Statistiek II

Prof. dr. A. Van Breedam

2e Kandidatuur TEW en HI
Faculteit TEW
Universiteit Antwerpen - RUCA

Inhoudsopgave

Lijst van tabellen

Lijst van figuren

Inleiding

Statistiek is een belangrijk ondersteuningsvak voor de studenten TEW en HIR.
De cursus is erop gericht om de student een aantal instrumenten aan te bieden waarmee hij statistische analyses kan uitvoeren. Een theoretische onderbouw van de analysemethoden is noodzakelijk om een beter begrip en inzicht te verwerven. Bovendien moet het de student in staat stellen om de geschikte analysemethode te kiezen om een statistisch probleem op te lossen.
Daarnaast wordt de toepasbaarheid van de statistische methode verduidelijkt aan de hand van vele voorbeelden en oefeningen.

Deze cursus behandelt de inferentiele statistiek. In hoofdstuk 1 wordt de basis van de inferentiele statistiek gelegd, de steekproeftheorie.
Vervolgens wordt in hoofdstuk 2 de eerste grote pijler van de inferentiele statistiek, de problematiek rond de schattingen beschouwd.
Hoofdstuk 3 heeft betrekking op de tweede grote pijler, de hypothesetoetsen. De keuze van een geschikte toets om een toetsingsprobleem op te lossen is de rode draad door dit zeer belangrijk hoofdstuk.
De procedure van de hypothesetoetsen wordt in hoofdstuk 4 verder uitgebreid naar de associatiematen.
Hoofdstuk 5 bevat nog een aantal aanvullende oefenigen m.b.t. de voorgaande hoofdstukken.

Hoofdstuk 1

Steekproeftheorie

1.1 Inleiding

Statistiek is de wetenschap van de gedeeltelijke waarneming. Dit wil zeggen, op grond van de uitkomsten van een steekproef gaan we een uitspraak doen over de populatie. Indien een steekproef representatief wordt geacht voor een populatie, dan kan men er van uit gaan dat een bepaalde karakteristiek die zich voordoet in de steekproef zich waarschijnlijk ook zal manifesteren in de populatie.

Dit hoofdstuk is de onderbouw van de inferentiele statistiek, die in deze cursus wordt behandeld in de twee volgende hoofdstukken i.v.m. schattingen en hypothesetoetsen.

1.2 Steekproeven

In wat volgt, zullen we steeds uitgaan van een populatie met N elementen en een steekproef met n elementen.
Het uitvoeren van een steekproef komt neer op het verzamelen van waarnemingen. Alvorens tot de gegevensverzameling over te gaan dient men toch twee belangrijke beslissingen te nemen:

1. Welk verschijnsel, welke *variabele* moet geobserveerd worden en wat zijn de mogelijke uitkomsten voor deze variabele;

2. Welke *steekproefmethode* zal gebruikt worden. De steekproefmethode omvat de manier waarop de steekproef zal worden samengesteld (bv. de wijze waarop de elementen uit de populatie worden geselecteerd voor de steekproef). Een aantal belangrijke steefproefmethoden zijn:

    - *aselecte steekproef:* alle elementen uit de populatie hebben een gelijke kans om geselecteerd te worden.
        - *met teruglegging:* een element kan meer dan éénmaal voorkomen in de steekproef. In theorie kan de steekproef zelfs groter worden dan de populatie ($n > N$). Bovendien is de trekking van elk element onafhankelijk. Zijn trekkingskans is gelijk aan $1/N$. Het totaal aantal mogelijke aselecte steekproeven met teruglegging is N^n (variatie, herhaling). Al deze steekproeven hebben een even grote waarschijnlijk, t.t.z. N^{-n}.

  - *zonder teruglegging:* elk element uit de popultatie kan hoogstens éénmaal voorkomen in de steekproef. Dit impliceert dat de steekproef hoogstens even groot kan zijn dan de populatie ($n \leq N$). De opeenvolgende trekkingen zijn niet meer onafhankelijk. Het totaal aantal mogelijke aselecte steekproeven zonder teruglegging bedraagt C_N^n (combinatie, geen herhaling). De trekkingskans voor elk element bij elke trekking is $1/N$.

- *quota-steekproef:* wordt vooral toegepast bij enquêtes. De elementen worden vrij (subjectief) geselecteerd, maar bepaalde verhoudingen (quota) moeten gerespecteerd worden; vb: verhouding man-vrouw 50%-50%.
- *gestratificeerde steekproef:* populatie wordt opgesplitst in deelpopulaties (strata) m.b.t. bepaalde kenmerken. Uit elk stratum wordt vervolgens een aselecte steekproef getrokken.
- *trosgewijze steekproef:* populatie wordt opgedeeld in clusters (trossen) en alle elementen van de door het toeval geselecteerde tros worden in de steekproef opgenomen.

Men spreekt van een *kanssteekproef* indien de steekproefmethode gebaseerd is op een toevalsmechanisme waardoor de insluitkans voor elk element van de populatie in de steekproef *bekend* en *positief* is.

De aselecte, de gestratificeerde en de trossteekproef zijn kanssteekproeven, de quota-steekproef niet.

1.3 Steekproefgrootheden

Conclusies voor de populatie zijn integraal gebaseerd op conclusies voor de steekproef. Daarom spreekt men ook van inferentiele statistiek. Anders gesteld: conclusies over kengetallen van de populatie zijn gebaseerd op kengetallen van de steekproef.
Doorgaans worden er drie belangrijke *steekproefkengetallen* onderscheiden: de waarde van het steekproefgemiddelde $\overline{x}$, de steekproeffractie p en de steekproefstandaarddeviatie s.

Bij een kanssteekproef is elk element op te vatten als een toevalsexperiment waarbij elke waarneming beschouwd kan worden als een realisatie van een toevalsvariabele. Stel een *populatievariabele* $\mathbf{X_t}$, dan kan het resultaat van de t-de trekking in de steekproef beschouwd worden als de realisatie van de *steekproefvariabele* X_t. De verdeling van elke steekproefvariabele X is dezelfde als die van de populatievariabele $\mathbf{X}$, voor zover de steekproef representatief is. Een steekproef van n getallen $x_1, x_2, ..., x_n$ wordt beschouwd als de reeks van realisaties van de n toevalsvariabelen $X_1, X_2, ..., X_n$.
Een *steekproefgrootheid* of *steekproefstatistiek* is een toevalsvariabele die een funktie van de steekproefvariabelen $X_1, X_2, ..., X_n$ is. Een *steekproefkengetal*, de waarde van de steekproefgrootheid in de steekproef, kan verschillende waarden aannemen, afhankelijk van de toevallige samenstelling van de steekproef. Een steekproefkengetal kan daarom beschouwd worden als de uitkomst van een toevalsvariabele, de steekproefgrootheid.

Zo is het *steekproefgemiddelde* $\overline{X}$ een variabele met als realisatie $\overline{x}$

$$\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t \qquad \overline{x} = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{1.1}$$

De *steekproefvariantie* en zijn realisatie worden gegeven door:

$$S_x^2 = \frac{1}{n-1}\sum_{t=1}^{n}(X_t - \overline{X})^2 \qquad s_x^2 = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \overline{x})^2 \tag{1.2}$$

De *steekproefstandaarddeviatie* wordt verkregen door de wortel uit de steekproefvariantie te trekken:

$$S_x = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(X_t - \overline{X})^2} \qquad s_x = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(x_t - \overline{x})^2} \tag{1.3}$$

Indien de beschouwde variabele $\mathbf{X}$ een dichotome (0-1) variabele is, komt de berekening van een steekproefgemiddelde neer op de berekening van een steekproeffractie P met waarde p:

$$P = \frac{1}{n}\sum_{t=1}^{n}X_t \qquad p = \frac{1}{n}\sum_{t=1}^{n}x_t \tag{1.4}$$

Naast deze drie meest gebruikte steekproefgrootheden, kunnen er nog een aantal andere gedefinieerd worden zoals de steekproefmediaan $\tilde{X}$, de steekproefcovariantie en het steekproefcorrelatiegetal. Deze laatste twee steekproefgrootheden kunnen slechts berekend worden wanneer de relatie tussen paren variabelen moet onderzocht worden. In voorkomend geval bestaat de steekproef uit paren waarnemingen (X_t, Y_t). De steekproefcovariantie S_{xy} en zijn waarde in de steekproef s_{xy} worden berekend als volgt

$$S_{xy} = \frac{1}{n-1}\sum_{t=1}^{n}(X_t - \overline{X})(Y_t - \overline{Y}) \qquad s_{xy} = \frac{1}{n-1}\sum_{t=1}^{n}(x_t - \overline{x})(y_t - \overline{y}) \tag{1.5}$$

De steekproefcorrelatie R_{xy} kan beschouwd worden als de gestandaardiseerde waarde van de steekproefcovariantie:

$$R_{xy} = \frac{S_{xy}}{S_x S_y} \qquad r_{xy} = \frac{s_{xy}}{s_x s_y} \tag{1.6}$$

De overeekomende *populatiegrootheden* zijn:

$$\mu = \frac{\sum \mathbf{X_t}}{N}$$

$$\sigma_x^2 = \frac{\sum(\mathbf{X_t} - \mu)^2}{N}$$

$$\sigma_x = \sqrt{\frac{\sum(\mathbf{X_t} - \mu)^2}{N}}$$

$$\pi = \frac{1}{N}\sum_{t=1}^{N}\mathbf{X_t}$$

$$\sigma_{xy} = \frac{1}{N}\sum(\mathbf{X_t} - \mu_x)(\mathbf{Y_t} - \mu_y)$$

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

1.4 Steekproevenverdelingen

Elke toevalsvariabele wordt gekenmerkt door een waarschijnlijkheidsverdeling. Steekproevenverdelingen zijn kansverdelingen van een steekproefgrootheid, die eveneens een toevalsvariabele is. Het gebruik van steekproevenverdelingen is tweeledig:

1. Het beantwoorden van probabiliteitsvragen i.v.m. de steekproefgrootheid.
2. Het vormen van de basis voor de statistische inferentie (intervalschattingen, hypothesetoetsen).

Definitie 1 *Een steekproevenverdeling is de verdeling van alle mogelijke waarden voor een bepaalde steekproefgrootheid, die berekend werden uit aselekte steekproeven van een zelfde omvang, getrokken uit een zelfde populatie.*

Voor het opzetten van een steekproevenverdeling voor een discrete, eindige populatie dienen de volgende stappen te worden gevolgd:

1. Trek alle mogelijke aselekte steekproeven van grootte n uit een discrete, eindige populatie van grootte N.
2. Bereken de waarde van de gewenste steekproefgrootheid (d.i. het steekproefkengetal) voor elke steekproef.
3. Voor elke mogelijke waarde van de steekproefgrootheid wordt de frekwentie van voorkomen bepaald in alle steekproeven.

Voor elke steekproevenverdeling kunnen de (gewone en centrale) momenten berekend worden (gemiddelde, variantie en hogere orde momenten) en de vorm van de verdeling bepaald worden.

Voor een *oneindige populatie* kan de steekproevenverdeling slechts benaderend worden opgesteld aangezien het niet mogelijk is om *alle* steekproeven (oneindig veel) te trekken uit een populatie. De steekproevenverdeling wordt dan benaderd op basis van een groot aantal steekproeven.

Nochtans heeft het empirisch opstellen van steekproevenverdelingen van steekproefgrootheden weinig praktische, maar vooral educatieve waarde. Daarom wordt in wat volgt de steekproevenverdelingen van een aantal steekproefgrootheden empirisch opgebouwd. Immers, steekproevenverdelingen kunnen mathematisch afgeleid worden. Deze afleidingen behoren tot het domein van de mathematische statistiek.
Bovendien, laat de praktische situatie ons niet toe een groot aantal aselekte steekproeven te trekken om een aldus een steekproevenverdeling op te stellen. We beschikken meestal slechts over één enkele steekproef uit een populatie en op basis van de informatie afkomstig van deze ene steekproef wordt de inferentie gedaan. Teneinde de inferentiele statistiek goed te kunnen vatten, is een goed begrip van steekproevenverdelingen echter onontbeerlijk.

1.4.1 Steekproevenverdeling van het steekproefgemiddelde

Het rekenkundig gemiddelde is een belangrijk maat die de centrale tendens aangeeft van een verzameling gegevens. Het gemiddelde van een populatie kan geschat worden op basis

$\overline{x}$	frekwentie n_i	relatieve frekwentie
1	1	0.04
1.5	2	0.08
2	3	0.12
2.5	4	0.16
3	5	0.20
3.5	4	0.16
4	3	0.12
4.5	2	0.08
5	1	0.04
Totaal	25	1.00

Tabel 1.1: Steekproevenverdeling van $\overline{X}$.

van de berekening van het gemiddelde van een steekproef. De validiteit van deze procedure is volledig afhankelijk van de kennis van de steekproevenverdeling van het steekproefgemiddelde.

In het volgende instructief voorbeeld wordt aangetoond hoe men een steekproevenverdeling empirisch kan opzetten voor een *eindige populatie* middels *aselecte steekproeven met teruglegging.*

Voorbeeld. Een populatie bestaat uit N=5 wagens van evenveel topmensen van een firma. De populatievariabele $\mathbf{X}$ die ons interesseert is het aantal dienstjaren van de wagen. De 5 waarden zijn achtereenvolgens

$$3, 2, 4, 5, 1.$$

Dit geeft de volgende waarden voor het populatiegemiddelde en de populatievariantie:

$$\mu = \frac{\sum \mathbf{x_i}}{N} = 3 \qquad \sigma^2 = \frac{\sum(\mathbf{x_i} - \mu)^2}{N} = 2$$

Veronderstel nu dat we de steekproevenverdeling van het steekproefgemiddelde $\overline{X}$ wensen te bepalen op basis van deze populatie. De volgende stappen zijn hiervoor vereist:

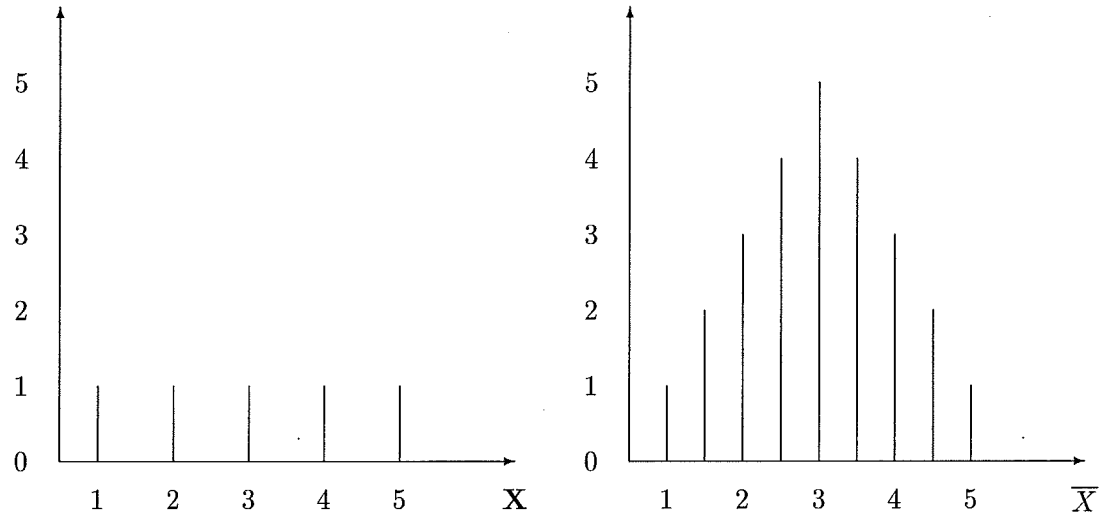
1. Indien we een steekproefgrootte van n=2 veronderstellen, kunnen er in het geheel $N^n = 5^2 = 25$ aselecte steekproeven met teruglegging getrokken worden uit de bovenstaande populatie.

2. Voor elk van deze steekproeven n=2 wordt $\overline{x}$, de waarde van de steekproefgrootheid $\overline{X}$ berekend. De 25 waarden $\overline{x}_i$ zijn:

$$1, 1.5, 2, 2.5, 3, 1.5, 2, 2.5, 3, 3.5, 2, 2.5, 3, 3.5, 4, 2.5, 3, 3.5, 4, 4.5, 3, 3.5, 4, 4.5, 5$$

3. De frekwentieverdeling van $\overline{X}$ wordt opgesteld.

De steekproevenverdeling is wel degelijk een waarschijnlijkheidsverdeling, aangezien de relatieve frekwentie van elke waarde positief is en de som van alle relatieve frekwenties 1



Figuur 1.1: Populatieverdeling van $\mathbf{X}$ en steekproevenverdeling van $\overline{X}$.

bedraagt.

Figuur 1.1 vergelijkt de oorspronkelijke populatiedistributie van de variabele $\mathbf{X}$ met de steekproevenverdeling van de steekproefgrootheid $\overline{X}$. De oorspronkelijke populatieverdeling is een uniforme verdeling. De steekproevenverdeling van $\overline{X}$ is een symmetrische verdeling met als meest voorkomende waarde 3, wat eveneens het populatiegemiddelde was.
Het rekenkundig gemiddelde van de steekproefgrootheid $\overline{X}$ wordt berekend uit

$$E(\overline{X}) = \mu_{\overline{X}} = \frac{\sum \overline{x}_i n_i}{N^n} = \frac{1 \cdot 1 + 1.5 \cdot 2 + ...}{25} = 3$$

Hierbij wordt $\mu_{\overline{X}}$ omschreven als het populatiegemiddelde van de steekproefgemiddelden.
Uit de bovenstaande resultaten blijkt dus

$$E(\overline{X}) = \mu_{\overline{X}} = \mu$$

Dit kan ook afgeleid worden door te stellen dat het gemiddelde van een steekproef als volgt uit zijn waarnemingen wordt berekend

$$\overline{X} = \frac{1}{n}(X_1 + X_2 + X_3 + ... + X_n). \tag{1.7}$$

Het nemen van de verwachtingswaarde van beide leden geeft:

$$E(\overline{X}) = \frac{1}{n}(E(X_1) + E(X_2) + E(X_3) + ... + E(X_n)).$$

Aangezien elke steekproefvariabele X_t dezelfde verdeling heeft als de oorspronkelijke variabele $\mathbf{X_t}$ geldt:

$$E(\overline{X}) = \frac{1}{n}(\mu + \mu + \mu + ... + \mu) = \frac{1}{n}(n\mu) = \mu.$$

De variantie van $\overline{X}, \sigma^2_{\overline{X}}$, wordt berekend als volgt

$$\sigma^2_{\overline{X}} = \frac{\sum(\overline{x}_i - \mu_{\overline{X}})^2 n_i}{N^n} = \frac{(1-3)^2 \cdot 1 + (1.5-3)^2 \cdot 2 + ...}{25} = 1$$

Dit betekent dat de variantie van het steekproefgemiddelde $\overline{X}$ niet overeenkomt met de variantie van de oorspronkelijke populatie. Niettemin, bestaat de volgende relatie tussen de beide:

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} = \frac{2}{2} = 1$$

Door de vierkantswortel te trekken van beide leden wordt de uitdrukking voor de steekproefstandaarddeviatie verkregen:

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

Een andere manier om dit te berekenen is door te variantie te nemen van uitdrukking 1.7.

$$\begin{aligned} var(\overline{X}) = \sigma^2_{\overline{X}} &= var(\frac{1}{n}(X_1 + X_2 + X_3 + ... + X_n)) \\ &= \frac{1}{n^2}(var(X_1) + var(X_2) + var(X_3) + ... + var(X_n)) \\ &= \frac{1}{n^2}(\sigma^2 + \sigma^2 + \sigma^2 + ... + \sigma^2) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n} \end{aligned}$$

Aangezien er getrokken wordt zonder teruglegging zijn de trekkingen onafhankelijk en geldt

$$var(X_i + X_j) = var(X_i) + var(X_j) + 2covar(X_i, X_j) = var(X_i) + var(X_j)$$

omdat bij onafhankelijke trekkingen voor X_i en X_j wordt verwacht dat $covar(X_i, X_j) = covar(X_j, X_i) = 0$.

$\sigma_{\overline{X}}$ wordt gemeenzaam de *standaard fout* genoemd. De standaard fout meet dus de variabiliteit tussen steekproeven en is aldus een maat voor de schattingsfout van de mogelijke steekproeven voor het populatiegemiddelde μ. M.a.w. de standaard fout is een maat voor de toevallige variatie in steefproefgemiddelden.

Aangezien er minder variabiliteit is tussen steekproefgemiddelden dan tussen de oorspronkelijke populatiewaarden, geldt steeds voor $n > 1$ dat

$$\sigma_{\overline{X}} < \sigma$$

1.4.1.1 Normaal verdeelde populaties

De inferentiele statistiek is grote mate gebaseerd op normaal verdeelde populaties. In de realiteit, en in het bijzonder de economische realiteit, zijn er slechts weinig verschijnselen waarvan de populatieverdeling perfect normaal zou zijn. In de praktijk veronderstellen we echter reeds een normaalverdeelde populatie indien deze de normaal verdeling voldoende dicht benaderd. De normaliteit van de oorspronkelijk populatie van $\mathbf{X}$ heeft echter wel gevolgen voor de steekproevenverdeling van $\overline{X}$:

- De verdeling van $\overline{X}$ is eveneens normaal, onafhankelijk van de steekproefomvang, zoals geïllustreerd wordt door figuur 1.2(a).

- Het gemiddelde van de verdeling van $\overline{X}$ is gelijk aan het populatiegemiddelde van waaruit de steekproeven werden getrokken: $\mu_{\overline{X}} = \mu$
- Voor de variantie van de verdeling van $\overline{X}$ geldt: $\sigma^2_{\overline{X}} = \sigma^2/n$
- De standaardnormale Z-waarde kan steeds berekend worden.

$$Z = \frac{\overline{X} - \mu}{\sigma_{\overline{X}}}$$

De Z-waarden maakt berekeningen gemakkelijk, aangezien de standaard normaal tabel rechtstreeks kan geraadpleegd worden om probabiliteiten geassocieerd met de Z-waarden af te lezen.

Indien men twijfelt aan het normale karakter van een populatie, is het aangewezen om een geschiktheidstoets (χ^2 of Kolmogorov-Smirnov) uit te voeren voor een normaal verdeling (zie verder).

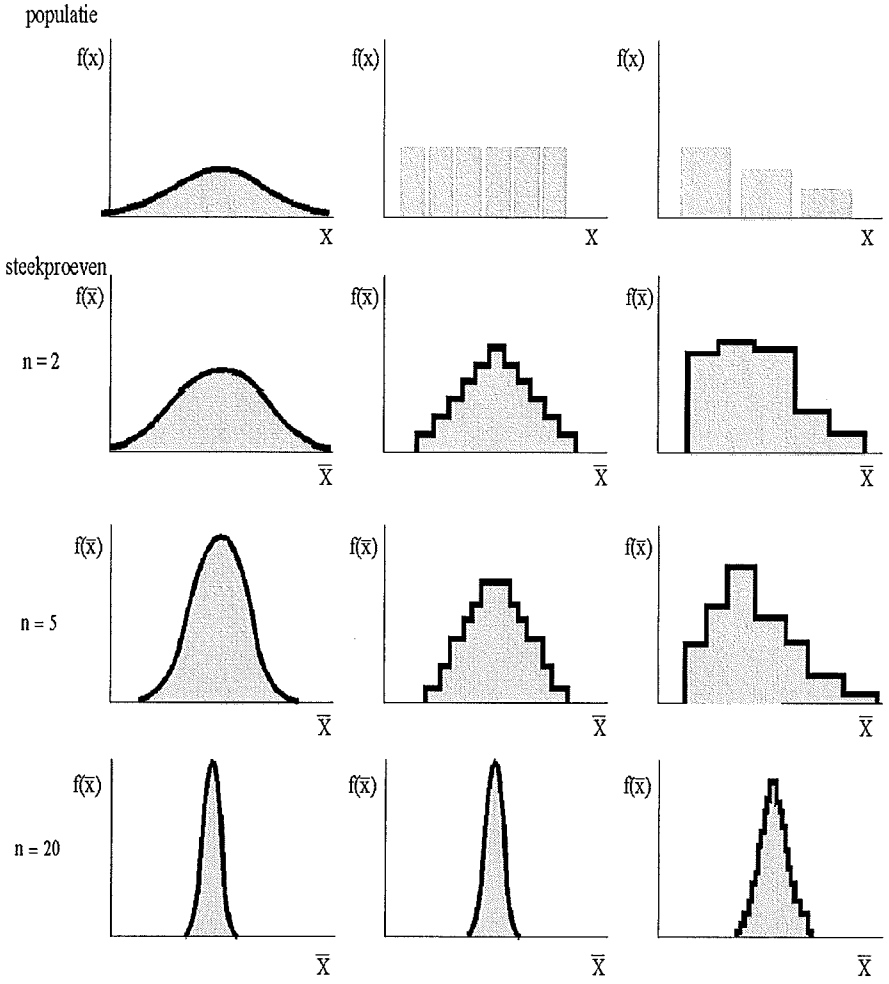
1.4.1.2 Niet-normaal verdeelde populaties

Indien de populatie geen normale verdeling volgt, dan lopen we het risico dat, indien de analyses een normaal verdeling veronderstellen, de resultaten misleidend zijn. Teneinde betrouwbare analyseresultaten te verkrijgen, moet de steekproevenverdeling van de steekproefgrootheid niettemin gekend zijn. Belangrijk daarbij is de *centrale limietstelling.*

Theorema 1 *De steekproevenverdeling van het steekproefgemiddelde $\overline{X}$ bepaald uit aselekte steekproeven met omvang n getrokken uit een niet-normaal verdeelde populatie met gemiddelde μ en standaarddeviatie σ, volgt benaderend een normaal verdeling met gemiddelde μ en standaarddeviatie $\sigma/\sqrt{n}$ indien de steekproefomvang n voldoende groot is.*

Dus, indien de omvang n van de steekproeven voldoende groot is, zal, onafhankelijk van de verdeling van de populatievariabele **X**, de steekproevenverdeling van het steekproefgemiddelde $\overline{X}$ steeds normaal zijn. Naarmate de n groter wordt, des te dichter zal de steekproevenverdeling de normaal verdeling benaderen. Dit wordt geïllustreerd door figuur 1.2 (b) en (c).
Wat is nu een voldoende grote steekproef? Een beschouwing hierbij is dat hoe meer de oorspronkelijke populatie afwijkt van de normaal verdeling, hoe groter de steekproef moet zijn. Nochtans wordt algemeen een steekproefomvang van 30 aanvaard als zijnde voldoende groot om normaliteit van de steekproevenverdeling te veronderstellen. De benadering wordt niettemin beter naarmate n stijgt.
Indien de oorspronkelijke populatie wel normaal verdeeld is, leiden alle steekproefgroottes n naar normaliteit voor de steekproevenverdeling van $\overline{X}$.

