

HOOFDSTUK II

UNIVARIATE BESCHRIJVENDE STATISTIEK

1. DOELSTELLINGEN

Op het einde van dit hoofdstuk zijn studenten in staat om onderzoeksgegevens samen te vatten, te analyseren en te interpreteren. Daarbij dienen zij rekening te houden met het meetniveau en de behoeften van de analyse. Studenten kunnen frequentietabellen en grafieken correct lezen en interpreteren.

2. FREQUENTIES2.1. Te onthouden kernbegrippen

Absolute frequentie		Het aantal keer dat een bepaalde waarde of categorie van een variabele voorkomt.
Cumulatieve frequentie of percentage		Kan pas berekend worden als je categorieën logisch kan ordenen van laag naar hoog. Bij elke frequentie of percentage tellen we de volgende op. De cumulatieve frequentie van de laatste waarde is steeds gelijk aan het totale aantal waarnemingen (idem voor het cumulatief percentage bij de laatste waarde). Hierdoor is het mogelijk te zien hoeveel waarnemingen kleiner dan of gelijk zijn aan een bepaalde waarde.
Frequentie		Het aantal keer dat een score of waarde voorkomt
Frequentieverdeling		Rangschikking van het aantal waarnemingen van laag naar hoog met vermelding van het aantal keer dat elke categorie van de variabele voorkomt.
Geldig percentage		Percentage berekend op diegenen die een geldig antwoord gaven op de vraag
Histogram		Grafische voorstelling van een frequentieverdeling van metrische variabelen. Lijkt op een staafdiagram maar de staafjes worden aan elkaar getekend omdat de waarden van de variabele elkaar opvolgen. De categorieën liggen op een continuüm.
Percentage		De verhouding tussen het absolute aantal keer dat een waarde voorkomt gedeeld door de totale steekproefgrootte (n) maal 100
Proportie		Absolute frequentie delen door het totaal aantal waarnemingen (n), getal dat gaat van 0 tot 1. De som van alle proporties is gelijk aan 1.
Relatieve frequentie		Geeft weer hoe vaak een waarde voorkomt ten opzichte van het totaal aantal waarnemingen. Kan worden uitgedrukt in

Staafdiagram

percentage op 100 of in proporties (of fracties) op 1 (ten opzichte van de geldige steekproefomvang)

De gegevens worden voorgesteld aan de hand van twee assen. Op de horizontale as (X-as) worden de verschillende categorieën van de variabele weergegeven. De verticale as (Y-as) geeft de aantallen weer onder de vorm van absolute frequenties of percentages. Bij elke categorie op de X-as tekent men een staafje en deze worden los van elkaar geplaatst. Geschikt voor grafische voorstelling van variabelen gemeten op het nominale en ordinale niveau.

Taartdiagram of cirkeldiagram

Een cirkelschijf verdeeld in sectoren op basis van de verschillende frequenties of percentages. Er wordt gebruik van gemaakt wanneer we met een beperkt aantal categorieën werken. Geschikt voor variabelen gemeten op het nominale en ordinale niveau.



! Vooraleer een univariate beschrijvende analyse uit te voeren, dien je jezelf volgende vragen te stellen:

1. *Wat is het gemeten kenmerk/wat is de variabele*
2. *Uit welke categorieën bestaat de variabele*
3. *Wat is het meetniveau van de bestudeerde variabele*
4. *welke beschrijvende onderzoeksvraag wil je beantwoorden: een vraag naar centraliteit, spreiding of vorm van een kenmerk*



! Leer goed het onderscheid tussen exacte frequenties, percentage en geldige percentages. We bespreken enkel de geldige percentages. Je dient te vermelden op basis van hoeveel geldige antwoorden (= aantal respondenten dat de vraag geldig beantwoordde) de onderzoeksvraag werd beantwoord. Vermeld ook steeds het totale aantal respondenten. Op die manier heeft de lezer inzicht in het aantal missings of het aantal respondenten dat de vraag niet of niet geldig beantwoordde.

3. MATEN VAN CENTRALITEIT

3.1. Te onthouden kernbegrippen

Deviatiescore	Afwijking ten opzichte van het gemiddelde. De som van de deviatiescores is steeds 0.
Kwantielen	Een maat die men gebruikt wanneer men een geordende rij elementen in om het even welk aantal groepen met een gelijk aantal elementen verdeelt. Ze kunnen berekend worden vanaf het ordinale niveau.
Kwartielen	De drie waarden (Q1, Q2 en Q3) die een geordende reeks uitkomsten in 4 gelijke stukken verdelen, die ieder een kwart van de uitkomsten bevatten.
Maten van centraliteit of centrummaten	Geven weer rond welke waarden de verdeling van een variabele gepositioneerd is. Maten van centraliteit geven weer welke de meest centrale waarde is van een verdeling. Welke maat van centraliteit je kan gebruiken, hangt af van het meetniveau van de variabele. Geven een antwoord op beschrijvende onderzoeksvragen die erop gericht zijn centrale tendensen te ontdekken.
Mediaan	Het midden van een statistische verdeling. Het is een centrummaat die het punt in de frequentieverdeling aangeeft waaronder 50% van de gevallen en waarboven de andere 50% van de gevallen liggen. Kortom, de frequentieverdeling wordt in twee gelijke stukken gedeeld. Om de mediaan te bepalen moeten de categorieën in oplopende volgorde gerangschikt zijn. Ze kan berekend worden vanaf het ordinale niveau.
modus	De modus is een centrummaat: de categorie van de variabele met de hoogste frequentie. Kan gehanteerd worden voor alle meetniveaus.
Rekenkundig gemiddelde	Een centrummaat die wordt verkregen door alle voorkomende waarden bij elkaar op te tellen en vervolgens het totaal te delen door het aantal respondenten. Ze wordt gebruikt bij variabelen gemeten op het metrische niveau.

