddelde. Variatie wordt dan gezien als % van gemiddelde
Let op: enkel bij ratio niveau van meting.

Variatiecoëfficiënt

Samenvatting :

Meetniveau	Centrummaat	Spreidingsmaat	Andere
Nominaal	**Modus**		
Ordinaal	Modus **Mediaan**	**Interkwartielafstand**	
Interval	Mediaan **Gemiddelde**	Bereik Interkwartielafstand **Standaardafwijking**	Gemiddelde absolute afwijking
Ratio	Mediaan **Gemiddelde**	Bereik Interkwartielafstand **Standaardafwijking**	Gemiddelde absolute afwijking

5.9. besluit

* Een spreidingsmaat geeft aan hoe de scores verschillen t.o.v. het rekenkundig gemiddelde
- variatiebreedte (range)
- interkwartielafstand
- gemiddelde absolute afwijking

- variantie
- standaardafwijking

* Een lineaire transformatie heeft geen invloed op de vorm van de verdeling, maar wel op het nieuwe rekenkundige gemiddelde en op de standaarddeviatie.

* Meest gebruikte transformatie is de omzetting in Z-waarden

OPGAVE 1
Van een frequentieverdeling worden alle scores door 4 gedeeld. Wat gebeurt er met de standaarddeviatie en met de variantie?

Antwoord
s wordt door 4 gedeeld
variantie gedeeld door 16

Van een verdeling van uitslagen worden alle scores verhoogd met 4. Wat gebeurt er met de standaarddeviatie en variantie?

Antwoord?
Blijft dezelfde

OPGAVE 2
Indien je bij een examen 130 hebt behaald, bij welke situatie heb je dan het beste resultaat? En het slechtste?
Gem=100 en s=15
Gem=100 en s=30
Gem=90 en s=15
Gem=90 en s=20

Standaardscores : individuele scores :
130 – 100 : 15 = 2
130 – 100 : 30 = 1 -> slechtste
130 – 90 : 15 = 2,7 -> beste
130 – 90 : 20 = 2

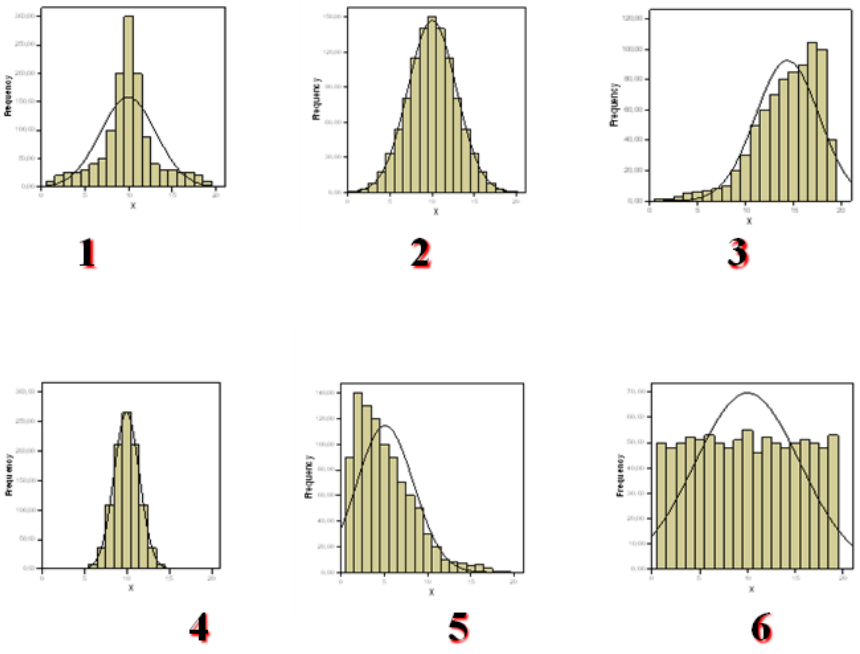
OPGAVE 3
A) Een student haalt een C-score van 7. Met welke IQ-score (met gemiddelde 100 en s gelijk aan 15) komt dit overeen? (opl : IQ =115)

B) Een student haalt een T-score van 70. Met welke IQ-score (met gemiddelde 100 en s gelijk aan 15) komt dit overeen? (opl. IQ 130)

Welk histogram hoort bij welke berekeningen?

	V1	V2	V3	V4	V5	V6
Mean	9,93	10,00	14,37	10,00	5,10	10,00
Median	10,00	10,00	15,00	10,00	4,00	10,00
Mode	10,00	10,00	17,00	10,00	2,00	10,00
Std. Deviation	3,20	2,86	3,32	1,49	3,28	5,48
Variance	10,27	8,16	11,02	2,22	10,75	29,98
Range	18,00	18,00	18,00	10,00	18,00	18,00
Percentiles 25	9,00	8,00	12,00	9,00	3,00	5,00
Percentiles 75	11,00	12,00	17,00	11,00	7,00	15,00

3 modus hoog , links scheve verdeling
6 variabiliteit = hier het grootste



Figuur 4 : kleine range (zo heel op elkaar)
figuur 1 : interkwartielafstand is kleiner als op figuur 2

HOOFDSTUK : 6 : de normaalverdeling

Doelstellingen:
De student kent de eigenschappen van een normaal verdeling;
De student kent de standaardnormaal verdeling en kan omgaan met Z-waarden;
De student kan – in een normale verdeling - het verband leggen tussen proporties en Z-waarden en omgekeerd;
De student kan deze principes toepassen in om het even welke normaalverdeling.

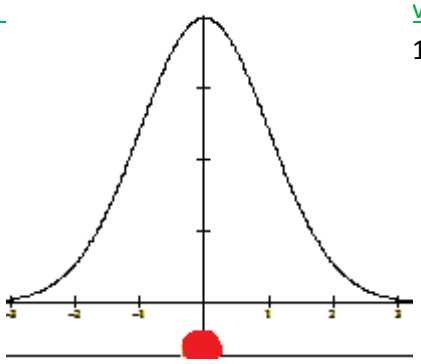
6.1. de normale verdeling

Een intervalwaarde die afhankelijke is van een oneindig aantal factoren, die los van elkaar inwerken, zal in de populatie een normaalverdeling vertonen (Gausscurve); bv. Intelligentie.

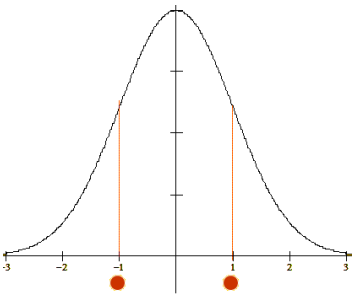
Gemiddelde in steekproef: $\bar{X}$
In μ bereikt de Gausscurve zijn maximum
Gemiddelde in de populatie: μ
Standaarddeviatie in de steekproef: s
Standaarddeviatie in de populatie: σ
De twee buigpunten van de Gausscurve bevinden zich op μ - σ en op μ + σ

Kenmerken van de normale verdeling

1) Dergelijke verdeling heeft 1 maximum
MO = Me = Gemiddelde



2) 2 Buigpunten

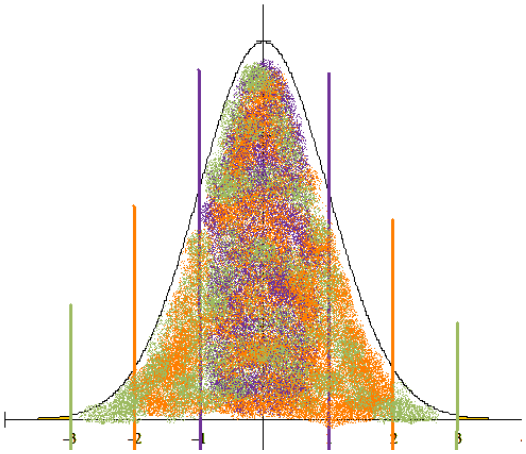


3) Symmetrie
De oppervlakte links en rechts van het gemiddelde zijn gelijk

4) De grafische weergave van de normale verdeling is klokvormig;
De uitslagen liggen vooral geconcentreerd rond het gemiddelde, naarmate scores afwijken t.o.v. het gemiddelde wordt de frequentie kleiner.

5) Als we van een normaalverdeling het gemiddelde en de standaarddeviatie kennen, is deze verdeling gedefinieerd

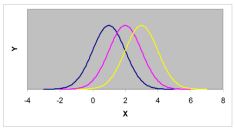
Verdeling in de Gausscurve !!!



* Tussen beide buigpunten bevinden zich (zie GRAFIEK) +/- 68% van de observaties;

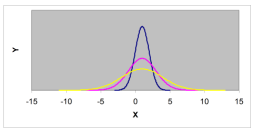
* Tussen μ - 2 σ en μ + 2 σ situeren zich +/- 95% van de observaties; (zie GRAFIEK)

* Tussen μ - 3 σ en μ + 3 σ bevinden zich +/- 99% van alle waarnemingen. (zie GRAFIEK)



verschillende μ, zelfde σ

Verschillende σ, zelfde μ



De speciale normaal verdeling
Een normale verdeling met gemiddelde nul en standaarddeviatie 1, noemen we een standaardnormaal verdeling

De standaard normaal verdeling

Vanuit de eigenschappen van de normale verdeling kunnen we vaste relaties vinden tussen de proportie van uitslagen en Z-waarden.

Welk is de kans om in een standaardnormaal verdeling een uitslag te vinden die groter of kleiner is een bepaalde Z-waarde?

Tabel van de standaard normale verdeling.

Voorbeeld: P(-∞<Z<1,77) = 0,9616

P(-∞<Z<z)	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7423	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7643	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8213	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9773	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9923	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9963	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000

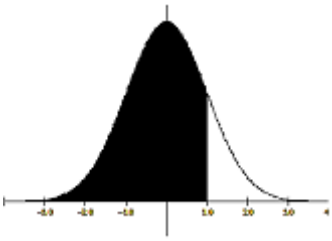
Zie apart blad, om af te printen, tabel voor oefeningen op te lossen vrij makkelijk

Gevallen te berekenen met de tabel

- Z is positief, hoeveel **% vd scores verwachten we beneden deze Z-waarde?**

  Percentage aflezen uit de tabel: Bv, Z=1

  P (Z<1)=0,8413



P(-∞<Z<z)	0,00	
0,0	0,5000	0,5
0,1	0,5398	0,5
0,2	0,5793	0,5
0,3	0,6179	0,6
0,4	0,6554	0,6
0,5	0,6915	0,6
0,6	0,7257	0,7
0,7	0,7580	0,7
0,8	0,7881	0,7
0,9	0,8159	0,8
1,0	0,8413	0,8
1,1	0,8643	0,8

- Hoeveel % van de observaties verwacht u in een normaalverdeling beneden Z=1,59

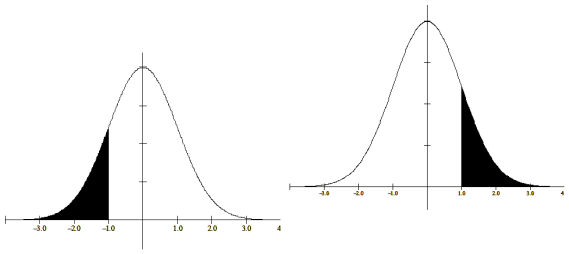
  P (Z<1,59)=0,9441

- De z-waarde is negatief, hoeveel % van de observaties liggen in de standaardnormaal verdeling onder de Z-waarde -1?