In het bovenstaand voorbeeld van pag. 6 was de oorspronkelijke populatievariabele **X** uniform verdeeld. Indien de steekproevenverdeling van $\overline{X}$ zou berekend worden op basis van een steekproefomvang n=4, dan zou een betere benadering van de normaal verdeling verkregen worden. Het totaal aantal mogelijke steekproeven met teruglegging van n=4 bedraagt dan al gauw 5^4=625.
Aangezien de relatie $\sigma_{\overline{X}} = \sigma/\sqrt{n}$ geldt en gezien σ=1 en n=4 is, bedraagt $\sigma_{\overline{X}}$=0.5. Dit betekent dat de standaarddeviatie van de steekproevenverdeling (de standaardfout) de helft



(a) (b) (c)

Figuur 1.2: Illustratie van de centrale limietstelling.

zo klein is als in het geval van n=2 ($\sigma_{\overline{X}}$=1). Bijgevolg zal de vorm van de steekproevenverdeling voor n=4 veel minder gespreid zijn dan die voor n=2 en aldus de typische klok-vorm beter benaderen.

1.4.1.3 Eindige populatie correctie

Het gebruik van de uitdrukking $\sigma_{\overline{X}} = \sigma/\sqrt{n}$ veronderstelt dat de populatie van waaruit de steekproeven getrokken worden een oneindig aantal elementen omvat of een eindige populatie met teruglegging is, zoals in het bovenstaand voorbeeld.
In de meeste praktische toepassingen worden steekproeven echter getrokken *zonder teruglegging*, zodat de opeenvolgende steekproeven die vereist zijn om de steekproevenverdeling op te zetten, *afhankelijk* zijn van de resultaten van de voorgaande trekkingen. Bijvoorbeeld, het tweemaal achter elkaar trekken van 5 kaarten uit een bundel van 52 kaarten. Voor de tweede steekproef van 5 kaarten zijn er nog 47 kaarten beschikbaar.
In geval van een eindige populatie moet de standaarddeviatie van de steekproevenverdeling aangepast worden in

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

waarbij

$$\sqrt{\frac{N-n}{N-1}}$$

de eindige populatie correctiefactor voorstelt.
Deze formule kan bepaald worden door de variantie te nemen van formule 1.7:

$$\begin{aligned} var(\overline{X}) = \sigma^2_{\overline{X}} &= var(\frac{1}{n}(X_1 + X_2 + X_3 + ... + X_n)) \\ &= \frac{1}{n^2}(var(X_1) + var(X_2) + var(X_3) + ... + var(X_n) \\ &\quad + covar(X_1, X_2) + covar(X_2, X_1) + ... + covar(X_{n+1}, X_n)) \end{aligned} \quad (1.8)$$

Voor een eindige populatie of trekkingen zonder teruglegging, zijn de opeenvolgende trekkingen niet meer onafhankelijk en is de $covar((X_i, X_j)$ verschillend van nul.

$$covar(X_i, X_j) = E[(X_i - \mu)(X_j - \mu)] = \frac{1}{N}\frac{1}{N-1}\sum_{i}^{N}\sum_{i \neq j}^{N}(x_i - \mu)(x_j - \mu)$$

Er geldt dat

$$\sum_{i}^{N}\sum_{j}^{N}(x_i - \mu)(x_j - \mu) = \sum_{i}^{N}\sum_{i \neq j}^{N}(x_i - \mu)(x_j - \mu) + \sum_{j}^{N}(x_j - \mu)^2$$

Aangezien

$$\sum_{i}^{N}\sum_{j}^{N}(x_i - \mu)(x_j - \mu) = \sum_{i}^{N}(x_i - \mu)\left[\sum_{j}^{N}(x_j - \mu)\right] = 0$$

verkrijgt men

$$\sum_{i}^{N}\sum_{i\neq j}^{N}(x_i-\mu)(x_j-\mu) = -\sum_{j}^{N}(x_j-\mu)^2$$

Zodat de formule voor de covariantie wordt:

$$covar(X_i,X_j) = -\frac{1}{N}\frac{1}{N-1}\sum_{j}^{N}(x_j-\mu)^2 = -\frac{\sigma^2}{N-1}$$

Invulling hiervan in formule 1.8 geeft

$$\begin{aligned}\sigma^2_{\overline{X}} &= \frac{1}{n^2}\left[n\sigma^2 - n(n-1)\frac{\sigma^2}{N-1}\right]\\ &= \frac{\sigma^2}{n}\frac{N-n}{N-1}\end{aligned}$$

De *eindige populatie correctiefactor* of *eindigheidscorrectie* is steeds kleiner dan 1. Indien de steekproefomvang n klein is in verhouding tot N, is de eindige populatie correctie ongeveer gelijk aan 1 en heeft bijgevolg geen invloed. De volgende vuistregel kan hierbij gehanteerd worden.

Regel 1 *De eindige populatie correctiefactor wordt toegepast op de berekening van* $\sigma_{\overline{X}}$ *indien* $n > 0.05N$.

Dit betekent dus dat een eindige populatie correctie slechts nodig is indien de steekproef meer dan 5% van de elementen van de populatie omvat.

Ook voor eindige populaties blijft de centrale limietstelling gelden. Nochtans heeft onderzoek uitgewezen dat strikt genomen zowel $n > 30$ als $N - n > 30$ moet zijn. In de praktijk is hieraan meestal wel voldaan.

Voorbeeld. Indien we in het voorgaande voorbeeld van pag. 6 van de leeftijden van de auto's, steekproeven zouden nemen zonder teruglegging, dan komt dit erop neer dat men de volgorde van de elementen in de steekproef niet beschouwd (bv. voor n=2 betekent dit dat de steekproeven met elementen 1,2 en 2,1 gelijk zijn. Dus het totaal aantal mogelijk steekproeven zonder teruglegging bedraagt dan C_N^n (een combinatie, want volgorde niet belangrijk). Voor het voorbeeld geeft dit

$$C_5^2 = \frac{5!}{2!(5-2)!} = 10$$

Voor elk van deze steekproeven n=2 wordt $\overline{x}$, de waarde van de steekproefgrootheid $\overline{X}$ berekend. De 10 waarden zijn:

$$1.5, 2.5, 2, 2.5, 3, 3.5, 3, 3.5, 4, 4.5$$

De frekwentieverdeling van $\overline{X}$ zit vervat in tabel 1.2
Het gemiddelde van deze 10 steekproefgemiddelden bedraagt:

$$\mu_{\overline{X}} = \frac{\sum \overline{x}_i n_i}{C_N^n} = \frac{1.5\cdot 1 + 2\cdot 1 + ...}{10} = 3$$

$\overline{x}$	frekwentie n_i	relatieve frekwentie
1.5	1	0.1
2	1	0.1
2.5	2	0.2
3	2	0.2
3.5	2	0.2
4	1	0.1
4.5	1	0.1

Tabel 1.2: Steekproevenverdeling van $\overline{X}$, steekproeven zonder teruglegging.

De variantie van $\overline{X}, \sigma^2_{\overline{X}}$, wordt berekend als volgt

$$\sigma^2_{\overline{X}} = \frac{\sum(\overline{x}_i - \mu_{\overline{X}})^2 n_i}{C^n_N} = \frac{(1.5-3)^2 \cdot 1 + (2-3)^2 \cdot 1 + ...}{10} = 0.75$$

Hierbij is de relatie tussen de standaardfout en de steekproefvariantie gegeven door

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} = \frac{2}{2} \cdot \frac{5-2}{5-1} = 0.75$$

Ten titel van overzicht wordt in figuur 1.3 het beslissingsproces voorgesteld bij keuze van de gepaste steekproevenverdeling van de steekproefgemiddelden $\overline{X}$.

Voorbeeld. Een vliegtuigbouwer heeft 750 gelijke vleugelstukken nodig, waarvan 700 vereist zijn voor installatie en 50 voor testen. De betreffende stukken zijn ontworpen om een treksterkte van gemiddelde van 18000 kg. met een standaarddeviatie van 3000 kg. te doorstaan. De treksterkte is normaal verdeeld. Veronderstel dat de 750 stukken de populatie uitmaken en dat alle volgende steekproeven afkomstig zijn uit deze populatie.
Vragen:

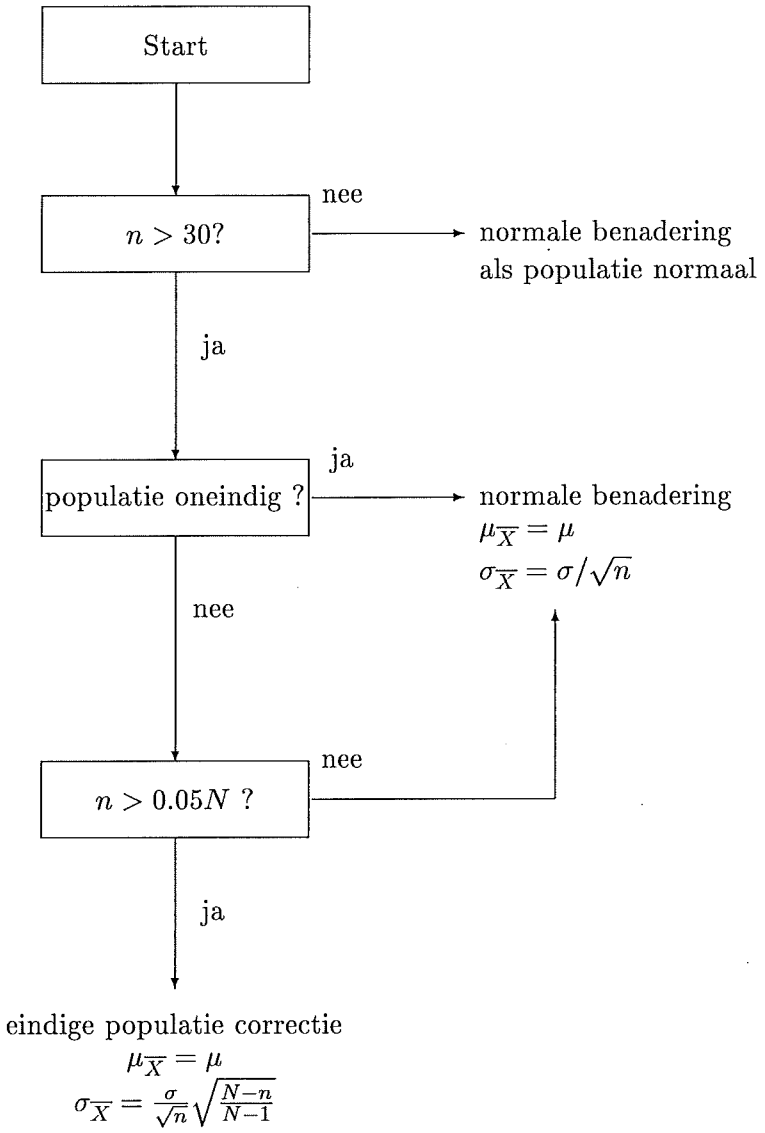
1. Indien 1 stuk toevallig geselecteerd wordt, wat is de kans dat zijn treksterkte minder is dan 17250 kg.
2. Indien 32 stukken toevallig geselecteerd worden, wat is de kans dat hun gemiddelde treksterkte minder is dan 17250 kg.
3. Indien 50 stukken toevallig geselecteerd worden, wat is de kans dat hun gemiddelde treksterkte minder is dan 17250 kg.

Oplossingen: (zie figuur 1.4)

1. De oorspronkelijke populatie is normaal verdeeld, bijgevolg zijn we op zoek naar het gebied < 17250 onder de normaalcurve $\mathbf{X} \sim N(18000, 3000)$:

$$z = \frac{x-\mu}{\sigma} = \frac{17250 - 18000}{3000} = -0.25$$

Volgens de tabel komt dit overeen met een kans van $P(Z < -0.25) = P(\mathbf{X} < 17250)$=0.4013.



Figuur 1.3: Beslissingen bij het beschouwen van een steekproevenverdeling voor $\overline{X}$.

2. $n > 30$ en $n/N = 0.04$: normale benadering zonder eindige populatie correctie. Gebruik van de steekproevenverdeling van de steekproefgemiddelden

$$z = \frac{\overline{x} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{x} - \mu_{\overline{X}}}{\sigma/\sqrt{n}} = \frac{17250 - 18000}{3000/\sqrt{32}} = -1.42$$

Volgens de tabel komt dit overeen met een kans van $P(Z < -1.42) = P(\overline{X} < 17250)$=0.0778.

3. $n = 50$ en $n/N > 0.05$: normale benadering met eindige populatie correctie:

$$z = \frac{\overline{x} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{x} - \mu_{\overline{X}}}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}} = \frac{17250 - 18000}{410} = -1.83$$

Volgens de tabel komt dit overeen met een kans van $P(Z < -1.83) = P(\overline{X} < 17250)$=0.0336.

1.4.2 Steekproevenverdeling voor het verschil tussen 2 steekproefgemiddelden

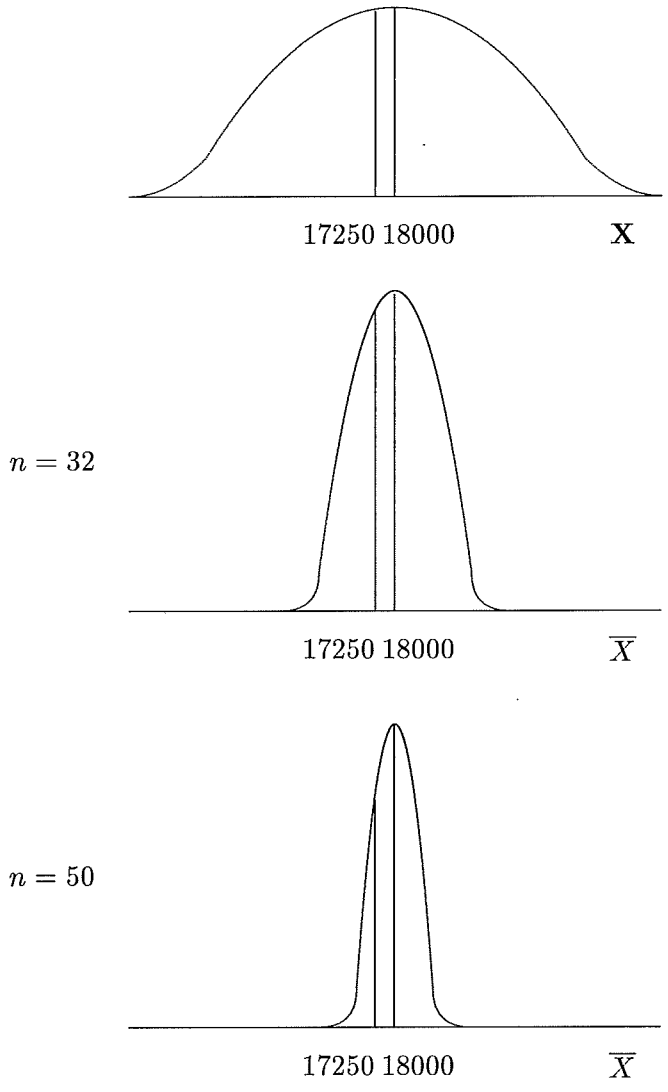
In vele praktische toepassingen bestaat er interesse voor het verschil tussen twee populatiegemiddelden. Het betreft hier vooral hypothesetoetsen om te zien of twee populaties significant verschillen, enz... Alvorens dergelijke inferentiele uitspraken te kunnen doen, is het noodzakelijk om de steekproevenverdeling van het verschil tussen de twee steekproefgemiddelden $\overline{X}_1 - \overline{X}_2$ te kennen.

De steekproevenverdeling van het verschil tussen de gemiddelden $\overline{X}_1 - \overline{X}_2$ van onafhankelijke steekproeven met omvang n_1 en n_2 getrokken uit twee normaal verdeelde populaties met gemiddelden μ_1 en μ_2 en standaarddeviaties σ_1 en σ_2 is eveneens normaal verdeeld met gemiddelde $\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$ en standaard deviatie $\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}$. Dit wordt geïllustreerd door figuur 1.5.
Twee steekproeven zijn *onafhankelijk* indien de selectie van de elementen uit de ene steekproef niet beïnvloed wordt door de elementenselectie uit de andere steekproef.

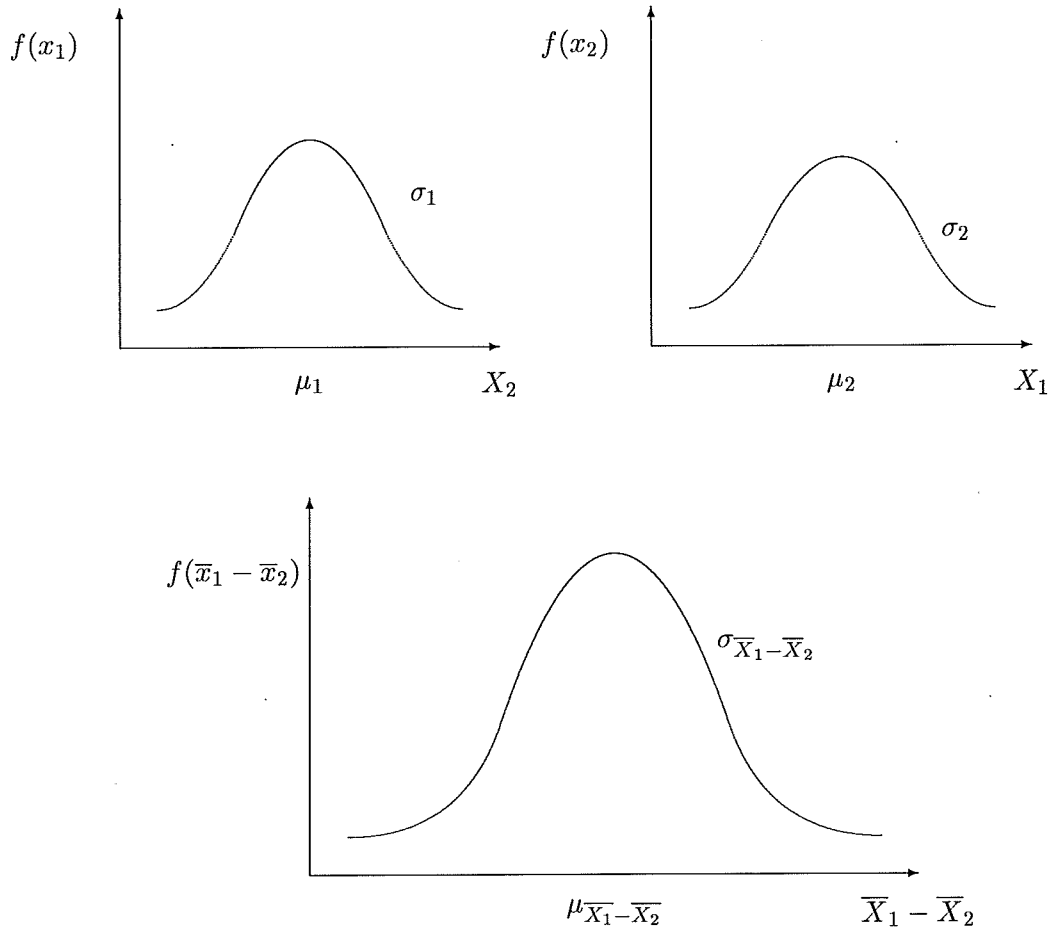
Net zoals we gedaan hebben voor het steekproefgemiddelde $\overline{X}$, kan hier ook hier de steekproevenverdeling van het verschil tussen 2 steekproefgemiddelden $\overline{X}_1 - \overline{X}_2$ opgezet worden op basis van 2 eindige populaties. Hiervoor worden alle $C_{N_1}^{n_1}$ steekproeven met omvang n_1 getrokken uit populatie 1 en wordt $(\overline{X})_1$ berekend. Hetzelfde wordt gedaan voor populatie 2. Vervolgens wordt het verschil tussen de twee gemiddelden berekend voor elk steekproefpaar van populatie 1 en 2. Vervolgens wordt de verdeling van al deze verschillen opgezet.

De formules blijven dezelfde zowel in het geval van eindige als oneindige populaties. In het geval van niet-normaal verdeelde populaties geldt de centrale limietstelling indien de steekproefgroottes $n_1, n_2 > 30$. Ook hiervoor is de steekproevenverdeling van het verschil van steekproefgemiddelden normaal met gemiddelde $\mu_{\overline{X}_1-\overline{X}_2} = \mu_1 - \mu_2$ en standaarddeviatie $\sigma_{\overline{X}_1-\overline{X}_2} = \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$.

Voorbeeld. Twee bedrijven maken eenzelfde hittebestendig produkt. Het produkt van bedrijf A is hittebestendig tot 505 K met een standaarddeviatie 10 K. Het produkt van



Figuur 1.4: Illustratie van het voorbeeld van de vleugelstukken.



Figuur 1.5: De originele populaties en de steekproevenverdeling van het verschil tussen de 2 steekproefgemiddelden.

bedrijf B is hittebestendig tot 475 K met een standaarddeviatie 7 K. Uit ervaring is bekend dat de verdeling van de hittebestendigheidstemperaturen normaal is. Een aselecte steekproef wordt genomen van 20 produkten van bedrijf A en 25 produkten van bedrijf B. Wat is de kans dat het verschil tussen de 2 steekproeven tussen 25 en 35 gelegen is?

Oplossing:
De steekproevenverdeling van $\overline{X}_A - \overline{X}_B$ is normaal verdeeld met gemiddelde $\mu_{\overline{X}_A - \overline{X}_B} = \mu_A - \mu_B$ en standaard deviatie $\sigma_{\overline{X}_A - \overline{X}_B} = \sqrt{\sigma_A^2/n_A + \sigma_B^2/n_B}$. Bijgevolg kan de z-statistiek berekend worden:

$$Z = \frac{\overline{X}_A - \overline{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$$

$$z_1 = \frac{25 - (505 - 475)}{\sqrt{\frac{10^2}{20} + \frac{7^2}{25}}} = -1.89 \qquad z_2 = \frac{35 - (505 - 475)}{\sqrt{\frac{10^2}{20} + \frac{7^2}{25}}} = 1.89$$

De beschouwde kans is

$$P(-1.89 \leq z \leq 1.89) = P(z \leq 1.89) - P(z \leq -1.89) = 0.9706 - 0.0294 = 0.9412$$

1.4.3 Steekproevenverdeling van de steekproeffractie

De steekproeffractie P kan beschouwd worden als het steekproefgemiddelde in het geval van een dichotome variabele $\mathbf{X_t}$, d.i. een variabele die slechts twee waarden kan aannemen. De steekproeffractie P stelt bijgevolg de fractie van het aantal "successen" in de steekproef voor. Indien van de N elementen in de populatie er $N_s(X = 1)$ "successen" en $N_m(X = 0)$ "mislukkingen" zijn, dan gelden de relaties

$$N_s + N_m = N \qquad \pi = N_s/N$$

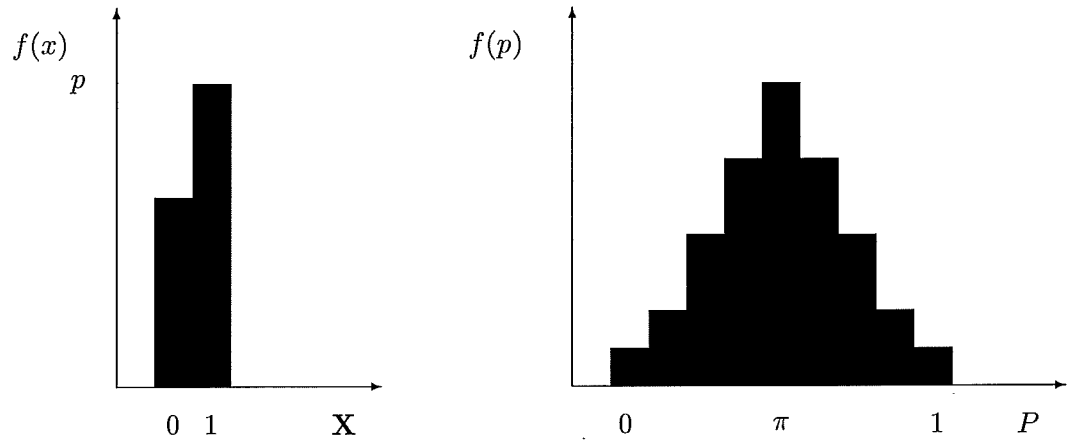
Veronderstel dat uit deze populatie een aselecte steekproef van omvang n wordt getrokken, en van elk element wordt er vastgesteld of het een succes dan wel een mislukking is. Dit experiment is een Binomiaal experiment met n onafhankelijke Bernouilli-experimenten, die de trekking van telkens 1 element omvatten. Het aantal successen in de steekproef $S = \sum_i X_i$ is bijgevolg binomiaal verdeeld $S \sim B(n, \pi)$ met gemiddelde en standaarddeviatie:

$$\mu_s = n\pi \qquad \sigma_s = \sqrt{n\pi(1 - \pi)}$$

De uiteindelijke steekproeffractie wordt verkregen door het aantal successen te delen door de steekproefgrootte: $P = S/n$

$$\mu_P = \pi \qquad \sigma_P = \sqrt{\frac{\pi(1 - \pi)}{n}}$$

De steekproevenverdeling van P kan volgens eenzelfde procedure als voor het steekproefgemiddelde $\overline{X}$ en het verschil tussen twee steekproefgemiddelden $\overline{X}_1 - \overline{X}_2$ samengesteld worden. Vertrekkende van eindige populatie worden alle mogelijke steekproeven van een bepaalde grootte n genomen. Voor elke steekproef wordt vervolgens P berekend. Op basis hiervan wordt de waarschijnlijkheidsverdeling voor P opgesteld, zoals voorgesteld door figuur 1.6.



Figuur 1.6: Dichotome populatie en steekproevenverdeling van de steekproefproportie P.

Indien de steekproef voldoende groot is, dan is de steekproevenverdeling van de steekproeffracties normaal verdeeld ten gevolge van de centrale limietstelling, met een gemiddelde $\mu_P = \pi$ en een standaarddeviatie $\sigma_p = \sqrt{\pi(1-\pi)/n}$, t.t.z. $P \sim N(\pi, \sqrt{\pi(1-\pi)/n})$.

