Hoofdstuk 9
Inferentiële statistiek en variantieanalyse

1. Waarom gebruiken we inferentiële statistiek?

Wanneer we vragen stellen over daders van bepaalde types criminaliteit, over het effect van roken op longkanker,… dan hebben die vragen meestal betrekking op de hele populatie. Maar het is per definitie onmogelijk om alle leden van een bepaalde populatie te onderzoeken. Zoiets is bovendien vaak te tijdrovend en te kostelijk, zelfs al kan men de populatie identificeren. Er zijn situaties waarin we over populatiegegevens beschikken, maar voor een beperkt aantal kenmerken. We mogen tot slot ook niet vergeten dat we in het domein van de criminologie ook in die gevallen waarin men beschikt over populatiegegevens, voorzichtig moet zijn. Populatiedata over criminele feiten weerspiegelen een fractie van de totale criminaliteit. Onze data moeten steeds met de nodige voorzichtigheid worden benaderd.

We verkiezen steeds steekproeven die een zo getrouw mogelijke afspiegeling zijn van de populatie en die, in theorie, een excellente basis vormen om besluiten naar de ganse populatie door te trekken. Een steekproef is steeds een subgroep van de populatie die moet bestudeerd worden. Vooraleer we kenmerken in steekproeven gaan bestuderen, is het belangrijk dat we ons de vraag stellen hoe goed de steekproef is. Is de steekproef werkelijk een getrouwe en representatieve afspiegeling van de populatie. Speelt dit dan zo een rol? Natuurlijk! Want als je een steekproef neemt waarbij een belangrijk segment van de populatie niet werd bevraagd, of heel erg ondervertegenwoordigd is, wat heb je dan aan zulke resultaten? De kennis over de steekproeftrekking komt in andere opleidingsonderdelen aan bod. In dit handboek besteden we aandacht aan de mechanismen die we gebruiken om verbanden door te trekken naar een populatie op basis van een steekproef en gaan er vanuit dat onze steekproeven voldoen aan de kwaliteitsvoorwaarden. Als een steekproef niet voldoet aan de essentiële kwaliteitsvoorwaarden, zoals random-selectie en dergelijke meer, dan is het in feite gevaarlijk om gebruik te maken van de inferentiële statistiek.

Maar nu terug naar onze steekproeven en de inferentiële statistiek. In het onderzoeksproces verzamelen we steeds een beperkt aantal kenmerken voor elke onderzoekseenheid die tot de steekproef behoort: dit noemen we de ruwe data of ruwe gegevens (raw data in het engels). Deze ruwe data worden gebruikt om variabelen van te maken die bruikbaar zijn voor criminologisch onderzoek. Criminologen proberen op basis van die partiële informatie afkomstig uit steekproeven uitspraken te doen over een gehele populatie. Om de resultaten op basis van deze steekproeven te kunnen veralgemenen naar de populatie, maakt men gebruik van de *inferentiële statistiek*, die gebaseerd is op de principes van het kansrekenen. Dankzij de inferentiële statistiek kunnen we uitspraken doen over een breder geheel aan eenheden dan diegenen die we bevraagd hebben. Dit doen we via het gebruik van *betrouwbaarheidsintervallen* en *significantietoetsen*. In dit deel staat *in de eerste plaats* het verwerven van inzicht in de basisprincipes van het infereren, i.e. het doen van uitspraken over een populatie op basis van een steekproef, centraal. *In de tweede plaats* staat de interpretatie van significantietoetsen en betrouwbaarheidsintervallen centraal.

2. De representativiteit van steekproeven

In dit deel van de cursus statistiek in de criminologie worden een aantal basisbeginselen van de inferentiële statistiek besproken. Het gaat hier slechts om een beperkte inleiding die ons toelaat om met enig inzicht te oordelen over het verwerken van gegevens en zonder te grote fouten een analyse uit te voeren. Deze inleiding volstaat echter niet om op eigen houtje een grondige en gefundeerde analyse van gegevens te kunnen uitvoeren.

Hiervoor hebben we reeds gesteld dat het doel van de *beschrijvende statistiek* bestaat in het op een overzichtelijke en synthetische wijze weergeven van alle gegevens die we voor een populatie of steekproef hebben verzameld. We ‘beschrijven’ deze gegevens van de steekproef of de populatie op verantwoorde wijze maar kunnen geen causale effecten (of oorzaak-gevolg) “bewijzen”. We kunnen wel aan de hand van beschrijvende statistiek mogelijke ideeën opdoen over de samenhang van kenmerken en hypothesen opstellen die dan via de *inductieve statistiek* moeten getoetst worden. Het doel van de inductieve statistiek bestaat uit het *veralgemenen* van de gegevens verzameld voor een steekproef naar de populatie waaruit ze getrokken werden. Inductie betekent dat we van het bijzondere (de steekproef) naar het algemene (de populatie) gaan. In onderzoek waar we stellingen uit theorieën toetsen vertrekken we vanuit algemene theorieën en willen die getoetst zien in een populatie. Dat is deductie: het afleiden van specifieke veronderstellingen uit algemene theorieën over criminaliteit. Als we dan de steekproefdata hebben verzameld en geanalyseerd, moeten we onze steekproefgegevens doortrekken naar de populatie en dat laatste is een inductief element. Dus: ook in deductief onderzoek zit altijd een stukje inductie.

Deze veralgemening van resultaten en het toetsen van hypothesen is echter niet eenvoudig. Er treedt vaak een aantal verstorende factoren op bij een onderzoek die een invloed uitoefenen. Zo kan men bv. door onprecieze metingen bepaalde onjuiste waarnemingen bekomen. Foutieve waarnemingen kunnen het gevolg zijn van slechte meetinstrumenten. Dat probleem bestaat ook in andere domeinen. Het is geen exclusief probleem van de sociale wetenschappen en de criminologie, echter, het is wel zo dat het in de sociale wetenschappen vaak iets moeilijker is om precieze metingen uit te voeren. Waarnemingen gedaan op een bepaald ogenblik in de tijd, kunnen later andere resultaten opleveren. We spreken van de ***meetfout***. Deze kan toevallig zijn of systematisch. Toevallige meetfouten zijn minder erg dan systematische meetfouten. Toevallige meetfouten heffen elkaar op, maar systematische meetfouten zorgen ervoor dat we de bal compleet mis slaan: we meten immers steeds onvolledig, omdat ons meetinstrument defect is. Vergelijk de situatie waarin je herhaaldelijk meet met een slecht afgestemd instrument. Bij herhaalde metingen heb je dezelfde uitkomst, maar deze is telkens fout. Zelfs indien we alle regels bij het trekken van steekproeven volgen, kunnen er vertekeningen optreden in de steekproef. Men spreekt van de ***steekproeffout***. Aangezien een steekproef slechts een deel van een populatie is, kunnen verschillen gemeten in de ene steekproef groter of kleiner zijn in de populatie. M.a.w. er zijn onzekerheden die inherent zijn aan het proces van het generaliseren naar velen op basis van weinigen. Statistische inferentie houdt zich met deze onzekerheden bezig door 2 functies te vervullen:

- **schatting**: het gebruik van steekproefkenmerken om ze naar de populatie te veralgemenen. bv. we voorspellen op basis van de meting in de steekproef die opleverde dat 49% vindt dat de regering haar werk goed doet, dat 49 % van de bevolking vindt dat de regering haar werk goed doet. Hoe precies is deze schatting? Steunend op de waarschijnlijkheidstheorie zullen we hier naar een antwoord zoeken.

- **hypothesen testen of toetsen**: kan men met voldoende zekerheid een bepaalde onderzoekshypothese aannemen of moet men ze op basis van de gegevens verwerpen?

In statistische inferentie zijn de vier vermelde begrippen (**populatie, steekproeven, schatting en toetsing**) cruciaal. We zullen ze daarom in dit deel één voor één toelichten en hier in eerste instantie blijven stilstaan bij het onderscheid populatie versus steekproef.

3. Steekproeven en populatie

De relevante populatie voor een studie noemen we **de set (groep/reeks) van personen of objecten waarin een onderzoeker geïnteresseerd is**. Soms kunnen we gegevens verzamelen voor alle individuen uit een populatie, maar vaak wordt slechts een deel van de populatie onderzocht: dit deel noemen we de steekproef. Een eerste stap in onderzoek is zich de vraag te stellen wat de populatie is die men wenst te onderzoeken. Deze populatie kunnen we afleiden uit de onderzoeksvraag.

- *Voorbeeld 1. Heeft de detentieduur een effect op het zelfbeeld van gedetineerden? Hier bestaat de populatie uit alle gevangenen.*
- *Voorbeeld 2. Wordt het zelfbeeld van personen die geïnterneerd zijn beïnvloed door processen van etikettering op basis van het type delict? Hier bestaat de populatie uit geïnterneerden.*
- *Voorbeeld 3. Wat zijn de belangrijkste demografische achtergrondkenmerken van slachtoffers van seksuele delinquentie? Hier bestaat de populatie uit alle slachtoffers van seksuele delinquentie.*
- *Voorbeeld 4. Genieten politiezones waar de "community policing" filosofie aanhangen wordt een hoger aanzien onder de bevolking dan politiezones waar dat niet het geval is. Hier is de populatie de bevolking. Als je dacht dat het de politiecommissariaten waren, dan heb je je laten misleiden. We zijn geïnteresseerd in de attitudes van burgers ten aanzien van politiezones. Community policing is een kenmerk van politiezones. De formulering heeft je wellicht misleid. Lees dus steeds goed de formulering.*
- *Voorbeeld 5. Worden werklozen sneller daders van criminaliteit dan werkenden? Hier bestaat de populatie uit werklozen en werkenden.*
- *Voorbeeld 6. Welke vormen van criminaliteit komen het meeste voor in België? Hier bestaat de populatie uit alle criminaliteit in België, dit wil zeggen alle feiten die een strafrechtelijke vervolging kunnen hebben. En dat zijn er heel wat! België heeft enorm veel bijzondere strafwetten. Sommigen stellen zich daarom de vraag of het systeem niet moet vereenvoudigd worden.*

Wanneer men een onderzoekspopulatie kent, moet men zich afvragen of het om een ***afbakenbare*** of een ***hypothetische*** *populatie* gaat.

Afbakenbare populaties

In vele gevallen kan de populatie onder studie fysiek in kaart worden gebracht. De gegevens opgeslagen in computers of in een kaartenbakkensysteem leveren ons lijsten. Zo worden alle mogelijke personen die mogen stemmen op lijsten bijgehouden, beschikt men over een aantal kenmerken van de bevolking via de bevolkingsregisters, houden gevangenissen lijsten bij van degenen die opgesloten worden, houden jeugdrechters dossiers bij van degenen die ze begeleiden en plaatsen, enz.

Bekijken we de eerder geformuleerde onderzoeksvragen dan blijken voorbeelden 1, 2, 4 en 5 over afbakenbare populaties te gaan.

Hypothetische populaties

In het geval van het bestuderen van de invloed van roken op longkanker, de criminaliteitscijfers, cijfers over slachtofferschap,… is de populatie veel minder duidelijk afgebakend. (alle rokers, alle daders of slachtoffers, ….) Men noemt ze **hypothetische populaties** omdat we ze niet op lijsten terugvinden of kunnen identificeren. De populatie kan zelfs elementen en mensen inhouden die nu nog niet bestaan; bv. het aantal kinderen dat volgend jaar zal gemaakt worden. Filosofisch gezien kunnen volledige populaties waarover we gegevens hebben, bijvoorbeeld alle Belgen die in 2010 geregistreerd staan, ook beschouwd worden als een “steekproef van alle Belgische populaties uit alle voorbije jaren”. Toegegeven, het is misschien wat ver gezocht. Sommige criminologen zien hierin een reden om ook op populatiegegevens statistische significanties toe te passen. De meningen zijn hier evenwel over verdeeld. Wij menen alvast dat het geen kwaad kan om ook op populatiegegevens ook inferentiële statistiek toe te passen. Waarom? Wel, het argument dat vaak gebruikt wordt is dat een statistische significantietoets meer ziet dan het blote oog alleen. Wat daarmee bedoeld wordt, zal duidelijk worden wanneer we de variantieanalyse uiteenzetten.

4. Steekproeven en het principe van toeval

Een steekproef moet zo goed mogelijk de relevante kenmerken van een populatie vertegenwoordigen; ze moet met andere woorden *representatief* zijn. In de praktijk zullen we steeds spreken over een *zeer goede afspiegeling* in plaats van een perfecte afspiegeling. Om een goede afspiegeling te krijgen zullen we de steekproef lukraak samenstellen (*‘random sample’ of toevalssteekproef*). Door statistieken te berekenen of te rapporteren voor een steekproef in plaats van voor de gehele populatie, treedt er een fout op die **steekproeffout** wordt genoemd. Deze steekproeffout wordt bepaald en onder controle gehouden door voldoende aandacht te besteden aan het steekproefkader, het steekproefontwerp en de implementatie ervan. Het steekproefkader geeft weer wie - administratief - deel uitmaakt van de te onderzoeken doelpopulatie, en bijgevolg kans heeft of moet hebben om in de steekproef opgenomen te worden. Zo beogen nationale enquêtes naar onveiligheidsbeleving en slachtofferschap in eerste aanleg representativiteit voor ‘de bevolking van 15 jaar en ouder’.