  P(Z<-1)= 1 – P(Z<1) = 0,1587

  Eerst kans bekijken kleiner dan 1 want is tegengestelde

  P(Z<1)= 0,8413



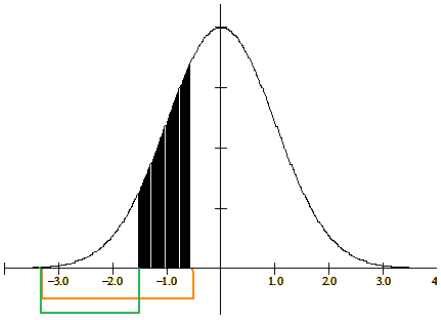
- Hoeveel % van de uitslagen situeren zich ie standaardnormaal verdeling beneden Z=-1,16?

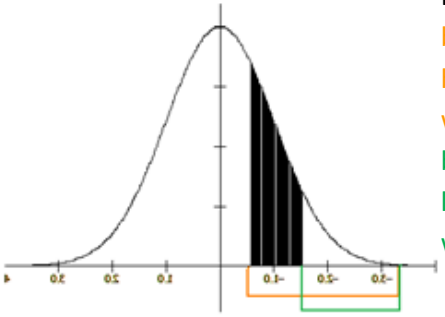
  Eerst kans bekijken kleiner dan 1,16 want is tegengestelde

  P(Z<1,16)= 0,8770

  P(Z<-1,16)= 1-P(Z<1) = 0,123, dus 12%

- **Hoeveel observaties situeren zich tussen 2 Z-waarden:**



- 

Principe: % beneden de hoogste waarde MIN % beneden de laagste waarde

- Beide Z-waarden zijn positief: Hoeveel % vd uitslagen verwacht je in een standaardnormaal verdeling tussen Z=0,5 en Z=1,5

  P(0,5<Z<1,5)

  = P(Z<1,5) – P(Z<0,5) = 0,9332 – 0,6915 = 0,2417

- Beide Z-waarden zijn negatief: Bepaal oppervlakte tussen Z=-1,53 en Z=-0,56

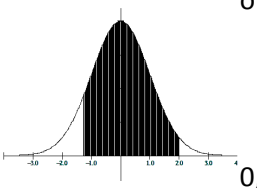
  P(-1,53<Z<-0,56)

= P(Z<-0,56) – P(Z<-1,53) = [1-0,7123] – [1-0,9370] = 0,2247

- 1 vd Z-waarden is negatief, ander positief: bepaal oppervlakte tussen Z=-1,26 en Z= 2,01

  P(-1,26<Z<2,01)

  = P(Z<2,01) – P(Z<-1,26) = 0,9778 – [1-0,8962] = 0,8740



Opdrachten nog doen?

Hoofdstuk 7 : vorm van de verdeling en invloed van transformaties

Doelstelling :
Na de studie van dit hoofdstuk…
kent u de betekenis van scheefheid en kurtosis van een verdeling;
kunt u een boxplot via SPSS maken en interpreteren;
begrijpt u het effect van transformaties op de vorm van de verdeling

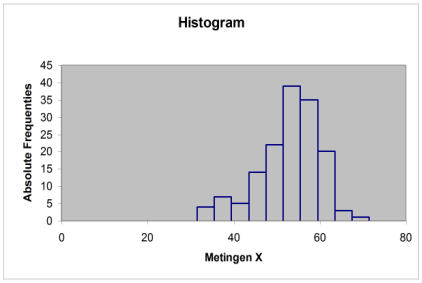
Veel frequentieverdelingen hebben niet de vorm van de normaal verdeling.
Ze vertonen niet de gelijkmatige welving van de klokfunctie, of zijn scheef.

Skewness of scheefheid (skw) en welving of kurtosis (kur)(gepletheid) kunnen via SPSS berekend worden. Deze skw en kur dient gerelateerd te worden aan de betreffende standaardfout. (de uitslag dient tweemaal zo groot te zijn om betekenisvol te zijn) dus de skewness is betekenisvol als ze 2X zo groot is als de kurtosis.

Formule niet kennen, maar via spss kunnen berekenen en op blad kunnen interpreteren

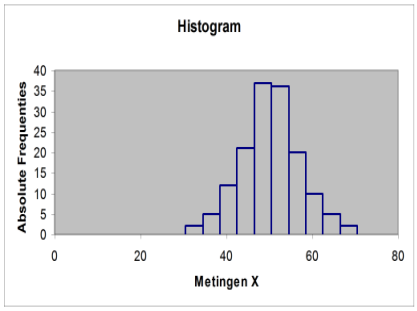
Scheefheid van de verdeling = SKEWNESS

1) skewness is kleiner dan 0 , skw < 0
wijst op een links scheve verdeling
bv. score voor een test met heel gemakkelijke items (zgn. plafondeffect)



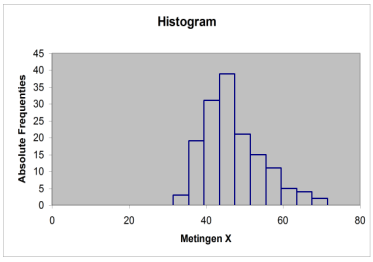
Skw = - 0,75

2) Skw = 0 wijst op een symmetrische verdeling



Skw = 0,02

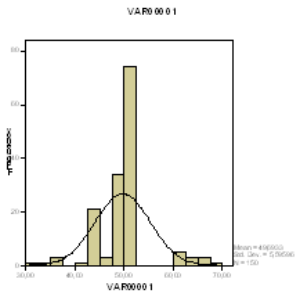
3) Skw > 0 wijst op een rechts scheve verdeling
bv. de scores op een test bestaande uit veel te moeilijke items (zgn. vloereffect)



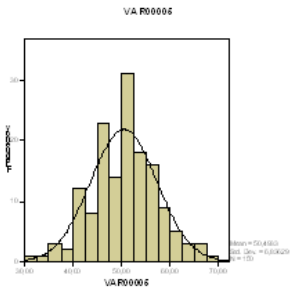
skw =0,75

KURTOSIS (welving, gepletheid, hoge piek?lage piek?)

1) Kurt > 0 wijst op een – in vergelijking met normaalverdeling - scherpe top

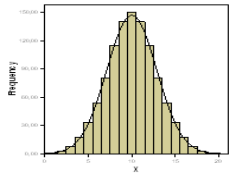


Kurt = 3.45

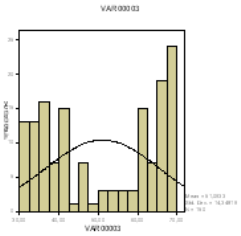


Kurt = 0,30

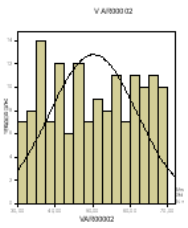
2) Kurt = 0 wijst op een welving die vergelijkbaar is met de normaalverdeling



3) Kurt < 0 wijst op een afgeplatte top



Kurt = - 1,69



Kurt = - 1,25

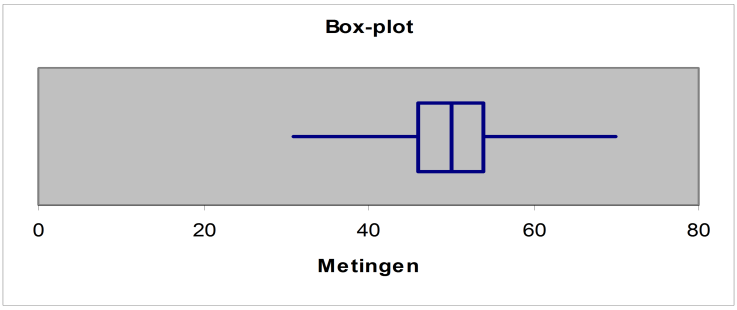
De boxplot

In de boxplot wordt in een doos de mediaan, het 25ste en het 75ste percentiel geplaatst, waardoor de doos in feite het interkwatielafstand voorstelt. Daarnaast worden de extreme en uiterst extreme waarden (=uitbijter) afgebeeld.

Uitbijter ligt op meer dan 1,5 dooslengte van het 25ste of 75ste percentiel

--> hoger of lagere scores dan de meeste, het einde van de streep + 1,5 doos lengte, of in het begin – 1,5 dooslengte

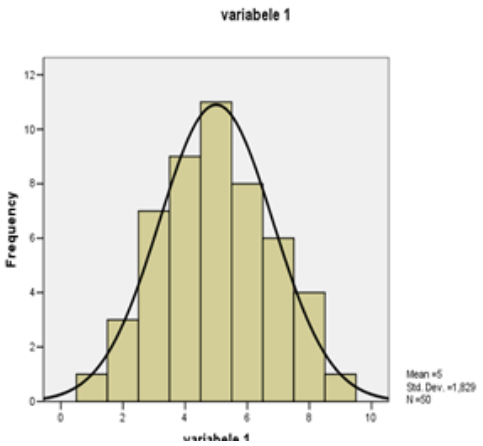
--> zelfde als hier boven ma dan ipv 1,5 doos lengte , 3 dooslengtes

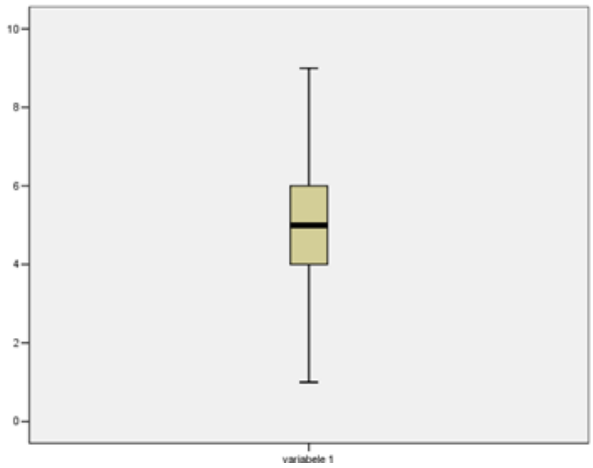


Variabele 1

variabele 1

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	1	2,0	2,0	2,0
	2	3	6,0	6,0	8,0
	3	7	14,0	14,0	22,0
	4	9	18,0	18,0	40,0
	5	11	22,0	22,0	62,0
	6	8	16,0	16,0	78,0
	7	6	12,0	12,0	90,0
	8	4	8,0	8,0	98,0
	9	1	2,0	2,0	100,0
	Total	50	100,0	100,0	



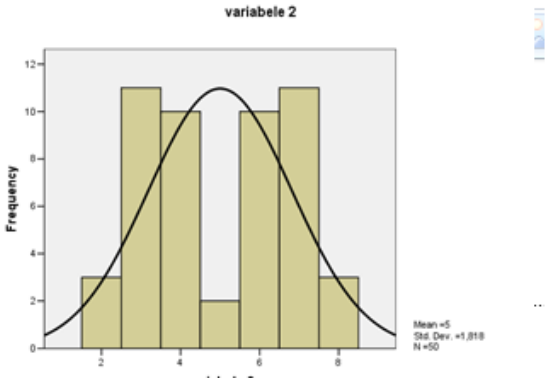


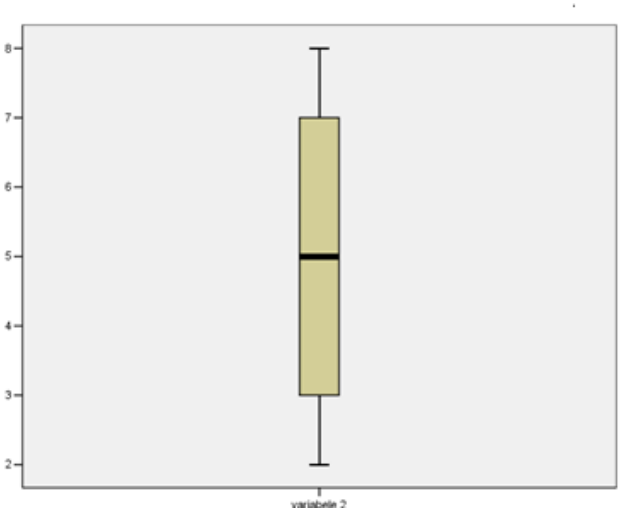
		Statistic	Std. Error
variabele	Mean	5,00	,259
	95% Confiden Lower Bound	4,48	
	Interval for Me Upper Bound	5,52	
	5% Trimmed Mean	5,00	
	Median	5,00	
	Variance	3,347	
	Std. Deviation	1,829	
	Minimum	1	
	Maximum	9	
	Range	8	
	Interquartile Range	2	
	Skewness	,062	,337
	Kurtosis	-,508	,662