Regel 2 *Deze normaal benadering voor de steekproevenverdeling van P is enkel toegestaan indien $np > 5$ én $n(1-p) > 5$.*

In het geval van *steekproeven zonder teruglegging* uit een eindige populatie, is de steekproevenverdeling eigenlijk de *hypergeometrische verdeling*, niet meer de binomiaal verdeling (eindige populatie met teruglegging of oneindige populatie). Voor voldoende grote steekproeven, vormt de binomiaal verdeling een goede benadering voor de hypergeometrische. De binomiaal verdeling wordt dan op haar beurt benaderd door de normaal verdeling.

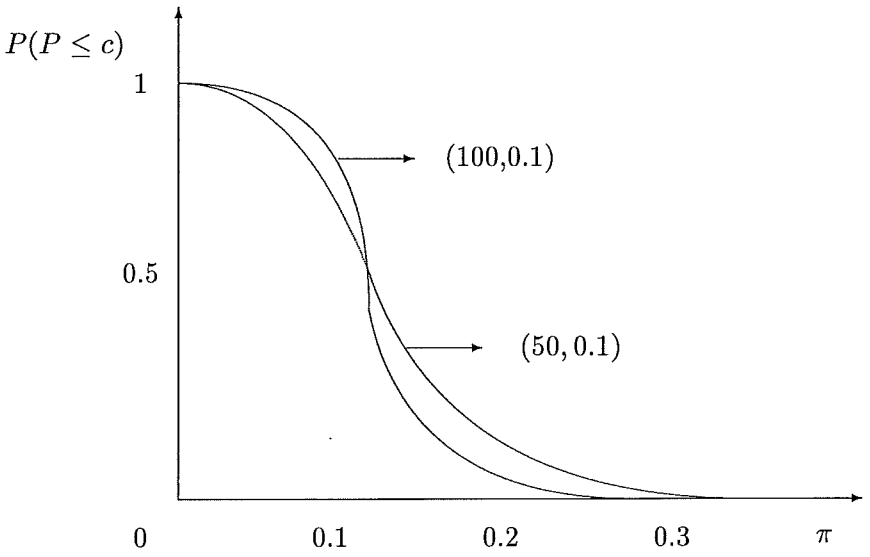
Net zoals voor de steekproevenverdeling van het steekproefgemiddelde, dient hier ook een *eindige populatie correctiefactor* te worden toegepast voor de berekening van de standaarddeviatie indien $n > 0.05N$:

$$\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}}\sqrt{\frac{N-n}{N-1}}$$

Toepassing. Vele toepassingen van steekproeffracties hebben betrekking op de *kwaliteitscontrole*. In vele bedrijven wordt aan de hand van een steekproef nagegaan of een inkomende partij goederen niet te veel defecte exemplaren bevat. Op basis van de fractie defecten uit de steekproef wordt dan beslist om een partij goederen al of niet te aanvaarden. Men spreekt in dat verband vooral van *(n,c)-keuringsschema's* waarbij n de steekproefomvang en c de fractie toelaatbare defecten in de steekproef voorstellen. De goedkeurkans wordt hierbij gegeven door

$$P_\pi(P \leq c)$$

Zo kunnen consument en producent afspreken om bijvoorbeeld een (100,0.1)-keuringsschema te hanteren. Hierbij wordt de partij goederen afgekeurd en op kosten van de leverancier



Figuur 1.7: Keuringskarakteristieken.

vervangen indien op een steekproef van n=100 er meer dan 10% defecten worden gevonden. In termen van het aantal defecten S betekent dit $S \sim B(100, \pi)$, waarbij π de werkelijke fractie defecten in de partij voorstelt. De goedkeurkans wordt hier gegeven door:

$$P_\pi(P \leq 0.1) = P_\pi(S \leq 10)$$

De keuringskarakteristiek geeft het verband weer tussen de goedkeurkans $P_\pi(P \leq c)$ en de werkelijke fractie defecten in de populatie π. Figuur 1.7 geeft de keuringskarakteristieken weer voor de $(50, 0.1)$ en $(100, 0.1)$ keuringsschema's.
Indien de steekproefomvang n=50, is bijvoorbeeld de goedkeurkans ongeveer 40% (of afkeurkans 60%) bij een werkelijke fractie defecten π=0.125. De keuringskarakteristiek met steekproefomvang n=100 is steiler: goede partijen worden sneller goedgekeurd en slechte partijen sneller afgekeurd omdat de beslissing gesteund is op een groter aantal waarnemingen.

Voorbeeld. Een producent van nagels heeft vastgesteld dat 3% van zijn nagels defecten vertonen. Een steekproef met omvang 300 wordt getrokken en geanalyseerd. Wat is de kans dat de fractie defecten gelegen is tussen 0.02 en 0.035 ?

Oplossing:
Het gebruik van de normaal benadering is toegestaan aangezien $n\pi = 300 \times 0.03 = 9 > 5$ en $n(1-\pi) = 300 \times 0.97 = 291 > 5$. Dit impliceert dat de steekproeffractie normaal verdeeld is met gemiddelde en standaarddeviatie gelijk aan

$$\mu_P = \pi = 0.03$$

$$\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.03(0.97)}{300}} = 0.0098$$

De overeenkomstige standaardnormale waarden zijn

$$z_1 = \frac{0.02 - 0.03}{0.0098} = -1.02 \qquad z_2 = \frac{0.035 - 0.03}{0.0098} = 0.51$$

De kans die we zoeken:

$$P(-1.02 \leq z \leq 0.51) = P(z \leq 0.51) - P(z \leq -1.02) = 0.6950 - 0.1539 = 0.5411$$

1.4.4 Steekproevenverdeling voor het verschil tussen 2 steekproeffracties

Net zoals voor de het verschil tussen 2 steekproefgemiddelden, kan er eveneens een steekproevenverdeling voor het verschil tussen 2 steekproeffracties opgesteld worden.
De steekproevenverdeling van $P_1 - P_2$ is eveneens benaderend normaal verdeeld met gemiddelde

$$\mu_{P_1-P_2} = \pi_1 - \pi_2$$

en standaarddeviatie

$$\sigma_{P_1-P_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

indien de onafhankelijke steekproeven met omvang n_1 en n_2 getrokken zijn uit 2 populaties met proporties π_1 en π_2. De normaal benadering vereist dat de omvang van de beide steekproeven voldoende is, t.t.z. dat $n_1p_1, n_2p_2, n_1(1-p_1), n_2(1-p_2) > 5$.

Voorbeeld. Er wordt beweerd dat 30% van de gezinnen in een randgemeente en 20% van de gezinnen in een stadskern minstens 1 kind hebben. Aselecte steekproeven met een omvang van 100 gezinnen in zowel de randgemeente als in de stadskern leveren een fractie van 0.34 en 0.13 op. Wat is de kans dat het verschil minstens zo groot is als wordt beweerd?

Oplossing:
Aangezien $100 \times 0.30 = 30 > 5$ en $100 \times 0.20 = 20 > 5$ is de steekproevenverdeling van het verschil tussen de steekproefproporties normaal verdeeld met gemiddelde

$$\mu_{P_1-P_2} = 0.3 - 0.2 = 0.1$$

en standaarddeviatie

$$\sigma_{P_1-P_2} = \sqrt{\frac{0.3(0.7)}{100} + \frac{0.2(0.8)}{100}} = 0.061$$

Het geobserveerde verschil tussen de proporties bedraagt:

$$P_1 - P_2 = 0.34 - 0.13 = 0.21$$

De standaardnormale waarde is dan

$$z = \frac{0.21 - 0.10}{0.061} = 1.83$$

De probabiliteit dat het verschil minstens zo groot is als beweerd wordt is $P(Z \leq 1.83)$=0.0336.

1.4.5 Steekproevenverdeling van de andere steekproefgrootheden

Voor de andere steekproefgrootheden, waaronder de steekproefvariantie S^2, de steekproefcovariantie S_{xy} en de steekproefcorrelatie R_{xy} is de steekproevenverdeling niet voor de hand liggend. Zo kan bijvoorbeeld de steekproevenverdeling van de steekproefvariantie S^2 bepaald worden via de steekproevenverdeling van $(n-1)S^2/\sigma^2$, die χ^2-verdeeld is met $n-1$ vrijheidsgraden.
Een ander voorbeeld is de steekproevenverdeling van de steekproefcorrelatie R_{xy}. Om inferenties met betrekking tot de correlatiecoëfficient te kunnen doen, wordt de steekproefgrootheid $R_{xy}\sqrt{\frac{n-2}{1-R_{xy}^2}}$ gebruikt die een t-verdeling volgt met $n-2$ vrijheidsgraden.

De steekproevenverdelingen van de steekproefgrootheden andere dan het steekproefgemiddelde en de steekproeffractie zullen besproken worden indien hun gebruik aangewezen is.

1.5 Samenvatting

1. Bij een aselecte (toevallige) steekproef heeft elk element van de populatie een even grote kans om geselecteerd te worden. Indien dit gebeurt met teruglegging zijn de trekkingen onafhankelijk, zonder teruglegging niet.

2. - populatievariabele: $\mathbf{X}$
   - steekproefvariabele: X
   - steekproefwaarneming: x
   - steekproefgrootheden: $\overline{X}, P, S^2, S_{xy}, R_{xy}$
   - steekproefkengetallen: $\overline{x}, p, s^2, s_{xy}, r_{xy}$

3. De voornaamste verdelingsparameters van de steekproevenverdelingen van het steekproefgemiddelde $\overline{X}$ en de steekproeffractie P.

populatie	oneindig (teruglegging)		eindig (geen teruglegging)	
steekproefgrootheid	$\overline{X}$	P	$\overline{X}$	P
verwachting	μ	π	μ	π
variantie	$\frac{\sigma^2}{n}$	$\frac{\pi(1-\pi)}{n}$	$\frac{\sigma^2}{n}\frac{N-n}{N-1}$	$\frac{\pi(1-\pi)}{n}\frac{N-n}{N-1}$

4. Het steekproefgemiddelde $\overline{X}$ fluctueert rond zijn doel μ met een standaardfout $\sigma/\sqrt{n}$.

5. Het verschil tussen 2 steekproefgemiddelden $\overline{X_1} - \overline{X_2}$ fluctueert rond zijn doel $\mu_1 - \mu_2$ met een standaardfout $\sigma_1/\sqrt{n_1} + \sigma_2/\sqrt{n_2}$.

6. Hoe groter de steekproefomvang n, hoe meer $\overline{X}$ zich concentreert rond μ en hoe beter de steekproevenverdeling de normaal verdeling benadert.

7. De steekproeffractie P fluctueert rond zijn doel π met een standaardfout $\sqrt{\pi(1-\pi)/n}$. De steekproeffractie is eigenlijk het steekproefgemiddelde voor een dichotome variabele.

8. Het verschil tussen 2 steekproeffracties $P_1 - P_2$ fluctueert rond zijn doel $\pi_1 - \pi_2$ met een standaardfout $\sqrt{\pi_1(1-\pi_1)/n + \pi_2(1-\pi_2)/n_2}$.

9. Voor eindige populaties of voor trekkingen zonder terugleggingen wordt de standaardfout $\sigma_{\overline{X}}$ gereduceerd met een factor $\sqrt{(N-n)/(N-1)}$.

10. Bij de normaal verdeling worden probabiliteiten berekend a.d.h.v. de standaardnormale Z-waarden.

11. Een aantal belangrijke vertalingen:

populatie	*population*
(on)eindige populatie	*(in)finite population*
steekproef	*sample*
aselecte steekproef	*random sample*
steekproefname	*sampling*
steekproevenverdeling	*sample distribution*
steekproefgrootheid	*sample statistic*
steekproefomvang	*sample size*
met of zonder teruglegging	*with or without replacement*
gemiddelde	*mean*
variantie	*variance*
fractie	*proportion*
standaarddeviatie	*standard deviation*

1.6 Oefeningen

1. Het gemiddelde en de standaarddeviatie van de wachttijd aan een loket van een bank bedraagt 150 sec. en 18 sec. Indien 36 klanten aselect gekozen worden, wat is dan de kans dat hun gemiddelde wachttijd tussen 150 sec. en 156 sec. begrepen is.

2. Een populatie mensen heeft een gemiddeld gewicht van 69 kg. en een standaarddeviatie van 3.2 kg.
   (a) Indien er veel aselecte steekproeven met omvang n=4 zouden getrokken worden en telkens het steekproefgemiddelde zou berekend worden, hoe zouden deze steekproefgemiddelden fluctueren?
   (b) 1 steekproef met een gemiddelde van 70 kg; is dit typisch of zeldzaam?
   (c) Wat is de kans dat het steekproefgemiddelde hoger is dan 70 kg.?
   (d) Beantwoordt vraag (a) opnieuw voor n=16.
   (e) Juist of fout: een verdubbeling van de steekproefomvang, verviervoudigt de juistheid waarmee $\overline{X}$ een schatting geeft voor μ.

3. Een lift is berekend op een totaal gewicht van 500 kg. en een capaciteit van 7 personen. Veronderstel dat het gewicht van alle personen die de lift gebruiken gemiddeld 70 kg. is met een standaarddeviatie van 10 kg. Wat is de kans dat een steekproef van 7 personen het maximaal gewicht van 500 kg. zou overstijgen? Moet het populatiegewicht hiertoe normaal verdeeld te zijn?

4. Wat is de kans dat er meer dan 10 jongens zijn op 15 kinderen? Los deze oefening op 2 manieren op (benaderend en exact).

5. De gemiddelde levensduurte van een boorstuk bedraagt 41.5 u. met een standaarddeviatie van 2.5 u. Wat is de kans dat een aselecte steekproef van 50 boorstukken een gemiddelde oplevert tussen 40.5 u. en 42 u.? Wat is de kans op een gemiddelde kleiner dan 39? Is de veronderstelling van normaliteit van de oorspronkelijke populatie noodzakelijk?

6. Een bedrijf stelt 1500 mensen tewerk. De gemiddelde maandelijkse bijdrage per werknemer voor "werken voor het goede doel" per werknemer bedraagt gemiddeld 25.75 BEF. met een standaarddeviatie van 5.25 BEF. Wat is de kans dat een steekproef van 100 mensen in het bedrijf een gemiddelde tussen 25 BEF. en 27 BEF. oplevert?
Hoe evolueert deze kans indien het aantal werknemers in het bedrijf met 1000 toeneemt.

7. Een bedrijf weet dat 35% van de documenten die door haar secretariaat worden opgesteld minstens 1 fout bevatten. Wat is de kans dat de proportie documenten met minstens 1 fout uit een steekproef van 20 documenten tussen 33.2% en 38% gelegen is. Bereken deze kans voor een totaal documentenaantal gelijk aan 200 en 500.

8. Een bepaald produktieproces kan volgens twee methoden uitgevoerd worden. De tijdsduur is de variabele die hierbij beschouwd wordt. De s.d. op de tijdsduur van methode A bedraagt 3 min., op die van B 3.46 min. Een aselecte steekproef van 35 werknemers die methode A gebruiken leverde een gemiddelde tijdsduur van 25 min. op. Een gemiddelde tijdsduur van 23 min. werd bekomen op basis van een aselecte steekproef van 35 werknemers die methode B gebruikten. Wat is de kans dat het verschil tussen de twee steekproefgemiddelden groter of gelijk is aan hetgeen hier bekomen werd, aangenomen dat er eigenlijk geen verschil mag zijn tussen de tijdsduren van beide methoden? Is normaliteit van de twee oorspronkelijke populaties vereist?

9. Een onderzoeksbureau heeft vastgesteld dat 16% van de bedrijven van type A meer marktonderzoek hebben gedaan. Voor de bedrijven van het type B was dit 9%.
(a) Wat is het gemiddelde en de standaarddeviatie van de steekproevenverdeling van het verschil tussen steekproefporporties, gebaseerd op onafhankelijke aselecte steekproeven van 100 bedrijven van elk type?
(b) Welke fractie van de steekproefverschillen $P_A - P_B$ ligt tussen 0.05 en 0.10?
(c) Veronderstel dat een aselecte steekproef van 100 wordt getrokken van elk bedrijfstype, wat is dan de kans dat het verschil kleiner of gelijk is aan 0.02?

10. Het werpen van een eerlijk muntstuk. Wat is de kans dat de fractie kruis tussen 40% en 60% gelegen.
(a) steekproefomvang n=10. (b) steekproefomvang n=100. (c) steekproefomvang n=1000.

11. Een populatie bestaat uit het aantal defecte transistoren per levering aan een assemblagefabriek. Het aantal defecte transistoren is 2 in de 1e, 4 in de 2e, 6 in de 3e en 8 in de 4e levering.
    (a) Bereken het gemiddelde en de standaardafwijking van de populatie.
    (b) Bepaal de steekproevenverdeling van $\overline{X}$ voor een steekproefomvang van 2.
    (c) Bepaal het gemiddelde en de standaardafwijking van het steekproefgemiddelde.

12. Een autobatterij van het merk A heeft een levensduur van 3.5 jaren gemiddeld en een standaarddeviatie van 0.4 jaren. De levensduur van eenzelfde batterij van merk B bedraagt gemiddeld 3.3 jaren met een standaardafwijking van 0.3 jaren. Wat is de kans dat de gemiddelde levensduur berekend uit een toevallige steekproef van 25 batterijen van merk A minstens 0.4 jaren langer is dan die berekend uit een aselecte steekproef van 36 batterijen van merk B?

Hoofdstuk 2

Schattingen

2.1 Inleiding

In de steekproeftheorie zijn we vertrokken vanuit een bekende populatie om conclusies te trekken m.b.t. de waar te nemen steekproefgrootheden. In de *inferentiele* of *analyzerende statistiek* vertrekt men van de steekproefgrootheden die waargenomen werden in de steekproef om conclusies te maken voor de onbekende populatieparameters. Belangrijk is echter hoe vaak onjuiste conclusies kunnen voorkomen en met welke gradaties. Veel hangt af van de methoden die gebruikt worden om conclusies te trekken. Het belangrijkste deel van de inferentiele statistiek heeft dan ook betrekking op de verschillende beschikbare methoden.

Definitie 2 *Statistische inferentie is de procedure waarbij een statistische conclusie over een populatie wordt gedaan op basis van de informatie afkomstig van een steekproef aselect getrokken uit de populatie.*

De *statistische conclusie* houdt meestal een uitspraak in over een onbekend kengetal (parameter) van de populatie, zoals het populatiegemiddelde μ, een populatiefractie π of een populatievariantie σ^2. De uitspraken zijn dan gebaseerd op het overeenkomstig steekproefkengetal berekend uit een steekproef: het steekproefgemiddelde $\overline{x}$, de steekproeffractie p en de steekproefvariantie σ^2.
In wat volgt, wordt het *onbekend populatiekengetal* algemeen voorgesteld door θ. Dit impliceert:

$$\theta = \mu \qquad \theta = \pi \qquad \theta = \sigma^2$$

Twee types van statistische inferentie kunnen onderscheiden worden: *schattingen* en *hypothese toetsen*. Dikwijls wordt schattingen nog verder opgedeeld volgens puntschattingen en intervalschattingen. Dit hoofdstuk beschrijft de puntschattingen en de intervalschattigen. De hypothesetoetsen worden in het volgende hoofdstuk behandeld.

2.2 Puntschattingen

Definitie 3 *Een puntschatting bevat één enkele waarde voor de populatieparameter berekend uit de steekproefgegevens en kan beschouwd als de beste gissing voor de overeenkomstige populatieparameter.*

Een *schatter* is de procedure of regel om de schatting te berekenen. Een voorbeeld hiervan is het steekproefgemiddelde $\overline{X} = \sum X_i/n$, waarvan de waarde kan gebruikt worden als schatting voor het populatiegemiddelde. Een puntschatter is bijgevolg een toevalsvariabele, bijvoorbeeld U, met als realisatie de waarde berekend uit de steekproef, de puntschatting, bijvoorbeeld u.

Samengevat kunnen we stellen dat het puntschatten een onderdeel is van de inferentiele statistiek waarbij de puntschatter de statistische procedure is en de puntschatting de statistische conclusie is.

2.2.1 Eigenschappen van een goede schatter

De kwaliteit van een schatter wordt meestal beoordeeld op basis van drie criteria: de onvertekendheid, de consistentie en de efficientie.

2.2.1.1 De onvertekendheid

De puntschatting u is 1 getal berekend uit de steekproef. De *schattingsfout* is dan het verschil tussen de puntschatting u en de onbekende waarde voor de populatieparameter θ

$$u - \theta$$

Schattingsfouten kunnen dus beschouwd worden als de uitkomsten van de toevalsvariabele $U - \theta$. Hoe beter de schatter hoe sterker de verdeling van deze toevalsvariabele rond nul geconcentreerd is. Dit impliceert dat de verwachting $E(U - \theta)$ in de nabijheid van nul dient te liggen. De verwachting $E(U - \theta)$ wordt de *vertekening* genoemd.

$$E(U - \theta) = E(U) - \theta = \mu_u - \theta$$

Een puntschatter is *zuiver* indien de vertekening nul is, t.t.z. dat bij het herhaaldelijk berekenen van de puntschatter U voor θ de werkelijke waarde van θ wordt gevonden.

Het steekproefgemiddelde is een zuivere schatter voor het populatiegemiddelde:

$$E(\overline{X}) = \mu$$

De steekproeffractie is een zuivere schatter voor de populatiefractie:

$$E(P) = \pi$$

De steekproefvariantie is een zuivere schatter voor de populatievariantie

$$E(S^2) = \sigma^2$$

indien de steekproefvariantie berekend wordt als:

$$S^2 = \frac{1}{n-1}\sum(X_t - \overline{X})^2$$

Voorbeeld. Verwijzend naar het voorbeeld op pag. 6 hadden we reeds aangetoond dat $E(\overline{X}) = \mu$. Om $E(S^2) = \sigma^2$ te verantwoorden, wordt $E(S^2)$ berekend door de 25 waarden voor de steekproefvariantie te berekenen:

$$0, 0.5, 2, 4.5, 8, 0.5, 0, 0.5, 2, 4.5, 2, 0.5, 0, 0.5, 2, 4.5, 2, 0.5, 0, 0.5, 8, 4.5, 2, 0.5, 0$$

$$E(S^2) = \frac{0 + 0.5 + ... + 0.5 + 0}{25} = 2 = \sigma^2$$

In het geval zonder teruglegging (zie pag. 12) zijn het de volgende 10 waarden van waaruit $E(S^2)$ wordt berekend:

$$0.5, 0.5, 2, 4.5, 8, 4.5, 2, 0.5, 2, 0.5$$

$$E(S^2) = \frac{0 + 0.5 + ... + 0.5}{10} = 2.5 \neq \sigma^2$$

Dus voor een eindige populatie zonder teruglegging is de steekproefvariantie geen goede schatter voor de populatievariantie maar wel voor de volgende populatieparameter σ'^2:

$$\sigma'^2 = \frac{\sum(\mathbf{x_t} - \mu)^2}{N - 1} = \frac{10}{4} = 2.5$$

Ter inlichting:

$$\sigma^2 = \frac{\sum(\mathbf{X_t} - \mu)^2}{N} \qquad \sigma'^2 = \frac{\sum(\mathbf{X_t} - \mu)^2}{N - 1}$$

In wat volgt zullen we echter steeds een oneindige populatie of een eindige populatie met teruglegging beschouwen.
Uit het voorgaande volgt dat de steekproefstandaarddeviatie een onvertekende schatter is voor de populatiestandaarddeviatie:

$$E(S) = \sigma$$

2.2.1.2 Efficientie

Een schatter U wordt efficient genoemd indien zijn variantie klein is, t.t.z. indien zijn kansverdeling sterk geconcentreerd is rond zijn verwachting. De variantie van een schatter U wordt berekend als:

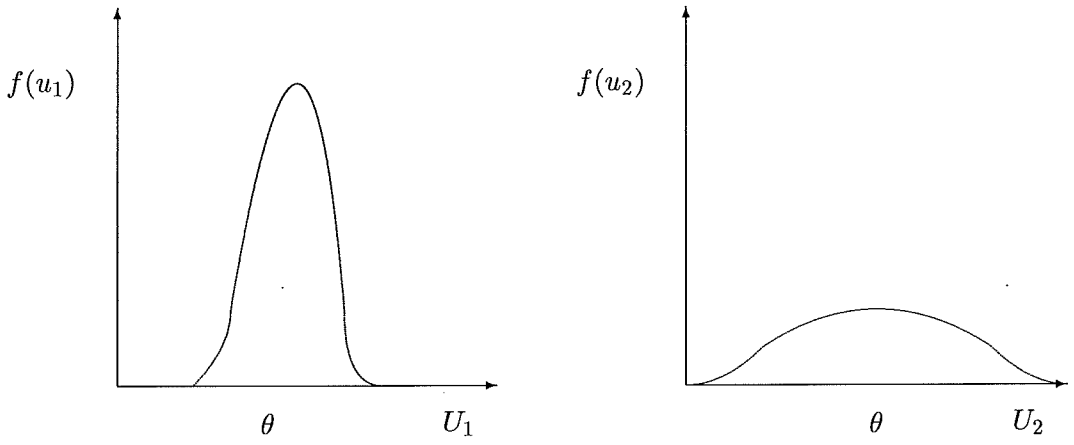
$$Var(U) = E(U - \mu_u)^2$$

De efficientie van een schatter kan slechts relatief bepaald worden ten opzichte van een andere schatter. Met andere woorden, de efficientie van een schatter U_1 t.o.v. U_2 wordt bepaald door de ratio $var(U_1)/var(U_2)$.
Zoals op figuur 2.1 wordt weergegeven, heeft de efficiente schatter de kleinste variantie. Een kleine variantie wijst op het beperkt karakter van de schattingsfout van de schatter.

Voorbeeld. Vergelijking van het steekproefgemiddelde en de steekproefmediaan voor de efficientste schatter te bepalen voor het schatten van het gemiddelde van een normale populatie. De variantie van het steekproefgemiddelde $\overline{X}$ bedraagt σ^2/n; de variantie van de steekproefmediaan $\tilde{X}$ als zuivere schatter voor het populatiegemiddelde bedraagt ongeveer $\pi\sigma^2/2n$, zodat de efficientie van de mediaan t.o.v. het gemiddelde gelijk is aan:

$$\frac{var(\overline{X})}{var(\tilde{X})} = \frac{\sigma^2/n}{\pi\sigma^2/2n} = \frac{2}{\pi} = 0.64$$

Dit betekent dat voor een gegeven steekproefomvang n, de variantie van het steekproefgemiddelde kleiner is dan de variantie van de steekproefmediaan. Bijgevolg is het steekproefgemiddelde $\overline{X}$ een efficientere schatter dan de steekproefmediaan $\tilde{X}$. De $var(\tilde{X}) = 1.57var(\overline{X})$ kan geïnterpreteerd worden als volgt: indien de steekproefmediaan een even efficiente schatter moet zijn als het steekproefgemiddelde, dan moet de steekproefmediaan berekend worden op



Figuur 2.1: Een efficiente schatter U_1 en een inefficiente schatter U_2

een steekproef die 57% groter is dan die voor de berekening van het steekproefgemiddelde.