Een steekproef is bovendien pas geslaagd indien zij op de relevante kenmerken waarin de criminoloog geïnteresseerd is, de populatie goed weerspiegelt. Men zegt dan dat de steekproef voor deze relevante kenmerken **representatief** is. ‘Representativiteit' betekent dat de vermelde kenmerken met de juiste aantallen aanwezig zijn in de steekproef. Het heeft geen enkele zin, en is zelfs misleidend, om een steekproef zomaar - zonder de kenmerken te vermelden - 'representatief' te noemen en zo de indruk te wekken dat de steekproef voor alle kenmerken een goede weergave is; immers, dat is een steekproef vrijwel nooit. In de meer populariserende ‘polls’ of opiniepeilingen (zoals bijvoorbeeld in het peilen naar het kiesgedrag van mensen) wordt dit voortdurend gedaan om in de media de indruk te wekken dat wat verteld wordt, ‘waar’ is.

In de praktijk kan de representativiteit van een steekproef vaak maar worden aangetoond voor een beperkt aantal kenmerken, zoals sekse, leeftijd, regio. In criminologisch onderzoek is dit niet zelden des te problematisch aangezien crimineel gedrag of slachtofferschap in belangrijke mate verborgen is. Slachtofferenquêtes en zelfrapportagestudies pogen beiden om een bepaald “***dark figure”***-problematiek op te heffen. Deze studies worden in hoofdzaak uitgevoerd net omwille van het ontbreken van een externe en betrouwbare bron met betrekking tot de fenomenen van daderschap en slachtofferschap. Het spreekt voor zich dat de representativiteit van de steekproef, hoe relatief dit vaak is, hierbij uiteraard van cruciaal belang is.

5. De theorie van toevalssteekproeven

Toevalssteekproeven zijn **aselecte** steekproeven. Het zijn steekproeven waarin elke elementaire eenheid uit de empirische populatie een **bekende** (lees: **berekenbare**) **kans** heeft om in de steekproef opgenomen te worden. Wanneer men het toeval laat spelen, kan de kans berekend worden aan de hand van de spelregels van de kansrekening, die reeds in deze syllabus aan bod kwamen. Het basismodel van de toevalssteekproef is een **enkelvoudige aselecte steekproef** of ‘**simple random sample’**. In de steekproeventheorie maakt men doorgaans een onderscheid tussen **steekproefgrootheden** en **populatieparameters**. Statistieken afkomstig uit steekproeven zoals het gemiddelde, de standaardafwijking, een richtingscoëfficiënt,… zijn steekproefgrootheden. Populatieparameters verwijzen naar het gemiddelde, de standaardafwijking, een richtingscoëfficiënt,… in de populatie. Die waarde is per definitie niet gekend. We vertrekken immers van de steekproef om iets te zeggen over de populatie.

Bij de louter beschrijvende statistiek wordt er geen onderscheid gemaakt tussen steekproef en populatie. Bij **inferentiële statistiek** is het verschil tussen steekproef en populatie essentieel. De eigenschappen van de steekproef zijn **bekend** (*gemiddelde, minimum, maximum, spreiding, enz*.). De steekproef is op zich niet interessant. We gebruiken hem om iets te weten te komen over de populatie. De eigenschappen van de populatie zijn **onbekend** (gemiddelde, minimum, maximum, spreiding enz.). We trekken een steekproef om informatie te krijgen over een onbekende populatie.

In de inferentiële statistiek doen we uitspraken over de populatie op basis van wat we vinden in een ***aselecte steekproef*** uit die populatie. Anders gezegd: we gebruiken steekproefkenmerken om iets te zeggen over populatiekenmerken.
De onbekende kengetallen van de populatie geven we aan met Griekse letters, bijvoorbeeld **μ** (mu), **σ** (sigma) en **π** (pi). Zulke kengetallen worden de parameters van de populatie genoemd.

	In de steekproef: steekproefkenmerk	**In de populatie: parameter**
gemiddelde	$\bar{x}$	μ
variantie	s^2	σ^2
standaardafwijking	s	σ
fractie	p	π
omvang	n	**N**

Zoals we daarnet al even kort aangehaald hebben, zijn er **twee hoofdactiviteiten** binnen de inferentiële statistiek:

- **Schatten**: we berekenen een steekproefkenmerk (bijvoorbeeld het aantal verschillende delicten waarvan men slachtoffer wordt) en gebruiken de waarde van dat kenmerk (bijvoorbeeld gemiddeld 2 delicten) om een uitspraak te doen over een populatiekenmerk (het gemiddeld aantal delicten waar men in de bevolking slachtoffer van wordt). Zo'n uitspraak is altijd gebaseerd op kansrekening. We schatten de parameter 'gemiddelde leeftijd' (symbool μ) met het steekproefkenmerk 'gemiddelde leeftijd' (symbool x).

We onderscheiden de begrippen **puntschatting** en **intervalschatting**. Onder puntschatting verstaan we de schatting van een kenmerk in de populatie op basis van steekproefgegevens. De schatting van het gemiddeld aantal delicten waarvan iemand slachtoffer wordt, is een puntschatting. Onder intervalschatting verstaan we de marges waarbinnen we met een zekere graad van onzekerheid een puntschatting inschatten. We spreken dan van ***betrouwbaarheidsintervallen*.**

- **Toetsen**: we veronderstellen dat er een verband bestaat tussen leefstijl en beroving op straat van geweld. Wie 's avonds vaker uitgaat, heeft meer kans om beroofd te worden op straat.
  (→ **hypothese** = er is een positieve correlatie tussen het aantal avonden in de week dat men zich in het nachtleven stort en het aantal keer dat men slachtoffer wordt van een beroving). We vinden een positieve correlatie in onze steekproef, maar dat betekent per definitie niet dat dit ook zo is in de populatie. Onze uitspraak is gebaseerd op **kansrekenen**. Kansrekening vormt zo de basis van de inferentiële statistiek. Als we op grond van een steekproefresultaat iets zeggen over een onbekende populatie is dat altijd een onzekere uitspraak. In dit deel van de syllabus leren we met deze onzekerheid om te gaan.

We kunnen de theorie die achter het schatten van de kans dat een steekproefgrootheid een goede weerspiegeling vormt van de onbekende populatieparameter (bijvoorbeeld de proportie jongens in een steekproef) duidelijk maken aan de hand van een **experiment,** waarbij we in omgekeerde volgorde te werk gaan. Dit kunnen we doen door uit een zekere populatie - stel alle criminologiestudenten die in de eerste bachelor aan de Ugent studeren - meerdere kleinere steekproeven van eenzelfde omvang - stel 100 eenheden - te trekken met bekende parameter, zoals het geslacht van de studenten. Het gaat hier om een werkelijke populatie. Via de studentenadministratie weten we dat de proportie mannelijke studenten 40% bedraagt. De verkregen steekproefuitkomsten (de proportie jongens in elke afzonderlijke steekproef) zijn telkens **frequentieverdelingen**. De toevalsvariabele of steekproefuitkomst waarin we geïnteresseerd zijn, is de waarde van de proportie jongens in verschillende steekproeven van 100 eenheden.

Indien we dergelijke oefening repetitief zouden uitvoeren, zou blijken dat naarmate het aantal toevalsteekproeven van dezelfde omvang (in ons voorbeeld is n = 100) groter wordt, de concentratie van de uitkomsten rond de werkelijke waarde toeneemt. De werkelijke waarde is de waarde in de populatie, die we kennen via de studentenadministratie. Het aantal steekproeven met uitkomsten aan de uitersten van de steekproevenverdeling neemt, relatief gesproken, af. Merk ook op dat de uitkomst van sommige particuliere steekproeven soms een heel eind verwijderd kan zijn van de populatiewaarde (de proportie in de originele steekproef die dienst deed als populatie).

Wat betekent deze informatie nu concreet? De vraag die zich stelt is immers of we met de informatie over de steekproevenverdeling in het achterhoofd tot een *betrouwbare schatting* kunnen komen van de - onder normale omstandigheden onbekende - populatiewaarde; kunnen we met andere woorden op basis van de steekproefverdeling van een singuliere toevalssteekproef een betrouwbare schatting maken van de prevalentie van een zeldzame ziekte, of van het herhaald slachtofferschap, of het aantal harde kernjongeren? Op basis van de kennis omtrent steekproevenverdelingen weten we dat de kans om heel ver van de populatiewaarde te zitten niet zo bijzonder groot is indien het toevalskarakter van onze steekproef werd gerespecteerd.

6. Kenmerken van steekproevenverdelingen

In de inferentiële statistiek spreken we steeds over een aantal theoretische kansverdelingen waarin berekend wordt welke kans met een bepaalde waarde van een stochastische variabele verbonden is. Een (theoretische) steekproevenverdeling is altijd een theoretische kansverdeling die de functionele relatie toont tussen de mogelijke waarden van een bepaalde statistiek, gebaseerd op een steekproef van n eenheden, en de kans (dichtheid) verbonden met elke waarde, voor alle mogelijke steekproeven van identieke omvang n die uit een specifieke populatie getrokken worden. Steekproevenverdelingen hebben enkele belangrijke eigenschappen waar onderzoekers gebruik van maken wanneer zij kwantitatief onderzoek doen op basis van steekproeven.

- **Ten eerste**: Indien de parameter die we willen schatten een gemiddelde is, dan weten we dat de verwachte waarde van het steekproevengemiddelde gelijk is aan het populatiegemiddelde. Dit geldt bij uitbreiding ook voor een percentage (of proportie) successen, aangezien dit eigenlijk een gemiddelde is van een variabele met code 1 (succes of kenmerk aanwezig) en code 0 (geen succes of kenmerk afwezig).

- **Ten tweede**: De variantie $\boldsymbol{\sigma_s^2}$ van de steekproevenverdeling van gemiddelden van onafhankelijke steekproeven met omvang n is steeds gelijk aan de populatievariantie $\boldsymbol{\sigma^2}$ gedeeld door de steekproefomvang n (dus $\boldsymbol{\sigma_s^2 = \sigma^2 / n}$). De vierkantswortel uit de variantie van de steekproevenverdeling - m.a.w. de standaardafwijking van het gemiddelde van de steekproevenverdeling -, wordt de **standaardfout** (of SE, 'Standard Error') genoemd.
  Dus: $\mathbf{SE} = \boldsymbol{\sigma_s} = \boldsymbol{\sigma} / \sqrt{n}$

- **Ten derde**: naarmate de steekproef groter wordt, neemt de kans toe dat het steekproefgemiddelde dichter bij het populatiegemiddelde komt. Dit is de wet van de grote getallen.

- **Ten vierde**: zowel het populatiegemiddelde als de populatievariantie kan worden geschat door het steekproefgemiddelde en de steekproefvariantie indien de steekproefomvang voldoende groot is.

- **Ten vijfde:** het gemiddelde en de standaardfout van de steekproevenverdeling kunnen bijgevolg ook geschat worden op basis van de parameters van een particuliere steekproef van voldoende omvang.

Deze eigenschappen stellen ons in staat om een schatting te maken van een onbekende populatieparameter en de betrouwbaarheid ervan volgens informatie over de parameters in de steekproef. Centraal hierbij staat echter dat het steeds om een schatting gaat met, inherent daaraan verbonden, een bepaalde mate van onzekerheid. Om echter gebruik te kunnen maken van bovenstaand theorema zouden we moeten weten, zonder zelf grote aantallen steekproeven te trekken, hoe de theoretische kansverdeling van een bij ons probleem passende steekproevenverdeling er uitziet.

In de inferentiële statistiek zijn een aantal theoretische steekproevenverdelingen met hun eigenschappen bekend en gedocumenteerd. De meest bekende is de standaard normale verdeling, waaraan we reeds in een afzonderlijk deel aandacht hebben besteed. Men weet dankzij de theoretische statistiek ook onder welke omstandigheden deze verdelingen van toepassing zijn. Zo kunnen bijvoorbeeld veel steekproevenverdelingen gebaseerd op steekproeven met een voldoende grote omvang n worden benaderd met de normale verdeling, ook al is de verdeling van het bestudeerde kenmerk in de populatie niet normaal. De criminoloog maakt hiervan gebruik bij het beantwoorden van probleemstellingen.

7. Het gebruik van de normale verdeling in de inferentiële statistiek

De normale verdeling is de 'koning onder de kansverdelingen', niet omdat zoveel kenmerken in de sociale werkelijkheid zo’n verdeling hebben. Waarom dan wel? Als we een reeks grote steekproeven trekken uit een populatie, dan weten we dat de steekproefgemiddelden normaal verdeeld zijn. Deze eigenschap van de *‘steekproefgemiddeldenverdeling’* is heel belangrijk in de inferentiële statistiek. Deze verdeling is symmetrisch en ééntoppig, gekenmerkt door 'klokvorm' en wordt volledig bepaald door de parameters **μ (populatiegemiddelde)** en **σ (standaardafwijking)**

We hebben eerder gezien dat de normale verdeling piekt en symmetrisch verdeeld is rond het gemiddelde. Om de normale verdeling toepasselijk te maken voor een veelheid aan empirische verdelingen, maakt men gebruik van transformaties waarbij elke score wordt uitgedrukt als een genormeerde afstand tot het gemiddelde van de verdeling. Dit zijn de **z-scores** waarbij men; zoals voorheen uiteengezet, het verschil neemt van een waarde en het gemiddelde en dit deelt door de standaardafwijking.