Variabele 2

variabele 2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2	3	6,0	6,0	6,0
	3	11	22,0	22,0	28,0
	4	10	20,0	20,0	48,0
	5	2	4,0	4,0	52,0
	6	10	20,0	20,0	72,0
	7	11	22,0	22,0	94,0
	8	3	6,0	6,0	100,0
	Total	50	100,0	100,0	



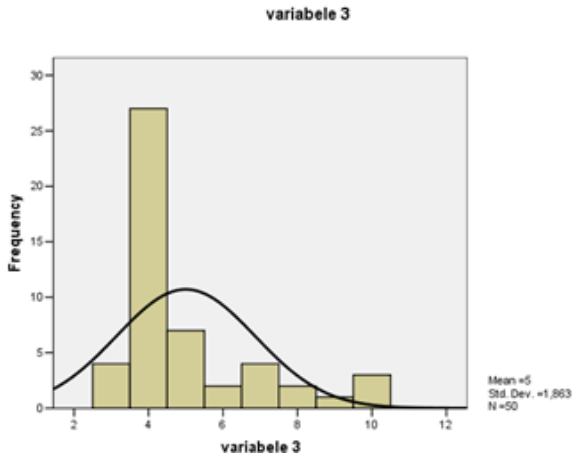


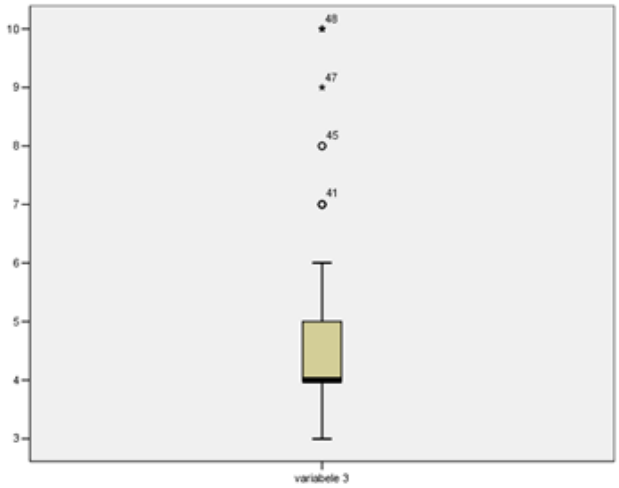
variabele 2	Mean		5,00	,257
	95% Confidence Interval for Mean	Lower Bound	4,48	
		Upper Bound	5,52	
	5% Trimmed Mean		5,00	
	Median		5,00	
	Variance		3,306	
	Std. Deviation		1,818	
	Minimum		2	
	Maximum		8	
	Range		6	
	Interquartile Range		4	
	Skewness		,000	,337
	Kurtosis		-1,382	,662

Variabele 3 vloereffect

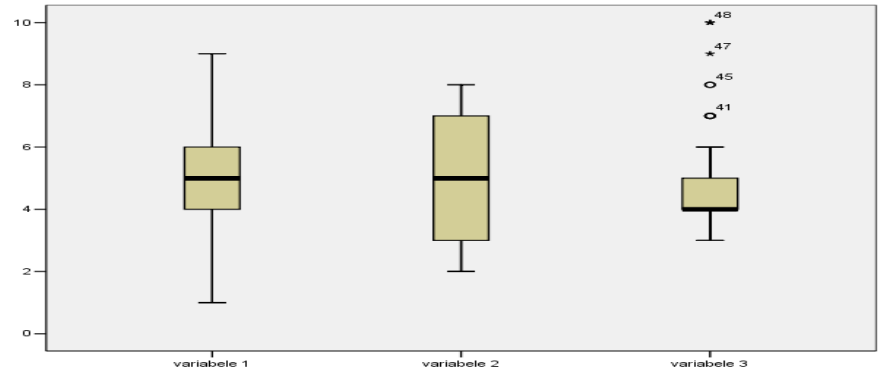
variabele 3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	3	4	8,0	8,0	8,0
	4	27	54,0	54,0	62,0
	5	7	14,0	14,0	76,0
	6	2	4,0	4,0	80,0
	7	4	8,0	8,0	88,0
	8	2	4,0	4,0	92,0
	9	1	2,0	2,0	94,0
	10	3	6,0	6,0	100,0
	Total	50	100,0	100,0	

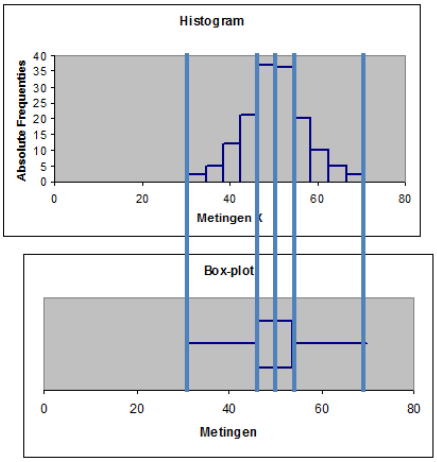




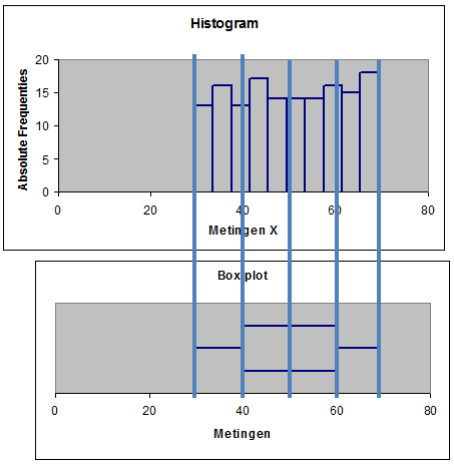
variabele 3	Mean		5,00	,263
	95% Confidence Interval for Mean	Lower Bound	4,47	
		Upper Bound	5,53	
	5% Trimmed Mean		4,83	
	Median		4,00	
	Variance		3,469	
	Std. Deviation		1,863	
	Minimum		3	
	Maximum		10	
	Range		7	
	Interquartile Range		1	
	Skewness		1,540	,337
	Kurtosis		1,503	,662



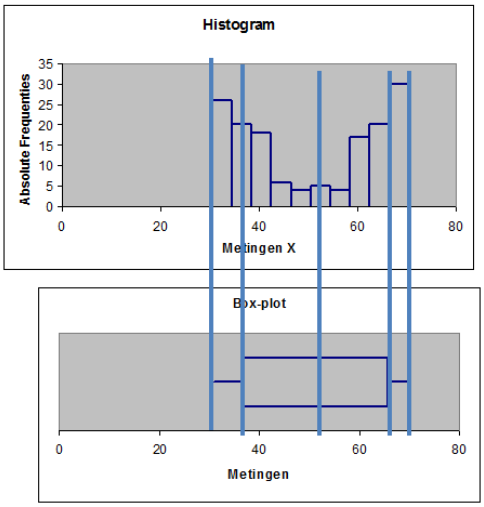
Normaal verdeling



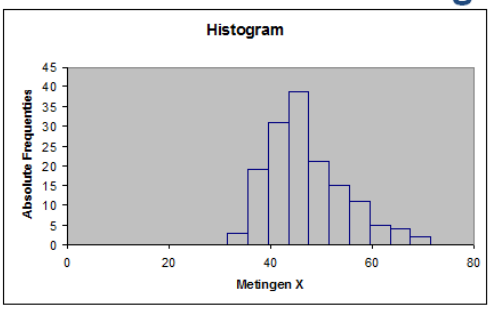
Uniforme verdeling

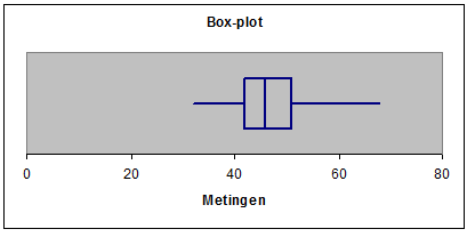


U-vormige verdeling

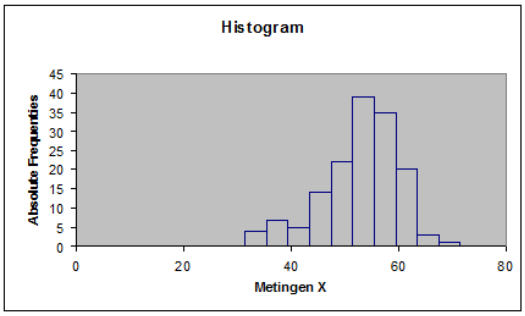


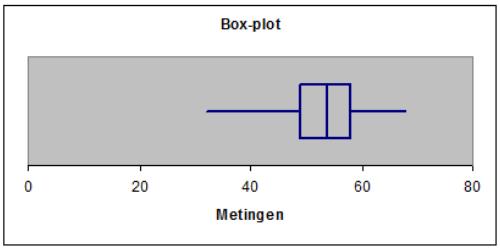
Rechtsscheve verdeling

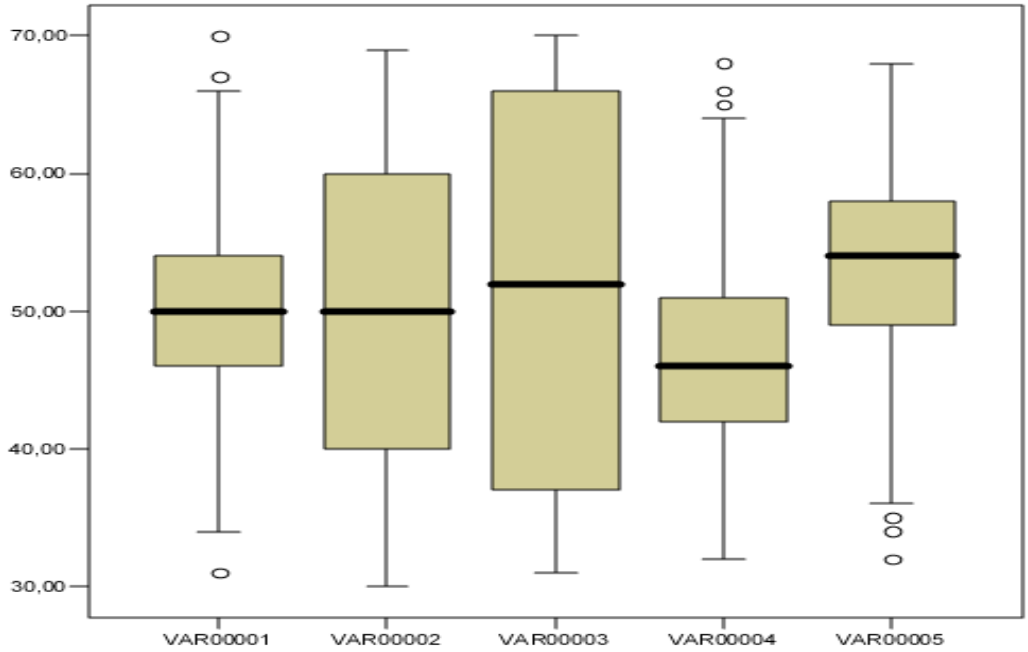




Linksscheve verdeling







Enkele vragen

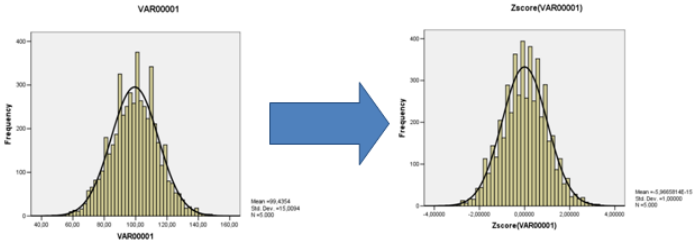
- Bij welke variabele is de mediaan het grootst?
  variabele 5

- Welke variabele heeft de grootste interkwartielafstand?
  variabele 3
- Welke variabele(n) is (zijn) ongeveer linksscheef?
  variabele 5
- Welke variabele(n) is (zijn) ongeveer rechtsscheef?
  variabele 4
- Welke variabele(n) is (zijn) ongeveer symmetrisch?
  variabele : 1,2,3
- Welke variabelen hebben outliers?
  variabele : 4,5,1

Lineaire transformaties : Y = a + b X

- Het gemiddelde wordt op dezelfde wijze getransformeerd.
- De standaarddeviatie wordt met |b| vermenigvuldigd, de variantie met b^2.
- De 'skewness' blijft onveranderd indien b>0.
- De kurtosis blijft onveranderd.