4. MATEN VAN SPREIDING

4.1. Te onthouden kernbegrippen

Gemiddelde absolute afwijking	Een metrische spreidingsmaat. Het is de som van de absolute waarden van de afwijkingen van elke waarde ten aanzien van het rekenkundig gemiddelde, gedeeld door het aantal waarnemingen.
Index van diversiteit (ID)	Nominale spreidingsmaat, gebaseerd op de relatieve frequenties van de categorieën. Er wordt rekening gehouden met de proportie van waarnemingen binnen iedere categorie. De waarde van deze parameter geeft een idee van de mate van concentratie van de waarnemingen over de categorieën van de variabele.
Interkwartielafstand (K3-K1)	Ordinale spreidingsmaat. De interkwartielafstand is het verschil tussen de derde en eerste kwartiel en bevat de helft van het totale aantal waarnemingen.
Spreidingsmaat	Geeft aan of waarden in een verdeling dicht bij elkaar liggen of juist ver uit elkaar
Standaardafwijking (s)	De vierkantswortel van de steekproefvariantie
Variatie (SS)	Metrische spreidingsmaat. Het is de som van de gekwadrateerde afwijkingen van elke waarde tegenover het gemiddelde.
Variantie (s^2)	De som van de gekwadrateerde afwijkingen van elke observatie tegenover het gemiddelde, gedeeld door $n-1$. De variantie is dus gelijk aan de variatie delen door $n-1$.
Variatiebreedte	Ordinale spreidingsmaat. De variatiebreedte is het verschil tussen de grootste en de kleinste waargenomen waarde.
Variatiecoëfficiënt (v)	Gestandaardiseerde spreidingsmaat. Wordt berekend door de standaardafwijking te delen door het rekenkundig gemiddelde. Niet afhankelijk van de meeteenheid en laat bijgevolg toe de spreidingen van verdelingen die worden uitgedrukt in verschillende meeteenheden met elkaar te vergelijken.
Variatieratio (VR)	Een nominale spreidingsmaat. De variatieratio is de proportie waarnemingen die niet tot de modale categorie behoren. Ze neemt de waarde nul aan indien alle waarnemingen tot de modale categorie behoren. Er is echter geen vaste bovengrens.

4.2. Statistische symbolen en formules

Mediaan	$(n + 1) / 2$
Rekenkundig gemiddelde	<div style="display: flex; align-items: center; justify-content: center;"> <div style="text-align: right; margin-right: 10px;"> het rekenkundig gemiddelde </div> <div style="text-align: center;"> $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$ </div> <div style="text-align: left; margin-left: 10px;"> <div style="margin-bottom: 10px;">de som van alle waarnemingen</div> <div>het aantal waarnemingen</div> </div> </div>
Som van de deviatiescores = 0	$\sum_{i=1}^n (x_i - \bar{x}) = 0$
Standaardafwijking	$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} \text{ of } s = \sqrt{s^2}$
Variatie	$SS = \sum (X - \bar{X})^2 \text{ of } SS = \sum_{i=1}^N (x_i - \bar{x})^2$
variantie	$s^2 = \frac{\sum (X - \bar{X})^2}{n - 1}$
variatiecoëfficiënt	$v = \frac{s}{\bar{X}}$

5. SAMENVATTENDE TABEL : maten van centraliteit en spreiding

	NOMINAAL	ORDINAAL	METRISCH
MATEN VAN CENTRALITEIT	Modus	Modus Mediaan kwartielen	Modus Mediaan Kwartielen Rekenkundig gemiddelde
MATEN VAN SPREIDING	Variatieratio (VR) Index van diversiteit	Variatiebreedte Interkwartielafstand (K3-K1)	Gemiddelde absolute afwijking Variatie (SS) Variantie (s ²) Steekproefstandaardafwijking (s) Variatiecoëfficiënt (v)

6. MATEN VAN VORM

6.1. Te onthouden kernbegrippen

Boxplot	Visuele weergave van centraliteit, spreiding en vorm van de verdeling (voor kenmerken vanaf ordinaal meetniveau).
Empirische coëfficiënt van Pearson	Parameter van vorm. Wordt berekend als volgt: 1) verschil tussen gemiddelde en mediaan en 2) deze waarde delen door de standaardafwijking. Positieve coëfficiënt duidt op een positief asymmetrische verdeling, Negatieve coëfficiënt duidt op een negatief asymmetrische verdeling.
Kurtosis of afplatting	Mate van afplatting van gegevens rondom het rekenkundig gemiddelde.
Negatief asymmetrische verdeling	Verdeling met een langere staart naar links: het rekenkundig gemiddelde is kleiner dan de mediaan. Lage waarden komen minder vaak voor dan hoge waarden.
Positief asymmetrische verdeling	Verdeling met een langere staart naar rechts: het rekenkundig gemiddelde is groter dan de mediaan. Hoge waarden komen minder vaak voor dan lage waarden
Symmetrische verdeling	Verdeling van gegevens ten opzichte van het rekenkundig gemiddelde. Een verdeling is symmetrisch als het rekenkundig gemiddelde en de mediaan aan elkaar gelijk zijn. Een symmetrische verdeling volgt een Gauss-curve, linker- en rechterhelft zijn elkaar spiegelbeeld.

6.2. Statistische symbolen en formules

Empirische coëfficiënt van Pearson	$S = \frac{\bar{X} - \tilde{x}}{s}$
------------------------------------	-------------------------------------

7. OEFENINGEN

1. Wanneer we scores rangschikken volgens hun waarde en frequentie, construeren we een zogenaamde *frequentieverdeling*. Hieronder vind je de gegevens over eerdere arrestaties van 100 bekende overtreeders:

14	0	34	8	7	22	12	12	2	8
6	1	8	1	18	8	1	10	10	2
12	26	8	7	9	9	3	2	7	16
8	65	8	2	4	2	4	0	7	2
1	2	11	2	1	1	5	7	4	10
11	3	41	15	1	23	10	5	2	10
20	0	7	6	9	0	3	1	15	5
27	8	26	8	1	1	11	2	4	4
8	41	29	18	8	5	2	10	1	0
5	36	3	4	9	5	10	8	0	7

Opgave: Maak een frequentietabel

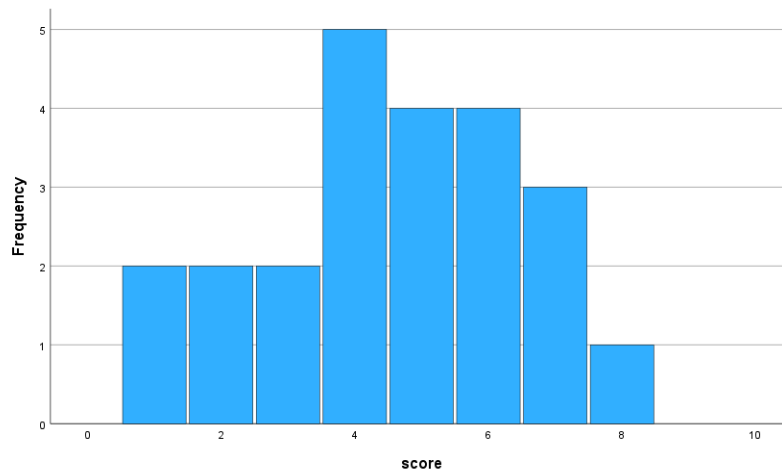
We groeperen eerst alle statistische eenheden met dezelfde waarde samen: zonder eerdere arrestaties, één eerdere arrestatie, twee eerdere arrestaties, enz. tot we alle scores in data hebben gegroepeerd. Vervolgens ordenen we deze scores in volgorde van grootte. Door op deze manier naar de data te kijken, kunnen we een idee krijgen van de aard van de verdeling van de scores. In de praktijk is het maken van een frequentieverdeling meestal de eerste stap die een onderzoeker neemt bij het analyseren van de resultaten van een onderzoek.