8. De centrale limietstelling

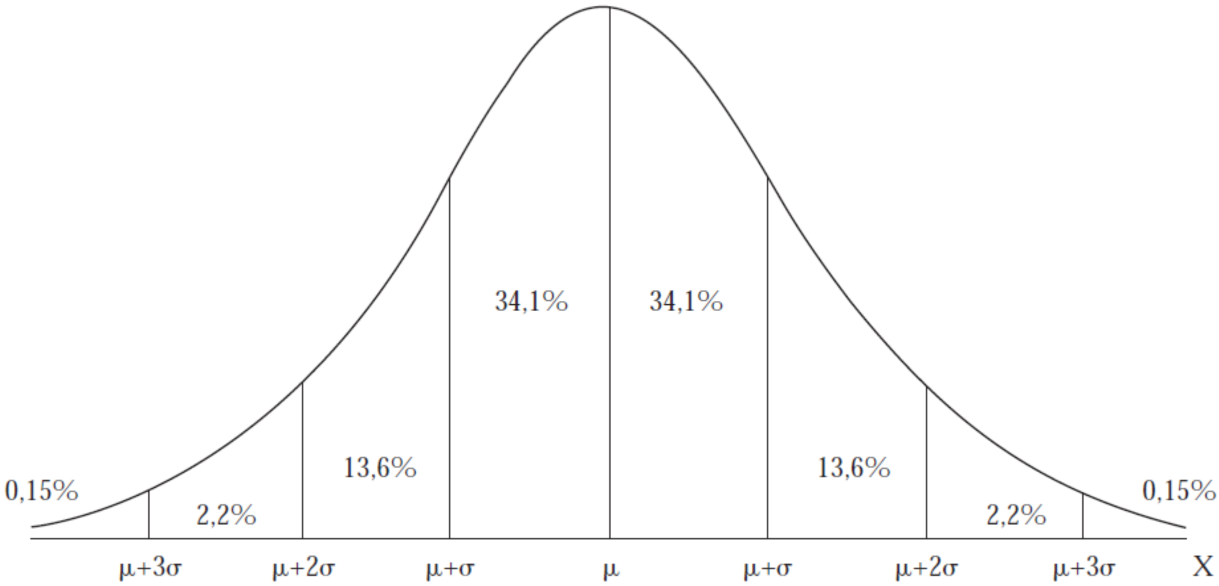
Voordat we de centrale limietstelling uitleggen herhalen we de drie bekende eigenschappen van de normale verdeling, met name de eerder besproken **68-95-99 regel**:

1. Ongeveer 68% van alle waarden valt binnen 1 standaardafwijking van het gemiddelde.
2. Ongeveer 95% van alle waarden valt binnen 2 standaardafwijkingen van het gemiddelde.
3. Ongeveer 99% van alle waarden valt binnen 3 standaardafwijkingen van het gemiddelde.

Het theorema van **de centrale limietstelling** is één van de belangrijkste begrippen uit de statistiek. Stel, je neemt heel veel steekproeven en van elke steekproef bepaal je het gemiddelde. Van al deze gemiddelden maak je een frequentieverdeling. Stel dat het gemiddelde van de hele populatie **μ** is, dan vormen al die steekproefgemiddelden (bij voldoende grote steekproeven) een normale verdeling rond **μ**. Dit is de centrale limietstelling.

Op basis hiervan kunnen we de kans berekenen dat alle steekproefgemiddelden in een bepaald interval rond μ liggen. De Centrale Limietstelling is een versterking en precisering van de Wet van de Grote Getallen. Hoe groter dit aantal, hoe dichter de verdeling bij een normale verdeling ligt.

Figuur: de normaalverdeling



Wanneer we de normaalverdeling toepassen op de steekproevenverdeling dan is een z-score:

$z_i = (x_i - \mu_s) / SE$

Hierin stelt xi elke mogelijke steekproefuitkomst voor.

Bij zo'n standaardnormaalverdeling (met gemiddelde 0 en standaardafwijking 1) ligt 99% van de oppervlakte onder de curve tussen -2,58z en +2,58z (eenheden standaardafwijking), 95 % tussen -1,96z en +1,96z en 90% tussen -1,64z en +1,64z eenheden standaardafwijking (de z-waarde). Met een kans van 5% op een vergissing ligt de werkelijke waarde met andere woorden tussen de waarden -1,96z en +1,96z. Met deze kans op vergissingen zijn sociale wetenschappers bereid te leven.

We kunnen steeds bepalen welke de kans is op een welbepaalde steekproefuitkomst indien we het populatiegemiddelde en de populatievariantie kennen. We kunnen dan immers de parameters van de steekproevenverdeling berekenen. In zo'n geval kunnen we ook met een vastgestelde waarschijnlijkheid de hypothese toetsen of een bepaalde steekproefuitkomst al dan niet uit die populatie afkomstig kan zijn. Indien we bovendien een voldoende grote toevalsteekproef getrokken hebben en we een populatieparameter of -statistiek (gemiddelde, proportie of percentage) willen schatten, kunnen we gebruik maken van de eigenschappen van de normaalverdeling.

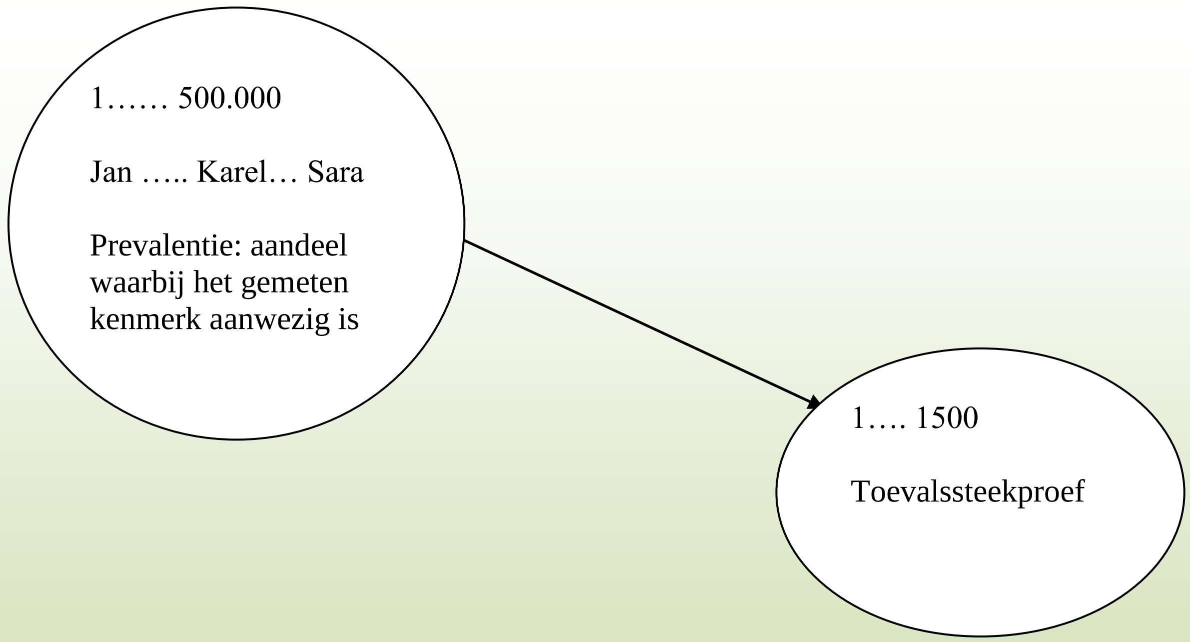
Deze procedure noemen we het *schatten en toetsen van parameters*. Met deze informatie is het mogelijk om, met een bepaalde waarschijnlijkheid en dus ook met een bepaalde kans op vergissing (bijvoorbeeld 5%), te bepalen binnen welk interval rond de steekproefuitkomst het populatiegemiddelde zal vallen. Zo'n interval wordt een **betrouwbaarbeidsinterval of -gordel** genoemd. In feite zal elke veralgemening van een steekproefwaarde naar de populatie rekening moeten houden met een interval dat qua breedte varieert naargelang de steekproefomvang, de mate van spreiding in de populatie (standaardafwijking), de gewenste betrouwbaarheid en het type van toevalsteekproef.

9. Puntschatting en intervalschatting

Uit het voorgaande bleek reeds dat statistische inferentie inhoudt dat men het resultaat van een steekproefonderzoek veralgemeent naar een volledige populatie. Belangrijke begrippen die men zeer goed dient te kennen zijn de begrippen *punt- en intervalschatting*. Zij vormen de basis van de statistische hypothesetoetsing en statistische inferentie, die verderop wordt uiteengezet. Met statistische inferentie kan men eigenschappen in populaties schatten met behulp van toevalsteekproeven. Veronderstel dat we geïnteresseerd zijn in herhaald slachtofferschap van geweldsdelicten in een populatie.

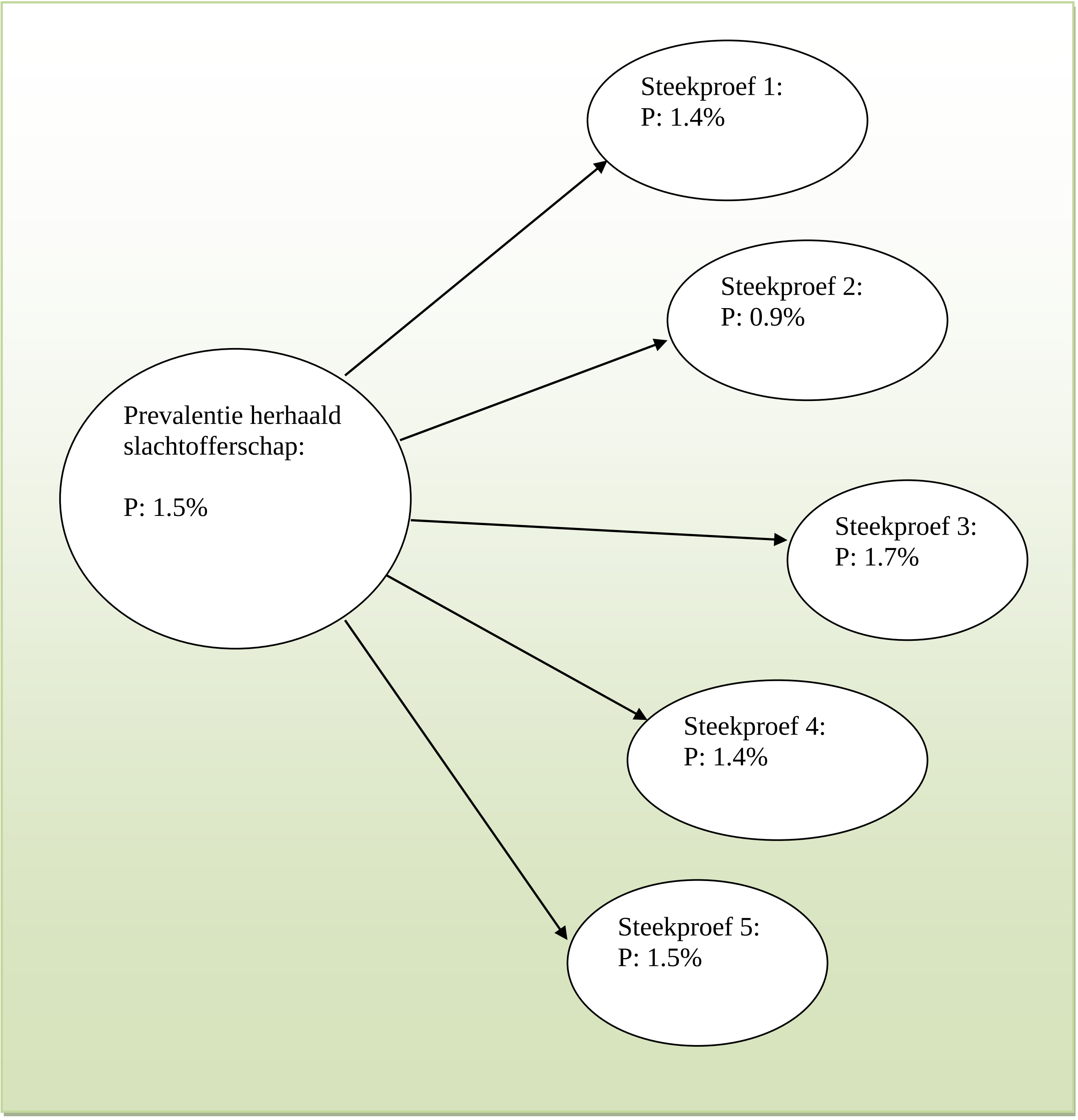
We gaan te werk in drie stappen.