Omzetting in Z waarden
(zie hoofdstuk 6 bij normaal verdeling omzetting in Z-waarden bij IQ)
--> (de X (wat gegeven wordt) – het gemiddelde) / sigma

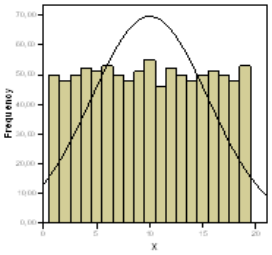


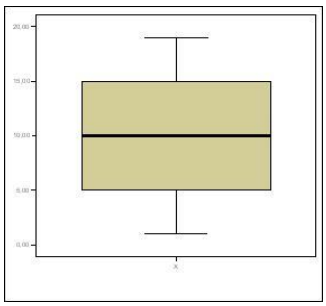
Statistics

		VAR00001	Zscore(VAR00001)
N	Valid	5000	5000
Mean		99,4354	,0000000
Median		99,0000	-,0290085
Mode		96,00	-,22888
Std. Deviation		15,00940	1,0000000
Variance		225,282	1,000
Skewness		,007	,007
Kurtosis		,034	,034
Range		107,00	7,12887
Minimum		43,00	-3,76000
Maximum		150,00	3,36886
Percentiles	25	89,0000	-,6952576
	75	110,0000	,7038656

<u>Heeft een omzetting in Z-waarden een invloed op de vorm van de verdeling?</u>

- Deze omzetting heeft GEEN invloed op de scheefheid en kurtosis van de verdeling; m.a.w. scheef blijft scheef.
- Deze omzetting heeft wel een invloed op het rekenkundig gemiddelde (altijd nul) en de s (altijd 1).

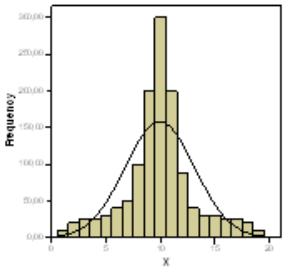


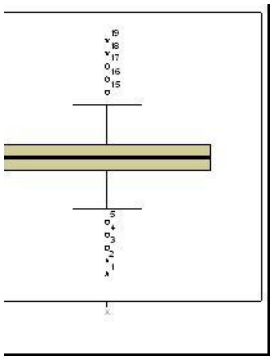


	V1	V2	V3	V4	V5	V6
Mean	9,93	10,00	14,37	10,00	5,10	10,00
Median	10,00	10,00	15,00	10,00	4,00	10,00
Mode	10,00	10,00	17,00	10,00	2,00	10,00
Std. Deviation	3,20	2,86	3,32	1,49	3,28	5,48
Variance	10,27	8,16	11,02	2,22	10,75	29,98
Skewness	0,04	0,00	-0,94	0,00	1,07	0,01
Kurtosis	1,10	-0,05	0,89	-0,04	1,21	-1,20
Range	18,00	18,00	18,00	10,00	18,00	18,00
Percentiles 25	9,00	8,00	12,00	9,00	3,00	5,00
Percentiles 75	11,00	12,00	17,00	11,00	7,00	15,00

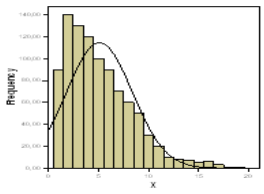
V1 :

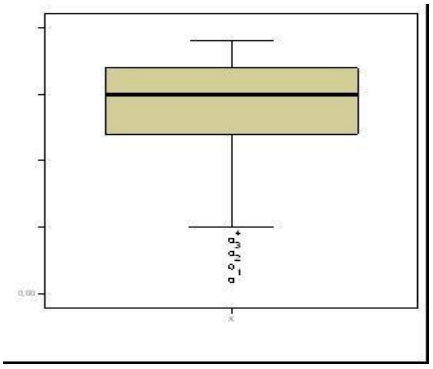
V2 :



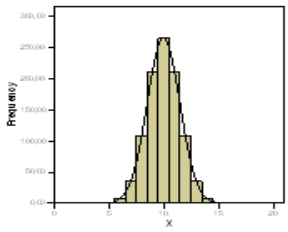


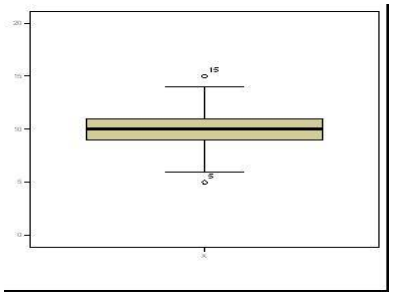
V3 :



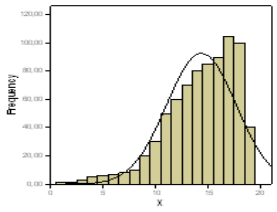


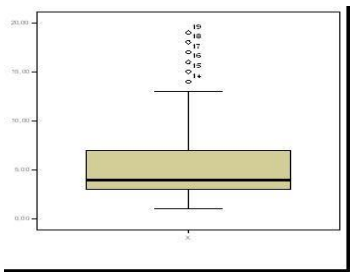
V4 :

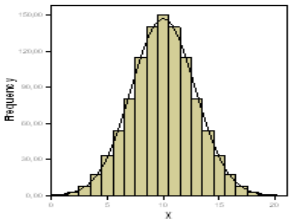


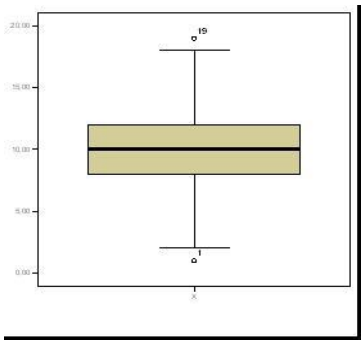


V5:









V6:

oefeningen

De gemiddelde afwijkingsscore ten opzichte van het gemiddelde is groter dan nul.

juist

De modus, de mediaan en het gemiddelde van deze verdeling zijn ongelijk aan elkaar.

juist

Het gemiddelde van deze verdeling is kleiner dan de mediaan.

fout

De modus van deze verdeling is kleiner dan de mediaan

juist (bij een positief scheve verdeling = rechtsscheef)

Van een verdeling is gegeven: gemiddelde = 26, mediaan = 22, en modus = 18, dit betekent dat:

rechtsscheve verdeling

Van een verdeling is gegeven: gemiddelde = 11, mediaan = 15 en modus = 18. Dit betekent dat:

links scheve verdeling

Welke van de volgende drie centrummaten is gevoelig voor uitbijters?

Het gemiddelde. nee

De mediaan. ja

De modus. nee

Een verdeling is links-scheef (dus top naar rechts).
Wat is dan de volgorde, qua grootte, van de drie maten van centrale tendentie?
Mo = 17, Me = 15, Gem = 13.

Na de lineaire transformatie Y = 2X + 10 is het nieuwe gemiddelde gelijk aan 100 en de nieuwe standaarddeviatie gelijk aan 10.
Wat waren het gemiddelde en de standaarddeviatie vóór de transformatie?
X = 45, S = 5.

een verdeling is de kurtosis 1,60 en de skewness = -1,40. De standaardfouten zijn respectievelijk 0,40 en 0,30.
Wat kan men opmerken over de vorm van de verdeling ten opzichte van de normale verdeling?
gepiekt en scheef naar links

DEEL 2 : BIVARIATE ANALYSE (2 (OF MEER) VARIABELE)

Hoofdstuk 8 : kruistabellen

Analyse van de samenhang

hoofdstuk

Vereiste: ***twee*** uitslagen per persoon

- Twee nominale variabelen stellen we voor in een kruistabel. Analyse met Chi-kwadraat en de associatiematen. 8
- Twee interval variabelen: gebruik de Pearson correlatiecoëfficiënt en de regressietechniek. 9
- Twee ordinale variabelen: gebruik de correlatiecoëfficiënt van Spearman. 10

dit is belangrijk : bij oef -> juiste oplossing aan de hand hiervan

Zeer belangrijk

Analyse van de kruistabel

- Bestaat er een betekenisvol verband tussen twee nominale variabelen?
  Gebruik de Chi-kwadraat test van onafhankelijkheid.

- Hoe sterk is dit verband?
  Phi-coëfficiënt (enkel in een vierveldentabel)
  Contingentiecoëfficiënt (kan nooit 1 worden)
  Cramér's V

Doelstellingen:

- De student kan een kruistabel opstellen
- De student kan de samenhang in een kruistabel onderzoeken

  - bestaat er een significant verband tussen twee nominale variabelen?
  - hoe sterk is dit verband?
- De student kan deze analyse uitvoeren zowel handmatig als via SPSS. De student kan SPSS output lezen en de resultaten interpreteren.

Voorbeeld van een onderzoek.

Bestaat er een verband tussen het al dan niet roken en het voorkomen van hart- en vaatziekten?

Operationalisering?

Wie gaan we bevragen?

Onafhankelijke/afhankelijke variabele?

Hoe samenhang berekenen?

Verband tussen roken en hart- en vaatziekten

1)

- Veronderstel een perfect verband:

	Hart- en vaatziekten ja	neen
• Roker	20	0
• Niet roker	0	100

Probleemstelling : - wat is roken (1 sigaret per dag? Een pakske? Discussie mogelijk)

- wat is precies hart- en vaatziekten

wie bevragen ?

onafhankelijke variabelen : roken / niet roken

afhankelijke variabelen : hart- en vaatziekten ja of nee?

2)

- Veronderstel dat er geen verband bestaat

	Hart- en vaatziekten		
	Ja	neen	
• Roker	1	19	20
• Niet roker	5	95	100
• Totaal:	6	114	120

feitelijke observaties

	Hart- en vaatziekten		
	Ja	neen	
Rokers	5	15	20
Niet rokers	1	99	100
Totaal:	6	114	120

Welke zijn de celfrequenties? Welke de randtotalen/marginale totalen?

voorstelling via percentages

- In functie van het totaal aantal ppn.
  Dit heeft geen zin
- Verticaal percenteren?
- Horizontaal percenteren?