- a. Hoeveel bedraagt totale n ?
- b. Wat is de modus?
- c. Wat is de minst frequente score?
- d. Hoeveel respondenten hebben de score van 8?
 - a. In absolute aantallen
 - b. In relatieve frequenties
- e. Hoeveel respondenten hebben de score van 26?
 - a. In absolute aantallen
 - b. In relatieve frequenties

Tabel*Frequentieverdeling van eerdere arrestaties van 100 bekende overtredders*

Waarde	Absolute frequentie	Relatieve frequentie	Cumulatieve frequentie

2. Bekijk onderstaande histogram en beantwoord de vragen.



- Wat is een histogram?
- Hoeveel bedraagt totale n ?
- Wat is de modus of de meest frequente score?
- Wat is de minst frequente score?
- Hoeveel respondenten hebben de score van 5?
- Hoeveel respondenten hebben de score van 2?

3. Aan studenten uit de faculteit Recht en Criminologie wordt gevraagd hoe vaak ze hard-drugs gebruiken. Onderstaande tabel geeft de absolute frequenties weer. Vul de tabel aan met n, cumulatieve frequentie, proportie en cumulatieve proportie.

Categorie	Frequentie	Cumulatieve frequentie	Proportie	Cumulatieve proportie
Nooit	276			
Zelden	8			
Af en toe	4			
Dikwijls	2			
Zeer vaak	1			
	n=			

- Hoeveel studenten nemen dikwijls hard-drugs?
- Hoeveel procent is dat?
- Hoeveel procent van de studenten neemt zelden of nooit hard-drugs?
- Welke proportie van de studenten neemt minstens “dikwijls” hard-drugs?
- Stel dat de 100 mensen die niet hebben geantwoord, dat niet gedaan hebben omdat ze niet willen toegeven dat ze “dikwijls” hard-drugs gebruiken. Wat gebeurt er met de verschillende waarden als die mensen wel hadden geantwoord?

4. Een groep van 20 gevangenen in een bepaald cellenblok werd getest op hun kennis van de regels van de instelling. De scores (totaal 70 punten) waren als volgt:

31	28	27	19	18	18	41	0	30	27
27	36	41	64	27	39	20	28	35	30

- Bereken de range of variatiebreedte.
- Verwijder de laagste en de hoogste score en bereken de variatiebreedte van de resterende scores.
- Hoe verklaar je het verschil tussen waarden van deze twee maten van spreiding?

5. Hieronder vind je drie frequentietabellen. Bepaal voor elke frequentietabel *de best passende maten* van centraliteit en maten van spreiding.

Wettelijke vertegenwoordiging voor witteboordencriminaliteit	Frequentie
Geen	20
Rechtshulp	26
Door rechtbank aangeduide vertegenwoordiger	92
Openbare aanklager	153
Privé-advocaat	380
Totaal	671

Attitude van studenten over openbare dronkenschap	Frequentie
Helemaal niet erg	73
niet erg	94
neutraal	27
erg	65
Heel erg	22
Totaal	281

Aantal eerdere arrestaties	Frequentie
0	4
1	1
2	2
4	3
5	3
7	4
8	2
10	1
Totaal	20

6. De toeristische dienst van de Vlaamse kust wilt weten hoe tevreden de toeristen zijn over hun daguitstapje naar de kust. Één van de vragen luidt: "Was u tevreden met de bediening in de horecazaken?" De verdeling van de toeristen over de antwoorden vind je terug in onderstaande frequentietabel.

ANTWOORD	AANTAL TOERISTEN
Zeer ontevreden	33
Ontevreden	84
Noch tevreden, noch ontevreden	102
Tevreden	63
Zeer tevreden	48

- Vul de tabel verder aan met de cumulatieve absolute frequenties, relatieve frequenties en cumulatieve relatieve frequenties.
- Wat is het meetniveau van de variabele?
- Bereken alle relevante maten van centraliteit. (hoe weet je welke maat van centraliteit het meest relevant is?)
- Bereken alle relevante maten van spreiding.

7. We vroegen aan 30 vrouwen en 30 mannen hoeveel minuten zij op een gewone avond in de week studeren. Hieronder volgen hun antwoorden:

VROUWEN						MANNEN				
180	120	180	360	240		90	120	30	90	200
120	180	120	240	170		90	45	30	120	75
150	120	180	180	150		150	120	60	240	300
200	150	180	150	180		240	60	120	60	30
120	60	120	180	180		30	230	120	95	150
90	240	180	115	120		0	200	120	120	180

- Bereken voor beide groepen de maten van centraliteit.

	VROUWEN	MANNEN
Modus		
Mediaan		
Gemiddelde		

- Beschrijf voor beide groepen de maten van centraliteit.