(1) We nemen een toevalssteekproef van 1500 individuen
(2) We bevragen de individuen in de steekproef en noteren hoeveel individuen herhaald slachtoffer werden van een geweldsdelict
(3) We schatten de prevalentie



Figuur: Illustratie van een puntschatting en intervalschatting

Uit de toevalssteekproef blijkt dat 18/1500 het kenmerk bezitten. Dit is 1.2%. De prevalentie van het kenmerk 'herhaald slachtofferschap' wordt dus op 1.2% geschat. Dit is een puntschatting van de prevalentie met een zekere mate van onzekerheid. Om deze onzekerheid te illustreren, volgt hieronder een computersimulatie. We nemen vijf toevalssteekproeven waarbij we de prevalentie schatten. De populatie bestaat uit 500.000 individuen, waarvan 1.5% herhaald slachtoffer zijn geworden van een geweldsdelict.



Figuur: Illustratie van een puntschatting en intervalschatting (vervolg)

Een schatter is een grootheid die berekend wordt aan de hand van een bepaalde functie (algoritme) op basis van de informatie van de steekproef en wordt uitgedrukt in een formule. Een schatter is het resultaat van de schatting. We schatten steeds onbekende parameters in een populatie en dit op basis van informatie uit de steekproef.

Een *zuivere schatter* of *onvertekende schatter* (Engels: *unbiased estimate*) is dus een statistische grootheid waarvan de verwachtingswaarde samenvalt met de te schatten populatieparameter.

De geschatte prevalenties variëren tussen 0.9% en 1.7%. Hoe groot de variatie van de geschatte prevalenties tussen de verschillende steekproeven is, hangt samen met hoe groot de steekproeven zijn. Grotere steekproeven leiden tot een kleinere variatie in de prevalentieschattingen. We krijgen dus een zekerder resultaat als de steekproefgrootte toeneemt. Dit is een belangrijke statistische wetmatigheid.

In de praktijk van het onderzoek nemen we uiteraard genoegen met één steekproef. Op basis van deze steekproef zal een *intervalschatting* worden gemaakt. De intervalschatting geeft de (on)zekerheid van onze uitkomst weer. De schatting van het interval gebeurt aan de hand van **betrouwbaarheidsintervallen**. In het voorbeeld hebben we de prevalentie van het fenomeen herhaald slachtofferschap van een geweldsdelict geschat op 1.2%. Deze schatting geeft ons nog geen idee over de nauwkeurigheid ervan. Hoeveel vertrouwen kunnen we nu hebben in dit resultaat? De schatting van een betrouwbaarheidsinterval biedt ons meer precieze informatie. Een betrouwbaarheidsinterval is een schatting +/- een foutenmarge. Dit interval wordt berekend uit de steekproefdata volgens een methode die een bepaalde kans heeft een interval op te leveren waarin de populatiewaarde ligt. We baseren ons dus op het kansrekenen. Intervalschatting komt in het voorbeeld dus neer op het schatten van een interval dat ligt rond de geschatte proportie herhaalde slachtoffers in onze steekproef, en dat met een zekere graad van waarschijnlijkheid ook de onbekende populatieproportie van slachtoffers bevat.

De breedte van het interval geeft informatie over de precisie. De breedte van het interval kan worden berekend aan de hand van onze kennis over de steekproevenverdelingen van schattingen. Het komt er met andere woorden op neer dat, willen we die 1.2% veralgemenen naar de populatie, dit enkel kan in een uitspraak als volgt: ‘met 95% zekerheid kunnen we stellen dat de prevalentie van herhaald slachtofferschap van een geweldsdelict in de populatie ligt tussen de …% en …%’; het eerste percentage is dan ‘1.2% min het betrouwbaarheidsinterval’, het tweede percentage ‘1.2% plus het betrouwbaarheidsinterval’.

10. Het berekenen van een betrouwbaarheidsinterval rond een parameter

BI = schatting +/- foutenmarge

= interval berekend uit steekproefdata volgens methode die bepaalde kans heeft een interval op te leveren waarin de populatiewaarde ligt.

Stel: we vinden een gemiddelde waarde x in een steekproef. We kennen de steekproefgrootte, we kiezen een niveau van betrouwbaarheid (bijvoorbeeld 95%, dus hebben we 5% kans op een verkeerde waarde), en een foutenmarge (deze noemen we α) waarmee we bereid zijn te leven (in geval van 95% betrouwbaarheid is α= 5%). Algemeen kunnen we hieruit berekenen tussen welke waarden de populatieparameter μ zal liggen en kan het betrouwbaarheidsinterval als volgt weergegeven worden:

$$\bar{x} - z^{*}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z^{*}\frac{\sigma}{\sqrt{n}}$$

Daarin staat x voor de steekproefuitkomst (bijvoorbeeld een gemiddelde, proportie of percentage). z_α is een score van de standaard normaalverdeling die overeenkomt met de zelf gekozen kans op fout (α). Deze score wordt vermenigvuldigd met de standaardafwijking in de populatie gedeeld door de wortel van n. De standaardafwijking van het bestudeerde kenmerk in de populatie kennen we natuurlijk niet, maar we weten dat bij voldoende grote steekproeven de standaardafwijking in de steekproef wel kan gebruikt worden als schatter voor deze onbekende.

Een betrouwbaarheidsniveau CI (*Confidence interval*) wordt gedefinieerd als 1-α (waarbij geldt dat α = de kans op een vergissing, bvb 5% of 10%).
Er zijn verschillende betrouwbaarheidsniveaus.

CI= 90% (z-score: 1.645)

CI= 95% (z-score:1.960)

CI= 99% (z-score: 2.576)

Als we een uitspraak willen doen met een waarschijnlijkheid van 95% en een kans van 5% op vergissing, dan is z gelijk aan 1,960. Willen we de betrouwbaarheid van onze uitspraken optrekken tot bijvoorbeeld 99 % - we permitteren ons dan slechts 1 % kans op een fout -, dan is z gelijk aan 2,576. Het bredere betrouwbaarheidsinterval compenseert in dit geval ons verlangen om met een grotere waarschijnlijkheid uitspraken te poneren.

Voorbeeld zelf uitrekenen betrouwbaarheidsintervallen:

Uit een enquête afgenomen bij 3709 studenten blijkt, dat de gemiddelde score op het einde van het 6^{de} middelbaar 72% bedraagt. De standaardafwijking in de populatie bedraagt 8.

a) Geef het 95% en 99% betrouwbaarheidsinterval
b) Bereken het 95% betrouwbaarheidsinterval voor dezelfde resultaten maar met n=10, n=100 en n = 1000

Oplossing:

Gegeven:

$x_{gem} = 72$

$\sigma = 8$

N = 3709

a)

95% betrouwbaarheid → z = 1,96

$72 – (1,96 * 8/\sqrt{3709}) < \mu < 72 + (1,96 * 8/\sqrt{3709})$

$71,74 < \mu < 72,26$

99% betrouwbaarheid → z = 2,575

$72 – (2,575 * 8/\sqrt{3709}) < \mu < 72 + (2,575 * 8/\sqrt{3709})$

$71,66 < \mu < 72,34$

b)

95% betrouwbaarheid

→ z = 1,96

- **Voor n= 10**

$72 – (1,96 * 8/\sqrt{10}) < \mu < 72 + (1,96 * 8/\sqrt{10})$

$72 – 4,9584 < \mu < 72 + 4,9584$

$67,04 < \mu < 76,96$

- **Voor n= 100**

$72 – (1{,}96 * 8/\sqrt{100}) < \mu < 72 + (1{,}96 * 8/\sqrt{100})$

$72 – 1{,}568 < \mu < 72 + 1{,}568$

$70{,}43 < \mu < 73{,}57$

- **Voor n= 1000**

$72 – (1{,}96 * 8/\sqrt{1000}) < \mu < 72 + (1{,}96 * 8/\sqrt{1000})$

$72 – 0{,}4959 < \mu < 72 + 0{,}4959$

$71{,}50 < \mu < 72{,}50$

Ook statistische verwerkingspakketten berekenen de betrouwbaarheidsintervallen en geven deze weer. Van de studenten wordt verwacht dat ze de basisprincipes die er achter schuilgaan kennen en dat ze deze intervallen naar waarde leren schatten. Studenten dienen deze ook zelf te kunnen berekenen. Betrouwbaarheidsintervallen zijn een manier om met onnauwkeurigheid om te gaan. De grootte van het interval wordt bepaald door de standaardafwijking van het kenmerk in de populatie en ook door de grootte van de steekproef. Bij kleine steekproeven zullen de intervallen dus steeds groter zijn dan bij grote steekproeven.

11. Statistische hypothesetoetsing

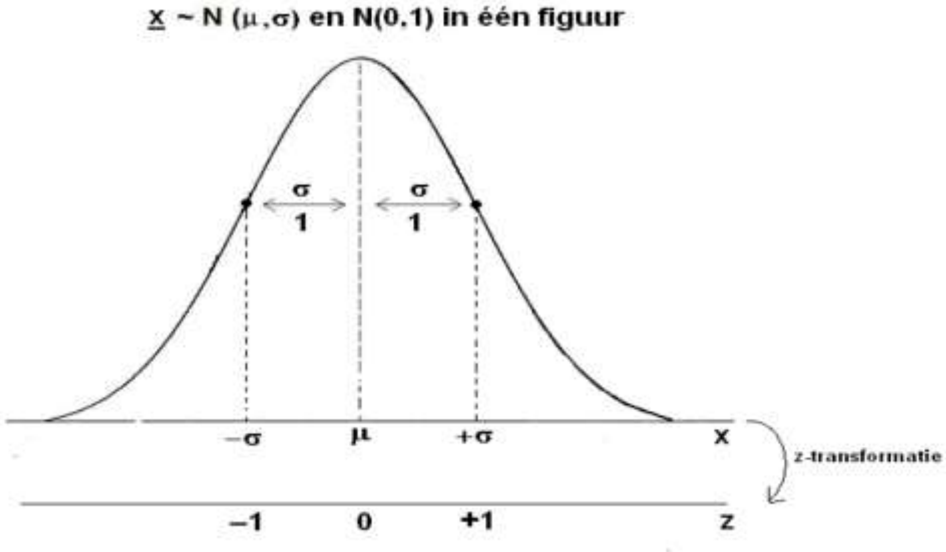
Nu we de theorie over de steekproevenverdeling en de schatting van populatieparameters op basis van steekproefuitkomsten uiteengezet hebben, gaan we verder in op de principes en de praktijk van het statistisch toetsen. Het komt er voor criminologen die in de praktijk onderzoek doen op basis van steekproeven, immers op neer goed te begrijpen wat een statistische toets ons kan leren. Met statistische hypothesetoetsing kan men testen hoe aannemelijk een bepaalde uitspraak over de populatie, op grond van de steekproef in kwestie, werkelijk is. Is een nieuwe behandeling van jeugdige delinquenten effectiever dan een oude behandelingsmanier? Is de proportie die herhaald slachtoffer wordt van een geweldsdelict werkelijk lager dan 1.4% van de bevolking?

Een *significantietoets* is een procedure om gegevens (zoals) uitkomsten van een steekproef te vergelijken met een vooropgestelde hypothese, die we de *nulhypothese* gaan noemen. Een hypothese is een bewering over parameters in een populatie. We kunnen bijvoorbeeld de hypothese stellen dat een parameter in de populatie niet verschilt van nul. In onze steekproef verschilt deze echter wel van nul. Kunnen we daarom aannemen dat dit in de populatie ook zo zal zijn? De uitkomst van een significantietoets leert ons meer. De uitkomst van een significantietoets wordt uitgedrukt in termen van een kans die aangeeft hoe goed data en hypothese met elkaar overeenkomen.

Vragen kunnen zijn: is een effect aanwezig van variabele X op Y? Is er een associatie tussen het roken en krijgen van kanker? Is er een relatie tussen leeftijd en onveiligheidsbeleving?
De hypothese die stelt dat het effect of een associatie niet bestaat noemen we de *nulhypothese* (H0). De *alternatieve hypothese* (Ha) wordt door de onderzoeker geformuleerd. Significantietoetsen worden uitgevoerd om de sterkte van het bewijs tegen de nulhypothese vast te stellen.

Hoe verder de waargenomen uitkomst verschilt van de nulhypothese, hoe onwaarschijnlijker dat H0 waar is, hoe sterker de indicatie voor Ha. De significantietoets meet de kans op het krijgen van een uitkomst die even extreem is, of nog extremer dan de waargenomen uitkomst. Dit noemen we de **overschrijdingskans** (p) van de toets.
Als p < is dan 0.05 betekent dit dat we, gesteld dat we een oneindig aantal steekproeven zouden trekken van dezelfde grootte, we slechts in 5 op 100 gevallen de gevonden steekproefuitkomst zouden uitkomen, terwijl deze in realiteit een waarde 0 heeft, of een andere vooropgestelde waarde.