  Maak een keuze in de richting van de onafhankelijke variabele.

Foutieve manier van voorstelling

	hart- en vaatziekten	
	ja	neen
Rokers	83%	13%
Niet rokers	17%	87%
Totaal	100%	100%

juiste voorstelling van zaken

	Hart- en vaatziekten		
	ja	neen	
Rokers	25%	75%	100%
Niet rokers	1%	99%	100%

Let op in welke richting percenteren.

vertrekt van de onafhankelijke variabelen

Samenhang tussen twee nominale variabelen

Is het verband betekenisvol?

Chi-kwadraat waarde: welk is de afstand tussen de geobserveerde waarden en de verwachte waarden in de veronderstelling dat er geen verband zou zijn. Is dit significant? --> dus ja of nee (een chi-kwadraat zegt enkel of er verband is , of er samenhang is ja of nee, meer ni, de sterkte word gemeten met : associatiematen is zwak of sterk)

0-hypothese : gaan ervan uit dat er geen samenhang is
alternatieve hypothese : er is wel samenhang
alternatieve zal 0-hypothese onderuit halen?

De chi-kwadraat

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

f_o : geobserveerde celfrequentie

f_e : verwachte celfrequentie

Een voorbeeld

- Er bestaan slechts drie politieke partijen:
  NOTAX, NOVA, en Nieuw groen.
  Welk is uw voorkeur?
- We bevragen de houding t.o.v. een belastingsvermeerdering voor milieuonvriendelijke producten. (voor/weet niet/tegen)

	Voor	Weet niet	Tegen	Totaal
Nieuw Groen	12	5	8	25
NOTAX	5	12	23	40
NOVA	12	7	1	20
Totaal:	29	24	32	85

- Bestaat er een betekenisvol verband?
  Gebruik de Chi-kwadraat waarde
  Opstellen van de verwachte frequenties
  Bereken de Chi-kwadraat waarde
  Test de significantie

- betekenis vol verband?

berekening : Fe
Voor de bepaling van de verwachte frequenties dient u het product van de overeenkomstige randtotalen te delen door het totaal aantal ppn.

$$\frac{29 \times 25}{85} = 8{,}53$$

alles zo vermenigvuldigen, dus waar het cijfer 7 staat zal ik de berekening : (24X20)/85,
dan kom je het volgende uit (zie vb Fe)
voorbeeld Fe

	Voor	Weet niet	Tegen
Nieuw Groen	8,53	7,05	9,41
NOTAX	13,65	11,29	15,05
NOVA	6,82	5,64	7,53

daarna komt volgende berekening

$$\frac{(12 - 8{,}53)^2}{8{,}53} = 1{,}41$$

dus de 12 komt van de eerste tabel , de 8,53 komt van de 2de tabel.
vb ; bij dat cijfertje 7 in de eerste tabel zou het zijn {(7-5,64)²/5,64}
dan alles uitgeteld , dit alles optellen :

Deze waarde is gelijk aan: 1,41 + 0,60 + 0,21 + 5,48 + 0,04 + 4,20 + 3,93 + 0,33 + 5,66 = 21,86

--> is dat cijfer groot genoeg?, wat is de maatstaf?-> rekening houdend met de grootte van de tabel

Toets vervolgens of deze waarde significant is, rekening houdend met het aantal vrijheidsgraden.
Vrijheidsgraden(= vrij te variëren cellen, randtotalen zijn ingevuld) ? (r-1)*(k-1)

Kritische waarde bij 5% niveau (df=4): 9,49. Derhalve bestaat er een significant verband tussen beide variabelen.

--> (aantal rijen – 1) X (aantal kolommen – 1) = vrijheidsgraden (= DF) (opgelet kan ook andere betekenis krijgen)

Kritische waarde bij 5% als 0-hypothese wordt verwerp kan ik er maximum 5% van de gevallen er naast zitten, kunt u dus vergissen

Handmatig : niet kennen uit het hoofd: wordt gegeven , zo wnr significant enz. wnr groter dan kritische water) chi-kwadraat, =SAMENHANG,
kleiner = GEEN SAMENHANG

Het begrip veiligheidsgraden (DF)

- Hoeveel waarden kunnen er in een kruistabel vrij variëren wanneer de randtotalen ingevuld werden?
  Dit is in een viervelden tabel slechts 1. Na het invullen van één van de celfrequenties liggen de overige drie cellenfrequenties ook vast.
- df: (r-1)*(k-1)

het begrip significantie?

- Wat is de kans om deze waarde van Chi-kwadraat te vinden indien de nulhypothese waar zou zijn?
- Deze nulhypothese is bij een Chi-kwadraat waarde: er is geen verband tussen beide variabelen. De alternatieve hypothese is er wel een verband.
- Deze kans is uiterst klein, vandaar dat we de nulhypothese verwerpen.
  Ofwel
  Deze kans is redelijk aanwezig, vandaar dat we de nulhypothese niet verwerpen.

Handmatige berekening van significantie

- Vergelijk de gevonden Chi-kwadraat met de kritische waarden.
- De vastgestelde waarde van 21,86 is groter dan de kritische waarde 9,49, derhalve is de kans om dergelijke Chi-kwadraat te vinden indien er geen samenhang was, kleiner dan 5%, dus weinig waarschijnlijk. Vandaar …

Kritische waarden van de Chi-kwadraat: maak gebruik van de tabel

df	0,05
1	3,84
2	5,99
3	7,82
4	9,49
5	11,07
…	….

--> ZIE SPSS

Chi –kwadraat beperkingen

meer proefpersonen -> sneller significantie ,
te weinig proefpersonen -> nooit significantie

- Gebruik enkel absolute getallen, geen proporties
- Verwachte frequentie mag in max. 20% van de gevallen kleiner zijn dan 5; geen enkele Fe waarde mag kleiner zijn dan 1.
- Niet gebruiken bij herhaalde meting;
  vb. voor- en nameting
  gebruik hiervoor de Mc Nemar test
- Is sterk afhankelijk van het aantal ppn.

	milieubesef	
	hoog	laag
Vrouw	14	6
Man	9	11

Chi-kwadraat = 2,56 (p=.110)

	milieubesef	
	hoog	laag
Vrouw	28	12
Man	18	22

Chi-kwadraat = 5,12 (p=.024)

Besluit de Chi-kwadraat is sterk onderhevig aan het aantal proefpersonen. Vandaar belang van associatiematen.
Gebruik Chi-kwadraat niet bij kleine proefgroepen.

Tip: controle via spss

Mc Nemar toets

- Kan enkel gebruikt worden voor herhaalde metingen, bv. voor en na de therapie.
- Let op welke celfrequenties u gebruikt als A en D versus C en B. A en D hebben betrekking op de proefpersonen die veranderen in functie van de behandeling

MC Nemar test, een voorbeeld

		NA	
		niet geslaagd	geslaagd
VOOR	geslaagd	5(A)	35(B)
	niet geslaagd	40(C)	20(D)

Mc Nemar test: $(|A-D|-1)^2/(A+D)$
Uitwerking: 14*14/25 = 7,84, hetgeen getest wordt als Chi-kwadraat waarde
Let op: enkel de groep van proefpersonen die van positie gewisseld zijn worden opgenomen in de formule.
Let op: enkel te gebruiken bij dichotome variabelen.

--> ZIE SPSS

Chi-kwadraat goodness-of-fit

- Voldoet een verdeling van nominale waarden aan bepaalde verwachtingen/verdeling?
  In dit geval hebben we niet te maken met een kruistabel !!!
- Gebruik de goodness-of-fit Chi-kwadraat test.

Goodness-of-fit test

- Onderzoek de aanhang bij 90 studenten voor drie politieke partijen

NOTAX	60
NOVA	5
Nieuw Groen	25
totaal:	90

chi-kwadraat goodness-of-fit test

Betekenisvolle voorkeur voor één van de drie partijen?

Fo	60	5	25
Fe	30	30	30
Verschil	30	25	5
Kwadraat	900	625	25
Kw/Fe	30	20,83	0,83

- Bestaat er een betekenisvolle voorkeur voor een van de drie partijen?

Chi-kwadraat = 51,66
kritische Chi-kwadraatwaarde =5,99

Dus er bestaat een significante voorkeur voor de partij NOTAX.

--> zie SPSS voor chi-kwadraat : goodness of fit test

Goodness of fit

- Wordt vaak gebruikt om aan te tonen dat de samenstelling van de steekproef een weerspiegeling is van de populatie.
- Let op: gebruik geen percentages, beide verdelingen dienen eenzelfde aantal ppn. te hebben. Eventueel downscalen.

Hoe sterk is dit verband?

Associatiematen, gebaseerd op de Chi-kwadraat

- De Cramér's V
- De Contingentiecoëfficiënt
- De Phi-coëfficiënt
  (enkel bij een 2X2 tabel)

De Cramér's V

Gebruik de Cramér's V
Deze index geeft de sterkte van het verband aan en
varieert van 0 tot 1. Kan in elk type kruistabel gebruikt worden.

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

voorbeeld uitwerking Cramér's V

In het voorbeeld van het verband tussen de politieke voorkeur en de houding t.o.v. milieutaks.

$$V = \sqrt{\frac{\chi^2}{N(k-1)}}$$

Concreet: V= 0,36

Hoe sterk is dit verband?

- Gebruik de contingentiecoëfficiënt
- in vergelijking met hetgeen max. haalbaar is in functie van de grootte van de tabel

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}} \qquad C_{max} = \sqrt{\frac{k-1}{k}}$$

k = ofwel het aantal rijen , ofwel het aantal kolommen (het kleinste van de 2)
dus 3 rijen en 2 kolommen , k = 2
of 6 rijen en 8 kolommen , k =6

De Contingentiecoëfficiënt

- In het voorbeeld van verband tussen politieke voorkeur en goedkeuring van milieutaks:
- Concreet: C = 0,45
- Maximale C-waarde: 0,82

$$C = \sqrt{\frac{\chi^2}{\chi^2 + N}}$$

Hoe sterk is dit verband

Gebruik de Phi-coëfficiënt (enkel bij een viervelden tabel)

$$\phi = \frac{ad - bc}{\sqrt{efgh}}$$

a,b,c, en d vormen de celfrequenties
e,f,g,h vormen de randtotalen

De Phi-coëfficiënt (enkel bij een 2X2 tabel)

Kan beschouwd worden als een standaardisering van de Chi-kwadraat waarde. (tweede wijze van berekening)

$$\varphi = \sqrt{\frac{\chi^2}{N}}$$

de phi-coëfficiënt :

- Kan enkel gelijk worden aan 1 indien verhouding tussen rijtotalen en kolomtotalen gelijk is, in het andere geval blijft deze waarde kleiner dan 1.
- Het teken van deze coëfficiënt speelt geen rol: we kunnen de waarden van plaats wisselen, waardoor het teken verandert.
  Het is niet zinvol te spreken over een negatief of positief verband bij nominale waarden.