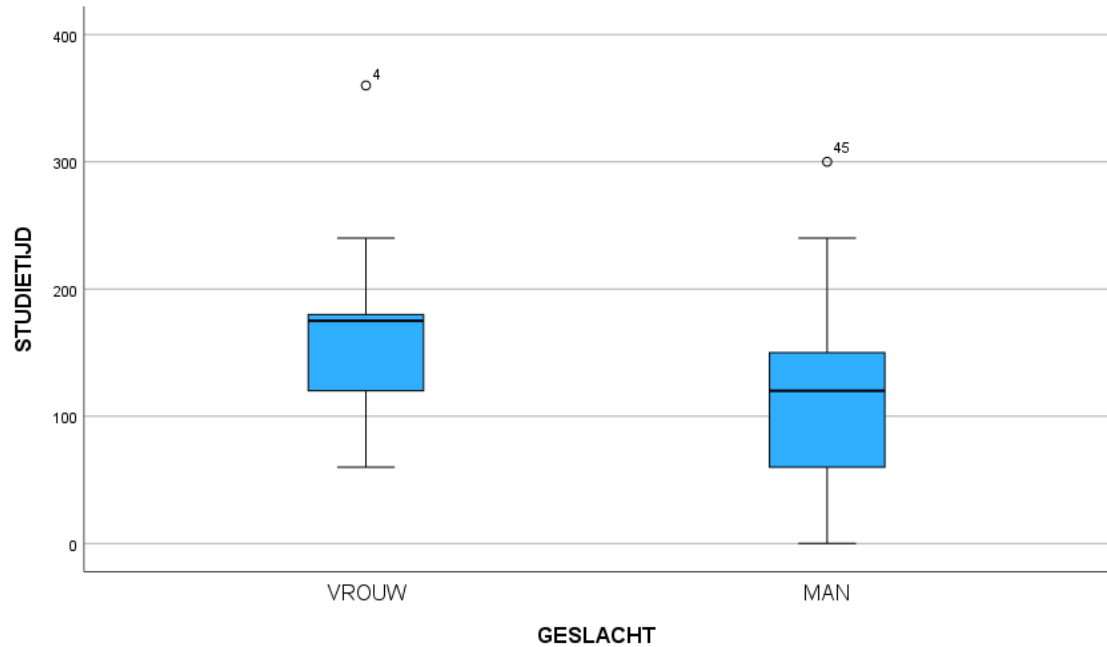
Modus

Mediaan

Gemiddelde

Hieronder vind je de visuele weergave van de verdeling van de variabele 'studietijd' voor de vrouwen en voor de mannen aan de hand van twee boxplots.

Beschrijf de vorm van de verdeling.



- Als je de boxplot van mannen en vrouwen met elkaar vergelijkt, wat kun je zeggen over:
 - Het centrum van de verdeling?
 - De spreiding van de variabele?

8. Van een groep van 20 mensen is bekend hoeveel uren ze per week studeren:

24 36 35 28 24 28 24 36 32 36 40 38 36 34 40 36 32 36 40 36

- a. Bereken de modus, mediaan en het gemiddelde.
- b. Bereken alle behandelde spreidingsmaten.
- c. Welke parameter van centraliteit en spreiding zou jij kiezen om deze variabele te beschrijven en waarom ?

9. Van de groep van 20 mensen uit vorige vraag is niet alleen het aantal uren bekend dat zij per week studeren maar ook hun geslacht:

Vrouwen : 24 36 35 28 24 28 24 36 32 36

Mannen : 40 38 36 34 40 36 32 36 40 36

- a. Bereken het gemiddelde en de standaardafwijking voor elke groep.
- b. Wat kun je zeggen over de verschillen tussen mannen en vrouwen in studieduur op basis van jouw berekeningen ?

10. De trainer van het universitaire zwemteam onderzoekt of er een effect is van het aantal keer trainen per week op hoe lang ze kunnen watertrappen. De gegevens die hij verzamelde, vind je terug in tabel 2 (te vinden onder documenten). Bereken de meest relevante maat van spreiding voor de afhankelijke variabele

Aantal minuten watertrappen	Aantal keer trainen per week
0	3
4	1
8	2
3	1
5	0
6	0
1	1
2	0
7	1

Maak uw keuze

- 7,5
- 2,74
- 2,8
- 7,85

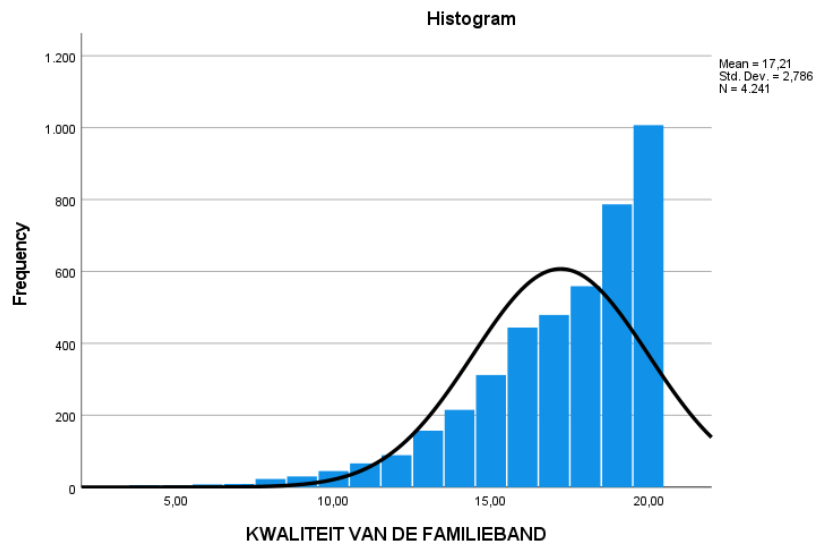
11. Variantie in bedragen van borgtocht voor een steekproef van 15 arrestanten

15 arrestanten	Bedrag van borgtocht (€)
1	500
2	1000
3	1000
4	1000
5	1200
6	1500
7	2500
8	2500
9	2500
10	2750
11	5000
12	5000
13	5000
14	7500
15	10000

Opdracht:

- Hoeveel bedraagt de standaardafwijking van borgtocht-bedragen in deze steekproef van 15 gearresteerde personen?
- Wat betekent het als de standaardafwijking gelijk is aan nul? Wat zou dit betekenen voor de scores van de variabele 'bedrag van borgtocht'?
- In welke eenheid is de standaardafwijking uitgedrukt?
- Zijn outliers een probleem voor de interpretatie van de standaardafwijking? Leg uit.

12. Bekijk onderstaande histogram en beantwoord de vragen.



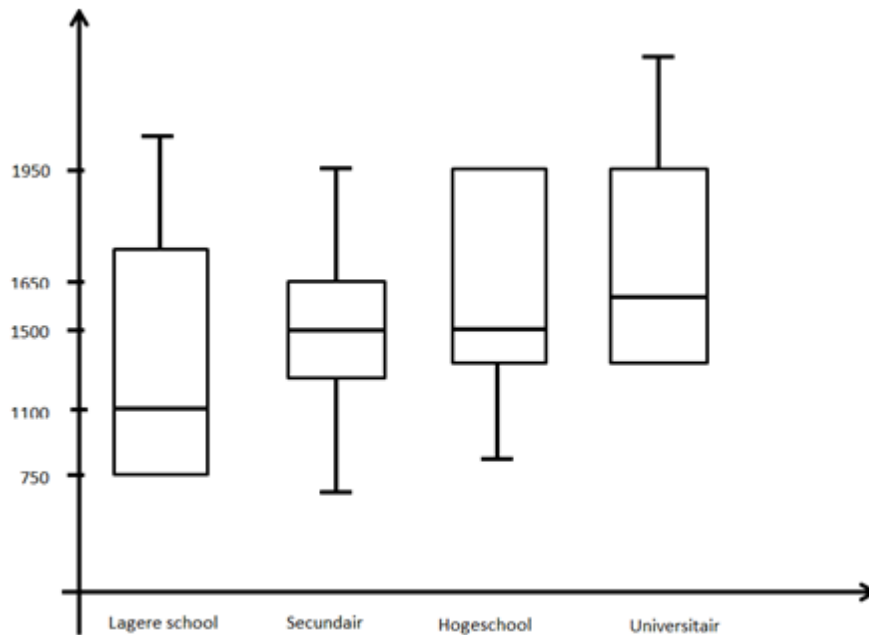
Noot: Hoge waarden verwijzen naar een goede kwaliteit van de familieband.