Laat ons een voorbeeld geven uit de praktijk van het onderzoek. De onderzoeksvraag is of fraudeurs een significant hoger IQ hebben dan de gemiddelde mens, dit wil zeggen dat hun IQ hoger is dan op grond van louter toeval mag verwacht worden. We zetten de IQ scores van fraudeurs om in z-scores waardoor het gemiddelde 0 wordt en de standaardafwijking 1. We kennen de eigenschappen van de normale verdeling (oppervlakte: 1 of 100%, ééntoppig en perfect symmetrisch) en kunnen opzoeken hoeveel procent van de fraudeurs een bepaald IQ heeft. Stel dat we uitgerekend hadden dat in onze steekproef de fraudeurs een gestandaardiseerd gemiddelde van 1.99 hebben. Boven de 1.99 bevindt zich slechts 2.33% van de z-scores van de normaalverdeling. De kans om dit resultaat te vinden, indien fraudeurs echt een gemiddeld IQ hebben is dus dermate klein, dat we de nulhypothese gaan verwerpen. Dit betekent dat we sowieso kunnen aannemen dat de fraudeurs uit onze steekproef significant afwijken van de gemiddelde Belg. Onze alternatieve hypothese (fraudeurs hebben een hoger IQ dan de gemiddelde Belg) wordt niet ontkracht. In statistisch jargon: onder de aanname dat er geen verschil is tussen fraudeurs en gemiddelde Belgen, is de kans op dit steekproefresultaat kleiner dan 5%. We verwerpen dus de aanname dat fraudeurs een gemiddeld IQ hebben en gaan ervan uit dat fraudeurs een significant hoger IQ hebben dan de gemiddelde Belg.



Statistische hypothesetoetsing kan in zekere zin vergeleken worden met de uitspraak die een strafrechter doet over een beklaagde. De nulhypothese is de hypothese die we willen verwerpen. De nulhypothese in het geval van een rechter is de uitspraak dat de beklaagde onschuldig is. De rechter moet schuld als bewezen achten en dus moet de nulhypothese van onschuld ontkracht worden. Parallel kan men stellen dat de nulhypothese veronderstelt dat een bepaalde uitkomst of een steekproefgrootheid uit onze steekproef, eigenlijk nul is. De alternatieve hypothese is dat de beklaagde wel schuldig is; de alternatieve hypothese is dat een steekproefgrootheid verschilt van nul.

Schema: Het oordeel van de strafrechter

	Vrijlaten van de beklaagde	**Veroordelen van de beklaagde**
De beklaagde is onschuldig (Nulhypothese)	De rechter laat een onschuldige beklaagde vrij. **Correct besluit**	De rechter veroordeelt een onschuldige beklaagde. **Fout besluit en type-I fout.**
De beklaagde is schuldig (Alternatieve hypothese)	De rechter laat een schuldige beklaagde vrij. **Fout besluit en type-II fout**	De rechter veroordeelt een schuldige beklaagde. **Correct besluit**

Een *type-I fout* maken houdt in dat men een correcte nulhypothese verwerpt. Een *type-II fout* maken betekent dat men een foute nulhypothese aanvaardt. De testvariabele is de variabele waarvan een waarde berekend wordt op basis van een observatie uit de steekproef. Het kan ook gaan om een samenhang tussen kenmerken (zoals een correlatiecoëfficiënt of een regressiecoëfficiënt). Bij hypothesetoetsing hoort een **p-waarde** (p = *probabiliteit*). De p-waarde kan worden gezien als een uitdrukking van de waarschijnlijkheid of het 'waarheidsgehalte' van een nulhypothese. Zo is bijvoorbeeld een hoge p-waarde een indicatie dat de in de steekproef aangetroffen samenhang tussen twee kenmerken weinig waarschijnlijk is en de nulhypothese of de veronderstelling dat er geen samenhang bestaat tussen beide kenmerken in de populatie, meer ondersteuning geniet

Of nog algemener kunnen we de mogelijke beslissingen uit steekproefinformatie als volgt voorstellen:

Beslissing uit een steekproef

	Aanvaarden van de nulhypothese	**Verwerpen van de nulhypothese**
Toestand A: De nulhypothese is juist	**Correct besluit** **1-α**	Verwerpingsfout. **Type-I fout.** **Vals alarm: de brandmelder gaat af, maar er is vals alarm**
Toestand B: De nulhypothese is fout	**Doorlatingsfout en type-II fout** **Niet ontdekken dat er iets bijzonders aan de hand is** **De brandmelder gaat niet af, terwijl er wel brand is**	**Correct besluit** **1-β**[15]

Een statistische test (gebaseerd op de verzamelde gegevens) start met het stellen van een nulhypothese ('Ho'). Men gaat na of deze nulhypothese waar of vals is, en of men de nulhypothese met andere woorden al dan niet dient te verwerpen. Terwijl de nulhypothese stelt dat een steekproefgrootheid niet significant verschilt van nul, is de **p-waarde** een indicatie van de waarschijnlijkheid van deze nulhypothese. Hoe weten we nu of een

[15] Dit gebied, $1 - \beta$ noemt men ook de power van de test. Het is het geheel van waarden waaronder een valse nulhypothese correct wordt verworpen en een juiste nulhypothese wordt behouden. De power van een test is de kans op het detecteren van een effect wanneer er in werkelijkheid één is, en het besluiten dat er niets aan de hand is, wanneer er werkelijk niets aan de hand is.
In de praktijk neemt men aan dat toetsen met een power die groter is dan .80 (of ß<=.2), krachtige toetsen zijn. In deze inleidende syllabus gaan we niet verder in op deze details.

nulhypothese waar of vals is? Indien de p-waarde kleiner is dan 0.05 (bij een α van 0.05 of een lagere α van 0.01, afhankelijk van de beslissing van de onderzoeker), dan is de kans dat de gestelde nulhypothese waar is, kleiner dan 5%. Dit is niet veel, en de meeste sociale wetenschappers zijn bereid deze grens als ondergrens te beschouwen: is de p-waarde met andere woorden kleiner dan 0.05, dan verwerpt men de nulhypothese. Softwarepakketten voor statistische analyse vermelden de exacte significantie in de output. Het is belangrijke te weten dat men die overschrijdingskans blijvend dient te zien als een kans en niet als echt bewijs.

In de samenvattende tabel hieronder geven we 4 voorbeelden van correlaties tussen X en Y uit een steekproef

Samenvattende tabel

	Aanvaarden van de nulhypothese	**Verwerpen van de nulhypothese**
H0 = waar	H0 = 0 r = 0.09 p: >. 05 aanvaarden H0 = JUISTE BESLISSING	**H0 = 0** **r = 0.20** **p: < .05** **verwerpen H0** **= TYPE 1 FOUT**
H0 = niet waar	**H0 ≠ 0** **r = 0.10** **p: > .05** **aanvaarden H0** **= TYPE 2 FOUT**	H0 ≠ 0 r = 0.32 p: < .05 verwerpen H0 = JUISTE BESLISSING

Samengevat:

<u>Stappen bij een significantietoets</u> :

Formuleer de nulhypothese H0 en de alternatieve hypothese Ha

Specificeer het significantieniveau α (bijvoorbeeld een foutenmarge van 5% Type I fout)

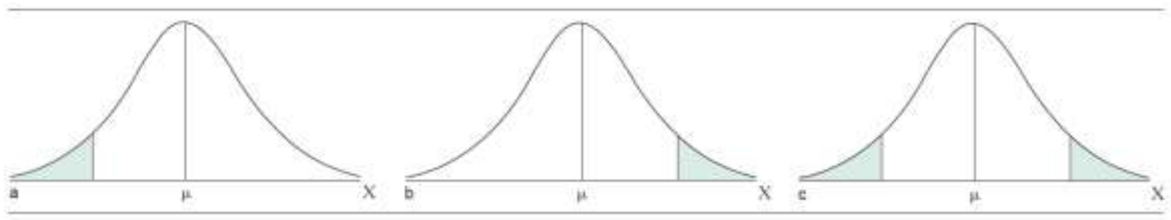
Bereken de steekproefgrootheid (bereken de correlatie, het gemiddelde, …)

Bepaal de bijhorende p-waarde of de overschrijdingskans. Is de p-waarde kleiner of gelijk aan α, dan is het toetsresultaat statistisch significant op niveau α

12. Eenzijdig of tweezijdig toetsen van een nulhypothese?

De formulering van de alternatieve hypothese bepaalt of we **eenzijdig** of **tweezijdig** toetsen. Een tweezijdige toets houdt in dat we stellen via de alternatieve hypothese (H1) dat de populatieparameter significant verschilt van nul, zonder een uitspraak te doen over de richting (positief verschillend van nul of negatief verschillend van nul). We spreken dan eigenlijk beter van toetsen met tweezijdig alternatief. Als de formulering daarentegen is, dat de populatieparameter positief verschilt van nul, dan wordt eenzijdig getoetst (beter: getoetst met eenzijdig alternatief). Het alternatief is rechtseenzijdig als de waarde van de populatieparameter een positieve waarde heeft en linkseenzijdig als de waarde van de populatieparameter een negatieve waarde heeft. Hoe we de alternatieve hypothese bij het toetsen formuleren en of we dus eenzijdig of tweezijdig toetsen hangt af van de onderzoeksvraag.

Figuur: eenzijdig toetsen (links of rechts) en tweezijdig toetsen



De p-waarde kan ook worden gezien als de kans dat een bepaald steekproefgemiddelde (of de correlatie) geheel bij toeval tot stand kwam. Hoe kleiner deze p-waarde, des te minder er sprake kan zijn van een toevallig resultaat. Het is echter van belang zich niet blind te staren op de significantie van resultaten alleen. De statistische significantie van een steekproefresultaat wordt sterk bepaald door de grootte van de steekproef. Men dient daarom ook naar de grootte van het verband of effect te kijken. Ook zeer geringe en inhoudelijk weinig relevante correlaties kunnen bij een voldoende grote steekproef significant zijn. Verder kan een nulhypothese ook altijd foutief verworpen worden. In het voorbeeld van het oordeel van de rechter wordt dan een onschuldige beklaagde veroordeeld. Men spreekt van een type-I fout. Behoudt men de nulhypothese ten onrechte, dan is er sprake van een type-II fout.

Wat is nu de ***"power"*** van een test? Dit is de kans dat de statistische test leidt tot een correcte verwerping van een valse nulhypothese. De statistische power van een test is dus de mate waarin een test er in slaagt een echt effect te detecteren als het effect ook echt bestaat, rekening houdende met zowel type-I als type II fouten. Als de statistische analyse van de power van een test een waarde van 0.80 of beter oplevert, dan wordt vaak aangenomen dat er voldoende power is. De waarde van de power varieert van 0 tot 1.

13. Andere belangrijke verdelingen

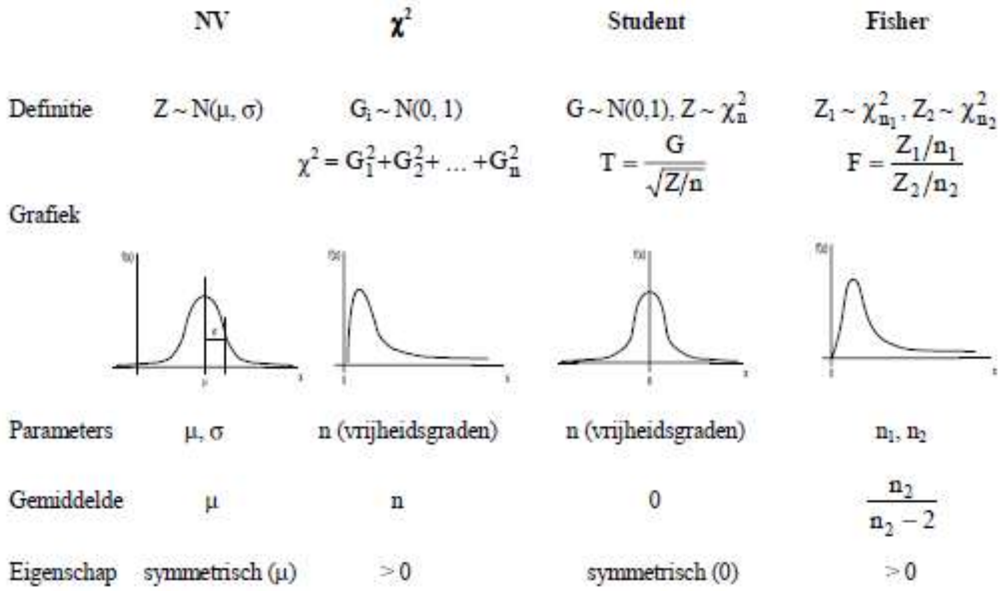
Niet alle kenmerken volgen een normale verdeling. Enkele andere belangrijke verdelingen, die eerder door wiskundige statistici werden beschreven, zijn belangrijk in het kader van de statistische inferentie. Deze andere verdelingen doen beroep op het begrip *vrijheidsgraden*, dat eerder al eens ter sprake kwam bij het uiteenzetten van de wijze waarop de steekproefstandaardafwijking wordt berekend. Wat onthouden dient te worden is dat van al deze andere verdelingen ook de oppervlakte onder de curve bekend is. We geven een overzicht van andere belangrijke statistische verdelingen. Sommige statistische significantietoetsen zijn gebaseerd op deze verdelingen.