Het steekproefgemiddelde is niet voor alle populatieverdelingen een efficientere schatter dan de steekproefmediaan, zoals aangetoond wordt in het volgende voorbeeld.

Voorbeeld. Als de populatieverdeling Laplace is, dan is $var(\tilde{X}) \approx 0.50\sigma^2/n$ zodat

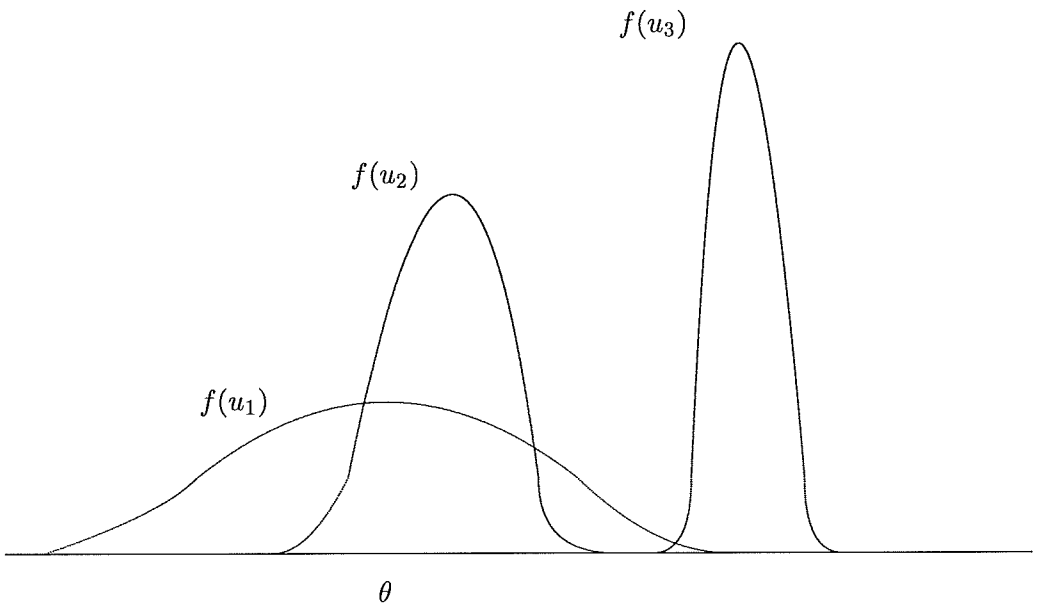
$$\frac{var(\overline{X})}{var(\tilde{X})} = \frac{\sigma^2/n}{0.50\sigma^2/n} = 2$$

De Laplace-verdeling wordt gekenmerkt door dikke staarten, waardoor het steekproefgemiddelde een dubbel zo grote variantie heeft als de variantie van de steekproefmediaan, die op zijn beurt ongevoelig is voor outliers.

Bij de bepaling van de efficientie van een zuivere schatter, wordt de keuze gedetermineerd door de minimale variantie. Indien er ook vergeleken wordt met *onzuivere schatters*, is het gebruik van de minimale variantie alleen niet meer voldoende als criterium, zoals geïllustreerd wordt door figuur 2.2.

Een goed criterium voor de kwaliteit van de schatter is daarom een combinatie van minimale vertekening en minimale variantie. Die combinatie wordt weergegeven door de *gemiddelde kwadratische afwijking* ($M(U)$), die de som is van de variantie en de vertekening in het kwadraat:

$$M(U) = E(U-\theta)^2 = E(U-\mu_u)^2 + (\mu_u - \theta)^2$$



Figuur 2.2: Schatter U_1 is zuiverste, U_2 heeft kleinste MSE en is de efficienste, U_3 heeft kleinste variantie.

Deze gemiddelde kwadratische afwijking is een maat voor de spreiding van de schattingen rond de waarde van de populatieparameter θ.
De relatieve efficientie van schatter U_1 t.o.v. schatter U_2 wordt bepaald door de ratio $M(U_1)/M(U_2)$. Indien de beide schatters zuiver zijn, wordt de gemiddelde kwadratische afkwijking $M(\cdot)$ beperkt tot de variantie. Wanneer één of beide schatters vertekend zijn, kan $M(\cdot)$ beschouwd worden als een algemene vorm van variantie, die zowel toepasbaar is op vertekende als onvertekende schatters.

Voorbeeld. Een reclamebureau heeft bij 100000 gezinnen van een stad een staaltje van een nieuw product voor kinderen verspreid. Een week na de verspreiding werden 200 gezinnen gedurende de werkuren opgebeld om te vragen hoe dikwijls zij het nieuw product reeds hadden gekocht. De verdeling voor **X**, het aantal gekochte eenheden van het product in 1 week werd samengesteld op basis van de 200 respondenten en weergegeven in tabel 2.1.

x_i	n_i	f_i
0	152	0.76
1	24	0.12
2	16	0.08
3	8	0.04
	200	1.00

Tabel 2.1: Verdeling van het aantal eenheden van het gekochte produkt per gezin op 1 week tijd op basis van een steekproefomvang n=200.

De vraag kan gesteld worden of de schatting van het gemiddeld aantal gekochte produkten per gezin, berekend op basis van de steekproef, $u = \overline{x}$, voldoende goed is voor het werkelijk aantal $\theta = \mu$ in de populatie van 100000 gezinnen. Uit de verkoopcijfers van de betreffende week in die stad, wist men dat het gemiddeld aantal gekochte produkten μ=0.90 was met een variantie van σ^2 van 0.64.

1. Bepaal de gemiddelde kwadratische afwijking van de schatter $\overline{X}$.
   Het gemiddeld aantal verkochte produkten per gezin op basis van de steekproef bedraagt:

$$E(\overline{X}) = \mu_X = 0 \cdot 0.76 + 1 \cdot 0.12 + 2 \cdot 0.08 + 3 \cdot 0.04 = 0.40$$

   Zodat de vertekening gegeven wordt door:

$$\mu_X - \mu = 0.40 - 0.90 = -0.50$$

   De variantie van de schatter $\overline{X}$ wordt berekend uit:

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} = \frac{0.64}{200} = 0.0032$$

   Dit geeft een gemiddelde kwadratische afwijking van schatter $\overline{X}$ gelijk aan:

$$M(\overline{X}) = \sigma^2_{\overline{X}} + (\mu_X - \mu)^2 = 0.0032 + 0.25 = 0.2532$$

2. Idem als (1), maar met een steekproefomvang die tweemaal zo groot is.
De variantie van de schatter $\overline{X}$ wordt dan

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} = \frac{0.64}{200 \times 2} = 0.0016$$

Dit geeft een gemiddelde kwadratische afwijking van schatter $\overline{X}$ gelijk aan:

$$M(\overline{X}) = \sigma^2_{\overline{X}} + (\mu_X - \mu)^2 = 0.0016 + 0.25 = 0.2516$$

Aangezien de vertekening dominant is voor de berekening van de gemiddelde kwadratische afwijking, heeft de steekproefomvang amper enige invloed op de efficientie van de schatter.

3. Een tweede telefonische enquête wordt afgenomen gedurende de avonduren met de verdeling van tabel 2.2 als resultaat.

x_i	n_i	f_i
0	10	0.50
1	4	0.20
2	4	0.20
3	2	0.10
	20	1.00

Tabel 2.2: Verdeling van het aantal eenheden van het gekochte produkt per gezin op 1 week tijd op basis van een steekproefomvang n=20.

$$E(\overline{X}) = \mu_X = 0 \cdot 0.50 + 1 \cdot 0.20 + 2 \cdot 0.20 + 3 \cdot 0.10 = 0.90$$

De vertekening bedraagt

$$\mu_X - \mu = 0.90 - 0.90 = 0$$

De variantie van de schatter $\overline{X}$:

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} = \frac{0.64}{20} = 0.032$$

Dit geeft een gemiddelde kwadratische afwijking van schatter $\overline{X}$ gelijk aan:

$$M(\overline{X}) = \sigma^2_{\overline{X}} + (\mu_X - \mu)^2 = 0.032 + 0 = 0.032$$

Deze tweede enquête werd echter fel bekritizeerd t.g.v. van de veel kleinere steekproefomvang. Nochtans is deze kleine steekproefomvang verdedigbaar aangezien de kwaliteit van de schatter duidelijk beter is (0.032 vs. 0.2532). Bovendien heeft het vergroten van de steekproef tot bv. n=100 enkel invloed op de variantie van de schatter en doet de vertekening slechts marginaal dalen tot 0.0064. Dus: *niet de kwantiteit, maar de kwaliteit van de steekproef telt!*

2.2.1.3 Consistentie

De consistentie van een schatter is een meer abstract begrip, want gedefinieerd als een limiet. Een schatter is *consistent* indien hij, bij een steeds toenemende steekproefomvang, een steeds dichtere band vormt rond de werkelijke populatieparameter, zoals geïllustreerd wordt door figuur 2.3.
Een voorwaarde die de consistentie van de schatter uitmaakt, is dat zijn gemiddelde kwadratische afwijking naar nul evolueert in de limiet ($n \to \infty$). Dit wil zeggen dat zowel zijn vertekening als zijn variantie naar nul evolueren.
Een schatter is een *asymptotisch onvertekende schatter* indien zijn vertekening sowieso naar nul gaat als de steekproefomvang n toeneemt.

Voorbeeld. Is $\overline{X}$ een consistente schatter voor μ?

Oplossing:
Er is geen vertekening:

$$\mu_{\overline{X}} - \mu = 0$$

en de variantie

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$$

gaat naar nul in de limiet voor een toenemende steekproefomvang. Dit betekent dat $\overline{X}$ een consistente schatter is voor μ.

Voorbeeld. Is P een consistente schatter voor π?

Oplossing:
Er is geen vertekening:

$$\mu_P - \mu = 0$$

en de variantie

$$\sigma^2_P = \frac{\pi(\pi - 1)}{n}$$

gaat naar nul in de limiet. Bijgevolg, P is een consistente schatter voor π.

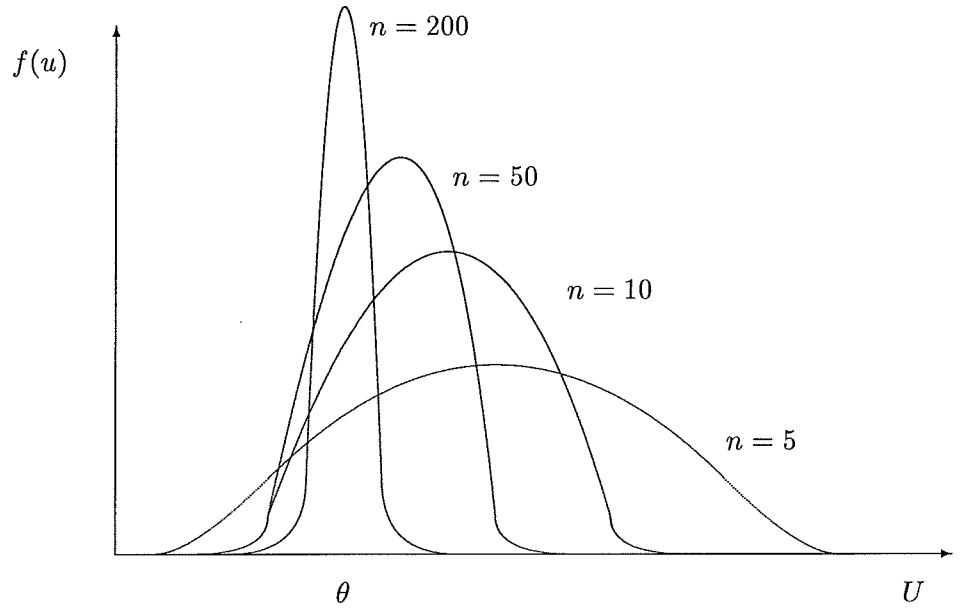
2.3 Intervalschatting

In tegenstelling tot een puntschatting, bestaat een *intervalschatting* uit een interval waarin de te schatten parameter θ vervat zit voor een gegeven betrouwbaarheid. Een *intervalschatter* voor een te schatten populatieparameter θ is een interval (U_o, U_b), dus met de steekproefgrootheden U_o en U_b als onder- en bovengrenzen. De intervalschatting wordt dan gegeven door de overeenkomstige steekproefkengetallen:

$$u_o < \theta < u_b$$

Hoe kleiner het interval (u_o, u_b), hoe beter de schatting, want hoe exacter dat we de waarde van de werkelijke populatieparameter zullen kunnen benaderen. Dit betekent dat de intervalschatter beter is naarmate

$$E(U_b - U_o)$$



Figuur 2.3: Een consistente schatter: U concentreert zich rond θ als n toeneemt.

kleiner is. Dit is een eerste kwaliteitsmaat voor een schatter. Bovendien dient de *vangkans*

$$P(U_o < \theta < U_b)$$

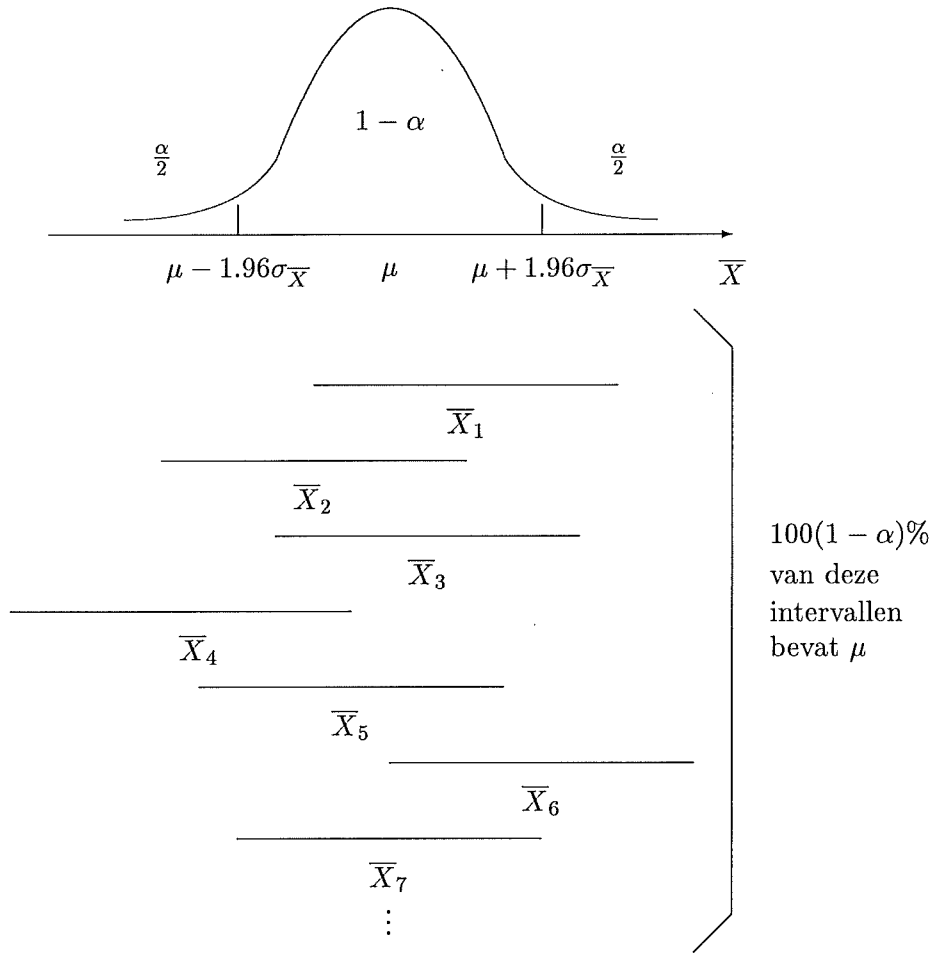
zo hoog mogelijk te zijn, t.t.z. de kans dat de werkelijke populatieparameter θ begrepen is in het interval dient voldoende hoog te zijn. De maximale vangkans is 100%, maar dit gaat gepaard met een interval van maximale breedte gaande van $-\infty$ tot ∞. Hiermee zijn we zeker dat de werkelijke populatieparameter θ in het interval begrepen is, maar de overeenkomende intervalschatter voldoet helemaal niet aan de eerste kwaliteitsmaat (een beperkte intervalbreedte). Dit toont aan dat er voor eenzelfde steekproefomvang een wisselwerking is tussen de intervalbreedte en de vangkans. Zoals we zullen zien, is de enige manier om een kleiner interval tezamen met een hogere vangkans te verkrijgen, het verhogen van de steekproefomvang n.
Aangezien aan elk interval een bepaalde betrouwbaarheid gekoppeld is, spreken we in dit verband ook van *betrouwbaarheidsinterval.* Voor de vangkans wordt er namelijk een drempelwaarde gekozen:

$$P(U_o < \theta < U_b) \geq 1 - \alpha.$$

Dit betekent dat de kans dat de waarde van de werkelijke populatieparameter θ niet in het interval (U_o, U_b) begrepen is, hoogstens gelijk is aan α. Bijgevolg wordt het interval (U_o, U_b) een $100(1-\alpha)\%$ of gewoon een $(1-\alpha)$ betrouwbaarheidsinterval genoemd. Een $100(1-\alpha)\%$ betrouwbaarheidsinterval kan dus beschouwd worden als het resultaat van de schatter (U_o, U_b) met een vangkans van minstens $(1-\alpha)$. Hierbij zijn u_o en u_b de betrouwbaarheidsgrenzen en kan het interval (u_o, u_b) de schatting genoemd worden.
In eenvoudige termen zou men kunnen stellen dat, indien men 100 steekproeven neemt van een populatie en op basis van elke steekproef het $100(1-\alpha)\%$ betrouwbaarheidsinterval berekent, dat dan de werkelijke te schatten populatiewaarde θ in $100(1-\alpha)$ van die 100 betrouwbaarheidsintervallen begrepen zal zijn. Dit wordt geïllustreerd door figuur 2.4 voor een $100(1-\alpha)\%$ betrouwbaarheidsinterval voor het populatiegemiddelde μ.

De keuze van de *drempelwaarde* $(1-\alpha)$ wordt in sterke mate beïnvloed door het gevolg van een onjuiste conclusie. Gebruikelijk wordt een drempel- of betrouwbaarheidswaarde van 95% gebruikt. Nochtans kan men in bepaalde situaties een hogere vangkans (bv. 99%) eisen, zoals in bepaalde toepassingen met betrekking tot de neveneffecten van geneesmiddelen, of met betrekking tot de hoge kosten resulterend uit de aanvaarding van een slechte partij goederen. De waarde van $(1-\alpha)$ kan dus geïnterpreteerd worden als een maatstaf voor het vertrouwen dat men heeft in de statistische conclusie.
Het dient opgemerkt te worden dat een 99% betrouwbaarheidsinterval tegelijk ook een 95%, 90%, ... betrouwbaarheidsinterval is. Immers, hoe hoger de betrouwbaarheid, hoe breder het interval.

De intervallen die we beschouwen zijn steeds *tweezijdig* van het type (u_o, u_b). Niettemin bestaan er ook linkseenzijdige (u_o, ∞) of rechtseenzijdige (∞, u_b) betrouwbaarheidsintervallen. Linkseenzijdige intervallen kunnen bijvoorbeeld nuttig zijn bij het bepalen van het gewicht van een bepaald product, waarbij enkel aan een minimumwaarde moet voldaan zijn. Rechtseenzijdige intervallen zijn nuttig indien enkel de bovengrenzen zinvol zijn, zoals bijvoorbeeld de fractie van patienten die de bijwerkingen van een medicijn ondervinden.



Figuur 2.4: Illustratie van het begrip betrouwbaarheidsinterval voor de parameter $\theta = \mu$

In wat volgt, worden de betrouwbaarheidsintervallen bepaald voor het populatiegemiddelde, het verschil tussen twee populatiegemiddelden, de populatieproportie, het verschil tussen tussen twee populatieproporties, de populatievariantie en de verhouding tussen twee populatievarianties.

2.3.1 Betrouwbaarheidsintervallen voor de schatting van het populatiegemiddelde μ

2.3.1.1 Normale verdeling met bekende populatievariantie

De bedoeling is om een $(1-\alpha)$ betrouwbaarheidsinterval op te stellen voor het onbekende populatiegemiddelde μ van een normaal verdeelde populatie $\mathbf{X}$ waarvan de variantie σ^2 bekend is, m.a.w.

$$\mathbf{X} \sim N(\mu, \sigma^2)$$

Het $(1-\alpha)$ betrouwbaarheidsinterval voor μ wordt bepaald op basis van een aselecte steekproef met omvang n: $X_1, X_2, ..., X_n$. Hiervan wordt het steekproefgemiddelde $\overline{X}$ bepaald. Aangezien het steekproefgemiddelde $\overline{X}$ normaal verdeeld is met gemiddelde $\mu_{\overline{X}} = \mu$ en variantie $\sigma^2_{\overline{X}} = \sigma^2/n$ kunnen we de volgende afleidingen maken. Zo weten we dat voor een normaal verdeling 95% van alle waarden van de verdeling binnen 1.96 standaardfouten van het populatiegemiddelde μ begrepen zijn (zie figuur 2.5). Dit betekent dat 95% van alle waarden van $\overline{X}$ berekend uit alle steekproeven van omvang n begrepen zijn in het interval

$$(\mu - 1.96\sigma_{\overline{X}}; \mu + 1.96\sigma_{\overline{X}})$$

of anders gezegd

$$P(\mu - 1.96\sigma_{\overline{X}} < \overline{X} < \mu + 1.96\sigma_{\overline{X}}) = 0.95$$

Aangezien μ onbekend is, en dus de te schatten parameter is, wordt deze uitdrukking getransformeerd in:

$$P(\overline{X} - 1.96\sigma_{\overline{X}} < \mu < \overline{X} + 1.96\sigma_{\overline{X}}) = 0.95$$

Het 95% betrouwbaarheidsinterval

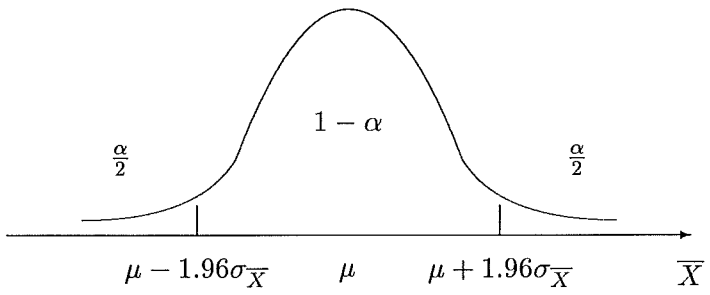
$$(\overline{X} - 1.96\sigma_{\overline{X}}; \overline{X} + 1.96\sigma_{\overline{X}}) \tag{2.1}$$

omvat het werkelijk populatiegemiddelde μ in 95% van de gevallen. Dit wordt eveneens geïllustreerd door figuur 2.4.
Voor alle duidelijkheid dient er toch opgemerkt te worden dat het onbekende populatiegemiddelde μ een constante is en dat de intervalschatter een variabele is, aangezien zijn centrum $\overline{X}$ een toevalsvariabele is.
De algemene vorm van een tweezijdig betrouwbaarheidsinterval is

$$\text{schatter} \pm \text{betrouwbaarheidsfactor} \times \text{standaardfout v.d. schatter}$$

Een tweezijdig betrouwbaarheidsinterval is steeds symmetrisch rond de schatter. De *betrouwbaarheidsfactor* wordt bepaald in functie van het gewenste betrouwbaarheidsniveau $1-\alpha$ (of het significantieniveau α). In het geval de steekproevenverdeling van de schatter normaal is, zoals voor het steekproefgemiddelde $\overline{X}$, dan is de betrouwbaarheidsfactor de standaardnormale Z-waarde, $z_{1-\alpha/2}$:

$$\overline{X} \pm z_{1-\alpha/2}\sigma_{\overline{X}}$$



Figuur 2.5: Steekproevenverdeling van $\overline{X}$ met aanduiding van het 95% betrouwbaarheidsinterval voor μ

Betrouwbaarheidsniveau $1-\alpha$	$z_{1-\alpha/2}$
0.90	1.645
0.95	1.96
0.99	2.58

Tabel 2.3: Betrouwbaarheidsniveau en overeenkomende betrouwbaarheidsfactor voor een tweezijdig betrouwbaarheidsinterval voor μ.

Voor een éénzijdig interval wordt de waarde $z_{1-\alpha}$ gebruikt.
Tabel 2.3 bevat de meest gebruikte waarden voor de betrouwbaarheidsfactor Z in functie van de gewenste betrouwbaarheid. Voor andere betrouwbaarheidsniveau's dient een tabel met standaardnormale waarden te worden geraadpleegd.
Zoals verwacht, verhoogt de waarde van de betrouwbaarheidsfactor en dus van het betrouwbaarheidsinterval naarmate de gewenste betrouwbaarheid hoger ligt.

Indien $n > 0.05N$ is een eindige populatiecorrectie vereist.

Voorbeeld. Bepaal een 95% en een 99% betrouwbaarheidsinterval voor de werkelijke gemiddelde treksterkte van staaldraad. Volgens de producent ervan is de treksterkte normaal verdeeld met een s.d. van 200 kg. Een steekproef van 16 draadexemplaren leverde een gemiddelde treksterkte van 6200 kg. op.

Oplossing:
Het 95% betrouwbaarheidsinterval:

$$\overline{X} \pm Z\sigma_{\overline{X}}$$

$$\overline{x} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$6200 \pm 1.96\frac{200}{\sqrt{16}}$$

$$(6102; 6298)$$

Het 99% betrouwbaarheidsinterval:

$$6200 \pm 2.58\frac{200}{\sqrt{16}}$$

$$(6071; 6329)$$

De *precisie* of *foutmarge* E van een schatter wordt gedefinieerd als het product van de betrouwbaarheidsfactor met zijn standaardfout. In het geval van een betrouwbaarheidsinterval voor het populatiegemiddelde μ wordt de precisie van schatter $\overline{X}$ gegeven door:

$$E = Z\frac{\sigma}{\sqrt{n}}$$

Des te hoger het betrouwbaarheidsniveau en dus ook Z, des te lager de precisie. Een hoge precisie wordt verkregen voor een lage waarde van het bovenstaand interval. Met andere woorden, een schatter met een hoge precisie varieert weinig rond zijn doel (hier het populatiegemiddelde) en dus is het interval niet breed. Een hogere precisie kan ook verkregen worden door de steekproefomvang te verhogen.