--> zie SPSS : Output van associatiematen in SPSS politieke voorkeur t.o.v. attitude milieutaks

Opmerking. Cohens Kappa coëfficiënt

- Deze is enkel te gebruiken als associatiemaat bij een symmetrische kruistabel, bv. twee mensen beoordelen beiden dezelfde objecten met dezelfde categorieën. (Instrument te gebruiken bij betrouwbaarheidsonderzoek)

- Kappa geeft de proportionele overeenkomst tussen beoordelaars aan, nadat de toevallige overeenkomst weggewerkt werd

Voorbeeld:

	eerste beoordelaar			
	A	B	C	D
Tweede A	14	1	-	-
B	3	10	-	2
C	-	-	10	-
D	3	4	-	3
	20	15	10	5

Kappa = (37 – 13,5)/(50-13,5) = .64

--> ZIE SPSS
(toepassing 1 bij spss bestand)

Toepassing 2

We onderzoeken of onze steekproef voor een onderzoek naar leerproblemen een weerspiegeling is van de werkelijke samenstelling van de populatie.In de steekproef hebben we 100 ASO leerlingen, 50 TSO en 50 uit het BSO. Uit de gegevens van het ministerie van onderwijs blijkt dat er in Vlaanderen volgende % voorkomen ASO 35%, TSO 30% en BSO 35%. Kunnen we stellen dat deze steekproef een weerspiegeling is van de populatie?

Steekproef:	
ASO	100
TSO	50
BSO	50
Populatie (Fe):	
ASO	70
TSO	60
BSO	70

Chi-kwadraat:
12,86 + 1,67 + 5,71 = 20,24

Toetsen t.o.v. de kritische waarde
bij df = 2: 5,99

Vandaar dat we kunnen stellen dat deze steekproef geen goede weerspiegeling biedt van de populatie.

- Controleer de uitkomst via SPSS

  invoer via de snelle invoer via weight cases
  analyse --- nonparametric – Chi-kwadraat

- Kunnen we ook associatiematen berekenen?

(pearson correlatie : als x en y waarde maal of gedeeld met een negatief getal , veranderd de correlatie.

Hoofdstuk 9 : correlatie en regressie (en SPSS toepassingen)

Doelstellingen

Handmatig : hoeft niet -> SPSS duh!!

- De studenten kunnen voor een eenvoudige set van gegevens de samenhang tussen twee intervalvariabelen bepalen en tevens de vergelijking van de regressielijn opstellen.
- De studenten kunnen deze berekeningen (handmatig en) via SPSS doen. De studenten kunnen de output van SPSS lezen en interpreteren.

Analyse van de samenhang

Vereiste: ***twee*** uitslagen per persoon

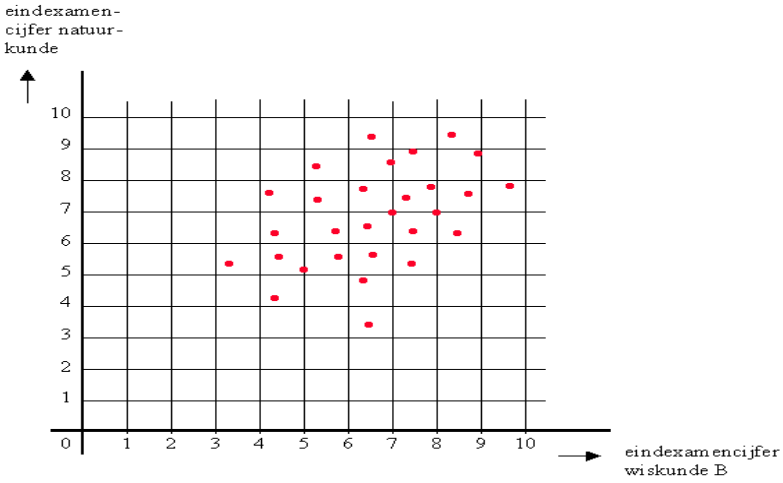
- Twee nominale variabelen stellen we voor in een kruistabel. Analyse met Chi-kwadraat en de associatiematen.
- Twee interval variabelen: gebruik de Pearson correlatiecoëfficiënt en de regressietechniek. Bv. samenhang tussen IQ en schooluitslag
- Twee ordinale variabelen: gebruik de correlatiecoëfficiënt van Spearman.

Samenhang tussen 2 interval variabelen

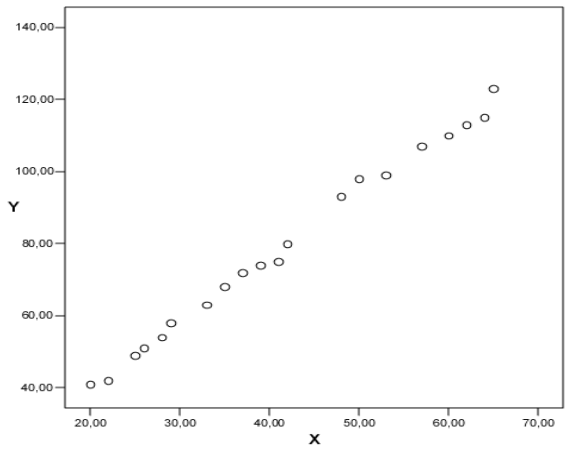
- Bestaat er een (lineair) verband?
  Gebruik de Pearson correlatiecoëfficiënt
- Hoe kunnen we de Y variabele voorspellen op grond van de X variabele?
  Gebruik de regressielijn van Y op X.

De correlatiecoëfficiënt

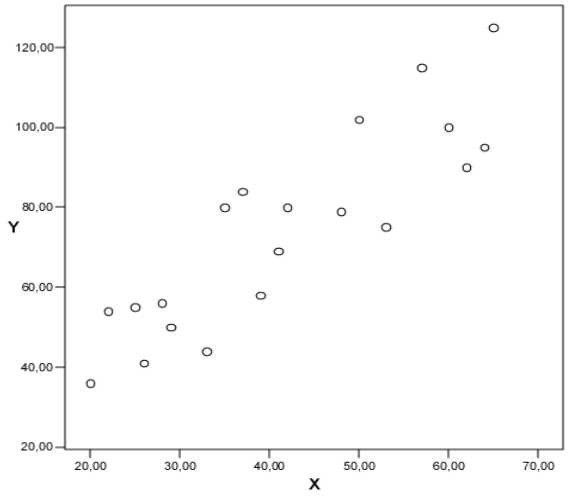
spreidingsdiagram : scatterplot



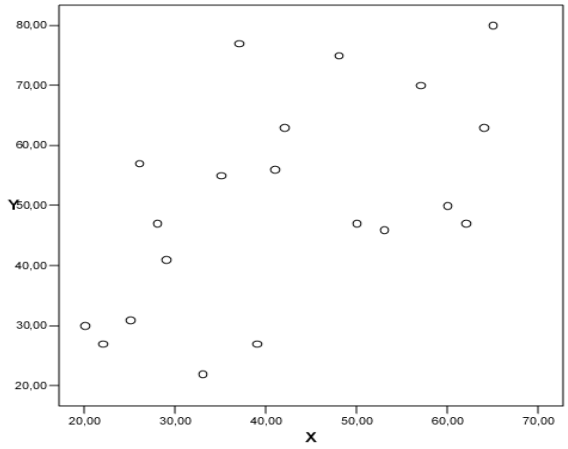
Voorbeeld van een zeer hoge positieve correlatie



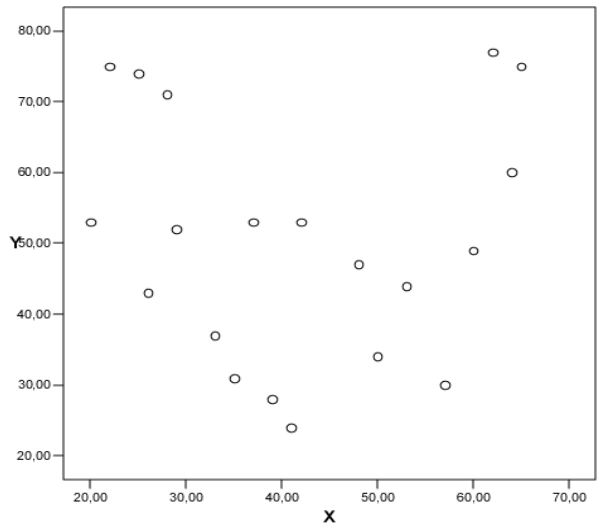
Voorbeeld van een hoge positieve correlatie



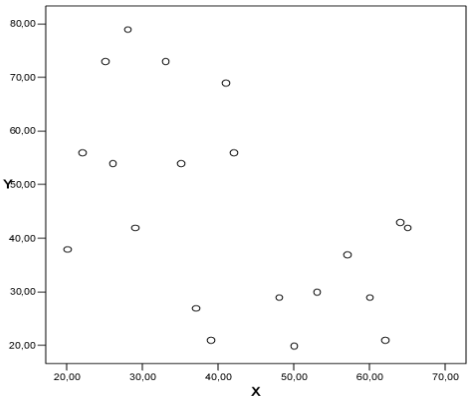
Voorbeeld van een geringe positieve correlatie



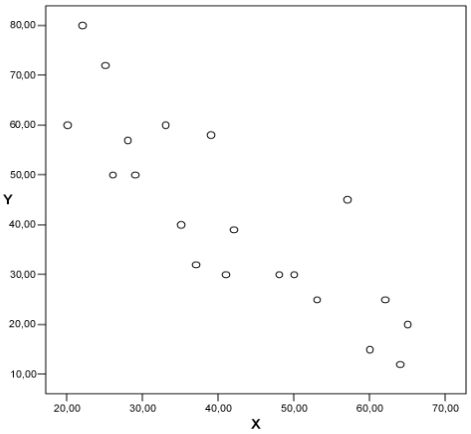
Voorbeeld van geen correlatie



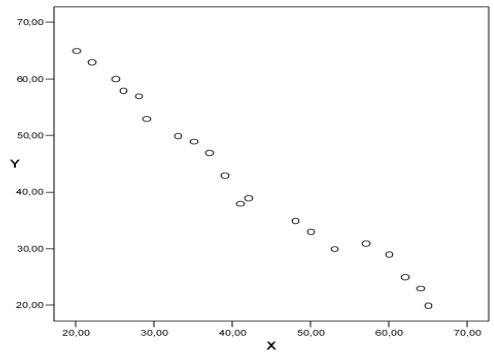
Voorbeeld van een negatieve matige correlatie



Voorbeeld van hoge negatieve correlatie



Voorbeeld van zeer hoge negatieve correlatie



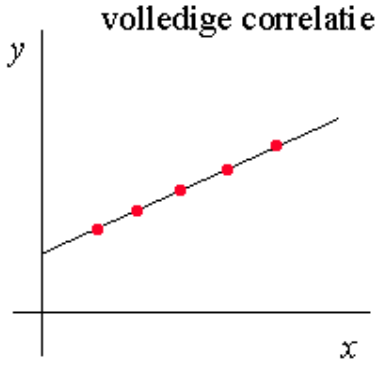
Correlatie

- Is het verband negatief? Of positief?
- Hoe sterk is het verband?

  -1 < r < 1
- Enkele concrete voorbeelden

Correlatiecoëfficient

voorbeeld 1 :



vb. : aantal juiste oplossingen en punt voor examen

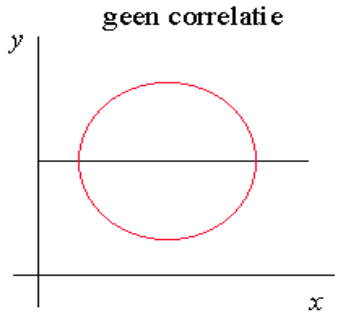
voorbeeld 2:

orrelatie

x

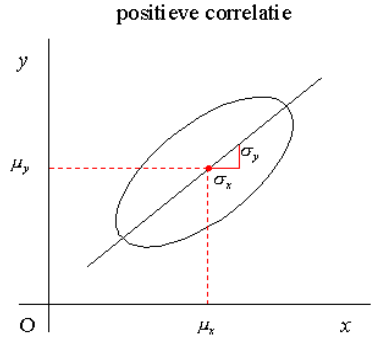
vb.: intelligentie en schooluitslag

voorbeeld 3 :



vb.: lichaamslengte en schooluitslag

Positieve correlatie :



Bv.: intelligentie en schooluitslag

negatieve correlatie :

bv.: faalangst en schooluitslag

Correlatie?