- a. Wat geeft een histogram weer ?

- b. Beschrijf de scheefheid van deze verdeling.

- c. Waar bevindt zich het gemiddelde ten opzichte van de mediaan in deze verdeling ?

13. De verdeling van het maandloon wordt voor vier opleidingscategorieën weergegeven in figuur 1. Welke uitspraak is **JUIST** ?



- Personen met een hogeschool diploma verdienen gemiddeld 1500 euro per maand; personen met een diploma van de lagere school verdienen gemiddeld 1100 euro per maand.
- 50% van de respondenten met een hogeschool diploma verdient tussen de 1500 en 1950 euro per maand; eenzelfde percentage van de respondenten met een diploma van de lagere school verdient tussen de 750 en 1100 euro per maand.
- Het maandloon varieert het meest bij personen met een secundair diploma: daar is de spreiding het grootst.
- Het percentage mensen met een secundair diploma dat tussen de 1500 en 1650 euro per maand verdienen, is kleiner dan het percentage mensen met een diploma hogeschool dat tussen 1500 en 1950 euro per maand verdienen

14. Het gemiddelde examenresultaat op het statistiekexamen bedraagt 9.3 op 20 met een variantie van 2.4.

Stel dat er een fout in het examen zit en dat de studenten elk 2 punten extra verdienen. Wat gebeurt er met het GEMIDDELDE ? Wat gebeurt er met de VARIANTIE ?

- Het gemiddelde en de variantie veranderen niet.
- Het gemiddelde wordt groter en de variantie wordt kleiner.
- Het gemiddelde wordt groter en de variantie blijft dezelfde.
- Het gemiddelde en de variantie worden allebei groter.

15. Het gemiddelde examenresultaat op het examen statistiek bedraagt 9.3 op 20 met een variantie van 2.4.

Stel dat de bekendmaking van de resultaten op 100 gebeurt. We moeten alle aangepaste scores vermenigvuldigen met een factor 5. Wat gebeurt er nu met het GEMIDDELDE ? En met de VARIANTIE ?

- Het gemiddelde en de variantie worden allebei 5x groter.
- Het gemiddelde wordt 5x groter en de variantie wordt 5x kleiner.
- Het gemiddelde wordt groter en de variantie blijft dezelfde.
- Het gemiddelde wordt 5x groter en de variantie wordt 5^2 keer groter

16. Bob tracht de vijf-getallen samenvatting te berekenen voor zijn examenscores. Zijn resultaten zijn als volgt:

Minimum = 30

Maximum = 90

Q1 = 50

Q3 = 80

Mediaan = 85

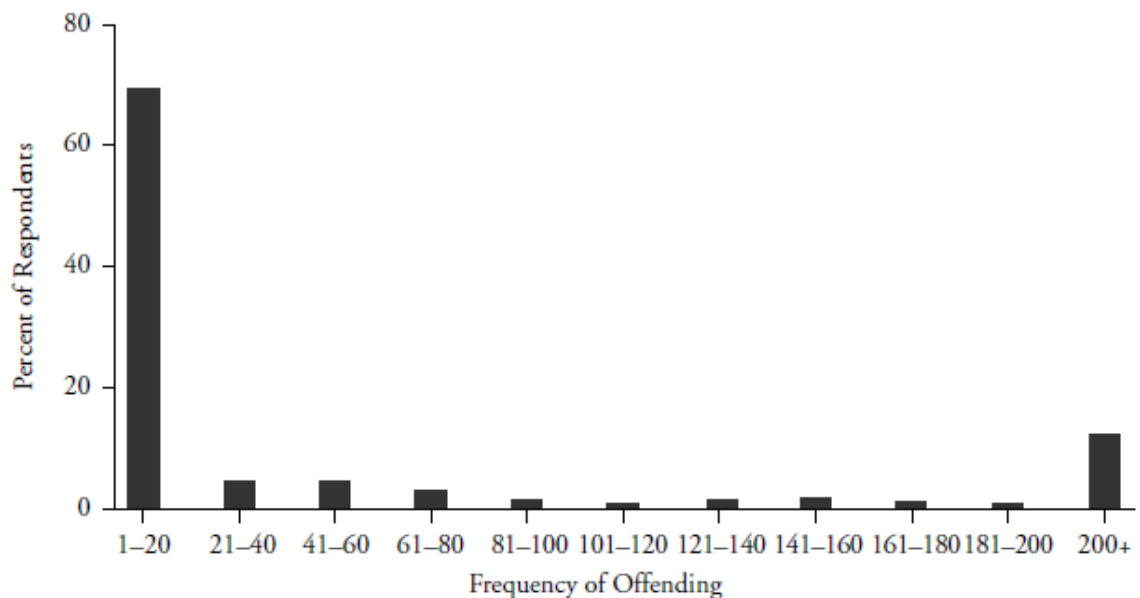
Vraag: Wat zit fout aan de vijf-getallen samenvatting van Bob?

17. Statistiek in de praktijk: vergelijking tussen mediaan en gemiddelde.

- De algemene regel is dat het gemiddelde de beste maat is voor de centrale tendens van een metrische variabele...

Leg uit.

- ... maar als de verdeling van een variabele sterk scheef is (skewed), dan geeft de median een betere schatting van de centrale tendens.
 - Wat betekent: "skewed"(scheef verdeeld)?
 - Wat is een 'negatieve skew'?
 - Wat is een 'positieve skew'?
- Bekijk onderstaande verdeling van zelfgerapporteerde criminaliteit van veroordeelde personen. Op de X-as staat het aantal eerdere delicten. Het gemiddelde in deze verdeling is 175 eerder gepleegde delicten.