Voorbeeld. Men wenst een schatting te maken voor de gemiddelde leeftijd van alle Amerikaanse vliegtuigen. Hiervoor neemt men een representatieve steekproef van 40 vliegtuigen. De gemiddelde leeftijd van de steekproef bedraagt 13.41 jaar. Uit onderzoek weet men dat de standaarddeviatie op de leeftijd van een commerciele vloot 8.28 jaar is. Bepaal de precisie van de schatter en een betrouwbaarheidsinterval voor μ met een betrouwbaarheidsniveau van 95%.

Oplossingen:
De precisie van schatter $\overline{X}$

$$z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} = 1.96\frac{8.28}{\sqrt{40}} = 2.57$$

Het 95% betrouwbaarheidsinterval:

$$\overline{x} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$13.41 \pm 2.57$$

$$(10.84; 15.98)$$

2.3.1.2 Niet-normale verdeling met bekende populatievariantie

Voor een voldoende grote steekproefomvang kan, volgens de centrale limietstelling, de steekproevenverdeling van $\overline{X}$ als benaderend normaal worden beschouwd, ook indien de oorspronkelijke populatie niet normaal is. Het gevolg hiervan is echter dat een $100(1-\alpha)\%$ betrouwbaarheidsinterval slechts als *benaderend* $100(1-\alpha)\%$ mag beschouwd worden.

Indien $n > 0.05N$ is een eindige populatiecorrectie vereist.

Voorbeeld. Een uitgever van een regionaal tijdschrift voor de derde leeftijd wenst de gemiddelde leeftijd van zijn lezers te bepalen. Hiervoor wordt een steekproef van 150 van de in het totaal 1000 lezers genomen. De gemiddelde leeftijd hiervan bedraagt 68 jaar. Uit vroeger onderzoek is gebleken dat de standaarddeviatie 10 jaar was. Uit datzelfde vroeger onderzoek is eveneens gebleken dat de populatie niet normaal verdeeld was. Bepaal het 99% betrouwbaarheidsinterval voor μ.

Oplossing:
$\frac{n}{N} = \frac{150}{1000} = 0.15 > 0.05 \rightarrow$ eindige populatie correctie
Het 99% betrouwbaarheidsinterval:

$$\overline{X} \pm Z\sigma_{\overline{X}}\sqrt{\frac{N-n}{N-1}}$$

$$\overline{x} \pm z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

$$68 \pm 2.58 \frac{10}{\sqrt{150}} \sqrt{\frac{1000-150}{1000-1}}$$
$$(66.05; 69.95)$$

Voor een populatiegrootte van $N = 10000$ wordt het volgend 99% betrouwbaarheidsinterval verkregen:

$$\overline{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$
$$68 \pm 2.58 \frac{10}{\sqrt{150}}$$
$$(65.88; 70.12)$$

2.3.1.3 Normale verdeling met onbekende populatievariantie

In de voorgaande paragraaf werd verondersteld dat de populatievariantie bekend was. Deze situatie doet zich in de praktijk slechts uiterst zelden voor. Het is veel realistischer om te veronderstellen dat de populatievariantie onbekend is. De populatievariantie of -standaarddeviatie is echter onontbeerlijk voor de berekening van een betrouwbaarheidsinterval voor de onbekende μ. In voorkomend geval wordt de steekproefstandaarddeviatie S gebruikt als schatter voor de populatiestandaarddeviatie σ. Niet enkel μ wordt geschat door $\overline{X}$, maar ook σ wordt geschat door S, hetgeen zorgt voor een bijkomende toevalsinvloed, waardoor de Z-waarde in

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

niet meer voldoende accuraat kan bepaald worden. In plaats van de Z-waarde, wordt een t-waarde gebruikt:

$$t = \frac{\overline{X} - \mu}{S_{\overline{X}}} = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

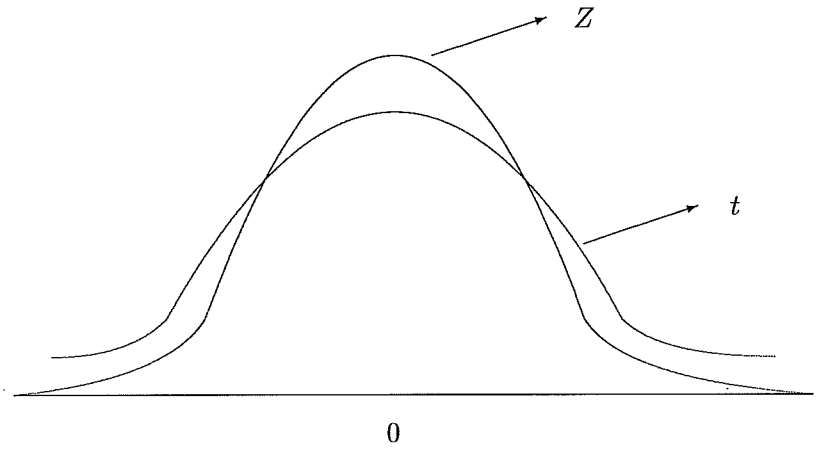
Door voor een groot aantal steekproeven van omvang n, telkens $\overline{X}$ en $S_{\overline{X}}$ en dan ook t te berekenen, kan met de verkregen waarde van t een verdeling opgesteld worden, de zogenaamde *Student's t verdeling.* Deze verdeling heeft de volgende kenmerken (zie figuur 2.6):

- gemiddelde gelijk aan 0 (zoals voor standaardnormale);
- symmetrisch rond het gemiddelde (zoals voor standaardnormale);
- variantie $\sigma > 1$, maar nadert 1 als n toeneemt (standaardnormale σ=1);
- $t \in]-\infty, \infty[$
- bij elke waarde voor de *vrijheidsgraden* (=$n-1$) behoort een andere t verdeling;
- de t verdeling is vlakker (grotere variabiliteit) dan de standaardnormaal verdeling, maar benadert deze laatste beter naarmate n toeneemt.

Een $(1-\alpha)$ betrouwbaarheidsinterval voor μ met een onbekende variantie σ wordt gegeven door

$$\overline{X} \pm t_{1-\alpha/2} \frac{S}{\sqrt{n}}.$$

De $t_{1-\alpha/2}$ kunnen in de tabel van de t-verdeling teruggevonden worden voor een gegeven betrouwbaarheidsniveau $(1-\alpha)$ en aantal vrijheidsgraden $n-1$.



Figuur 2.6: De standaardnormale Z verdeling en de Student's t verdeling.

Definitie 4 *Het aantal* vrijheidsgraden *kan omschreven worden als de hoeveelheid informatie vereist om de steekproefstandaarddeviatie S te berekenen.*

Voor de berekening van S is er een verzameling van $n-1$ stukken informatie vereist:

$$S = \sqrt{\frac{\sum(X_t - \overline{X})^2}{n-1}}$$

of anders gezegd, om S te berekenen kan men vrijuit waarden toekennen aan $n-1$ waarnemingen van X, enkel de $n$de waarneming kan dan niet meer vrij gekozen worden.
Voor een steekproefomvang $n > 100$ zijn de waarden van t en z benaderend gelijk voor eenzelfde betrouwbaarheidsniveau.

Voorbeeld. De volgende 10 scores vertegenwoordigen de leeftijden van een steekproef van Amerikaanse vliegtuigen:

$$3.2, 22.6, 23.1, 16.9, 0.4, 6.6, 12.5, 22.8, 26.3, 8.1$$

Indien de leeftijdsverdeling benaderend normaal is, bereken dan een 95% betrouwbaarheidsinterval voor de gemiddelde leeftijd van de gehele vloot.

Oplossing:
Het gemiddelde en de standaarddeviatie van de steekproef zijn:

$$\overline{x} = 14.25; s = 9.35$$

Het 95% betrouwbaarheidsinterval:

$$\overline{X} \pm t_{1-\alpha/2}\frac{S}{\sqrt{n}}$$

$$14.25 \pm t_{0.975}\frac{9.35}{\sqrt{10}}$$

$$14.25 \pm 2.262\frac{9.35}{\sqrt{10}}$$

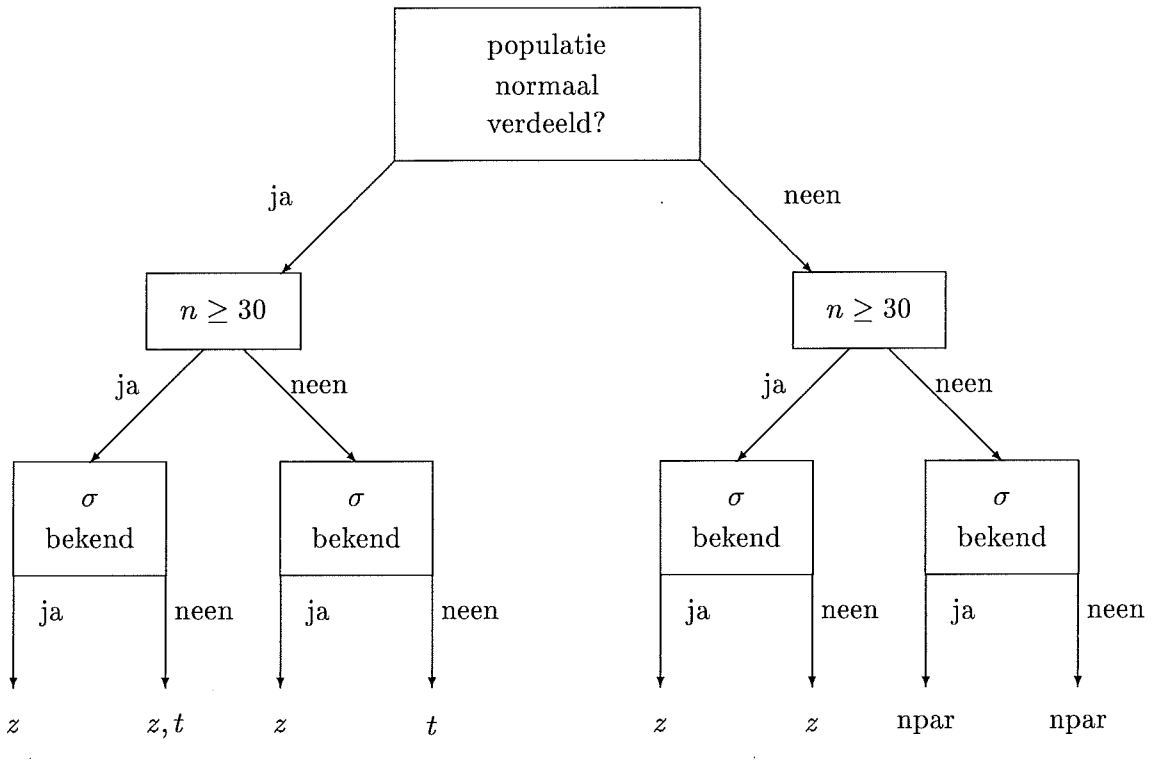
$$(7.56; 20.94)$$

2.3.1.4 Niet-normaal verdeelde populaties en onbekende populatievariantie

Indien de oorspronkelijke populatie niet-normaal verdeeld is, en indien de steekproefomvang voldoende groot is ($n > 30$), dan is volgens de centrale limietstelling de steekproevenverdeling van $\overline{X}$ normaal.
Een $(1-\alpha)$ betrouwbaarheidsinterval voor μ wordt dan gegeven door:

$$\overline{X} \pm z_{1-\alpha/2}\frac{S}{\sqrt{n}}.$$

Dit impliceert dat de steekproefomvang eerder dan het al of niet bekend zijn van de variantie als criterium wordt gebruikt om te kiezen tussen een t of een z statistiek (zie figuur 2.7).



Figuur 2.7: Keuze tussen t of z. Letterwoord *npar* staat voor niet-parametrische methode.

Indien $n > 0.05N$ is een eindige populatiecorrectie vereist.

Voorbeeld. Een distributiebedrijf wenst een nieuw filiaal te openen is daarom geïnteresseerd in de gemiddelde omvang van vergelijkbare handelszaken. Een steekproef van 50 vergelijkbare handelszaken levert een gemiddelde omvang van 10000 m^2 en een standaarddeviatie van 4800 m^2 op. Bepaal een 95% betrouwbaarheidsinterval voor μ.

Oplossing:
n=50 $\rightarrow$ grote steekproef $\rightarrow z$.

$$\overline{X} \pm z_{1-\alpha/2}\frac{S}{\sqrt{n}}$$

$$10000 \pm 1.96\frac{4800}{\sqrt{50}}$$

$$(8669; 11331)$$

2.3.2 Betrouwbaarheidsintervallen voor de schatting van het verschil tussen 2 populatiegemiddelden, niet gepaarde steekproeven

2.3.2.1 Bekende populatievariantie

We weten dat indien de twee populaties normaal verdeeld zijn, de steekproevenverdeling van $\overline{X}_1 - \overline{X}_2$ berekend vanuit twee onafhankelijke steekproeven, eveneens normaal is.
Een $(1-\alpha)$ betrouwbaarheidsinterval voor $\mu_1 - \mu_2$ wordt dan gegeven door:

$$(\overline{X}_1 - \overline{X}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Voorbeeld. Een bedrijf heeft 2 productiehuizen waar dezelfde vezel wordt vervaardigd. Om te bepalen of de 2 sites een uniforme kwaliteit halen, wordt een steekproef van 25 specimen van bedrijf 1 en 16 van bedrijf 2 genomen. Het resultaat hiervan is een gemiddelde treksterkte van 22 kg. voor de vezel van bedrijf 1 en 20 kg. voor die van bedrijf 2. De variantie op de treksterkte is vooraf bekend voor de beide en bedraagt 10 kg^2. Beide populaties zijn normaal verdeeld. Bepaal het 95% betrouwbaarheidsinterval voor het verschil van de gemiddelde treksterkten:

$$(\overline{X}_1 - \overline{X}_2) \pm z_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(22 - 20) \pm 1.96\sqrt{\frac{10}{25} + \frac{10}{16}}$$

$$(0.0; 4.0)$$

Aangezien 0 in dit interval begrepen is, kan men stellen dat de productiekwaliteit in de 2 sites niet significant verschillend is.

2.3.2.2 Onbekende populatievarianties

Het geval van onbekende populatievarianties leunt veel nauwer aan met wat we in de praktijk tegenkomen. Hierbij kunnen we echter twee situaties onderscheiden: gelijke of ongelijke populatievarianties.

Gelijke populatievarianties. Hierbij wensen we een betrouwbaarheidsinterval voor het verschil tussen 2 populatiegemiddelden te bepalen met onbekende maar weliswaar gelijke populatievarianties. De berekende varianties uit de steekproef worden gecombineerd tot een *gepoolde variantie* s_p^2 die de schatting is voor de populatievariantie σ^2. Deze gepoolde variantie wordt aldus verkregen:

$$s_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Het $(1 - \alpha)$ betrouwbaarheidsinterval wordt gegeven door:

$$(\overline{X}_1 - \overline{X}_2) \pm t_{1-\alpha/2}\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

Aangezien de berekening van de gepoolde variantie s_p^2, $n_1 + n_2 - 2$ vrijheidsgraden vereist, is de $t_{1-\alpha/2}$-waarde diegene die overeenkomt met $n_1 + n_2 - 2$ vrijheidsgraden.

Voorbeeld. Eenzelfde test werd afgenomen bij twee grote groepen managers. Van de eerste groep werd een steekproef van 4 genomen met als resultaten op 100: 64, 66, 89 en 77. Van de tweede groep werd een steekproef van 3 genomen met als resultaten op 100: 56, 71 en 53. Bepaal het 95% betrouwbaarheidsinterval voor het verschil tussen de twee groepgemiddelden $\mu_1 - \mu_2$.

Oplossing:
Berekening van het steekproefgemiddelde en de variantie in elke groep:

$$\overline{x}_1 = 74 \qquad s_1^2 = 398/3 = 132.7$$
$$\overline{x}_2 = 60 \qquad s_2^2 = 186/2 = 93$$

Er wordt verondersteld dat in de beide groepen (populaties) de varianties gelijk zijn. Bijgevolg wordt een gepoolde variantie berekend:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{398 + 186}{3 + 2} = 117$$

Het 95% betrouwbaarheidsinterval is dan:

$$(\overline{X}_1 - \overline{X}_2) \pm t_{1-\alpha/2}\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

$$(74 - 60) \pm 2.57\sqrt{\frac{117}{4} + \frac{117}{3}}$$

$$(-7; 35)$$

Niet-gelijke populatievarianties. In vele situaties gaat de veronderstelling van gelijke populatievarianties niet meer op. In voorkomend geval wordt de t-statistiek met $n_1 + n_2 - 2$ vrijheidsgraden onbruikbaar, maar dient een nieuwe $t'_{1-\alpha/2}$ te worden berekend:

$$t'_{1-\alpha/2} = \frac{\frac{S_1^2}{n_1} t_{(1-\alpha/2, n_1-1)} + \frac{S_2^2}{n_2} t_{(1-\alpha/2, n_2-1)}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Zodat een $(1-\alpha)$ betrouwbaarheidsinterval wordt gegeven door:

$$(\overline{X}_1 - \overline{X}_2) \pm t'_{1-\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

Voorbeeld. Veronderstel in het vorige voorbeeld met de twee groepen managers dat de populatievarianties ongelijk zijn.

Oplossing:
Berekening van $t'_{1-\alpha/2}$:

$$t'_{1-\alpha/2} = \frac{\frac{s_1^2}{n_1} t_{(1-\alpha/2, n_1-1)} + \frac{s_2^2}{n_2} t_{(1-\alpha/2, n_2-1)}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \frac{\frac{132.7}{4} 3.182 + \frac{93}{3} 4.303}{\frac{132.7}{4} + \frac{93}{3}} = 3.72$$

Het 95% betrouwbaarheidsinterval is dan:

$$(\overline{X}_1 - \overline{X}_2) \pm t'_{1-\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$$(74 - 60) \pm 3.72 \sqrt{\frac{132.7}{4} + \frac{93}{3}}$$

$$(-15.8; 43.8)$$

In dit interval is 0 begrepen, waardoor men mag stellen dat er geen significant verschil is tussen de testresultaten van de twee groepen.

2.3.2.3 Niet-normaal verdeelde populaties, onbekende populatievarianties

Indien zowel n_1 als n_2 voldoende groot zijn, is de centrale limietstelling van toepassing. Bijgevolg kan de Z-statistiek als betrouwbaarheidsfactor worden gebruikt. Het $(1-\alpha)$ betrouwbaarheidsinterval is dan:

$$(\overline{X}_1 - \overline{X}_2) \pm z_{1-\alpha/2} \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

2.3.3 Betrouwbaarheidsintervallen voor de schatting van het verschil tussen 2 populatiegemiddelden, gepaarde steekproeven

Gepaarde observaties zijn onlosmakelijk verbonden met alle vormen van voor- en na experimenten. De twee even grote steekproeven van observaties waarover we in voorkomend geval

beschikken worden in een eerste stap teruggebracht tot één steekproef met de gepaarde verschillen. Dus voor elk individu (element) i uit de beide steekproeven wordt het verschil berekend:

$$D_i = X_{1i} - X_{2i}$$

Aldus wordt er één steekproef met de verschillen D verkregen. Hiervan kan dan het gemiddelde en de standaarddeviatie berekend worden:

$$\overline{D} = \frac{\sum_i D_i}{n}$$

$$S_D = \sqrt{\frac{\sum(D_i - \overline{D})^2}{n-1}}$$

Het steekproefgemiddelde van de verschillen $\overline{D}$ is een schatter voor het populatiegemiddelde van de verschillen μ_D. Hierbij is $\mu_D = \mu_1 - \mu_2$.

Een $(1-\alpha)$ betrouwbaarheidsinterval voor μ_D wordt gegeven door

$$\overline{D} \pm t_{1-\alpha/2}\frac{S_D}{\sqrt{n}}.$$

Het aantal vrijheidsgraden voor $t_{1-\alpha/2}$ bedraagt $n-1$.
Indien $n > 0.05N$ is een eindige populatiecorrectie vereist.

Voorbeeld. Men is geïnteresseerd in het effect van een trainingscursus op het resultaat van een bepaalde test. Hiertoe wordt een steekproef van 4 mensen uit een grote groep getrokken. De resultaten voor de cursus waren: 57, 57, 73 en 65; de resultaten erna in respectieve volgorde waren: 64, 66, 89 en 77. Is er op een betrouwbaarheidsniveau van 95% een significant effect van de cursus?

Oplossing:
De steekproef van de verschillen: -7, -9, -16 en -12. Dit geeft:

$$\overline{D} = \frac{\sum_i D_i}{n} = -11$$

$$S_D = \sqrt{\frac{\sum(D_i - \overline{D})^2}{n-1}} = 3.91$$

Het 95% betrouwbaarheidsinterval wordt dan:

$$\overline{D} \pm t_{1-\alpha/2}\frac{S_D}{\sqrt{n}}$$

$$-11 \pm 3.182\frac{3.91}{\sqrt{4}}$$

$$(-17.22; -4.78)$$

De waarde 0 is niet begrepen in dit interval. Dit betekent dat voor een betrouwbaarheid van 95% kan gesteld worden dat de trainingscursus wel degelijk invloed heeft gehad.

2.3.4 Betrouwbaarheidsintervallen voor de schatting van de populatieproportie

Voor de schatting van de populatieproportie wordt op een analoge manier te werk gegaan als voor de schatting van het populatiegemiddelde. Voor een steekproef van n wordt de steekproefproportie P bepaald als een puntschatter voor de werkelijke populatieproportie π. Indien $nP > 5$ en $n(1-P) > 5$ dan is de steekproevenverdeling van P benaderend normaal met gemiddelde π en standaarddeviatie $\sqrt{\pi(1-\pi)/n}$, die op haar beurt kan geschat worden door de standaarddeviatie $\sqrt{P(1-P)/n}$. Bijgevolg wordt een $(1-\alpha)$ betrouwbaarheidsinterval gegeven door:

$$P \pm z_{1-\alpha/2}\sqrt{\frac{P(1-P)}{n}}$$

Voor $n > 0.05N$ is een eindige populatiecorrectie vereist.

Voorbeeld. Het fabrikageproces van geheugenchips. De kans dat een chip afgekeurd wordt door de kwaliteitscontrole is 0.4. Indien 10000 chips worden onderzocht, bepaal dan het 95% betrouwbaarheidsinterval voor het aantal goedgekeurde chips en het 95% betrouwbaarheidsinterval voor de fractie goedgekeurde chips.

Oplossingen:
Het 95% betrouwbaarheidsinterval voor het aantal goedgekeurde chips:
Aantal goedgekeurde chips $\mathbf{S} \sim B(10000; 0.6)$

$$\overline{S} = n\pi = 6000 \qquad \sigma_S = \sqrt{n(1-\pi)\pi} = 49$$

Normale benadering van de binomiaal verdeling omdat n voldoende groot is.

$$\overline{S} \pm z_{1-\alpha/2}\sigma_S$$

$$6000 \pm 1.96 \cdot 49$$

$$5904; 6096$$

Het 95% betrouwbaarheidsinterval voor de fractie goedgekeurde chips:

$$P \pm 1.96\sqrt{\frac{P(1-P)}{n}}$$

$$0.6 \pm 1.96\sqrt{\frac{0.6 \cdot 0.4}{10000}}$$

$$0.5904; 0.6096$$

2.3.5 Betrouwbaarheidsintervallen voor de schatting van het verschil tussen 2 populatieproporties

In sommige gevallen kan het interessant zijn om een idee te hebben over de omvang van het verschil tussen 2 populatieproporties, zoals bijvoorbeeld bij het vergelijken van man en vrouw, van twee bedrijven, ...

Het verschil tussen de steekproeffracties $P_1 - P_2$ wordt gebruikt als schatter voor het verschil tussen de populatiefracties $\pi_1 - \pi_2$. Indien voldaan is aan $n_1P_1, n_2P_2, n_1(1-P_1), n_2(1-P_2) > 5$, dan is de steekproevenverdeling van $P_1 - P_2$ benaderend normaal. Bijgevolg wordt een $(1-\alpha)$ betrouwbaarheidsinterval voor $\mu_1 - \mu_2$ gegeven door:

$$(P_1 - P_2) \pm z_{1-\alpha/2}\sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$$

Voorbeeld. Een reclamebureau doet een onderzoek naar de effectiviteit van een advertentie in 2 tijdschriften. Voor tijdschrift 1 blijkt dat op een steekproef van 500 lezers er 200 de advertentie opgemerkt hebben. Voor de tijdschrift 2 zijn dat er 300 op 500. Bepaal een 98% betrouwbaarheidsinterval voor het verschil tussen de 2 populatieporporties.

Oplossing:

$$p_1 = 200/500 = 0.4 \qquad p_2 = 300/500 = 0.6$$

$$(P_1 - P_2) \pm z_{1-\alpha/2}\sqrt{\frac{P_1(1-P_1)}{n} + \frac{P_2(1-P_2)}{n}}$$

$$(0.4 - 0.6) \pm 2.33\sqrt{\frac{0.4(0.6)}{500} + \frac{0.6(0.4)}{500}}$$

$$-0.2 \pm 2.33(0.03)$$

$$(-0.27; -0.13)$$

Dit interval bevat 0 niet, wat betekent dat er een verschil in effectiviteit van de advertentie is, afhankelijk van het tijdschrift.