- Is de mate waarin elk individu eenzelfde relatieve positie inneemt op de twee variabelen.
- Is positief als hoge score voor een variabele samengaat met hoge score voor tweede variabele.

Is negatief als hoge score voor een variabele samengaat met lage score voor tweede variabele

Correlatie

Y-as		
	$Z_x<0$ en $Z_y>0$	$Z_x>0$ en $Z_y>0$
	$Z_x<0$ en $Z_Y<0$	$Z_x>0$ en $Z_y<0$

X-as

Pearson correlatie

Is het gemiddelde product van de bij X en Y horende z-scores. (productmomentcorrelatie van Karl Pearson)

$R_{XY} = \sum Z_X * Z_Y / N$ D.w.z. de correlatie is symmetrisch

$R_{YX} = \sum Z_X * Z_Y / N$

(gemiddelde * het product van Zwaarden)/ het aantal

Covariantie

= niet gestandaardiseerde maat van samenhang tussen twee interval variabelen.

Gemiddelde product van de afwijking t.o.v. het rekenkundig gemiddelde

pearson correlatie

=een gestandaardiseerde maat van samenhang, varieert van – 1 tot + 1

Correlatie kan gedefinieerd worden als de covariantie van de twee variabelen gedeeld door het product van de bijbehorende standaarddeviaties

Pearson correlatie: berekening I

X	Y	$X-X_{gem}$	$Y-Y_{gem}$	Product
165	37	-13	-4	52
167	38	-11	-3	33
170	39	-8	-2	16
172	42	-6	1	-6
175	39	-3	-2	6
175	42	-3	1	-3
180	40	2	-1	-2
189	44	11	3	33
192	44	14	3	42
195	45	17	4	68
Tot1780	410			239
Gem178	41			23,9
SD 10,09	2,65			

Uitwerking voorbeeld van covariantie

Cov(X,Y) = 23,9

Uitwerking correlatie, berekening I

$R_{XY} = cov(X,Y)/SD_X*SD_Y = 23,9/10,09*2,65$

$= 0,895$

Pearson correlatie: berekening II

- Werkformules: p. 163
- Hiervoor hebben we nodig:
  som X, som Y, som X^2, som Y^2 en som YX
- Op dezelfde wijze zijn we eveneens in staat handmatig de correlatie te berekenen.

Uitwerking correlatie berekening II

X	Y	X*X	Y*Y	X*Y
165	37	27225	1369	6105
167	38	27889	1444	6346
170	39	28900	1521	6630
172	42	29584	1764	7224
175	39	30625	1521	6825
175	42	30625	1764	7350
180	40	32400	1600	7200
189	44	35721	1936	8316
192	44	36864	1936	8448
195	45	38025	2025	8775
1780	410	317858	16880	73219

Uitwerking correlatie berekening 2

- Hiervoor hebben we nodig:
  som X, som Y, som X^2, som Y^2 en som Y*X
- Teller: 10*73219 – (1780)*(410)
- Noemer: wortel uit
  (10*317858 – 3168400)(10*16880 – 168100)
- r = 0.895

Pearson correlatie :

- Blijft constant als de X en/of de Y waarden vermenigvuldigd, gedeeld worden door een bepaald getal. Let wel op het teken van de correlatie (toevoegen aan Van Peet, pg.164)
- Blijft constant als de X en/of de Y waarden opgeteld of verminderd worden met een bepaald getal.
- Dus r is *invariant* onder lineaire transformaties. (afgezien van het teken)

invariant van de correlatie

Invariant van de correlatie

X	Y	(X+3)/4	(Y+2)/6
1	5	1,00	1,17
2	2	1,25	0,67
3	3	1,50	0,83
4	4	1,75	1,00
5	1	2,00	0,50

r = - 60

r = - 60

Invariant van de correlatie ?

X	Y	(X+3)/4	-(Y+2)/6
1	5	1,00	- 1,17
2	2	1,25	- 0,67
3	3	1,50	- 0,83
4	4	1,75	- 1,00
5	1	2,00	- 0,50

r = -.60 r = .60

Correlatie bij lineaire transformatie

$$X'_i = a + b \cdot X_i \quad \text{en} \quad Y'_i = c + d \cdot Y_i$$

dan is $$r_{X,Y'} = \frac{b \cdot d}{|b| \cdot |d|} \cdot r_{X,Y}$$

Pearson correlatie :

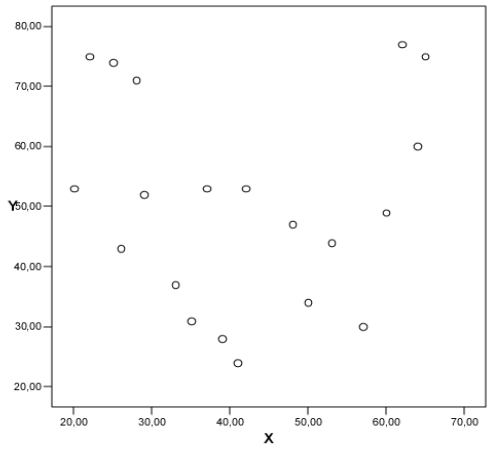
- Is niet invariant voor niet-lineaire transformaties, zoals bv. omzetting in percentielscores, of bv. worteltrekking of kwadratering.
- Niet lineaire transformaties wijzigen de vorm van de verdeling en tevens de correlatie met andere variabelen.

Lage correlatie?

- De variabelen hangen niet met elkaar samen (bv. lichaamslengte en schooluitslag) geen samenhang = 0-correlatie
- Het verband tussen de beide variabelen is niet lineair (bv. relatie tussen angst en prestaties) (spss kan alleen lineair samenhang zien)
- Er is sprake van 'restriction of range'. (gebrek van verschillen resulteert in lage covarianties en lage correlaties) -> hbp 167

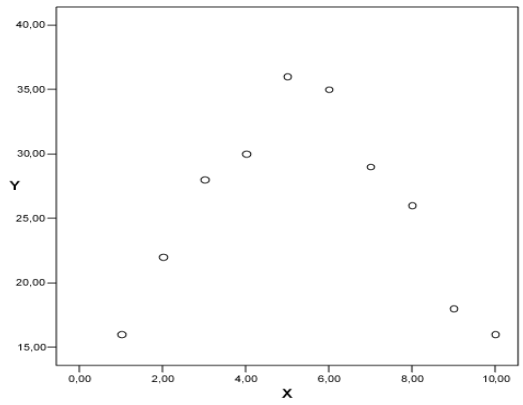
Lage correlatie :

1. Geen verband



$$r_{X,Y} = -0,01$$

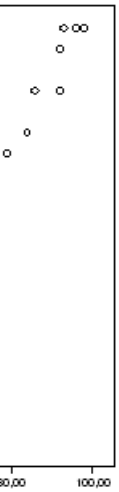
2. niet-lineair verband

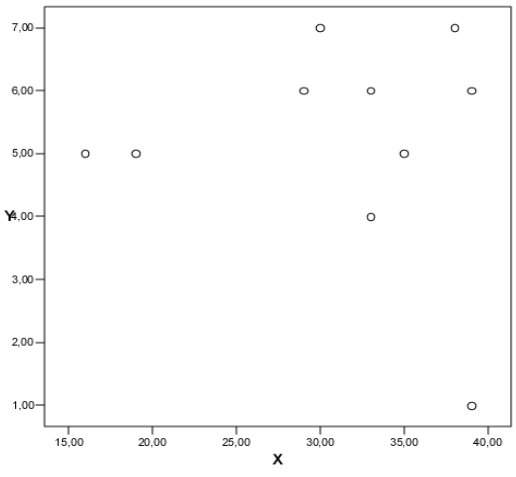


$$r_{X,Y} = -0,10$$

Pearson correlatie =0

3. Restriction of range



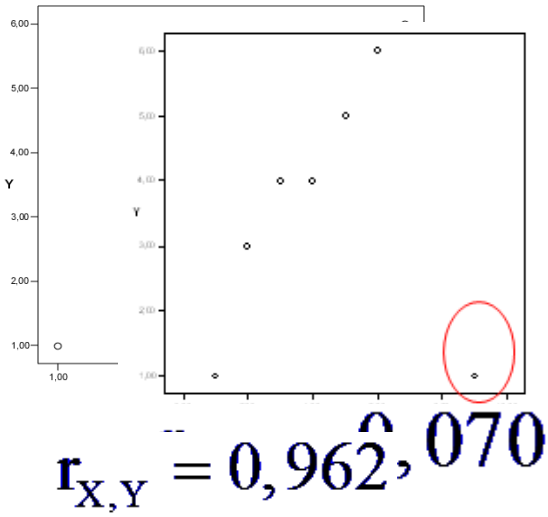


-->

959 --> $r_{X,Y} = -0,113$

We zien dat de correlatie veranderd , een oplossing hiervoor is voldoende proefpersonen gebruiken!

Lage correlatie?
Let op bij gering aantal metingen



-->

$r_{X,Y} = 0,962, 070$

-->

Bijvoegen van 1 outlier --> correlatie veranderd
oplossing : zorg voor voldoende proefpersonen , dan heeft het toevoegen of weglaten van proefpersonen minder impact op de correlatie

Correlatie en causaliteit
Als er een samenhang bestaat tussen twee variabelen, betekent dit een causaal verband?
de ene heeft invloed op andere, of er is misschien een andere factor wat x en y doet samenhangen.
Misschien
- X veroorzaakt Y
- Y veroorzaakt X
- Z veroorzaakt X, maar ook Y
- andere…

bv. medewerkerstevredenheid en productiviteit
bv. ooievaarsnesten en aantal geboorten
samenhang bij ooievaarsnesten en aantal geboorten? Uitleg : er is een factor Z : industrialisering : minder natuur = minder ooievaarsnesten , in die tijd ook anticonceptiemiddelen = minder geboortes. hier spreekt men dus van een externe factor : factor Z

Voorbeeld :
Het onderzoeksbureau 'Reason Foundation' publiceerde een opmerkelijke studie. Drinkers verdienen ruim 10% meer dan geheelonthouders. Iemand die buitenshuis zijn pintje drinkt, verdient op zijn beurt meer dan een thuisdrinker. De reden lijkt logisch: mensen die drinken, onderhouden meestal meer contacten. Contacten – en dus netwerking – kunnen zorgen voor een nieuwe of betere baan en snellere loonsverhogingen. Vrouwelijke drinkers verdienen gemiddeld 14% meer dan vrouwelijke niet drinkers. Het verschil bij de mannen bedraagt maar 10%, maar bij hen kan een regelmatig toogbezoek daar nog 7% aan toevoegen. Vanzelfsprekend geldt voor dit onderzoek ook de bekende slogan: overdaad schaadt
waarschijnlijk geen verband (causaal) . het ligt waarschijnlijk aan het feit dat wie meer verdient , vaker weg gaat en op café zal gaan

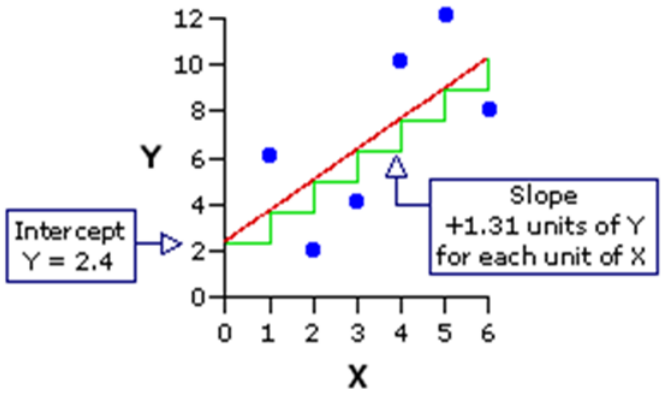
handberekening niet kennen: maar wel via spss
--> ZIE SPSS