Beschrijf in eigen woorden de vorm van deze verdeling. Welke maat van centraliteit zou jij verkiezen om de centrale tendens te beschrijven?

18. Een klein boekhouder kantoor betaalt aan ieder van zijn vijf administratieve medewerkers €25.000, aan twee junior-boekhouders ieder €60.000, en aan de eigenaar van de firma €225.000.

- Hoe groot is het door deze firma betaalde gemiddelde salaris?
- Hoeveel werknemers ontvangen minder dan het gemiddelde?
- Hoe hoog ligt het mediale salaris?
- Dit jaar geeft de firma enkel aan de eigenaar een salarisverhoging tot €455.000. Welk effect heeft deze wijziging op het gemiddelde?
- Welk effect is er voor de mediaan?

19. Een onderzoeker controleerde de reactietijden van de politie op tien noodoproepen telefoongesprekken. In de onderstaande gegevens staat het aantal minuten dat verstreken tussen het einde van het telefoongesprek en het tijdstip waarop de politie arriveerde:

24 26 14 27 198 22 27 17 19 29

- a. Bereken de modus, de mediaan en het gemiddelde
- b. Welke van deze centraliteitsmaten is het meest geschikt? Verklaar jouw antwoord.

20. In een paper over extreem weer schrijft een student het volgende:

“In de meeste staten van de US komen soms orkanen voor. Wanneer een orkaan toeslaat, zijn de verliezen catastrofaal. De gemiddelde jaarlijkse verliezen zijn geen zinvolle maat voor schade voor deze zeldzame, potentieel catastrofale gebeurtenissen”

Leg uit waarom student gelijk heeft.

21. UITSPRAKEN

Uitspraak	Juist	Fout
Als de skewness nul bedraagt, wijkt de verdeling van een kenmerk niet af van de normaalverdeling.		
De som van alle gekwadrateerde afwijkingen van het gemiddelde is nul.		
De mediaan is het middelpunt van de verdeling, namelijk de waarde van de variabele die toelaat de waarnemingen in twee gelijke delen op te splitsen.		
Bij de berekening van het rekenkundig gemiddelde worden de waarden van alle geldige waarnemingen gebruikt.		
Het rekenkundig gemiddelde kan enkel worden berekend op interval niveau.		
De mediaan kan zowel berekend worden voor een variabele gemeten op het nominaal als op het ordinaal meetniveau.		
Het 50ste percentiel is gelijk aan de mediaan of aan het vijfde deciel.		
Bij nominale data is er geen sprake van rangorde, de waarde die een variabele kan aannemen heeft alleen de betekenis van een naam, een categorie. Onderzoekseenheden worden onderverdeeld in categorieën. Daarbij gelden twee belangrijke voorwaarden: exclusiviteit en exhaustiviteit. LEG UIT. Exclusiviteit = Exhaustiviteit =		

EXTRA OEFENINGEN

1. Binnen de Ugent wordt aan een nieuw griepvaccin gewerkt. Tijdens de tweede fase van het onderzoek naar de doeltreffendheid van dit vaccin wordt het middel uitgetest op een beperkt aantal proefpersonen. De Ugent vond enkele studenten bereid om aan het onderzoek deel te nemen. Van die proefpersonen werd de voornaam en de leeftijd genoteerd.

Voornaam	Leeftijd
Bart	23
Leen	27
Marijke	29
Johan	31
Dirk	36
Marcel	44
Sofie	28
Anne-Marie	33
Anneleen	22
Hans	27
Jeroen	24
Sonja	38
Herman	46
Tom	21
Mieke	30
Luc	37
Sarah	38
Patrick	45
Rita	48
Nathalie	26

- Hoe groot is N ?
- Wat is de modus?
- Bereken het rekenkundig gemiddelde.
- Bereken de mediaan.
- Hoeveel bedraagt de standaardafwijking?
- Is de verdeling van dit kenmerk symmetrisch, positief scheef of negatief scheef?

2. Gent werd voor de gelegenheid van een criminologische studie verdeeld in 4 gebieden: Noord, Zuid, Oost en West. Een student vergelijkt het aandeel van de inbraken van elk gebied. De relatieve frequentieverdeling van deze vier gebieden vind je in onderstaande tabel.

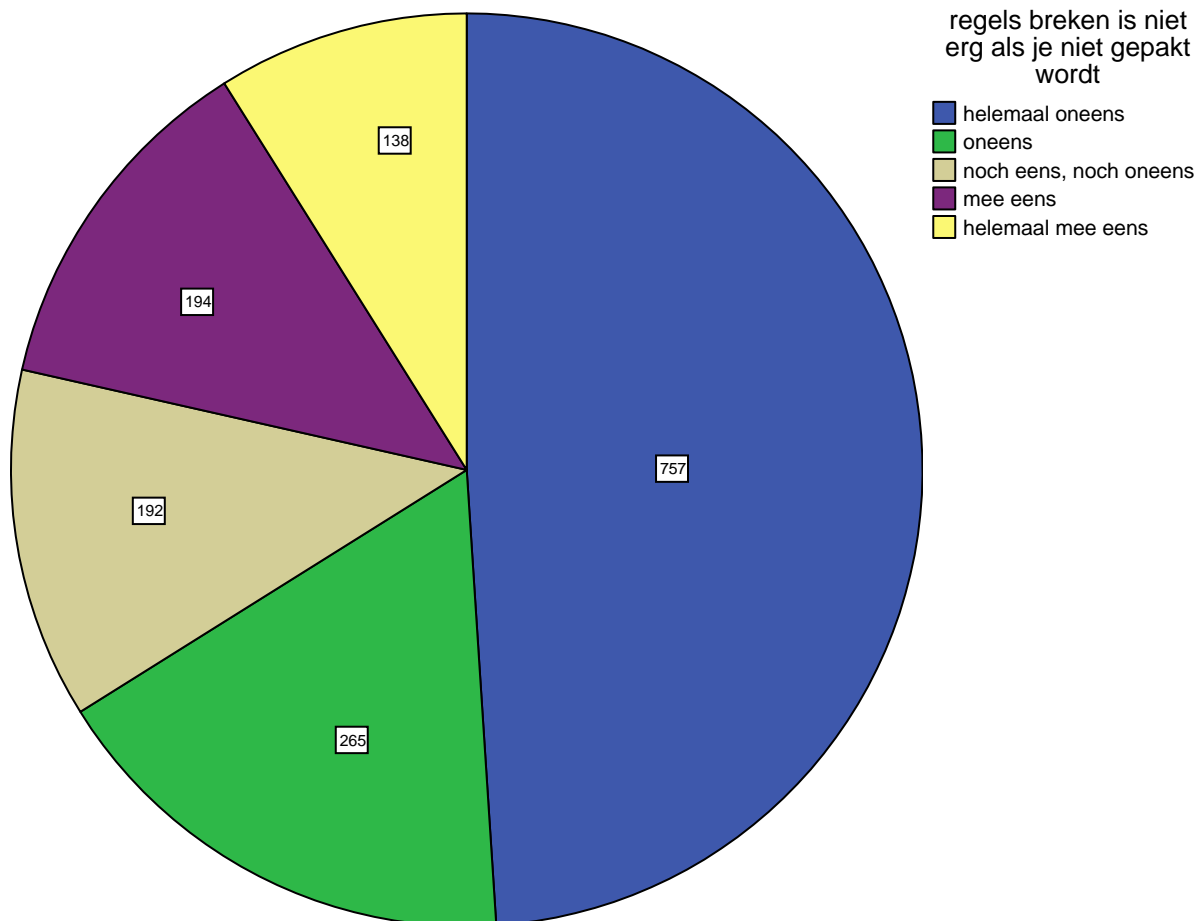
Gebied	Relatieve frequentie
Gent-Oost	0,18
Gent-Noord	0,22
Gent-Zuid	0,40
Gent-West	