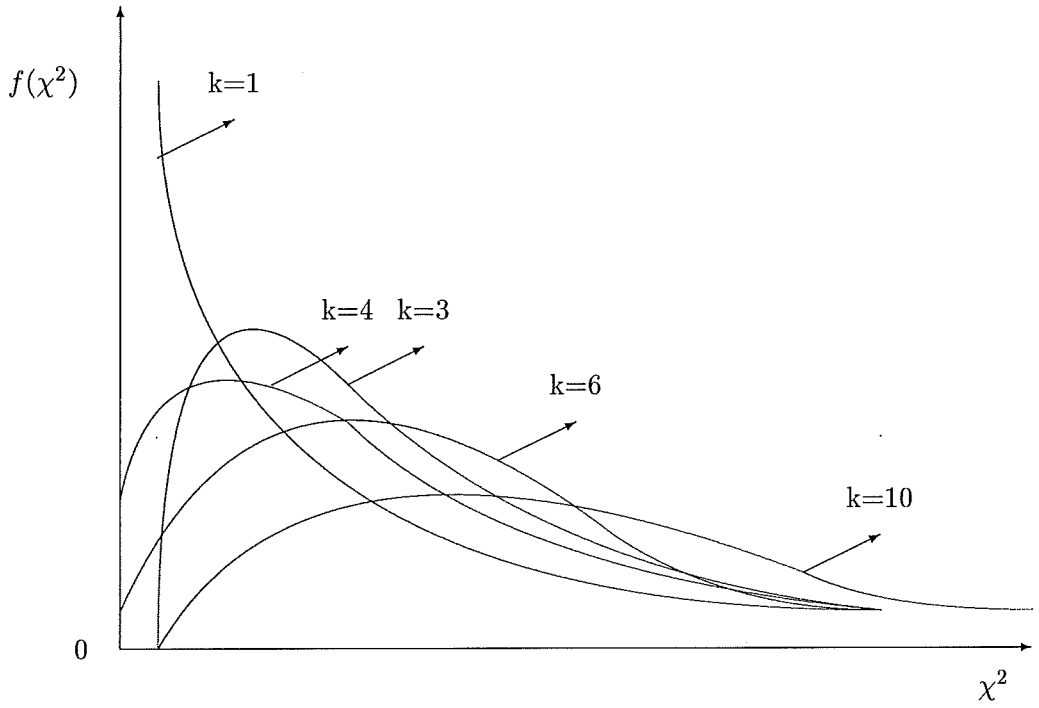
2.3.6 Betrouwbaarheidsintervallen voor de schatting van de populatievariantie.

De kennis van de populatievariantie kan nuttig zijn voor sommige toepassingen. Voorbeelden hiervan zijn een producent die de variantie (variabiliteit) in de kwaliteit van zijn producten wenst te determineren om aldus zijn garantieperiode te bepalen, een producent van farmaceutische producten die de variabiliteit in doeltreffendheid van zijn geneesmiddel op patienten wenst te bepalen,...
Aangezien deze variabiliteit vooraf niet gekend is, moet deze geschat worden vanuit een steekproef. De steekproefvariantie S^2 is hierbij een goede puntschatter voor de populatievariantie σ^2, zoals bepaald in het begin van dit hoofdstuk.
Teneinde een schattingsinterval voor de populatievariantie σ^2 te bepalen, moet de steekproevenverdeling van

$$\frac{(n-1)S^2}{\sigma^2}$$

(en niet van S^2) gebruikt worden. Het is mogelijk om de steekproevenverdeling van $(n-1)S^2/\sigma^2$ zelf op te zetten middels de volgende procedure:

1. Neem een groot aantal steekproeven van omvang n uit een normaal verdeelde populatie met bekende variantie σ^2.



Figuur 2.8: χ^2-verdelingen voor verschillende vrijheidsgraden.

2. Bereken voor elke steekproef s^2 en vervolgens $(n-1)s^2/\sigma^2$.

3. Ontwikkel de frekwentieverdeling van $(n-1)s^2/\sigma^2$.

De resulterende frekwentieverdeling is een goede benadering van de χ^2-*verdeling* met $n-1$ vrijheidsgraden.
De χ^2-verdeling is een familie van asymmetrische verdelingen, met een verdeling voor elke waarde van de vrijheidsgraden $n-1$, zoals geïllustreerd wordt door figuur 2.8.
Voor het bepalen van het $(1-\alpha)$ betrouwbaarheidsinterval, wordt uitgegaan van de volgende probabiliteit:

$$P\left[\chi^2_{\alpha/2} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{1-\alpha/2}\right] = 1-\alpha$$

Dit interval kan getransformeerd worden in een interval rond σ^2:

$$P\left[\frac{\chi^2_{\alpha/2}}{(n-1)S^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{1-\alpha/2}}{(n-1)S^2}\right] = 1-\alpha$$

$$P\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}}\right] = 1-\alpha$$

Het $(1-\alpha)$ betrouwbaarheidsinterval voor σ^2 wordt bijgevolg gegeven door:

$$\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}}$$

of voor de populatiestandaarddeviatie:

$$\sqrt{\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}}} < \sigma < \sqrt{\frac{(n-1)S^2}{\chi^2_{\alpha/2}}}$$

De normaliteit van de oorspronkelijke populatie is hierbij een bindende voorwaarde.

Voorbeeld. In het kader van de kwaliteitscontrole wenst een bedrijf de variantie van het versnijdingsproces van golfkarton te bepalen. Op basis van een steekproef van 51 versneden stukken werd een gemiddelde lengte van 5.55 cm. en een standaarddeviatie van 0.145 cm gemeten. Bepaal het 95% betrouwbaarheidsinterval voor de populatievariantie. De verdeling van de lengte is normaal.

Oplossing:
$n-1=50; \quad \chi^2_{1-\alpha/2}=71.42 \quad \chi^2_{\alpha/2}=32.36$

$$\frac{(n-1)S^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{\alpha/2}}$$

$$\frac{(50)(0.145)^2}{71.42} < \sigma^2 < \frac{(50)(0.145)^2}{32.36}$$

$$(0.0147; 0.0325)$$

2.3.7 Betrouwbaarheidsintervallen voor de schatting van de ratio van populatievarianties.

Het schatten van de ratio van populatievarianties van 2 *normaal verdeelde populaties* is vooral nuttig bij het vergelijken van twee populatievarianties. Zo kan het bijvoorbeeld interessant zijn om de kwaliteitsvariabiliteit te vergelijken van eenzelfde product afkomstig van twee verschillende productieplaatsen. Het vergelijken van de variabiliteit kan door de ratio σ_1^2/σ_2^2 te beschouwen. Aangezien de populatievarianties niet gekend zijn, worden deze geschat middels de steekproefvarianties.
Om een betrouwbaarheidsinterval voor de ratio σ_1^2/σ_2^2 te bepalen, wordt uitgegaan van de steekproevenverdeling voor

$$\frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$

Hierbij wordt verondersteld dat S_1^2 en S_2^2 berekend werden op basis van 2 onafhankelijke steekproeven met omvang n_1 en n_2 elk getrokken uit een normaal verdeelde populatie. Indien hieraan voldaan is, dan volgt de bovenstaande steekproevenverdeling een *F-verdeling.*
Deze steekproevenverdeling kan op de volgende wijze empirisch opgezet worden:

1. Trek een groot aantal steekproeven van omvang n_1 uit een normaal verdeelde populatie met gekende variantie σ_1^2. Analoog voor steekproeven van omvang n_2 uit een populatie met σ_2^2.

2. Bereken s_1^2 en s_2^2 uit steekproef.

3. Bereken voor elk paar s_1^2 en s_2^2 de ratio $(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$.

4. Ontwikkel de frekwentieverdeling van $(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$.

De resulterende steekproevenverdeling is een F-verdeling. De F-verdeling is een asymmetrische verdeling die bepaald wordt door vrijheidsgraden in de teller en vrijheidsgraden in de noemer. Bij elke combinatie van vrijheidsgraden in de teller en de noemer behoort een F-verdeling, zoals geïllustreerd door figuur 2.9.

Voor het bepalen van het $(1-\alpha)$ betrouwbaarheidsinterval, wordt uitgegaan van de volgende probabiliteit:

$$P\left[F_{\alpha/2} < \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} < F_{1-\alpha/2}\right] = 1-\alpha$$

Dit interval kan getransformeerd worden in een interval rond σ_1^2/σ_2^2:

$$P\left[\frac{F_{\alpha/2}}{S_1^2/S_2^2} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{F_{1-\alpha/2}}{S_1^2/S_2^2}\right] = 1-\alpha$$

$$P\left[\frac{S_1^2/S_2^2}{F_{1-\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2/S_2^2}{F_{\alpha/2}}\right] = 1-\alpha$$

Het $(1-\alpha)$ betrouwbaarheidsinterval voor σ_1^2/σ_2^2 wordt bijgevolg gegeven door:

$$\frac{S_1^2/S_2^2}{F_{1-\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2/S_2^2}{F_{\alpha/2}}$$

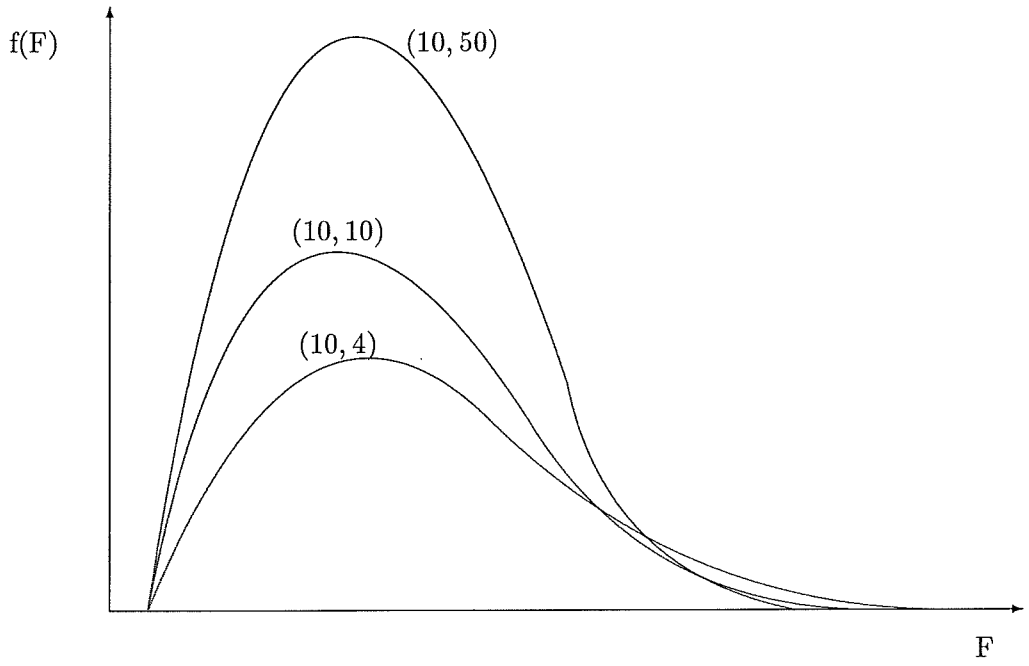
Hierbij is de keuze van s_1^2 en s_2^2 is niet zondermeer vrij te bepalen; s_1^2 moet de grootste van de 2 steekproefvarianties zijn.

Zoeken in de tabel:

- $F_{1-\alpha/2}$ met n_1-1 v.g. in teller en n_2-1 v.g. in noemer;
- $F_{\alpha/2} = 1/F_{1-\alpha/2}$ met n_2-1 v.g. in teller en n_1-1 v.g. in noemer.

Voorbeeld. Een bedrijf moet kiezen tussen twee productiewijzen voor een bepaald product. De maatstaf is de levensduurte van het product. Een aselecte steekproef van 16 stukken geproduceerd met methode 1 levert een gemiddelde levensduurte van 155 u. met een variantie 1200 u^2. Met methode 2 bedragen deze cijfers 178 u. en 800 u^2 op basis van een steekproef met 21 stukken. Kan men met een betrouwbaarheid van 95% stellen dat de variabiliteit van methode 1 significant groter is dan die van methode 2.

Oplossing:
n_1-1=16-1=15 v.g. in de teller; n_2-1=21-1=20 v.g. in de noemer.
$F_{0.975,(15,20)} = 2.57$
$F_{0.025,(15,20)} = 1/F_{0.975,(20,15)} = 1/2.76 = 0.36$



Figuur 2.9: F-verdelingen voor verschillende vrijheidsgraden in teller en noemer.

Het 95% betrouwbaarheidsinterval

$$\frac{S_1^2/S_2^2}{F_{1-\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2/S_2^2}{F_{\alpha/2}}$$

$$\frac{1200/800}{2.57} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{1200/800}{0.36}$$

$$(0.58; 4.14)$$

Aangezien 1 tot dit interval behoort, kan men niet zeggen dat de variabiliteit van methode 1 significant hoger is dan die van methode 2.

2.4 Bepaling van de steekproefomvang

2.4.1 Steekproefomvang voor het schatten van gemiddelde

Indien we een populatiegemiddelde moeten schatten op basis van het steekproefgemiddelde, dan kunnen we ons afvragen hoe groot de steekproef moet zijn om een bepaald betrouwbaarheidsniveau en precisie te halen.

Een betrouwbaarheidsinterval voor het populatiegemiddelde was van de vorm:

$$\overline{X} \pm Z\frac{\sigma}{\sqrt{n}}$$

Hierin was de foutmarge

$$E = Z\frac{\sigma}{\sqrt{n}}$$

gelijk aan de helft van het betrouwbaarheidsinterval. De foutmarge is een maatstaf die aangeeft hoe dicht de schatting bij het populatiegemiddelde ligt. Indien we de bovenstaande formule oplossen naar n verkrijgen we:

$$n = \frac{Z^2\sigma^2}{E^2}$$

In geval van $n > 0.05N$ dient de eindige populatiecorrectie te worden toegepast:

$$n = \frac{NZ^2\sigma^2}{E^2(N-1) + Z^2\sigma^2}$$

Het grote probleem hierbij is dat in de meeste gevallen de populatievariantie σ^2 niet bekend is, maar moet verkregen worden a.d.h.v. één van de volgende methoden:

1. Berekenen van de variantie van een pilootsteekproef. Deze pilootsteekproef kan dan als een deel van de finale steekproef meetellen.

2. Er zijn schattingen voor σ^2 beschikbaar uit vroegere of vergelijkbare studies.

3. Indien de oorspronkelijke populatie benaderend normaal verdeeld is, dan is:

   $$\sigma \approx R/6$$

   waarbij R de range voorstelt, d.i. het verschil tussen de grootste en de kleinste waarde in de populatie (als bekend).

Voorbeeld. Men wenst het CO-gehalte in een gasmengsel te bepalen. Op basis van een proefmonster wordt een standaarddeviatie op het gemiddelde CO-gehalte gevonden ten belope van 4%. Indien men het 95% betrouwbaarheidsinterval wenst te bepalen voor het gemiddeld CO-gehalte, hoe groot moet dan de steekproefomvang zijn, als er een foutmarge van 1% wordt toegestaan. Hoe evolueert de steekproefomvang indien een betrouwbaarheid van 90% of 99% vereist wordt?

Oplossingen:
95% betrouwbaarheid

$$n = \frac{Z^2\sigma^2}{E^2}$$

$$n = \frac{1.96^2 0.04^2}{0.01^2}$$

$$n = 62$$

90% betrouwbaarheid

$$n = \frac{1.645^2 0.04^2}{0.01^2}$$

$$n = 44$$

99% betrouwbaarheid

$$n = \frac{2.58^2 0.04^2}{0.01^2}$$

$$n = 107$$

Bovenstaande methode ter bepaling van de steekproefomvang voor schatting van het populatiegemiddelde kan uitgebreid worden naar het geval met 2 steekproeven. Hierbij wordt verondersteld dat beide steekproeven voldoend groot zijn om de centrale limietstelling toe te passen. We vertrekken vanuit de formule voor de foutmarge:

$$E = Z\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Indien we stellen dat $n = n_1 = n_2$ verkrijgen we

$$E^2 = Z^2 \frac{\sigma_1^2 + \sigma_2^2}{n}$$

en dus

$$n = \frac{Z^2(\sigma_1^2 + \sigma_2^2)}{E^2}$$

De populatievarianties kunnen geschat worden volgens een van de hiervoor aangehaalde methoden.

Voorbeeld. Een bedrijf wenst te weten of er een verschil is tussen het aantal gepresteerde dienstjaren van mannen en vrouwen op het moment van oppensioenstelling. Men eist hiervoor een betrouwbaarheidsinterval van 1 jaar en betrouwbaarheid van 95%. Een pilootsteekproef leverde varianties van 5 en 7 op. Hoe groot moet de steekproefomvang zijn indien deze even groot moet zijn voor mannen als voor vrouwen.

Oplossing:

$$n = \frac{Z^2(\sigma_1^2 + \sigma_2^2)}{E^2}$$
$$n = \frac{1.96^2(5+7)}{0.5^2}$$
$$n = 185$$

2.4.2 Steekproefomvang voor het schatten van proporties

Het bepalen van de steekproefomvang voor het schatten van proporties is vergelijkbaar met dat voor het schatten van gemiddelden. Onderstaande formules gelden indien de normaal benadering geldt voor de steekproevenverdeling van de steekproeffractie P.

$$n = \frac{Z^2(1-\pi)\pi}{E^2}$$

Indien $n > 0.05N$ is de eindige populatiecorrectie vereist.

$$n = \frac{NZ^2(1-\pi)\pi}{E^2(N-1) + Z^2(1-\pi)\pi}$$

Het probleem is hier opnieuw π, de te schatten onbekende populatieproportie. Een schatting voor π kan verkregen worden door:

- de fractie P uit een pilootsteekproef te berekenen;
- een goede schatting, omdat men goed vertrouwd is met de situatie;
- $\pi = 0.5$ te nemen; deze conservatieve reflex resulteert in de grootst mogelijke steekproefomvang.

Voorbeeld. Men wil te weten komen welke fractie van de bevolking voorstander is van een referendum. Hoe groot moet een steekproef zijn om met een nauwkeurigheid van 3% en een betrouwbaarheid van 95% die fractie te bepalen?

Oplossing:
π=? $\rightarrow \pi = 0.5$

$$n = \frac{Z^2(1-\pi)\pi}{E^2}$$
$$n = \frac{1.96^2(0.5)0.5}{0.03^2}$$
$$n = 1068$$

In geval van 2 steekproeven verkrijgen we de volgende formule:

$$n = Z^2 \frac{(1-\pi_1)\pi_1 + (1-\pi_2)\pi_2}{E^2}$$

Hierbij wordt verondersteld dat de steekproevenverdeling van het verschil tussen twee steekproefproporties normaal is en dat $n_1 = n_2 = n$.

Voorbeeld. Een reclamebureau wenst het verschil tussen twee proporties te kennen. De proporties hebben betrekking op de hoeveelheid klanten van twee verschillende groepen die een bepaald product hebben geconsumeerd. Men wenst een 95% betrouwbaarheidsinterval en een intervalbreedte van 0.10. De schattingen voor π_1 en π_2 zijn 0.20 en 0.25. Hoe groot moet de omvang van beide steekproeven zijn?

Oplossing:

$$n = Z^2 \frac{(1-\pi_1)\pi_1 + (1-\pi_2)\pi_2}{E^2}$$

$$n = 1.96^2 \frac{(0.20)0.80 + (0.25)0.75}{0.05^2}$$

$$n = 534$$

Opmerking. Bepaling van de steekproefomvang voor een gestratificeerde steekproef. Hiervoor kan de totale steekproefomvang berekend worden via de reeds beschouwde formules voor n. De verdeling van de totale steekproefomvang over de strata kan dan gebeuren aan de hand van de volgende formule:

$$n_A = \frac{nN_A\sigma_A}{N_A\sigma_A + N_B\sigma_B + N_C\sigma_C + ...}$$

Hierbij is N_A de populatiegrootte van stratum A, σ_A de standaarddeviatie van de populatie van stratum A.

2.4.3 Steekproefomvang voor het schatten van varianties

Het bepalen van de steekproefomvang voor het schatten van varianties is vooral vereist in toepassingen van kwaliteitscontrole. Nochtans is het bepalen van de steekproefomvang voor de schatting van σ^2 een veel complexere materie dan deze die we gezien hebben voor het schatten van gemiddelden en proporties. Bijgevolg zullen we ons hier beperken tot een aantal benaderende tabellen (2.4 en 2.5) die daarvoor kunnen gebruikt worden. Voor een gegeven foutmarge van de schatter S^2 en een gegeven betrouwbaarheidsniveau wordt de steekproefomvang uit de tabel gelezen.

Voorbeeld. Er wordt gevraagd om een schatting voor σ^2 te doen die binnen 20% van de werkelijke populatiewaarde ligt en met een betrouwbaarheidsniveau van 99%. Veronderstel hierbij dat de populatie normaal verdeeld is.

Oplossing:
Tabel 2.5: n=368.

Foutmarge%	Steekproefgrootte
1	77210
5	3150
10	806
20	210
30	97
40	57
50	38

Tabel 2.4: Steekproefomvang voor schatting van σ^2 bij een 95% betrouwbaarheid.

Foutmarge%	Steekproefgrootte
1	133362
5	5454
10	1400
20	368
30	172
40	101
50	67

Tabel 2.5: Steekproefomvang voor schatting van σ^2 bij een 99% betrouwbaarheid.

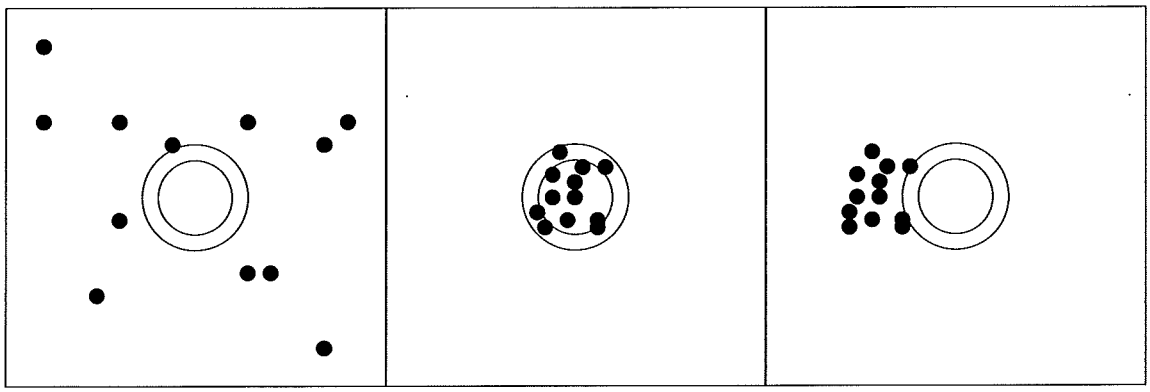
2.5 Samenvatting

1. Het steekproefgemiddelde $\overline{X}$ is een goede schatter voor het populatiegemiddelde μ.
2. De steekproefvariantie S^2 is een goede schatter voor de populatievariantie σ^2.
3. De steekproeffractie P is een goede schatter voor de populatiefractie π.
4. Een schatter is onvertekend (zuiver) als deze gemiddeld samenvalt met zijn doel in de populatie. Een schatter is efficient indien deze de kleinste variantie heeft.
5. De gemiddelde kwadratische afwijking is de meest geschikte maatstaf voor zowel vertekendheid als variantie van een schatter te meten.
6. Een consistente schatter haalt beter zijn doel in de populatie naarmate de steekproefomvang toeneemt.
7. Op basis van de normaal verdeling is in 95% van de gevallen $\overline{X}$ begrepen binnen 1.96 standaardfouten van μ.
8. Voor voldoende grote steekproeven is het betrouwbaarheidsinterval voor een proportie of het verschil tussen twee proporties eveneens gebaseerd op normaliteit van de steekproevenverdeling.
9. Als de populatievariantie σ onbekend is, dan moet deze met S^2 geschat worden en wordt de vlakkere t-verdeling gebruikt als steekproevenverdeling.
10. Het bepalen van het betrouwbaarheidsinterval voor het verschil van 2 populatiegemiddelden kan gebeuren in de veronderstelling dat de populatievarianties gelijk of ongelijk zijn.
11. Het bepalen van betrouwbaarheidsintervallen voor de populatievariantie of voor de ratio tussen twee populatievarianties is gebaseerd respectievelijk op de χ^2 en de F-steekproevenverdeling.
12. De steekproefomvang voor het schatten van gemiddelden en proporties kan bepaald worden via de foutmarge van de schatter. De steekproefomvang voor het schatten van varianties kan op basis van tabellen gerealiseerd worden.
13. Een aantal belangrijke vertalingen:

schatter	*estimator*
schatting	*estimation*
vertekening	*bias*
zuiver	*unbiased*
betrouwbaarheidsinterval	*confidence interval*
foutmarge	*error margin*

2.6 Oefeningen

1. Drie pistolen worden getest, elk met 12 schoten. Pistool A werd onvoldoende vastgemaakt op het statief. Pistool B werd vastgemaakt in een positie die naar links afwijkte. Pistool C werd correct vastgemaakt.
   (1) Welk van de onderstaande schietresultaten behoort bij welk pistool? (2) Welke pistolen zijn vertekend; welke hebben de kleinste variantie; welke hebben de grootste gemiddelde kwadratische afwijking; welke is de meest efficiente; welke is de minst efficiente?



2. Aan de hand van een meettoestel wenst men een schatting te maken van de gemiddelde hoeveelheid droge stof in een oplossing. De standaarddeviatie op een meting bedraagt 1.2 mg. Zes metingen hebben de volgende hoeveelheden droge stof opgeleverd:

   $$14.7, 13.2, 15.8, 16.1, 14.3, 15.0$$

   Indien verondersteld mag worden dat de metingen normaal verdeeld zijn, bepaal dan een 90%, 95% en 99% betrouwbaarheidsinterval voor de werkelijke gemiddelde hoeveelheid droge stof in de oplossing.

3. Twee soorten meststoffen worden getest op wortelen. Elke meststof wordt op een ander veld gebruikt. Een steekproef van 25 wortelen wordt genomen uit het veld met meststof 1, met als resultaat een gemiddeld gewicht van 44.1 gr. en een s.d. van 6 gr. Uit veld met meststof 2 wordt een aselecte steekproef met omvang 12 genomen met als resultaat een gemiddeld gewicht van 31.7 gr. en een variantie van 44 gr.^{2}. De beide populaties van de gewichten zijn normaal verdeeld. Bepaal een 95% betrouwbaarheidsinterval voor het verschil van de populatiegemiddelden $\mu_1 - \mu_2$, zowel in de veronderstelling van gelijke als ongelijke populatievarianties.

4. Aan een steekproef van 10 bedrijven werd gevraagd naar het bedrag aan opleiding per werknemer dat afgelopen jaar en tien jaar geleden werd uitgegeven. Kan men met 95%

Bedrijf	1	2	3	4	5	6	7	8	9	10
jaar t-1	12	14	8	12	8	10	8	9	10	10
jaar t-10	10	11	8	7	9	6	10	9	7	9

Tabel 2.6: Bedragen aan opleiding per werknemer in 1000 BEF.

betrouwbaarheid zeggen dat er een beduidend verschil is tussen de bedragen vorig jaar en die 10 jaar geleden?