De chi-kwadraat (χ2)-verdeling is een verdeling van het kwadraat van normaal verdeelde variabelen. Deze wordt onder andere gebruikt bij de verdeling van steekproefvarianties. De chi-kwadraat-verdeling is asymmetrisch en afhankelijk van de graden van vrijheid. Bij de steekproefvariantie zijn deze graden van vrijheid (degrees of freedom, afgekort Df) gelijk aan de steekproefgrootte (n) minus 1. De chi-kwadraat verdeling is in criminologisch onderzoek vooral belangrijk bij de analyse van kruistabellen om te weten of een percentageverschil statistisch significant is. De chi-kwadraat waarde die berekend wordt bij de analyse van kruistabellen volgt een chi-kwadraat verdeling.

De **t-verdeling** (**Student's t**) is een verdeling die uiterlijk heel erg lijkt op de normale verdeling. De frequentieverdeling (door wiskundigen *kansdichtheidsfunctie* genoemd) heeft dezelfde symmetrische klokvorm als die van de normale verdeling. De breedte van de klokvorm varieert (in tegenstelling tot bij de normale verdeling) met het aantal graden van vrijheid oftewel de "***degrees of freedom***" (**DF** = **de steekproefgrootte (n) minus 1**). De Student t-verdeling wordt gebruikt in de inferentiële statistiek waarbij we op basis van informatie uit één steekproef veralgemeningen willen bekomen naar de populatie toe. In de t-tabel staan zowel voor eenzijdige situaties met kans α in de staart (het oppervlak onder de curve rechts van t), als voor tweezijdige situaties met twee keer de kans ½ α in de staarten (de som van de oppervlakten links van –t en rechts van t) de t-waarden. De tabel geeft dus zowel voor $P(T < t) = (1- \alpha)$ als $P(-t < T < t) = (1- \alpha)$ de t-waarden. De t-verdeling wordt altijd gebruikt bij de toets van richtingscoëfficiënten.

De F-verdeling (genoemd naar de statisticus Fisher) is een quotiënt van twee chi-kwadraat verdeelde grootheden. Deze wordt onder andere gebruikt bij het quotiënt van twee steekproef varianties uit twee normaal verdeelde populaties. De F-verdeling is net als de chi-kwadraat verdeling asymmetrisch. Bovendien is de vorm afhankelijk van graden van vrijheid (df = n-1) in de teller en in de noemer van het quotiënt van variabelen dat samen F-verdeeld is. De F-verdeling wordt onder andere gebruikt in situaties waarin men wil weten of een determinatiecoëfficiënt die gevonden wordt in een steekproef statistisch significant verschilt van nul. De F-verdeling wordt gebruikt bij de toets van een ***determinatiecoëfficiënt***.

Figuur: enkele statistische verdelingen



De *binomiale verdeling* of **Bernoullie-verdeling** is genaamd naar wat het Bernoullie experiment genoemd wordt. Een Bernoullie experiment is een experiment met slechts twee mogelijke uitkomsten, die we "*succes*" (S) en *"mislukking*" (M) noemen. De binomiale verdeling wordt gebruikt bij de inferentie van *odds* en *odds-ratio's*. Dit valt echter buiten het bestek van deze inleidende cursus.

In bijlage bij dit handboek vindt de lezer de tabellen of tafels waarbij de opervlakte onder de curves vermeld staan. In de praktijk van het criminologisch onderzoek zullen we zien dat het volstaat om de kritische waarden die we per verdeling vinden te vergelijken met de gevonden waarden in onze steekproef.

14. De variantieanalyse als toets voor verschillen tussen groepen inzake metrische kenmerken

Variantieanalyse (Engels: ***analysis of variance***, **ANOVA**) is een toets voor de relatie tussen een ***nominale*** en een ***metrische*** variabele (bijvoorbeeld de relatie tussen verstedelijking en criminaliteit op basis van een gemeentetypologie). De berekeningswijze is gebaseerd op de **variaties van de steekproeven**. We geven een historisch voorbeeld: we doen een onderzoek naar criminaliteit in Engeland in de 19e eeuw onder invloed van industrialisatie en verstedelijking. We trekken drie steekproeven van acht gemeenten: industriële, handels- en rurale gemeenten. We hebben gegevens over het aantal moorden per 10,000 inwoners. De vraag is nu: zijn de verschillen in het moordniveau tussen de verschillende typen van gemeenten statistisch significant? We stellen de hypothese dat verstedelijking criminaliteit in de hand werkt omdat de anonimiteit in steden groter is dan op het platteland.

Variantieanalyse geeft antwoord op de vraag of de nominale variabele 'type gemeente' (industrieel, commercieel, ruraal) statistisch significant van invloed is op de ratiovariabele 'moordniveau'.

Onder de aanname dat de standaardafwijkingen in de populatie gelijk zijn, kun je op basis van de steekproeven de variantie op twee manieren schatten:

- Het gewogen gemiddelde van de variantie binnen iedere groep gemeenten: de ***binnengroepsvariatie (within-groups)***.
- De variantie van de gemiddelden van de drie groepen gemeenten rondom het algemeen gemiddelde: de ***tussengroepsvariatie (between-groups)***.

Als de gemiddelden in de populatie hetzelfde zijn, leveren beide berekeningswijzen een identiek getal op. Als de gemiddelden niet hetzelfde zijn, zal de tweede schatting een groter getal opleveren dan de eerste. **Hoe groter de spreiding tussen de groepen ten opzichte van de spreiding binnen de groepen, hoe meer de groepen onderling verschillen en hoe sterker het verband tussen de nominale en de interval variabele.**

Bij variantieanalyse worden beide schatters van de variantie in de populatie vergeleken door ze op elkaar te delen. Als de binnengroeps- en de tussengroepsvariatie gelijk zijn, is de breuk, de ***F-ratio***, gelijk aan 1. Als de tussengroepsvariantie groter is dan de binnengroepsvariatie, is de breuk groter dan 1. Naarmate de verschillen in de gemiddelden tussen de groepen groter zijn in vergelijking met de verschillen binnen iedere groep, is de breuk groter.

Natuurlijk kan het verschil toevallig zijn, maar hoe groter F (het verschil tussen de varianties) is, hoe kleiner de kans dat dit door het toeval komt. Bij iedere F-waarde kan de kans (p-waarde, *probability*) dat deze door het toeval is bepaald, berekend worden. Als de p-waarde kleiner is dan 0,05 is de kans dat de waarde aan het toeval te wijten is, kleiner dan 5%. Het verschil tussen de gemiddelden is dan significant, ofwel: de nominale variabele heeft een significante invloed op de interval-variabele.

De berekening van de F-ratio

De berekening van de variaties vindt plaats aan de hand van de som van de gekwadrateerde afwijkingen van het gemiddelde (Engels: ***Sum of Squares***, *SS*). De opdeling van de totale variatie in een tussengroepsvariatie is niet nieuw. Een soortgelijke redenering hebben we eerder behandeld bij de regressieanalyse, met name toen we spraken over de determinatiecoëfficiënt en de opdeling van de totale variantie in een afhankelijke variabele in verklaarde variantie en onverklaarde variantie.

$$SS_{Totaal} = SS_{Groepen} + SS_{Error}$$

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{k} n_i(\overline{X}_i - \overline{X})^2 + \sum_{i=1}^{k}\sum_{i=1}^{n}(X_i - \overline{X}_i)^2$$

Total SS = Within Groups SS + Between Groups SS

Within Groups SS = som van de gekwadrateerde afwijkingen van de individuele scores van hun eigen groepsgemiddelde: de 'niet-verklaarde' afwijkingen.

Between Groups SS = som van de gekwadrateerde afwijkingen van de groepsgemiddelden van het algemene gemiddelde: de 'verklaarde' afwijkingen.

In het ideale geval (perfect verband) is alle variatie toe te schrijven aan de verschillen tussen de groepen en niets aan het verschil binnen de groepen. Je krijgt de geschatte variantie door de gekwadrateerde afwijkingen te delen door het aantal vrijheidsgraden (Engels: *Degrees of Freedom, DF*).

Total DF = Within DF + Between DF

$$\mathbf{F} = \frac{\textbf{Verklaarde variantie}}{\textbf{Niet-verklaarde variantie}} = \frac{\textbf{\textit{Between SS / DF}}}{\textbf{\textit{Within SS / DF}}}$$

15. Zelf uitrekenen van een variantieanalyse

Moorden per 10.000 inwoners in 24 Engelse gemeenten (19^de^ eeuw)

	Homicide rate Industriesteden	**Homicide rate Handelssteden**	**Homicide rate Agrarische nederzettingen**	**Totaal**
	4,3	5,1	12,5	
	2,8	6,2	3,1	
	12,3	1,8	1,6	
	16,3	9,5	6,2	
	5,9	4,1	3,8	
	7,7	3,6	7,1	
	9,1	11,2	11,4	
	10,2	3,3	1,9	
Som	**68,6**	**44,8**	**47,6**	**161**
Gemiddelde	**8,575**	**5,6**	**5,95**	**6,708333**
N	**8**	**8**	**8**	**24**

Hieronder illustreren we hoe een variantieanalyse dient uitgerekend te worden. We zien in het voorbeeld 24 steden. Er zijn 8 industriesteden, 8 handelssteden en 8 agrarische nederzettingen. 3*8 = 24 steden. Nu kan je deze tabel bekijken en al met het blote oog vaststellen dat deze drie groepen verschillen van elkaar: de gemiddelden verschillen van elkaar. Als we geen rekening houden met deze typologie, dan is het grote gemiddelde gelijk aan 6.7 moorden per 10.000 inwoners. Echter, voor de industriesteden is dit gemiddelde 8.575 moorden per 10.000 inwoners. Voor handelssteden bedraagt het gemiddelde 5.6 moorden per 10.000 inwoners en voor agrarische nederzettingen bedraagt het gemiddelde 5.95 moorden per 10.000 inwoners. Er zijn wel degelijk verschillen, zo veel is duidelijk. De vraag alleen is natuurlijk: hoe significant zijn deze verschillen. Kunnen we zeggen dat deze groepen statistisch significant van elkaar verschillen? Of zijn deze verschillen louter aan het toeval te wijten? **We zetten ons even aan het rekenen om dit te weten te komen. Welke informatie hebben we nodig? We moeten voor elke groep van steden inzicht krijgen in de variabiliteit rond het eigen groepsgemiddelde en we moeten inzicht krijgen in de verschillen tussen de drie groepen. Als de verschillen tussen de groepen groter is dan de verschillen tussen de steden, dan is er een significant verschil.**

Dit heb je nodig:

- ***SS: Sum of Squares*** = som van gekwadrateerde afwijkingen tegenover het gemiddelde
- ***DF: Degrees of Freedom*** = vrijheidsgraden = aantal waarnemingen of groepen - 1 (bij steekproeven)
- ***Between Groups SS***: Tussengroepsvariatie of variatie tussen de gemeentetypes
- ***Within Groups SS***: Binnengroepsvariatie of variatie binnen de gemeentetypes
- ***Total SS*** *= Between Groups SS + Within Groups SS*
  Total DF *= Between Groups DF + Within Groups DF*
- ***Mean Squares*** = Variantie = SS / DF

Variantie tussen de groepen = 'verklaarde variantie' in de afhankelijke variabele (Engels: *Dependent*) door de onafhankelijke variabele (Engels: *Factor*); wordt soms ook '*Model*' genoemd in SPSS.

Variantie binnen de groepen = 'niet-verklaarde variantie'; wordt ook '*Error*' genoemd in SPSS.

F-Ratio = **variantie tussen groepen / variantie binnen groepen (= verklaarde / niet-verklaarde variantie). Dus: hoe hoger de F, hoe groter de verschillen tussen de groepen in verhouding tot de verschillen binnen de groepen.**

1) We beginnen onze analyse met de studie van de variatie binnen elke groep uit te rekenen. We beginnen met de industriesteden

a) Binnengroepsvariatie voor de *industriesteden*:

($\bar{x} = 8.57$)

x	$\boldsymbol{x - \bar{x}}$	$(\boldsymbol{x - \bar{x}})^2$
4.3	-4.27	18.23
2.8	-5.77	33.29
12.3	3.73	13.91
16.3	7.73	59.75
5.9	-2.67	7.13
7.7	-0.87	0.76
9.1	0.53	0.28
10.2	1.63	2.66
	variatie=	136.01

We starten met de individuele afwijkingen tegenover het gemiddelde en kwadrateren in de kolom daarnaast en maken de som. De variatie voor de industriesteden bedraagt 136.01.

b) **Binnengroepsvariatie voor de *handelssteden:***

($\bar{x} = 5.6$)

x	$\boldsymbol{x - \bar{x}}$	$(\boldsymbol{x - \bar{x}})^2$
5.1	-0.5	0.25
6.2	0.6	0.36
1.8	-3.8	14.44
9.5	3.9	15.21
4.1	-1.5	2.25
3.6	-2	4
11.2	5.6	31.36
3.3	-2.3	5.29
	variatie=	73.16

Nu berekenen we volledig op dezelfde manier de variaties voor de handelssteden. We starten met de individuele afwijkingen tegenover het gemiddelde en kwadrateren in de kolom daarnaast en maken de som. Deze bedraagt 73.16.

c) **Binnengroepsvariatie voor de *agrarische nederzettingen***

$(\bar{x} = 5.95)$

x	$\boldsymbol{x - \bar{x}}$	$(\boldsymbol{x - \bar{x}})^2$
12.5	6.55	42.90
3.1	-2.85	8.12
1.6	-4.35	18.92
6.2	0.25	0.06
3.8	-2.15	4.62
7.1	1.15	1.32
11.4	5.45	29.70
1.9	-4.05	16.40
	variatie=	122.04

Tenslotte berekenen we volledig op dezelfde manier de variaties voor de agrarische nederzettingen. We starten met de individuele afwijkingen tegenover het gemiddelde en kwadrateren in de kolom daarnaast en maken de som. Deze bedraagt 122.04.