Lineaire regressie
hbp 168
Lineaire regressie (enkelvoudige= er is 1 onafhankelijke variabelen)

- Welke rechte past het best bij een gevonden spreidingsdiagram? Welke rechte lijn biedt een zo goed mogelijke samenvatting van de trend in de puntenwolk?
- Zoeken de vergelijking van deze rechte, op basis waarvan we op grond van een X waarde de Y waarde kunnen voorspellen

Hoe dichter bij elkaar de puntjes van de puntenwolk hoe fijner voor lineaire regressie, dan kan je makkelijker uw Y bepalen

De regressielijn Y = 2,4 +1,31X



constante = y als X = 0

a = y = 0X = constante
constante is de waarde van y als X = 0
B waarde(slope) : als x met 1 eenheid toe neemt , neemt y toe met 1,31

Regressielijn

Y = a + bX

Waarbij:
X = de onafhankelijke variabele ('oorzaak') (horizontale as)
Y = de afhankelijke variabele ('gevolg') (verticale as)
a = de constante, die het snijpunt (*intercept*) met de Y-as vormt
b = de hellingscoëfficiënt (*slope*, of richtingscoëfficiënt)
OV op de x-as
AV op de y-as

De regressielijn een eenvoudig voorbeeld: het salaris
zie voor dit voorbeeld : SPSS
er staan daar heel wat slides uitgelegd
we gaan nu hier terug verder met :

Uitgaven restaurant en inkomen :

- Wat is het verband tussen het inkomen en de uitgaven aan restaurantbezoek?
- Hoe zal uitgaven voor restaurantbezoek toenemen in functie van inkomensverandering?

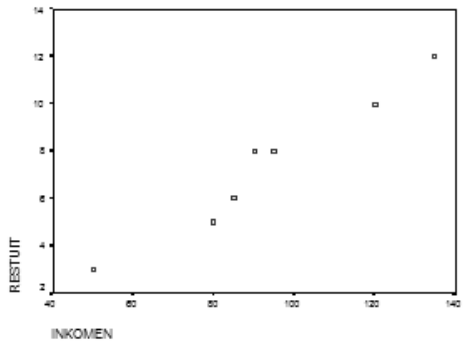
Restaurantuitgaven gegevens

Ppn	uitg. rest	inkomen
1	10	120
2	5	80
3	6	85
4	3	50
5	12	135
6	8	90
7	8	95

Afhankelijke/onafhankelijke variabele?

AV : restaurantuitgaven
OV : inkomen

Is er een lineair verband?



- Eerste inzicht in de relatie tussen de twee variabelen via de puntenwolk (spreidingsdiagram)
- Bestaat er een rechtlijnig verband tussen beide variabelen?

er is duidelijk een hoge samenhang (lineair) sterke samenhang : punten liggen dicht bij elkaar positieve samenhang (van links onder naar rechts boven)

Pearson correlatiecoëfficiënt

Correlations

		RESTUIT	INKOMEN
RESTUIT	Pearson Correlation	1,000	,978**
	Sig. (2-tailed)	.	,000
	N	7	7
INKOMEN	Pearson Correlation	,978**	1,000
	Sig. (2-tailed)	,000	.
	N	7	7

**. Correlation is significant at the 0.01 level (2-tailed).

- = Covariantie gestandardiseerd
- Geeft een aanduiding van de sterkte en de richting van het verband tussen twee variabelen.
- Significantietoets: zie verder inductieve statistiek

pearson kan alleen lineaire verbanden zien .

Output regressieanalyse

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-2,657	1,000		-2,657	,045
	INKOMEN	,108	,010	,978	10,457	,000

a. Dependent Variable: RESTUIT

Restuit = -2,66 + 0,11*Inkomen

a = -2,657 = constante (= -2,66)
b = 0,108 = slope / hellingscoëfficiënt (=0,11)
(als X = O) (= inkomen)

Output regressieanalyse

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,978[a]	,956	,948	,6980

a. Predictors: (Constant), INKOMEN

- R square is de determinatiecoëfficiënt.
- Geeft de verhouding aan tussen de verklaarde variantie en de totale variantie.

R Square : 0,96 -> we kunnen 96% voorspellen en 4% niet
dit is een goede voorspelling
standaard fout van estimate = 0,7
op 2/3de van de getallen maak ik een fout op het voorspellen voor y en x die kleiner is dan 0,7
in 1/3de van de gevallen maak ik een fout die groter is dan 0,7

De regressievergelijking

- Restaurantuitgaven =
  -2,66 + 0,11 * Inkomen

- Bijvoorbeeld: inkomen is 200, welk is dan de restaurantuitgave?
  -2,66 + 0,11 * 200 = 19,34.

  En als X = 0 ?

- Wat is de waarde van deze vergelijking?
  Dekt het model de werkelijkheid?
  - 96% van de verschillen in de restaurantuitgaven kunnen verklaard worden door de verschillen in inkomen
  - De standaardfout van estimatie bedraagt: 0,70. Dwz dat in 2/3 van de gevallen de fout in de voorspelling kleiner zal zijn dan 0,70

Meervoudige regressie

- In dit geval zijn er meerdere X variabelen, op grond waarvan de Y variabele geschat wordt.
- Niet meer mogelijk via handmatige uitwerking, enkel via SPSS.
- Veel gebruikte procedure om aan te geven hoe diverse onafhankelijke variabelen gezamenlijk een invloed uitoefenen op de afhankelijke variabele. Diverse onafhankelijke variabelen worden t.o.v. mekaar uitgespeeld.
- De afzonderlijke betacoëfficiënten bieden een inzicht in het impact van elke onafhankelijke variabele, onder constant houding van de overige variabelen.

in SPSS doen : ervoor zorgen dat ze allemaal in scale niveau.

opdrachten :
Uit het werkboek: 9.2; 9.3; 9.4; 9.5; 9.6; 9.8;9.9;9.10; 9.11; 9.12; 9.14; 9.15; 9.16; 9.18; 9.19; 9.20

Hoofdstuk 10 : Rangcorrelatie, partiële en multiplecorrelatie

--> zie tekst ! (niet in het handboek)

10.1. De rangcorrelatiecoëfficiënt

Doelstellingen

- De student kan de samenhang bepalen tussen twee ordinaal geschaalde variabelen. De student kan dit handmatig en via SPSS berekenen.
- De student kan deze resultaten interpreteren

(spss vraag op het examen : een amateur zou alle items opstellen zo worden ook de missing values mee geteld , maar wij experten doen het via mean.

Rangcorrelatiecoëfficiënt

- Het betreft het verband tussen twee ordinale variabelen.
- Correlatiecoëfficiënt van Spearman
- Formule:

$$r_s = 1 - \frac{6\sum D_i^2}{N \cdot \left(N^2 - 1\right)} \quad \text{met} \quad D_i = \left(X_{i,R} - Y_{i,R}\right)$$

Deze correlatie varieert van -1 tot + 1

deze correlatie varieert dus van -1 tot +1 / -1 = perfect negatieve samenhang / 0= geen samenhang/ +1= perfect positieve samenhang

Een voorbeeld

- Welk is het verband tussen intelligentie en leiderschap bij kinderen?
- Geef voor elke ppn. een rangorde voor beide variabelen
- Dit geeft volgende beeld

Ppn	X_r	Y_r	D	D^2
F	1	8	-7	49
D	2	5	-3	9
B	3	7	-4	16
H	4	6	-2	4
E	5	4	1	1
G	6	2	4	16
C	7	3	4	16
A	8	1	7	49
Totaal:			0	160

$$r_s = 1 - \frac{6\sum D_i^2}{N \cdot (N^2 - 1)}$$

$r_s = 1\text{-}6*160/8(64\text{-}1) = -0{,}90$

Besluit?

x = intelligentie : 1 = de beste , maar leiderschap (Yr) = 8 dat is het zwakste
D = het verschil : 1-8 = -7 dan $-7^2 = 49$ (D^2)
dit we dan $1 - \frac{160 \cdot 6}{8 \cdot (64 - 1)} = -0{,}90$ (= spearman)
doen we bij F,D,B?H,E,G,C,A
tellen al de kwadraten op (D^2) = 160

N = 8 = het aantal F,D,B,H,E,G,C,A
N^2 = dus $8^2 = 64$

--> zie SPSS

Probleem: Knopen

- Wanneer er binnen X of Y twee of meer ppn een gelijke uitslag hebben
- Wat dan?

- Aan deze gelijke uitslagen wordt hetzelfde rangnummer toegekend, nl. het gemiddelde van de nummers.

bv.: 3 leerlingen , scoren allemaal 1 dus dat is een gemiddelde van 2 , we geven ze dan alledrie een 2

10.2. De Puntbiseriële coëfficiënt

Doelstellingen

- De student kan deze samenhang berekenen voor een beperkt aantal observaties

- De student kan deze resultaten interpreteren

De punt-biseriële coëfficiënt

- Betreft verband tussen natuurlijk dichotome variabele en een interval variabele.

- Nut?

- Voor de berekening van item-test relatie in het kader van item validatie.

--> niet handmatig kunnen – wel via spss
natuurlijke dichotomie : 2 niveaus die niet gemaakt zijn door de onderzoekers, maar die er al waren
bv.: goed =1 , fout = 0
betrouwbaarheid : stabiliteit van mening
via SPSS :
-> hoe zal de betrouwbaarheid veranderen als er een item uitgehaald wordt
wanneer er minder waardevolle items eruit halen -> stijgt de betrouwbaarheid
zie SPSS

Doelstellingen

- De student kan de samenhang tussen meerdere interval geschaalde variabelen bepalen

- De student kan deze resultaten interpreteren

De multiple correlatiecoëfficiënt

- Verband tussen schooluitslag en intelligentie en studietijd
  bv. $r_{YX1} = 0,65$
  $r_{YX2} = 0,35$
  $r_{X1X2} = -0,40$

  via toepassing van de formule.
  $R_{Y.X1X2} = 0,930$
  Wat stellen we vast?

- Deze coëfficiënt is afhankelijk van de correlaties tussen Y en X_1 en X_2 maar ook van de correlatie tussen X_1 en X_2.

- Naarmate de correlatie tussen X_1 en X_2 kleiner is zal de multiple correlatie meer opgedreven worden.

De partiële correlatiecoëfficiënt

- Wordt vaker gebruikt dan de multiple correlatiecoëfficiënt.

- We trachten het verband te bestuderen van intelligentie en schooluitslag met constant houding van de bestede studietijd.

- Op gelijkaardige wijze bepalen we het verband tussen studietijd en schoolresultaat met constant houding van de intelligentie.
- En het verband tussen intelligentie en studietijd, met constant houding van schoolresultaat.
- Wat stellen we vast?

Zie SPSS

Samenhang tussen meerdere interval variabelen

- Multiple correlatie wordt weinig gebruikt, tenzij in relatie met multiple regressietechniek
- Partiële correlatiecoëfficiënt des te meer.
- Andere technieken (niet meer handmatig te berekenen):

  Factoranalyse. Uitgaande van de samenhang tussen de variabelen onderzoeken we de basisdimensies.

  Multiple regressielijn. Wat is de invloed van meerdere onafhankelijke variabelen op de afhankelijke variabele? De hoogte van de significante B-coëfficienten geeft een idee van de impact op de afhankelijke variabele.