5. Een personeelsdirecteur wilt weten hoeveel mensen in het bedrijf reeds voor een ander bedrijf hebben gewerkt. Hij neemt daarvoor een steekproef van 500 mensen. Hiervan bleken er 76 reeds in een ander bedrijf te hebben gewerkt. Bepaal een 95% betrouwbaarheidsinterval voor het aantal mensen en een 99% betrouwbaarheidsinterval voor de fractie van de mensen in het bedrijf die reeds in een ander bedrijf hebben gewerkt.

6. Een steekproef van 1000 mannen en 1000 vrouwen leert dat 60% van de vrouwen en 52% van de mannen trouwt met hun eerste partner. Kan men met 96% zekerheid stellen dat er een significant verschil is tussen mannen en vrouwen?

7. Het reservoir met antivriesvocht van een bepaalde wagen heeft een inhoud van 3785 ml. De verantwoordelijke voor de kwaliteit wil zeker zijn dat de standaarddeviatie niet meer dan 30 ml. is. Om dit te controleren wordt een steekproef van 19 vaten genomen. Het resultaat is een gemiddelde van 3787 ml. met een standaarddeviatie van 55.4 ml. Bepaal een 95% betrouwbaarheidsinterval voor de populatievariantie σ. Suggereert dit betrouwbaarheidsinterval dat de variabiliteit van het productieproces aanvaardbare blijft? Veronderstel dat de verdeling van de inhoud van de reservoirs normaal is.

8. Een aselecte steekproef van 16 lokale radiostations wordt gebruikt om de gemiddelde prijs per minuut voor een reclameboodschap met 95% betrouwbaarheid te schatten. Het resultaat is een gemiddelde prijs van 1550 BEF met een standaarddeviatie van 800 BEF. We veronderstellen dat de prijs globaal gezien normaal verdeeld is.

9. Een onderzoeksbureau wenst de steekproefomvang te bepalen om de fractie van de 500 gezinnen in een gemeente te bepalen die een bepaalde advertentie in een lokale krant hebben gelezen. De geschatte proportie moet binnen 4% van de werkelijke proportie liggen en een betrouwbaarheidsniveau van 90% halen. In een pilootsteekproef met 20 gezinnen werd een fractie van 35% opgetekend.

10. Een bedrijf wenst het effect van een bepaalde promotiecampagne te meten op de verkopen. Hiervoor worden 16 verkoopspunten geselecteerd, waarvan de verkopen van vorige maand (zonder promotie) vergeleken worden met die van deze maand (met promotie). Heeft de promotiecampagne een significant effect op de verkopen gehad, indien een betrouwbaarheidsniveau van 95% wordt gehanteerd?

Winkel	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
maand t-1	44	61	46	55	49	50	45	64	40	62	53	54	57	55	61	52
maand t	47	62	56	39	45	51	56	47	48	40	44	52	51	60	61	57

Tabel 2.7: Verkopen in 1000 BEF.

11. Een job-tevredenheidstest wordt afgenomen van een groep mensen die een opleiding heeft gevolgd en van een andere groep die er geen heeft gevolgd. De gemiddelde score van 50 mensen met de opleiding is 4.5, die van de groep zonder de opleiding is 3.75. De twee populaties kunnen benaderend normaal verdeeld worden verondersteld m.b.t. de testresultaten. Is er een significant effect van de opleiding op een betrouwbaarheidsniveau van 95%? De populatievariantie van de opgeleiden is 1.8 en die voor de niet-opgeleiden is 2.1.

12. Voor het bepalen van de gemiddelde leeftijd van zijn werknemers doet een bedrijf een steekproef bij 50 van haar 3123 werknemers. Het resultaat was een gemiddelde leeftijd van 36 jaar met een standaarddeviatie van 12 jaar.
(a) Bepaal het 95% betrouwbaarheidsinterval voor de gemiddelde leeftijd van de 3123 werknemers.
(b) Stel dat je het 95% betrouwbaarheidsinterval wilt terugbrengen tot ± 2 jaar, wat is de steekproefomvang daarvoor nodig? (c) Stel dat je het 95% betrouwbaarheidsinterval wilt terugbrengen tot ± 1 jaar, wat is de steekproefomvang daarvoor nodig?

13. Een bedrijf wenst de gemiddelde leeftijd van zijn werknemers te bepalen. Op basis van een steekproef van 60 werknemers wordt een gemiddelde leeftijd van 23.67 jaar verkregen. Uit vergelijkbare studies is geweten dat de leeftijd van de populatie niet normaal verdeeld is en dat de standaarddeviatie 15 jaar bedraagt. Bepaal een 99% betrouwbaarheidsinterval.

14. De borstomtrek van de 21 kandidaten voor de Mister Body Building verkiezing was 155 cm met een variantie van 6 cm^2. Die van de 17 kandidaten voor de Mister Fitness was 145 cm met een variantie van 4 cm^2. Indien we normaliteit van de borstomtrek in beide populaties veronderstellen, is de variabiliteit in de borstomtrek van een Mr. Body Building kandidaat hoger dan die van een Mr. Fitness, voor een 99% betrouwbaarheid.

15. De ziekteverzekering is geïnteresseerd in de leeftijden van ziekenhuispatienten in 2 bepaalde ziekenhuizen. In ziekenhuis A werd een steekproef van 22 patienten genomen, in B één van 20. De resultaten voor A zijn een gemiddelde leeftijd van 54.5 jaar met een standaarddeviatie van 16 jaar; voor B zijn deze waarden respectievelijk 39.4 jaar en 14.14 jaar. De beide populaties zijn normaal verdeeld met betrekking tot de leeftijd. Is er een significant verschil in gemiddelde leeftijd tussen de patienten van ziekenhuis A en die van ziekenhuis B op een betrouwbaarheidsniveau van 95%? Veronderstel: (a) gelijke populatievarianties, (b) ongelijke populatievarianties.

16. Een bedrijf met 2500 werknemers wenst de gemiddelde reistijd van zijn werknemers te kennen van en naar het bedrijf. Het beheer eist dat de schatting voor de gemiddelde reistijd binnen 1 min. van het werkelijk gemiddelde ligt. Een pilootsteekproef heeft een variantie van 25 min^2 opgeleverd. Wat is de vereiste steekproefomvang?

17. In een opiniepeiling waren 320 van de 400 ondervraagden tegen de regeringspolitiek inzake partijfinanciering.
    (a) Bepaal een 95% betrouwbaarheidsinterval voor de werkelijke proportie tegenstanders.
    (b) Hoeveel bedraagt de foutmarge voor een 99% betrouwbaarheidsniveau.
    (c) Wat is de kans dat de proportie tegenstanders tussen 77% en 83% ligt?

18. Een industriele psycholoog wenst het verschil in proportie te meten van arbeiders en bedienden die te maken hebben gehad met ongewenste intimiteiten. Hij wenst hiervoor een betrouwbaarheidsniveau van 90% en een intervalbreedte van 14%. Hoeveel werknemers van elke groep zal hij hiervoor moeten selecteren, indien verondersteld wordt dat beide steekproeven even groot moeten zijn? Schattingen voor P_1 en P_2 zijn 0.10 en 0.28.

Hoofdstuk 3

Hypothesetoetsen

3.1 Inleiding

Het toetsen van hypothesen vormt naast het intervalschatten de andere belangrijke tak van de statistische inferentie. De bedoeling van het toetsen van hypothesen is om een beslissing te ondersteunen omtrent een populatie op basis van steekproefgegevens.

Definitie 5 *Een toets is een statistische procedure die toelaat om te kiezen tussen twee tegenstrijdige hypothesen.*

Een *hypothese* of theorie is een bewering omtrent 1 of meerdere populaties. Bv. nagaan of het bijwonen van de les significant betere examenresultaten oplevert. Indien men zou beschikken over de gehele populatie, dan kan deze hypothese gemakkelijk aanvaard of verworpen worden. Het probleem is echter om deze hypothese te aanvaarden of te verwerpen op basis van een beperkte steekproef. Wegens het beperkt karakter van de steekproef moeten de conclusies met een zekere mate van onzekerheid beschouwd worden.

3.2 Stappen in een hypothesetoets

Bij het opzetten van een hypothesetoets, zullen we steeds de volgende 7-staps procedure aanhouden.

1. Bepalen van de hypothesen.
2. Identificatie van de toetsstatistiek.
3. Bepalen van het significantieniveau.
4. Beslissingsregel.
5. Berekeningen.
6. Statistische beslissing.
7. Conclusie.

In wat volgt zullen de onderscheiden stappen in detail belicht worden en verder geïllustreerd worden a.d.h.v. het onderstaand voorbeeld.

Voorbeeld. Een vereniging ter promotie van de fiets beweert dat de fiets van de Vlaming hoogstens 10 jaar oud is. Nochtans wordt op basis van een bevraging bij 40 Vlamingen een gemiddelde leeftijd van 13.41 jaar verkregen met een standaarddeviatie van 8.28 jaar. We kunnen ons hierbij afvragen of het verschil tussen de hypothetische waarde 10 en het steekproefgemiddelde 13.41 jaar het gevolg is van het toeval of niet. Of m.a.w. kunnen we op basis van het steekproefresultaat besluiten dat het werkelijk populatiegemiddelde hoger is dan 10 jaar?

3.2.1 Bepalen van de hypothesen

Bij elke toets worden twee tegenstrijdige hypothesen geëvalueerd: de nulhypothese H_0 en de alternatieve hypothese H_1. Beide hypothesen zijn disjunct:

$$H_0 \cap H_1 = \emptyset$$

De nulhypothese krijgt steeds het voordeel van de twijfel; H_0 wordt aanvaard tenzij de waarnemingen voldoende sterke aanwijzingen bevatten al zou H_0 onjuist zijn.

Regel 3 *De nulhypothese is de hypothese die getest wordt.*

De nulhypothese kan beschouwd worden als de hypothese van geen verandering, de hypothese van gelijkheid. Daarom bevat de nulhypothese steeds een gelijkheidsteken. De hele statistische analyse voor een hypothesetoets wordt trouwens uitgevoerd in de veronderstelling dat de nulhypothese waar is. Indien de hypothesetoets wordt gebruikt om na te gaan of θ, de onbekende populatieparameter (gemiddelde, proportie, variantie,...), al of niet gelijk is aan een vooropgestelde waarde θ_0, dan zijn de hypothesen de volgende: De alternatieve hypothese H_1 bevat de bewering die we willen aantonen. Bv. je wil toetsen of je steekproefgegevens kunnen aantonen of een populatiegemiddelde μ verschillend is van een bepaalde μ_0. De hypothesen zijn dan:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

In voorkomend geval spreekt van een tweezijdige toets, aangezien verwacht wordt dat de populatieparameter ofwel groter ofwel kleiner kan zijn dan de hypothetische waarde μ_0. Meer algemeen:

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

In het geval van een eenzijdige toets, doen er zich twee mogelijke toetsen voor. De eerste is die waarbij men zich afvraagt of op basis van de steekproefgegevens mag besloten worden dat de werkelijke populatieparameter θ groter is dan een hypothetische waarde θ_0. De beide hypothesen zijn dan:

$$H_0 : \theta \leq \theta_0$$
$$H_1 : \theta > \theta_0$$

In het andere geval vraagt men zich af of de werkelijke populatieparameter θ kleiner is dan een hypothetische waarde θ_0.

$$H_0 : \theta \geq \theta_0$$
$$H_1 : \theta < \theta_0$$

Hieruit blijkt nogmaals dat de nulhypothese steeds de gelijkheid, de niet-verandering in zich draagt, terwijl de alternatieve hypothese hetgeen bevat dat men wil aantonen op basis van de steekproefresultaten.
Zoals reeds aangehaald, wordt de statistische analyse uitgevoerd in de veronderstelling dat de nulhypothese niet verworpen wordt. Indien de toets leidt naar een *verwerping* van de nulhypothese, dan is dit voldoende sterk om te kunnen beweren dat de alternatieve hypothese juist is. Wanneer H_0 niet verworpen wordt, dan betekent dit daarom nog niet dat H_0 juist zou zijn; immers, de steekproefresultaten tonen niet aan dat H_0 juist zou zijn. Dus, men vermijdt best de uitdrukking "aanvaarden van de nulhypothese".
Ter verduidelijking het volgend voorbeeld. Een fabrikant van een product om vlekken te verwijderen beweert dat zijn product alle vlekken kan verwijderen. Om deze verklaring te toetsen wordt een steekproef genomen van 25 soorten vlekken. Alle 25 vlekken werden verwijderd met het product. Dit laat ons echter nog niet toe om de nulhypothese (alle vlekken kunnen verwijderd worden) als juist te evalueren. Het is voldoende dat een 26ste soort vlek niet meer verwijderbaar is om de nulhypothese te verwerpen.

Voorbeeld. In het voorbeeld met betrekking tot de leeftijd van de fiets van de Vlaming, wenst men, op basis van de steekproefresultaten, aantonen dat het werkelijk populatiegemiddelde hoger ligt dan 10 jaar. Dit is meteen de alternatieve hypothese. De nulhypothese omvat de bewering van de vereniging al zou de gemiddelde leeftijd van de fiets de 10 jaar niet overstijgen.

$$H_0 : \mu \leq 10$$
$$H_1 : \mu > 10$$

3.2.2 Identificatie van de toetsstatistiek

De toetsstatistiek kan beschouwd worden als de beslissingsnemer, aangezien de beslissing om de nulhypothese al of niet te verwerpen gebaseerd is op de numerieke waarde van de toetsstatistiek. Die numerieke waarde wordt berekend uit de steekproefgegevens en vergeleken met een kritische waarde, die de grens vormt tussen het aanvaardings- en het verwerpingsgebied.

De toetsstatistiek is afhankelijk van de steekproefgegevens. Indien de gegevens metrisch of intervalgeschaald zijn, dan is het gebruik van de klassieke t en Z statistieken aangewezen. Verder in dit hoofdstuk zullen we echter zien dat in vele situaties de gegevens niet meer metrisch, maar ordinaal of zelfs nominaal geschaald zijn. In dergelijke gevallen dringt het gebruik van een niet-parametrische toets met bijgaande toetsstatistiek zich op.

Voorbeeld. In het voorbeeld van de fietsenleeftijd hebben we te maken met metrische gegevens. Dit betekent dat het gebruik van een Z of t statistiek aangewezen is. Aangezien de steekproefomvang $n > 30$, geldt de centrale limietstelling en is de Z statistiek de aangewezen

toetsstatistiek:

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

3.2.3 Significantieniveau

Het significantieniveau is een probabiliteit die de grens vormt tussen de toevallige en significante verschillen.

	Statistische conclusie	
In populatie	H_0 *niet verwerpen*	H_0 *verwerpen*
H_0 *juist*	juist $(1-\alpha)$	type I fout (α)
H_0 *fout*	type II fout (β)	juist $(1-\beta)$

Tabel 3.1: Mogelijke situaties bij een hypothesetoets.

Tabel 3.1 geeft de situaties weer die zich kunnen manifesteren bij een hypothesetoets. Twee type fouten kunnen zich voordoen. In het geval van een fout van de eerste soort (type I) wordt de nulhypothese onterecht verworpen (*false positive*). Men begaat een fout van de tweede soort (type II) indien men de nulhypothese onterecht niet verwerpt (*false negative*). Bij een hypothesetoets is er steeds een bepaalde kans dat men een van beide fouten zal begaan. De waarschijnlijkheid dat men een type I fout begaat is α; de waarschijnlijkheid van een type II fout is β. Hoe groter α (β), hoe groter de kans dat men een type I (type II) fout zal begaan. Het probleem is echter dat α en β samenhangen, t.t.z. indien men de kans α wil verminderen, verhoogt men β en vice-versa. De enige manier om beide kansen te verminderen is door een verhoging van de steekproefomvang.

Zoals geweten is de conclusie om H_0 te verwerpen de meest definitieve. De H_0 wordt pas verworpen als de resultaten van de steekproef voldoende doorslaggevend zijn. Het niet verwerpen van H_0 is veel minder definitief; in voorkomend geval is het daarom niet zeker dat H_0 juist is.
Indien de nulhypothese verworpen wordt, dan gaat men er van uit dat H_0 onjuist is. Als H_0 toch juist blijkt te zijn, dan maakt men een fout van de eerste soort. Het niet verwerpen van H_0 is minder definitief en bijgevolg zal een fout van de tweede soort overwegend minder ernstig zijn dan één van de eerste soort. Andersom: door de ernst van een fout van eerste soort zal H_0 niet gauw verworpen worden; men laat liever het voordeel van de twijfel.

Een vergelijkbaar probleem is dat van de rechtspraak, zoals geïllustreerd wordt met tabel 3.2. Hierbij is de nulhypothese dat de verdachte onschuldig is. In onze rechtspraak dient de fout van type I in elk geval vermeden te worden: het veroordelen van een onschuldige. Het gevolg van een type II fout is dus minder ernstig. Voor bepaalde totalitaire regimes ligt dit probleem anders: je bent schuldig tot je je onschuld kan bewijzen. In voorkomend geval wordt de kans op een type II fout geminimaliseerd.

	Uitspraak	
Verdachte is	*Verdachte onschuldig*	*Verdachte schuldig*
onschuldig	juist	type I fout
schuldig	type II fout	juist

Tabel 3.2: Mogelijke situaties bij een rechterlijke uitspraak.

	Conclusie	
Partij is	*Partij aanvaarden*	*Partij verwerpen*
goed	juist	type I fout
niet goed	type II fout	juist

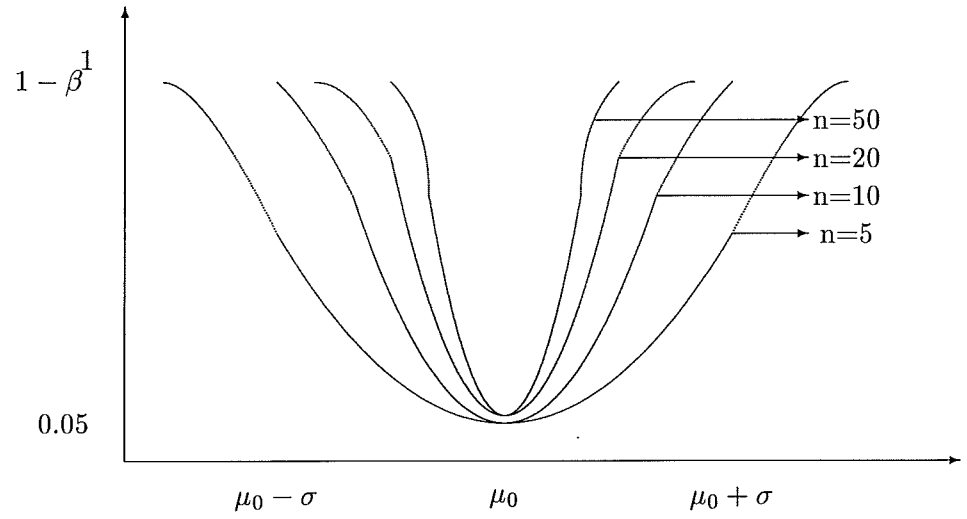
Tabel 3.3: Mogelijke situaties bij de kwaliteitscontrole van een partij goederen.

De aard van de toepassing bepaalt grotendeels welk type van fout dient geminimaliseerd te worden. Een ander voorbeeld is dat van de kwaliteitscontrole, waarbij een steekproef wordt getrokken uit een partij geleverde goederen. Op basis van de fractie defecten wordt een beslissing genomen om de partij goederen al of niet te aanvaarden. De nulhypothese is dat de partij goederen goed is.
Tabel 3.3 reflecteert de mogelijke situaties die kunnen ontstaan bij de kwaliteitscontrole.
De manager dient de kans van de twee fouttypes aan te passen met betrekking tot de kosten geassocieerd met elk van deze foute beslissing. Indien bijvoorbeeld het aanvaarden van een slechte partij goederen (type II fout of het consumentenrisico) meer kosten veroorzaakt dan het verwerpen van een goede partij (type I fout of het producentenrisico), dan moet de kans op een type II fout geminimaliseerd worden.

Het *significantieniveau* α wordt gebruikt om de kans op een type I fout vast te leggen. In principe kunnen hiervoor alle waarden tussen 0 en 1 gebruikt worden, maar traditioneel worden meestal de waarden $\alpha = 0.05$ en $\alpha = 0.01$ verkozen. Het kiezen van een waarde voor het significantieniveau impliceert eveneens het vastleggen van het betrouwbaarheidsniveau $\beta = 1 - \alpha$.
De nulhypothese wordt verworpen indien de probabiliteit die overeenkomt met de waarde van de toetsstatistiek kleiner of gelijk is aan α. In voorkomend geval wordt de toetsstatistiek significant genoemd.

Het *verwerpingsgebied* omvat de waarden van de toetsstatistiek die onwaarschijnlijk zijn in het geval de nulhypothese waar is. De overige waarden van de toetsstatistiek vormen het *aanvaardingsgebied*, of correcter het "niet-verwerpingsgebied". De *kritieke waarden* vormen de grens tussen het verwerpings- en het aanvaardingsgebied. De waarde van het significantieniveau α, de hypothetische waarde en de relevante steekproevenverdeling bepalen de lokatie van het aanvaardings- en het verwerpingsgebied.

De probabiliteit $1 - \beta$ wordt ook het *onderscheidingsvermogen* van de toets genoemd. Het



Figuur 3.1: Onderscheidingsvermogen voor een tweezijdige toets met $\alpha = 0.05$

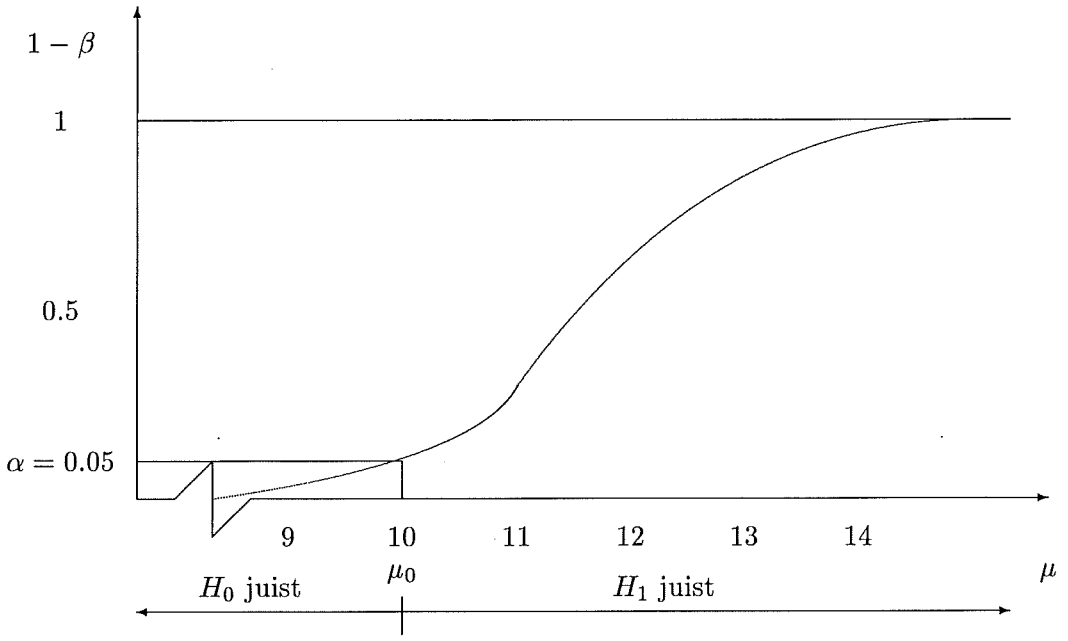
vertegenwoordigt de kans dat H_0 verworpen wordt indien H_0 werkelijk vals is. Hierbij is het interessant om de evolutie van het onderscheidingsvermogen van de toets weer te geven in functie van verschillende waarden voor het werkelijk populatiegemiddelde. Hiermee kan men eveneens de evolutie van de beide fouttypes meevolgen in functie van een veranderende waarde van μ.
Figuur 3.1 geeft weer hoe de kans op het maken van een type II (β) fout evolueert in functie van de steekproefomvang n voor een tweezijdige test op het gemiddelde. Het gemiddelde onder H_0 wordt voorgesteld door μ_0. Uit de figuur is eveneens af te leiden dat als het populatiegemiddelde $\mu = \mu_0$ (d.i. als H_0 waar is) er een kans van $\alpha = 0.05$ bestaat dat H_0 verworpen wordt. Duidelijk is dat het onderscheidingsvermogen van een toets verhoogt naarmate de steekproefomvang toeneemt; d.w.z. dat de toets een duidelijkere beslissing tussen aanvaarden en verwerpen van H_0 kan nemen.

Voorbeeld. In het voorbeeld met de gemiddelde fietsenleeftijd wordt geen significantieniveau of geen betrouwbaarheidsniveau gespecificeerd. Traditioneel geldt dan $\alpha = 0.05$.

Indien we voor deze eenzijdige toets de evolutie van het onderscheidingsvermogen en de beide fouttypes in functie van de waarde van het populatiegemiddelde willen voorstellen, dan dienen we vooreerst de kritieke waarde van $\overline{X}$ te berekenen die de grens vormt tussen het aanvaardings- en het verwerpingsgebied. Voor een significantieniveau van $\alpha = 0.05$ is dit de waarde van $\overline{X}$ die overeenkomt met de waarde $z = 1.645$.

$$1.645 = \frac{\overline{x} - 10}{1.31}$$

$$\overline{x} = 12.15$$



Figuur 3.2: Onderscheidingsvermogen van het voorbeeldprobleem voor steekproefomvang $n = 40$

Berekening van de kans op type II fout voor een aantal alternatieve waarden van μ geeft:

$$z = \frac{12.15 - 9}{1.31} = 2.40 \qquad 1 - \beta = 0.0082$$

$$z = \frac{12.15 - 11}{1.31} = 0.88 \qquad 1 - \beta = 0.1894$$

$$z = \frac{12.15 - 12}{1.31} = 0.11 \qquad 1 - \beta = 0.4562$$

$$z = \frac{12.15 - 13}{1.31} = -0.65 \qquad 1 - \beta = 0.7422$$

$$z = \frac{12.15 - 14}{1.31} = -1.41 \qquad 1 - \beta = 0.9207$$

Figuur 3.2 bevat de grafische weergave van de evolutie van het onderscheidingsvermogen. Hieruit blijkt duidelijk dat naarmate het werkelijke populatiegemiddelde μ lager wordt dan de hypothetische waarde $\mu_0 = 10$, de kans op een fout van de eerste soort (α) afneemt. Hoe hoger μ t.o.v. $\mu_0 = 10$ wordt, hoe hoger het onderscheidingsvermogen ($1 - \beta$). Immers, hoe dichter μ en μ_0 bij elkaar liggen, hoe moeilijker het wordt om de juiste keuze te maken tussen de twee hypothesen.