- a. Wat is de relatieve frequentie van het gebied Gent-West?
 - b. Het absolute aantal inbraken bedraagt 800 door de politie geregistreerde gevallen. Wat is de absolute frequentie van Gent-Zuid?
3. Een vragenlijst gaf als antwoorden 68 maal ja, 42 maal nee en 10 maal weigering. Hieronder volgen twee uitspraken. Zijn deze juist of fout?
- a. N bedraagt $68+42+10$.
 - b. Valid N bedraagt $68 + 42$.
4. Hieronder vind je de ruwe data van het aantal e-mails dat een professor effectief moet beantwoorden per dag in een kleine faculteit gedurende een periode van twee weken.

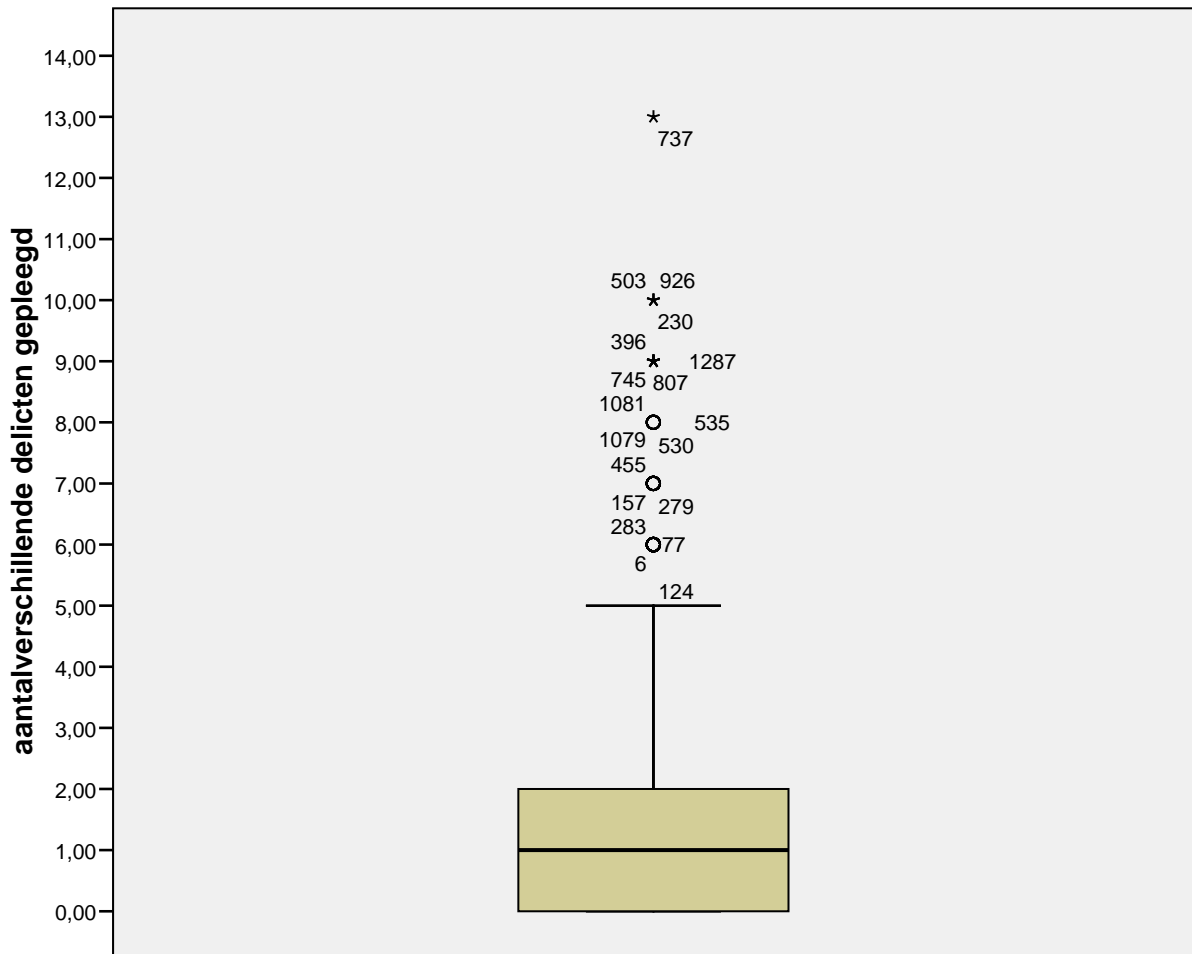
18 10 15 13 17 15 12 15 18 16

- a. Bereken de modus, de mediaan, het gemiddelde
- b. Bereken de standaardafwijking en de variatiecoëfficiënt.

5. Hieronder zie je een taartdiagram. De statisticus heeft echter de absolute waarden weergegeven.
- Geef de percentages.
 - Wat is de modus?
 - Bereken ook de variatieratio, de Index van diversiteit en de spreidingsmaat 'd' en interpreteer!



6. Hieronder zie je een box-plot. Bestudeer deze goed en beantwoord de uitspraken erover.



Uitspraak	Juist	Fout
Het eerste deciel valt samen met het eerste kwartiel.		
Er zijn twaalf outliers.		
Het derde kwartiel komt overeen met de waarde twee.		
De variabele is negatief asymmetrisch verdeeld.		