Nu we de drie afzonderlijke variaties hebben uitgerekend, kunnen we de totale binnengroepsvariatie berekenen. Dit is niet meer dan de optelsom van de afzonderlijke binnengroepsvariaties. De totale binengroepsvariantie is de totale binnengroepsvariatie gedeeld door het aantal vrijheidsgraden. Het aantal vrijheidsgraden is hier 21. Dat is het gevolg van het feit dat er in elke groep een element niet kon variëren aangezien elk groepsgemiddelde vastlag. Aangezien we drie groepen hebben, bedraagt het aantal vrijheidsgraden hier 24-3 en dus 21.

De totale binnengroepsvariantie =

(binnengroepsvariatie a + binnengroepsvariatie b + binnengroepsvariatie c) / aantal vrijheidsgraden = $\frac{(136.01+73.16+122.04)}{(24-3)} = 15.77$

2) Variatie tussen de groepen:

Om de variatie tussen groepen te kennen moeten we ons baseren op de afwijkingen van elk groepsgemiddelde tegenover het grote gemiddelde. We zien een formule met een teller en een noemer. In de teller tellen we de gekwadrateerde verschillen op, dit wil zeggen we tellen de verschillen op tussen elk groepsgemiddelde en het grote gemiddelde, vermenigvuldigd met het aantal eenheden in elke groep. Omdat het aantal eenheden nu gemakshalve even groot is voor de drie groepen, kunnen we die groepsgrootte buiten de haakjes plaatsen.

In de noemer komt het aantal vrijheidsgraden. Er zijn drie groepen, en dat wil zeggen dat er twee groepen vrij kunnen variëren. Het aantal vrijheidsgraden is immers n-1, dus 3-1 = 2. De idee is dat er altijd één eenheid geen vrij variërende waarde kan hebben als het gemiddelde vastligt.

8*[(gemiddelde industriesteden - totaal gemiddelde)^2+ (gemiddelde handelssteden - totaal gemiddelde)^2+ (gemiddelde agrarische nederzettingen- totaal gemiddelde)^2]/ aantal vrijheidsgraden

De tussengroepsvariantie bedraagt

$$\frac{8*\left[(8.57-6.71)^2+(5.6-6.71)^2+(5.95-6.71)^2\right]}{3-1} = \frac{8*(3.46+1.23+0.58)}{2} = 21.08$$

We zien in elk geval in deze analyse dat de verschillen tussen de groepen groter zijn dan de verschillen binnen de groepen. Maar hoe significant zijn deze verschillen? De F-toets leert het ons.

Laat ons eerst de F-waarde berekenen: F is de verhouding tussen de tussengroepsvariantie (21.08) en de totale binnengroepsvariantie (15.77). Deze breuk is heel eenvoudig te berekenen. F = 21.08 / 15.77 = 1.34.

Opgelet! De gevonden F-waarde dien je nu op te zoeken in de tafel van F-waarden en je moet de waarde vergelijken met de kritische F-waarde. Wat is de kritische F-waarde? Dat is de waarde die past bij een gegeven aantal vrijheidsgraden in teller en noemer en een bepaald significantieniveau. Je F-waarde moet groter zijn dan de kritische gegeven een bepaald significantieniveau en vrijheidsgraden in de teller en noemer. Deze tabel vind je in bijlage. Op basis daarvan kan besloten worden of de gemiddelde scores tussen groepen significant van elkaar verschillen.

We hebben 21 vrijheidsgraden in de noemer en 2 vrijheidsgraden in de teller. Dit zoeken we op in de F tabel bij het 95% significantieniveau. Onze F waarde (F = 1.34) is duidelijker lager dan de kritische waarde 3.47. Het verdict is duidelijk: het door ons onderzochte verband is NIET statistisch significant. We kunnen de nulhypothese dat de steden niet verschillen van het algemene gemiddelde niet verwerpen.

Een determinatiecoëfficiënt voor de variantieanalyse

In een variantieanalyse bestaat er een equivalent voor de determinatiecoëfficiënt uit een regressieanalyse, met name **eta-kwadraat**. Eta-kwadraat is steeds *de verhouding tussen de tussengroepsvariatie en de totale variatie in Y*.

Toegepast op het voorbeeld bekomen we volgende waarde voor eta-kwadraat:
42.16/373.33= 0.1129 → 11.29 % (tussengroepsvariatie/totale variatie. Totale variatie = tussengroepsvariatie PLUS binnengroepsvariatie of 42.16 + 331.17).

16. Voorbeelden van statistische inferentie in andere analysetechnieken

We hebben verschillende analysetechnieken besproken en bij elke analyse hebben we aandacht gehad voor de beschrijvende resultaten. In dit hoofdstuk is het tijd om nu eens te kijken naar de interpretatie van de statistische significantie van de beschrijvende analyses die werden gepresenteerd.

De toets op significantie van regressieparameters

Coefficients[a]								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	1.613	.039		41.488	.000	1.537	1.689
	cumulative negative events	.455	.040	.184	11.319	.000	.376	.533

a. Dependent Variable: totale frequentie van regelovertredend gedrag

We bespreken eerst de regressiecoëfficiënten, i.e. de parameters van de best passende rechte, het intercept en de richtingscoëfficiënt en daarna de significantietoets van de determinatiecoëfficiënt. Hierboven presenteren we de resultaten van een regressieanalyse waarbij de afhankelijke variabele het totale aantal zelf-gerapporteerde delicten is en de onafhankelijke variabele het aantal negatieve levensgebeurtenissen. Sommige theorieën legden nogal veel klemtoon op de rol van negatieve levensgebeurtenissen (scheiding ouders, verlies of dood van een partner of familielid, …). Deze theorieën benadrukten dat het aantal negatieve levensgebeurtenissen waaraan men blootgesteld wordt een goede predictor is voor de frequentie van criminaliteit. Welnu, in de internationale zelfrapportagestudie hebben we de kans gehad om deze veronderstelling na te gaan en de resultaten zie je in de tabel hierboven. De afhankelijke variabele is het aantal regelovertredingen (totale frequentie de afgelopen 12 maanden) en de onafhankelijke variabele is het totaal aantal negatieve gebeurtenissen. De tabel geeft het intercept weer en de richtingscoëfficiënt. We hebben eerder reeds behandeld hoe je die coëfficiënten dient te interpreteren, maar nu gaan we in op de statistische significantietoets. Hiertoe is het belangrijk de juiste informatie te lezen. De tabel bevat naast het intercept en de richtingscoëfficiënt (B0 enB1) ook de standaardfout voor de betrokken parameter. Hoe weten we nu of de waarde die we in onze steekproef hebben gekregen ook statistisch significant is? We kunnen dit gemakshalve aflezen uit de laatste kolom, waar "sig" de afkorting is voor significantie en waar je het exacte significantieniveau kan aflezen. Je ziet dat beide coëfficiënten statistisch significant verschillen van nul. Op basis van de ongestandaardiseerde parameters en de standaardfout (de standard error) kan de t-waarde berekend worden. De t-waarde moet een waarde hebben die groter is dan 1.96 om te kunnen zeggen dat een verband statistisch significant is. Je moet die t-waarde niet met de hand zelf kunnen uitrekenen, maar je dient deze wel te kunnen interpreteren. Deze t-waarde kan opgezocht worden in de tabel in bijlage en de t-waarde dient eveneens groter te zijn dan de kritische t-waarde die je kan opzoeken voor elk gewenst significantieniveau en aantal vrijheidsgraden. Hier is de t-waarde duidelijk significant. Maar dat wist je al omdat we er op wezen dat t-waarden die groter zijn dan 1.96 zeker significant zijn op het 0.05 niveau.

Vervolgens kijken we naar de analyse van de determinatiecoëfficiënt.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.184[a]	.034	.034	2.349

a. Predictors: (Constant), cumulative negative events

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	706.673	1	706.673	128.113	.000[b]
	Residual	20150.038	3653	5.516		
	Total	20856.711	3654			

a. Dependent Variable: **totale frequentie van regelovertredend gedrag**

b. Predictors: (Constant), cumulative negative events

De determinatiecoëfficiënt bedraagt slechts 3.4%. Dit wil zeggen dat slechts 3.4 procent van de individuele verschillen in delinquent gedrag kan verklaard worden door de negatieve levensgebeurtenissen. Dat is heel weinig, maar toch verschillend van nul. IS dit nu significant verschillend van nul? Ja, zo blijkt, want de ANOVA tabel bevat de *regression sum of squares* en *residual sum of squares* met het gegeven aantal vrijheidsgraden. Voor de regression sum of squares is er maar één vrijheidsgraad omdat er slechts een onafhankelijke variabele in het model betrokken is. De residual sum of suares bedraagt 3653. Je kan de breuk berekenen: tussen elke mean square en dan bekom je de F waarde. Deze F-waarde kan je opnieuw gaan opzoeken in je kritische tabel, met het aantal vrijheidsgraden in teller en noemer, en dan merk je dat de determinatiecoëfficiënt significant verschilt van nul.

De toets op significantie van een correlatiecoëfficiënt

Alle correlatiecoëfficiënten (Pearson, Spearman, Kendal, Gamma, …) worden op hun significantie getoetst in SPSS. De output vermeldt steeds de exacte p-waarde. Als je een p-waarde tegenkomt die lager is dan 0.05 is het verband significant op het niveau 0.05. Doorgaans rapporteert men ook het niveau 0.01 en 0.001. Is een verband echter hoger is dan

0.05 is het verband strikt genomen niet meer statistisch significant. De vraag die zich stelt, en waarop het perfecte antwoord niet bestaat, is de vraag wat te doen bij een randgeval. Ook op nominaal niveau vind je in de output van SPSS de exacte significantie voor parameters zoals Phi en Cramer's V en Chi-kwadraat. De interpretatie is telkens analoog.

Randsignificantie en bedrog?

Wetenschappers worden vaak geconfronteerd met niet-perfecte data en resultaten. Als een p-waarde nu 0.06 bedraagt. Wat doe je dan? De conventie wil dat we verbanden tussen 0.05 en 0.10 als **randsignifican**t gaan beschouwen. We toetsen immers steeds tweezijdig en vaak is men vrij zeker van een verband. Als je bijvoorbeeld een verband wil onderzoeken tussen de sociale bindingen van iemand en de criminaliteit die iemand pleegt, dan is de hypothese normaal gezien dat hoge mate van sociale bindingen gepaard gaat met lagere betrokkenheid bij criminaliteit. Echter, dat is niet steeds zo. Dus: voorzichtigheid is geboden. Je zal wellicht al in de krant gelezen hebben dat sommige onderzoekers betrapt werden op het mooier maken van de gegevens. Dat is puur bedrog en gebeurt om significante resultaten nog significanter te maken. Bij mijn weten is het in de criminologie nog niet gebeurd, maar in de farmacie is het al gebeurd dat onderzoekers hun p-waarde hebben opgelapt, speciaal om de tests er beter te laten uitzien. Waarom zou men dat doen? Nou, als er veel geld mee gemoeid is, bijvoorbeeld een nieuw medicijn tegen een moeilijk te overwinnen ziekte, dan is het voor diegene die het medicijn heeft uitgevonden wel relevant om het op de markt te krijgen. Met een rand-significante p-waarde wordt zoiets alvast moeilijk. Zo zie je maar: het is niet de statistiek die liegt, maar de gebruiker.

17. Testvragen

Hieronder vinden Je enkele uitspraken over de bivariate statistiek en inferentiële statistiek. Deze vragen kan je gebruiken om je parate kennis te toetsen over de basiskennis die tot nog toe werd meegegeven. **In gewijzigde vorm kunnen dergelijke theorievragen ook op het examen voorkomen.** Deze vragen zijn afkomstig uit vroegere examens. De correcte antwoorden zijn rechtstreeks uit de cursus afleidbaar en worden vanuit didactisch oogpunt niet meegegeven. Dit is niet meer dan een test om te zien of je mee bent met de leerstof. Als je op deze test faalt, dan is het hoogdringend tijd om in actie te schieten.

1. **De Y-as wordt ook wel het ordinaat genoemd**

    - Deze uitspraak is juist
    - Deze uitspraak is verkeerd

2. **De X-as wordt ook wel de abscis genoemd**

    - Deze uitspraak is juist
    - Deze uitspraak is verkeerd

3. **Een scatterplot kan gebruikt worden om de relatie tussen een ratio-variabele en een interval-variabele te presenteren**

    - Deze uitspraak is juist
    - Deze uitspraak is verkeerd

4. **Een scatterplot kan gebruikt worden om de relatie tussen een ordinale variabele en een interval-variabele te presenteren**

    - Deze uitspraak is juist
    - Deze uitspraak is verkeerd

5. **Een spreidingsdiagram kan gebruikt worden om de relatie tussen een nominale en een ordinale variabele te presenteren**

    - Deze uitspraak is juist
    - Deze uitspraak is verkeerd

6. **Het (x,y) coördinaat van het rekenkundig gemiddelde van x en het rekenkundig gemiddelde van y is het bivariate zwaartepunt van het spreidingsdiagram**

    - Deze uitspraak is juist
    - Deze uitspraak is verkeerd

7. **De ongestandaardiseerde regressiecoëfficiënt (b1) uit een bivariate regressieanalyse is een**
    - Symmetrische maat
    - Asymmetrische maat

8. **De bivariate gestandaardiseerde regressiecoëfficiënt uit de regressie van Y op X is gelijk aan de gestandaardiseerde covariantie tussen x en y**
    - Deze uitspraak is juist
    - Deze uitspraak is fout

9. **De regressie van Q op S wil zeggen dat we**
    - Q als afhankelijke variabele hebben en S als onafhankelijke variabele
    - S als afhankelijke variabele hebben en Q als onafhankelijke variabele

10. **Het percentageverschil is een symmetrische associatiemaat**

    - Deze uitspraak is juist
    - Deze uitspraak is verkeerd

11. **Hierna vind je een tabel waarbij de relatie tussen twee nominale variabelen, X en Y wordt voorgesteld. We veronderstellen dat R de afhankelijke variabele is en dat S een causale invloed uitoefent op R.**

	Variabele S		Totaal
Variabele R	A	B	E
	C	D	F
Totaal	G	H	I

Welke van de uitspraken is juist:

- E en F noemen we kolommarginalen
- E en F noemen we rijmarginalen (ik twijfel!)
- A / G geeft de proportie van A
- A / C geeft de odds op A

- Het juiste percentageverschil moet berekend worden op basis van de vergelijking van de verschillen tussen

12. Chi-kwadraat is een associatiemaat die zeer gevoelig is aan N

- Deze uitspraak is juist
- Deze uitspraak is verkeerd

13. Als je N vermenigvuldigt met 2, dan wordt de waarde van chi-kwadraat twee keer zo groot

- Deze uitspraak is juist
- Deze uitspraak is fout

14. Chi-kwadraat moet worden berekend op basis van de ruwe scores

- Deze uitspraak is juist
- Deze uitspraak is verkeerd

15. Chi-kwadraat kan ook worden berekend op basis van de proporties

- Deze uitspraak is juist
- Deze uitspraak is verkeerd

16. In een 2*2 tabel is de waarde van Phi gelijk aan de waarde van V

- Deze uitspraak is juist
- Deze uitspraak is verkeerd

17. Bekijk volgende tabel

Slachtofferschap afgelopen vijf jaar	**Scholingsgraad**		
	Laag	**Midden**	**Hoog**
Nooit	A	B	C
Een maal	D	E	F
Twee maal of meer	G	H	I

Welke uitspraken over deze tabel zijn juist

- (E*F) is een consistent paar
- (F*H) is een inconsistent paar
- Gamma is gebaseerde op de verhouding tussen de consistente paren min de inconsistente paren en de consistente paren plus de inconsistente paren
- Gamma is een associatiemaat die monotoniciteit veronderstelt (weet ik niet of dit nu ook nog juist is of niet)
- Gamma is een asymmetrische maat (en symmetrisch)
- Gamma wordt gebruikt op nominaal niveau (zeker ordinaal)

18. Welke uitspraken zijn juist?

- Spearman's Rho is een rangcorrelatiecoëfficiënt
- Pearson's r is een rangcorrelatiecoëfficiënt (= product-moment correlatiecoëfficient)
- Kendall's Tau- is een rangcorrelatiecoëfficiënt
- Rangcorrelaties zijn asymmetrische maten (symmetrisch)

19. Welke uitspraken over Pearson's r zijn juist?

- De bivariate correlatiecoëfficiënt van Pearson veronderstelt lineariteit (zoniet kan Pearson's r leiden tot verkeerde conclusies)
- De bivariate correlatiecoëfficiënt is een gestandaardiseerde covariantie
- De bivariate correlatiecoëfficiënt wordt berekend uit niet gekwadrateerde deviatiescores

20. **De bivariate correlatiecoëfficiënt tussen ouderlijk toezicht en criminaliteit en y is negatief en de gestandaardiseerde rico uit de regressie van criminaliteit op toezicht bedraagt -0.35.**
    - Daaruit volgt dat toezicht een remmend effect heeft op criminaliteit
    - Daaruit volgt dat toezicht nefast is voor de criminaliteit want meer toezicht, resulteert in meer criminaliteit
    - Daaruit volgt dat meer toezicht ongerelateerd is aan criminaliteit

21. **De covariatie is**
    - De kruisproductensom
    - De som van de deviatiescores van x + de deviatiescores van y

22. **Hieronder volgen een aantal uitspraken over de lineaire regressieanalyse**
    - De lineaire regressie is een asymmetrische techniek
    - De lineaire regressie (het basismodel) is gebaseerd op de kleinste kwadratenoplossing
    - De lineaire regressie is niet geschikt voor kwadratische curvilineaire relaties (de andere twee weet ik niet zo goed, maar in fb document staan ze alle 3 wel aangeduid als juist

23. **Het residu is**
    - Het verschil tussen geobserveerde en verwachte waarde
    - Het verschil tussen verwacht en geobserveerde waarde

24. **De som van de gekwadrateerde residuelen bedraagt nul**
    - Deze uitspraak is juist
    - Deze uitspraak is fout

25. **De variatie in Y kan ook uitgedrukt worden als de som van de regressie sum of square minus de residual sum of square**
    - Deze uitspraak is juist
    - Deze uitspraak is fout (niet minus, maar plus!)

26. **Als de punten uit een puntenwolk heel dicht bij de regressierechte liggen, dan kunnen we vermoeden dat**
    - De model fit zeer hoog zal zijn
    - De model fit zeer laag zal zijn

27. **Het is perfect mogelijk dat twee regressiecoëfficiënten een zelfde beta-waarde hebben, maar een verschillende aliënatiecoëfficiënt hebben**
    - Deze uitspraak is juist
    - Deze uitspraak is fout

28. **De variantieanalyse wordt gebruikt wanneer**
    - De onafhankelijke variabele nominaal is en de afhankelijke variabele metrisch
    - De afhankelijke variabele nominaal is en deon afhankelijke variabele metrisch

29. **Eta-kwadraat komt qua interpretatie overeen met een determinatiecoëfficiënt**
    - Deze uitspraak is juist
    - Deze uitspraak is fout

30. **De toets van de significantie van de determinatiecoëfficiënt gebeurt aan de hand van**
    - De F-toets
    - De T-toets

31. **De toets van het intercept en de rico gebeurt aan de hand van de**
    - De F-toets
    - De T-toets
    - De Z-toets

32. **Intervalschatting is een centraal element uit de**
    - Inferentiële statistiek
    - Descriptieve statistiek

33. De steekproevenverdeling van X is

- De verdeling van X
- De verdeling van alle gemiddelde waarden voor X in een reeks van steekproeven met gelijke omvang
- De verdeling van alle gemiddelde waarden voor X in een reeks van steekproeven met ongelijke omvang
- Geen van de voorgaande beweringen is juist

34. Een betrouwbaarheidsinterval van met een alfa (kans op vergissing) van 10 %

= Z-score van 90

- Komt overeen met een z-score van 1.645
- Komt overeen met een z-score van 1.960
- Komt overeen met een z-score van 2.576

35. Ik doe criminologisch onderzoek naar belastingontduiking en stel vast dat het gemiddeld aantal veroordelingen onder belastingontduikers (mean = 3) lager is dan onder inbrekers (mean =5). De betrouwbaarheidsintervallen van de beide gemiddeldes blijken elkaar te overlappen.

Hieruit besluit ik dat

- De gemiddeldes niet significant van elkaar verschillen
- De gemiddeldes wel significant van elkaar verschillen

36. Een type I fout betekent :

- De nulhypothese is juist en ik verwerp ze foutief
- De nulhypothese is verkeerd en ik behoud ze foutief

37. Een type II fout betekent :

- De nulhypothese is juist en ik verwerp ze foutief
- De nulhypothese is verkeerd en ik behoud ze foutief

38. **De power van een test berekenen is belangrijk want hierdoor houdt men rekening met zowel type I als type II fouten**
    - Deze uitspraak is juist
    - Deze uitspraak is fout (de power van de test is de kans dat een toets geen type-II fout maakt of de kans op het terecht verwerpen van een foute nulhypothese)

39. **De tussengroepsvariantie is de tussengroepsvariatie gedeeld door het aan tal vrijheidsgraden tussen groepen**
    - Deze uitspraak is juist
    - Deze uitspraak is fout

40. **Als de tussengroepsvariantie groter is dan de binnengroepsvariantie, en de F-toets geeft een waarde die hoger is dan de kritische waarde, is het verband tussen x en y significant**
    - Deze uitspraak is juist
    - Deze uitspraak is fout

41. **Als de tussengroepsvariantie groter is dan de binnengroepsvariantie, en de F-toets geeft een waarde die lager is dan de kritische waarde, is het verband tussen x en y significant**
    - Deze uitspraak is juist
    - Deze uitspraak is fout

18. Leerdoelen

Dit deel van het handboek behandelt de praktijk van het schatten en toetsen, twee uiterst belangrijke procedures in de inferentiële statistiek. Criminologen doen hierop heel vaak beroep wanneer zij steekproeven trekken en bevolkingsbevragingen organiseren omtrent slachtofferschap, onveiligheidsbeleving, attitudes tegenover het strafrechtssysteem,… Het is van belang de uitkomsten van steekproeven naar waarde te schatten en dit kan via de principes van schatten en toetsen. De centrale begrippen die verband houden met de inferentiële statisitiek, zoals schatten, toetsen, p-waarden, nulhypothese, alternatieve hypothese, het betrouwbaarheidsinterval, variantieanalyse, F-waarden dienen te kunnen worden toegepast. De steekproevenverdeling werd geïntroduceerd in dit hoofdstuk als een bijzondere verdeling van steekproefuitkomsten (gemiddelden, standaardafwijkingen, correlaties,…).

Het principe van de centrale limietstelling dient zeer goed begrepen te worden. In de oefensessies voorzien we hierin simulaties om deze abstracte kennis in de praktijk toe te passen. We verwachten dat studenten betrouwbaarheidsintervallen ook zelf met de hand kunnen uitrekenen en dat deze geïnterpreteerd kunnen worden wanneer resultaten van wetenschappelijk onderzoek onder de vorm van output van statistische verwerkingspakketten worden voorgelegd aan studenten. Studenten worden tijdens de oefensessies getraind om vragen te beantwoorden aan de hand van output van het verwerkingspakket SPSS. Het is verder van belang te weten dat bij het toetsen niet enkel de normale verdeling wordt gebruikt, maar ook andere verdelingen. De principes zijn echter dezelfde. Op basis van deze verdelingen kunnen we kansen nagaan dat een bepaalde parameter in de populatie waaruit de steekproef afkomstig is effectief bestaat. Ook de variantieanalyse dient door de studenten zelf uitgerekend te kunnen worden. Hieronder voorzien we nog in een samenvattende tabel van toetsingsprocedures voor associaties categorisch en metrisch niveau.

Samenvattende tabel: inferentie op de geziene associatiematen en analysetechnieken

VERBANDEN TUSSEN 2 VARIABELEN: niet-dependente associatiematen (samenhang)			
	Nominaal	**Ordinaal**	**Metrisch**
verbanden tussen 2 variabelen	Controleer de Chi-kwadraat toets op Phi, Cramer's V	**Controleer de significantieniveaus van Spearman's rho en gamma**	**Controleer het significantieniveau van de correlatiecoëfficiënt van Pearson**

Samenvattende tabel: inferentie op de geziene dependente analysetechnieken (causatie en predicitie)

Onafhankelijke variabele	**Afhankelijke variabele**	**ANALYSETECHNIEK**	**Toetsen**
Interval/Ratio	**Interval/Ratio**	**lineaire regressieanalyse**	**T-toets op de regressiecoëfficiënten** **F-toets op determinatiecoëfficiënt** **P < 0.05 of beter!**
Nominaal	**Interval/Ratio**	**Een-factor variantieanalyse**	**F-toets op** **Sum of Squares between Groups** **Sum of Squares within Group**