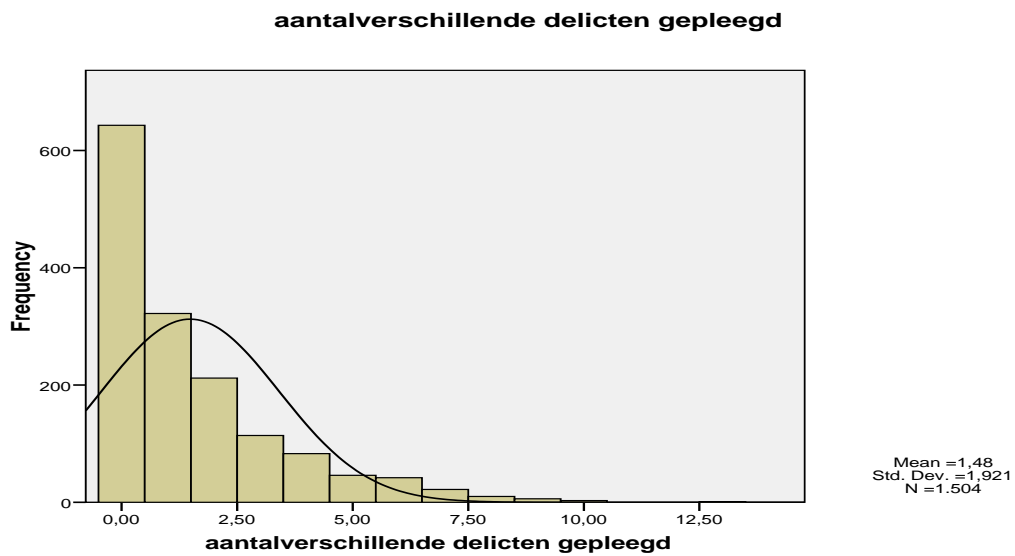
7. Bestudeer onderstaande frequentietabel aandachtig. Wat valt op het eerste zicht al op?

aantalverschillende delicten gepleegd

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid ,00	643	41,4	42,8	42,8
1,00	322	20,7	21,4	64,2
2,00	212	13,6	14,1	78,3
3,00	114	7,3	7,6	85,8
4,00	83	5,3	5,5	91,4
5,00	46	3,0	3,1	94,4
6,00	42	2,7	2,8	97,2
7,00	22	1,4	1,5	98,7
8,00	10	,6	,7	99,3
9,00	6	,4	,4	99,7
10,00	3	,2	,2	99,9
13,00	1	,1	,1	100,0
Total	1504	96,8	100,0	
Missing System	50	3,2		
Total	1554	100,0		

Statistics

	aantalverschillende delicten gepleegd
N Valid	1504
Missing	50
Skewness	1,679
Std. Error of Skewness	,063
Kurtosis	3,006
Std. Error of Kurtosis	,126



- Beschrijf de scheefheid en de platheid van de variabele *aantal delicten gepleegd*
- Wijkt de verdeling van deze variabele af van de normale verdeling ? Zo ja, hoe ?
Is deze verdeling met andere woorden positief of negatief asymmetrisch ?

8. Hieronder vind je de SPSS output van univariate beschrijvende statistieken. Bespreek de parameters waarvan je weet dat je deze mag bespreken op basis van het meetniveau. Opgepast, er zitten een aantal addertjes onder het gras!

Statistics				
		hoeveel van je vrienden hebben al eens gespijbeld?	regels breken is niet erg als je niet gepakt wordt	lid jeugdbende
N	Valid	1546	1546	1554
	Missing	8	8	0
Mean		,3234	2,1533	,06
Median		,0000	2,0000	,00
Mode		,00	1,00	0
Std. Deviation		,51783	1,37727	,247
Variance		,268	1,897	,061
Skewness		1,446	,833	3,533
Std. Error of Skewness		,062	,062	,062
Kurtosis		2,148	-,702	10,493
Std. Error of Kurtosis		,124	,124	,124
Range		3,00	4,00	1
Minimum		,00	1,00	0
Maximum		3,00	5,00	1
Percentiles	25	,0000	1,0000	,00
	50	,0000	2,0000	,00
	75	1,0000	3,0000	,00

- Variabele *hoeveel van je vrienden hebben al eens gespijbeld*
- Variabele *regels breken is niet erg als je niet gepakt wordt*
- Variabele *lid jeugdbende*

9. Vul de ontbrekende output verder aan.

Onderstaande tabellen uit een onderzoek van de Vlaamse regering geven de resultaten weer van een peiling naar het vertrouwen van Vlamingen ouder dan 18jaar in de politie en het gerecht. De respondenten zijn een steekproef van de Vlaamse bevolking. Beantwoord de volgende vragen:

VERTROUWEN POLITIE

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ZEER WEINIG	56	3,8	3,8	3,8
	WEINIG	191	12,9	13,0	16,8
	NIET VEEL EN NIET WEINIG	528	35,7	36,0	52,8
	VEEL	609	41,2	41,5	94,3
	ZEER VEEL	84	5,7	5,7	100,0
	Total	1468	99,4	100,0	
Missing	Weet niet	7	,5		
	Geen	2	,1		
	Total	9	,6		
Total		1477	100,0		

Vertrouwen Gerecht

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	ZEER WEINIG	154	10,4	10,8
	WEINIG	433	29,3	30,3
	NIET VEEL EN NIET WEINIG	512	34,7	35,8
	VEEL	299	20,2	20,9
	ZEER VEEL	32	2,2	2,2
	Total	1430	96,8	100,0	
Missing	Weet niet	44	3,0		
	Geen antwoord	3	,2		
	Total	47	3,2		
Total		1477	100,0		

- a. Wat zijn de statistische eenheden ?
- b. Wat zijn de variabelen en wat is het meetniveau ?
- c. Hoeveel respondenten hebben deze vragen ingevuld (bespreek per vraag)?
- d. Wat is het verschil tussen het 'percent' en het 'valid percent'?
- e. Vul op de stippellijnen de cumulatieve percentages aan? Wat geven de cumulatieve percentages weer?
- f. Welke zijn hier de relevante parameters van centraliteit?
- g. Bespreek deze relevante parameters van centraliteit.
- h. Hebben de respondenten over het algemeen meer vertrouwen in de politie of in het gerecht?
- i. Op welke manier zou je deze frequentietabellen grafisch voorstellen