3.2.4 Beslissingsregel

De beslissingsregel is de richtlijn die wordt gevolgd bij het al of niet verwerpen van de nulhypothese.

Regel 4 *Indien de waarde van de toetsstatistiek berekend uit de steekproef in het verwerpingsgebied valt, dan wordt H_0 verworpen. Valt de waarde van de toetsstatistiek binnen het aanvaardingsgebied, dan wordt H_0 niet verworpen. Komt de waarde van de toetsstatistiek overeen met de kritieke waarden, dan wordt H_0 verworpen.*

Dezelfde regel kan eveneens in probabiliteitstermen gesteld worden.

Regel 5 *In de veronderstelling dat H_0 waar is, dan wordt H_0 verworpen indien de waarschijnlijkheid om een waarde van de toetsstatistiek te verkrijgen die minstens zo ver van de hypothetische waarde verwijderd is als de berekende waarde van de toetsstatistiek, lager of gelijk is aan α.*

Voorbeeld. In het voorbeeld met de gemiddelde fietsenleeftijd geldt de volgende beslissingsregel:

$$\text{Als } Z \geq 1.645 \text{ dan } H_0 \text{ verwerpen}$$

Hierbij is 1.645 de kritieke waarde $z_{1-\alpha}$ voor een eenzijdige toets met een significantieniveau α. Het feit dat er rechtseenzijdig getest wordt heeft alles te maken met de formulering van de alternatieve hypothese:

$$H_1 : \mu > 10$$

en dus wordt H_1 aanvaard als

$$Z \geq 1.645$$

3.2.5 Berekeningen

De berekeningen zijn gebaseerd op gegevens afkomstig van een aselecte steekproef.

Voorbeeld. De berekening van de Z-waarde

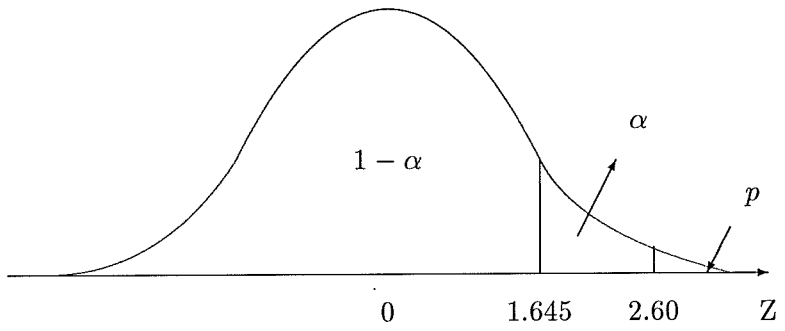
$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

$$z = \frac{13.41 - 10}{8.28/\sqrt{40}} = 2.60$$

3.2.6 Statistische conclusie

De berekende waarde van de toetsstatistiek wordt vergeleken met de kritieke waarde in het licht van de beslissingsregel.

Voorbeeld. Aangezien $z = 2.60 > 1.645$, bevindt de berekende waarde van Z zich in het verwerpingsgebied. H_0 wordt dus verworpen.



Figuur 3.3: Grafische weergave van de hypothesetoets op de leeftijden van de fietsen.

3.2.7 Conclusie

Indien de nulhypothese verworpen wordt, dan is de conclusie dat de alternatieve hypothese waar is. Het niet verwerpen van de nulhypothese laat ons enkel toe te besluiten dat de nulhypothese waar zou kunnen zijn.

Bovendien kan ook de zogenaamde *p-waarde* berekend worden. De p-waarde vat in 1 probabiliteitswaarde samen in hoeverre er overeenkomst bestaat tussen de steekproefgegevens en de nulhypothese. Het wordt ook wel eens het *minimale significantieniveau* genoemd, want de p-waarde is eigenlijk het kleinst mogelijk significantieniveau waarbij de nulhypothese verworpen kan worden. Indien bijvoorbeeld $p = 0.01$ dan betekent dit dat H_0 verworpen kan worden tot op een significantieniveau van 0.01. Uiteraard impliceert dit dat H_0 eveneens verworpen wordt voor elke waarde van het significantieniveau groter of gelijk aan 0.01.
Het gebruik van de p-waarde als enige resultaat van een hypothesetoets is voldoende. De waarde draagt in zich een aanwijzing waarmee aangegeven wordt hoe sterk de H_0 al of niet verworpen werd.
De berekening van de p-waarde is afhankelijk van de toetsstatistiek, de berekende waarde van de toetsstatistiek en het één- of tweezijdig karakter van de toets.

Voorbeeld. Aangezien de nulhypothese verworpen wordt, is de conclusie dat de gemiddelde leeftijd van de Vlaamse fiets niet kleiner of gelijk is aan 10 jaar.
De p-waarde wordt berekend door in de standaardnormale tabel de kans op te zoeken die overeenkomt met de berekende waarde $z = 2.60$. Dit is $p = 0.0047$. Deze waarde is kleiner dan $\alpha = 0.05$ waardoor men dadelijk kan zien dat de H_0 duidelijk verworpen werd. Figuur 3.3 geeft de hypothesetoets grafisch weer.

3.3 Classificatie van hypothesetoetsen

3.3.1 Inleiding

Voor het oplossen van een statistisch probleem, moet men op zoek gaan naar een gepaste statistische techniek. De keuze van de geschikte statistische toets is een probleem op zich. Als leidraad kan men opteren voor de toets met het grootste onderscheidingsvermogen, die met de kleinste kans op een type I fout,...
Elke toets heeft een bepaald onderscheidingsvermogen. Indien het onderscheidingsvermogen van een toets doorslaggevend is, dan wordt de toets met het grootste *onderscheidingsvermogen* gekozen, t.t.z. de toets die resulteert in een kleine kans om H_0 te verwerpen indien H_0 waar is en in een grote kans om H_0 te verwerpen indien H_0 vals is.
Het onderscheidingsvermogen is de leidraad van de classificatie die hiernavolgend wordt voorgesteld. De keuze van de geschikte toets wordt hierbij bepaald op basis van twee criteria: het aantal steekproeven waarop de toets betrekking heeft, en de *meetschaal* van de waarnemingen van de toets.

3.3.2 Meetschalen

Het belangrijkste classificatiecriterium voor hypothesetoetsen is de meetschaal. Het definieren van meetschalen impliceert het gebruik van *metingen.*

Definitie 6 *Metingen houden verband met het toekennen van symbolen (niet enkel getallen) aan bepaalde objecten of gebeurtenissen aan de hand van bepaalde regels.*

Afhankelijk van de regels die worden gebruikt om symbolen toe te kennen aan objecten ontstaan er verschillende *meetschalen.* Vier categorieën meetschalen worden onderscheiden: nominaal, ordinaal, interval en ratio. De nominale en de ordinale meetschaal vertegenwoordigen de categorische gegevens, terwijl de interval en de ratio schaal metrische gegevens representeren.

1. **Nominale schaal.** Indien getallen of symbolen enkel gebruikt worden om objecten te classificeren, dan spreekt men van een nominale schaal. Dit is de zwakste schaal. Getallen of namen worden enkel gebruikt om de verschillende objecten te onderscheiden. Bv.: de variabele die de kleur van ogen vertegenwoordigt in een vragenlijst heeft de volgende uitkomsten: (1) blauw, (2) bruin, (3) groen, (4) andere. Men kan hierbij niet zeggen dat blauwe ogen beter zijn dan bruine. Alle uitkomsten zijn ekwivalent.

2. **Ordinale schaal.** Indien de objecten niet enkel meer geclassificeerd worden, maar bovendien ook nog in relatie staan tot elkaar, dan spreekt men van een ordinale schaal. Typische relaties tussen de objecten zijn: "groter dan", "moeilijker dan", "beter geschikt dan",... Het is echter onmogelijk te zeggen hoe groot het verschil is tussen de onderscheiden, geordende klassen.
   Bv.: de variabele die de smaak van een frisdrank representeert heeft de volgende realisaties: (1) zeer slecht, (2) slecht, (3) onverschillig, (4) goed, (5) zeer goed. Men kan dus zeggen dat "zeer goed" beter is dan "goed", zonder daarbij te kunnen zeggen hoeveel beter.

3. **Interval schaal.** Deze schaal omvat alle karakteristieken van de ordinale schaal met het bijkomende kenmerk dat de afstand tussen de categorieën kwantificeerbaar is. D.w.z.

dat de intervallen tussen de klassen meetbaar zijn. Een andere karakteristiek van een interval schaal is dat het nulpunt arbitrair gekozen is, en geen absoluut nulpunt voorstelt. Bv.: metingen van de temperatuur in graden Celsius: 3,7,5.3,6.7,8.5,4.5,... Het verschil tussen 3 en 5 is even groot als dat tussen 7 en 9.

4. **Ratio schaal.** Een ratio schaal kan beschouwd worden als een interval schaal met een absoluut nulpunt. Gewicht, lengte, temperatuur in graden Kelvin zijn voorbeelden van een ratio schaal. Een persoon met lengte 0 cm. bestaat niet...

Het dient wel opgemerkt te worden dat de meetschaal in sommige gevallen niet enkel bepaald wordt op basis van de intrinsieke kenmerken, maar ook op basis van subjectieve criteria, afhankelijk van de context. Bv. In een studie wordt gevraagd naar het onderwijsniveau: "humaniora", "hoger onderwijs", "universiteit". Indien het een studie betreft inzake de relevantie van een vooropleiding, dan kan het onderwijsniveau als ordinaal geschaald worden beschouwd, aangezien individuen moeten kunnen worden gerangschikt op basis van hun opleiding. In een andere studie is het voldoende het onderwijsniveau te beschouwen als nominaal geschaald.

Het spreekt voor zich dat de keuze van de gepaste statistische techniek om een probleem op te lossen afhankelijk is van de schaal van de variabele(n) betrokken bij het probleem. De interval en de ratio schalen zijn de sterkste schalen. De statistische toetsen die daarop van toepassing zijn hebben ook het grootste onderscheidingsvermogen en zijn vooral parametrische toetsen.

Een *parametrische toets* is een toets waarvan het model aan een aantal voorwaarden moet voldaan m.b.t. de parameters van de populatie van waaruit de steekproef werd getrokken. De belangrijkste en de sterkste (grootste onderscheidingsvermogen) parametrische toetsen zijn de Z, t en F toetsen, die reeds in het vorige hoofdstuk werden ingeleid. De meeste van de *voorwaarden* die voldaan moeten zijn om valide toetsresultaten te verkrijgen, kennen we reeds uit het vorige hoofdstuk:

1. De waarnemingen van de steekproef werden *onafhankelijk* van elkaar verkregen, op basis van een aselecte steekproef.

2. De waarnemingen werden getrokken uit een *normaal verdeelde populatie* (enkel voor t en F toetsen). Bij grote steekproeven kan deze voorwaarden afgezwakt worden ten gevolge van de centrale limietstelling (Z-toets).

3. In het geval van meerdere steekproeven, dienen de oorspronkelijke populaties *homoscedastisch* (zelfde variantie) te zijn (enkel voor t en F toetsen).

4. De betrokken variabelen moeten minstens *interval geschaald* zijn.

Het probleem ontstaat echter wanneer aan deze veronderstellingen duidelijk niet wordt voldaan. In voorkomend geval verliest de parametrische toets een onberekenbaar deel van zijn onderscheidingsvermogen. Wellicht is het gebruik van een niet-parametrische toets dan aangewezen.

Een *niet-parametrische toets* is een toets waarvan het model geen voorwaarden oplegt aan de parameters van de populatie van waaruit de steekproef getrokken werd. De beperkte en weinig stringente *voorwaarden* waaraan de meeste niet-parametrische toetsen moeten voldoen, zijn:

Meetschaal	*Geschikte toets*
Nominaal	Niet-parametrisch
Ordinaal	Niet-parametrisch
Interval	Parametrisch + Niet-parametrisch
Ratio	Parametrisch + Niet-parametrisch

Tabel 3.4: De 4 meetniveau's en hun geschikte toets

1. De waarnemingen van de steekproef werden *onafhankelijk* van elkaar verkregen, op basis van een aselecte steekproef.
2. De betrokken variabelen dienen een onderliggende continuiteit te bezitten.

De meeste niet-parametrische toetsen vereisen geen interval meetschaal; meestal is een ordinale of soms zelfs een nominale meetschaal reeds voldoende.
Een parametrische toets heeft een sterker onderscheidingsvermogen dan een niet-parametrische indien aan alle voorwaarden van de parametrische toets werd voldaan. Nochtans kan voor eenzelfde probleem het onderscheidingsvermogen van een niet-parametrische toets totdat van een parametrische toets verhoogd worden door de *steekproefomvang* te vergroten.
Een aantal *voordelen* van niet-parametrische toetsen.

- Probabiliteiten die resulteren van niet-parametrische toetsen zijn meestal *exact*, onafhankelijk van de populatieverdeling van waaruit de aselecte steekproef getrokken werd.
- Voor kleine steekproeven uit *onbekende populatieverdelingen* zijn enkel niet-parametrische toetsen van toepassing.
- Niet-parametrische toetsen bieden de mogelijkheid om toetsen uit te voeren op steekproeven uit *verschillende populaties*, zonder dat daarbij soms onrealistische veronderstellingen (zoals homoscedasticiteit) dienen gemaakt te worden.
- Niet-parametrische toetsen kunnen toegepast worden op ordinale of zelfs nominale (*categorische*) gegevens.
- Niet parametrische toetsen zijn *gemakkelijk* aan te leren en toe te passen (quick and dirty).

De belangrijkste *nadelen* van niet-parametrische toetsen.

- Indien aan alle voorwaarden voor een parametrische toets voldaan zijn, heeft een niet-parametrische toets een lager onderscheidingsvermogen.
- In geval van grote steekproeven kunnen berekeningen vrij complex worden.

Niet-parametrische toetsen kunnen dus vooral gebruikt worden in het geval van nominaal en ordinaal geschaalde variabelen, maar eveneens indien de variabelen interval geschaald zijn, maar niet of slechts gedeeltelijk voldoen aan de voorwaarden van de parametrische toetsen (zie tabel 3.4).

3.3.3 Aantal steekproeven

Het tweede classificatiecriterium voor hypothesetoetsen houdt verband met het aantal steekproeven. Met betrekking tot dit criterium kunnen de volgende 5 gevallen onderscheiden worden:

1. **1 steekproef.** In dit geval wordt 1 aselecte steekproef uit de populatie getrokken. Hierbij zijn mogelijke toetsen: toets op verschil tussen steekproef- en populatiegemiddelde, tussen steekproef- en populatieproportie, tussen steekproef- en populatievariantie, tussen geobserveerde en verwachte frekwenties; toets om na te gaan of een steekproef afkomstig is van een welbepaalde distributie; toets om te zien of een steekproef werkelijk toevallig is.

2. **2 verwante steekproeven.** Het geval van twee verwante steekproeven wordt vooral gebruikt indien men het effect van een "behandeling" (cursus, geneesmiddel, reclame,...) wil toetsen. Belangrijk hierbij is de "matching" die bestaat tussen elk paar objecten van elke populatie. De hypothesetoetsen voor 2 verwante steekproeven zijn verwant met die voor 1 steekproef, aangezien beide verwante steekproeven gereduceerd kunnen worden tot 1 enkele steekproef van gepaarde verschillen. Nietttemin zijn er naast deze conceptuele gelijkenis nog voldoende verschillen om een scheiding te verantwoorden. Beide steekproeven moeten even groot zijn.

3. **2 onafhankelijke steekproeven.** Twee onafhankelijke steekproeven kunnen op 2 manieren verkregen worden: (1) als aselecte steekproeven uit 2 populaties, of (2) door het toevallig toedienen van 2 behandelingen aan de objecten van eenzelfde steekproef, waarvan de oorsprong willekeurig is. Bijgevolg hoeven beide steekproeven niet even groot te zijn. De meeste toetsen met 2 onafhankelijke steekproeven zijn gericht op het evalueren van significante verschillen op basis van de 2 gemiddelden, de 2 proporties of de 2 varianties

4. **k verwante steekproeven.** In geval we drie of meer verwante steekproeven wensen te toetsen op significante verschillen, dienen we te beschikken over hypothesetoetsen die de omnibus nulhypothese toetsen, t.t.z. die globaal vaststelt of er al dan niet significante verschillen bestaan tussen de steekproeven. Enkel indien de globale toets een verwerping van nulhypothese geeft, dan kunnen de verschillen tussen elk paar steekproeven door toetsen voor 2 verwante steekproeven gelokaliseerd worden. Hierbij dient men wel voor ogen te houden dat de kans op een type I fout stijgt bij de paarsgewijze hypothesetoetsen.

5. **k onafhankelijke steekproeven.** De drie of meer onafhankelijke steekproeven dienen niet noodzakelijk meer dezelfde lengte te hebben. Ook hier kan een toets gebruikt worden om de globale nulhypothese te toetsen en kunnen de toetsen voor 2 onafhankelijke steekproeven verder gebruikt worden voor 2 aan 2 vergelijkingen om de significante verschillen te lokaliseren.

In wat volgt, wordt een overzicht gegeven van de hypothesetoetsen, opgedeeld per aantal steekproeven en daarbinnen per meetniveau. Het overzicht is zeker niet exhaustief, maar bevat vast en zeker de meest gebruikelijke hypothesetoetsen. Tabel 3.5 bevat het volledige overzicht van de behandelde hypothesetoetsen.

	Meetschaal		
Steekproeven	**Nominaal**	**Ordinaal**	**Interval**
1 steekproef	Z toets voor proportie (3.4.6) χ^2 1 steekproef toets (3.4.5)	Kolmogorov-Smirnov toets (3.4.3) 1 steekproef Runs toets (3.4.4)	Z of t toets voor gemiddelde (3.4.1) χ^2 toets op variantie (3.4.2)
2 verwante steekproeven	McNemar toets (3.5.4)	Teken toets (3.5.2) Wilcoxon rangtekentoets (3.5.3)	t-toets voor gepaarde waarnemingen (3.5.1)
2 onafhankelijke steekproeven	χ^2 2 steekproeven toets (3.6.4) Z toets op verschil in proporties (3.6.5)	Mann-Whitney toets (3.6.3)	Z of t toets voor verschil van gemiddelden (3.6.1) F toets voor verschil van varianties (3.6.2)
k verwante steekproeven	Cochran Q toets (3.7.3)	Friedmann toets (3.7.2)	F toets (3.7.1)
k onafhankelijke steekproeven	χ^2 toets (3.8.3)	Kruskal-Wallis toets (3.8.2)	F toets (3.8.1)

Tabel 3.5: Classificatie van hypothesetoetsen.

3.4 Eén steekproef

3.4.1 Interval: Z of t toets voor gemiddelde

Doel. Toetsen van de significantie van het verschil tussen een hypothetisch populatiegemiddelde μ_0 en het populatiegemiddelde μ op basis van het steekproefgemiddelde $\overline{X}$ (toets op centrale tendens).

Voorwaarden.

- Populatie is normaal verdeeld. Indien de steekproefomvang voldoende groot is ($n > 30$), dan geldt de centrale limietstelling, waardoor normaliteit van de oorspronkelijke populatie niet meer vereist is.
- Minstens interval schaal gegevens.

Methode. Een steekproef met omvang n en gemiddelde $\overline{X}$ wordt aselect getrokken uit een populatie met hypothetisch gemiddelde μ_0. De hypothesen zijn in geval van een éénzijdige toets:

$$H_0 : \mu \geq \mu_0 \qquad H_0 : \mu \leq \mu_0$$
$$H_1 : \mu < \mu_0 \qquad H_1 : \mu > \mu_0$$

en in geval van een tweezijdige toets:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

De toetsstatistiek is afhankelijk van het geval:

1. *Normaal verdeelde populatie met bekende variantie.* Toetsstatistiek Z:
$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$
2. *Normaal verdeelde populatie met onbekende variantie.* Toetsstatistiek t met $n-1$ vrijheidsgraden:
$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$
3. *Niet-normaal verdeelde populatie.* Toetsstatistiek Z (centrale limietstelling):
$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$
of
$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

Indien $n/N > 0.05$ dient een eindige populatiecorrectie te worden toegepast.

Voorbeeld. Een bedrijf vervaardigt TV beeldbuizen die een gemiddelde levensduurte hebben van 1200 u. met een standaarddeviatie van 300 u. Een nieuw productieproces werd uitgetest op 100 beeldbuizen met een resulterende gemiddelde leeftijd van 1265 u. Is dit nieuw proces toevallig of significant beter dan het oude? Zijn de beide processen significant verschillend?

1. *Hypothesen*:

$$H_0 : \mu \leq 1200 \quad H_0 : \mu = 1200$$
$$H_1 : \mu > 1200 \quad H_1 : \mu \neq 1200$$

2. *Toetsstatistiek*: centrale limietstelling ($n > 30$) in beide gevallen

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*:
significant verschillend (2-zijdig): als $z \leq -1.96$ en $z \geq 1.96$ dan H_0 verwerpen.
significant beter (1-zijdig): $z \geq 1.645$ dan H_0 verwerpen.

5. *Berekening*:

$$z = \frac{1265 - 1200}{300/\sqrt{100}} = 2.17$$

6. *Beslissing*:
significant verschillend: $2.17 > 1.96 \rightarrow H_0$ verwerpen.
significant beter: $2.17 > 1.645 \rightarrow H_0$ verwerpen.

7. *Conclusie*:
Er is een significant verschil tussen de beide processen; $p = 2 \times 0.015 = 0.030$.
Nieuw proces is significant beter dan oude; $p = 0.015$.

Voorbeeld. Een autobandenfabrikant beweert dat een autoband gemiddeld meer dan 25000 km. kan meegaan. Een toevallige steekproef van 15 banden levert een gemiddelde op van 27000 km. met een standaarddeviatie van 5000 km. Veronderstel dat de levensduur van een band benaderend normaal verdeeld is. Is de bewering van de fabrikant juist? Wat als de fabrikant beweert dat de levensduur gelijk is aan 25000 km.

1. *Hypothesen*:

$$H_0 : \mu \leq 25000 \quad H_0 : \mu = 25000$$
$$H_1 : \mu > 25000 \quad H_1 : \mu \neq 25000$$

2. *Toetsstatistiek*:

$$t = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} \qquad 14 \text{ v.g.}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*:
   significant verschillend (2-zijdig): als $t \leq -2.15$ en $t \geq 2.15$ dan H_0 verwerpen.
   significant beter (1-zijdig): $t \geq 1.76$ dan H_0 verwerpen.

5. *Berekening*:
$$t = \frac{27000 - 25000}{5000/\sqrt{15}} = 1.55$$

6. *Beslissing*:
   significant verschillend: $1.55 < 2.15 \rightarrow H_0$ niet verwerpen.
   significant beter: $1.55 < 1.76 \rightarrow H_0$ niet verwerpen.

7. *Conclusie*:
   Levensduur van autoband is niet significant meer dan 25000 km.;
   $0.05 < p < 0.1$.
   Levensduur van autoband is niet significant verschillend van 25000 km.;
   $2 \times (0.05 < p < 0.1) = 0.10 < p < 0.20$.

3.4.2 Interval: χ^2 toets voor variantie

Doel. Toetsen van het verschil tussen de populatievariantie σ^2 en een hypothetische populatievariantie σ_0^2 op basis van de steekproefvariantie S^2.

Voorwaarden.

- Normaal verdeelde populatie.
- Minstens interval gegevens.

Methode. De hypothesen zijn

$$\begin{array}{ll} H_0 : \sigma^2 \geq \sigma_0^2 & H_0 : \sigma^2 \leq \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 & H_1 : \sigma^2 > \sigma_0^2 \end{array}$$

voor een eenzijdige toets, en

$$\begin{array}{l} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{array}$$

voor een tweezijdige toets.
De steekproefvariantie S^2 wordt berekend uit de steekproef met omvang n.

$$S^2 = \frac{\sum (X_i - \overline{X})^2}{n-1}$$

De toetsstatistiek is de χ^2 statistiek met $n-1$ vrijheidsgraden:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

Voorbeeld. Een luchtvaartbedrijf bouwt hoogtemeters met fouten van gemiddelde 0 m. en een s.d. van 43.7 m. Uit een steekproef van 26 hoogtemeters wordt een s.d. van 54.7 m. gevonden. Is deze gevonden waarde significant verschillend van 43.7 m.? Veronderstel dat de foutenverdeling normaal is.

1. *Hypothesen*:
$$H_0 : \sigma = 43.7$$
$$H_1 : \sigma \neq 43.7$$

2. *Toetsstatistiek*: χ^2 statistiek met $n - 1 = 25$ vrijheidsgraden:
$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $\chi^2 \leq 13.12(\alpha/2 = 0.025)$ of $\chi^2 \geq 40.65(1 - \alpha/2 = 0.975)$

5. *Berekening*:
$$\chi^2 = \frac{(26-1)54.7^2}{43.7^2} = 39.17$$

6. *Beslissing*: $13.12 < 39.17 < 40.65 \rightarrow H_0$ niet verworpen.

7. *Conclusie*: De werkelijke standaarddeviatie is niet significant verschillend van 43.7 m. $2(0.025 < p < 0.05)$ of $0.05 < p < 0.1$.

3.4.3 Ordinaal: Kolmogorov-Smirnov 1 steekproef toets

Doel. Toets op het significant verschil tussen de geobserveerde verdeling van een verzameling steekproefwaarden en een bepaalde (theoretische) verdeling (toets voor een verdelingsfunctie).

Voorwaarden.

- Minstens ordinale gegevens.

Methode. Enerzijds wordt op basis van de steekproef de cumulatieve verdeling $S_X(x)$ opgezet. Anderzijds wordt de cumulatieve verdeling $F_X(x)$ bepaald, d.i. de cumulatieve verdelingsfunctie die verwacht wordt onder de nulhypothese. De hypothesen hierbij zijn:

H_0 : geen significant verschil tussen geobserveerde en theoretische verdeling

H_1 : significant verschil tussen geobserveerde en theoretische verdeling

Elke categorie van $S_X(x)$ wordt gepaard met de corresponderende categorie van $F_X(x)$. Vervolgens wordt het maximale verschil D bepaald tussen de beide cumulatieve verdelingen.

$$D = \max |F_X(x) - S_X(x)|$$

De steekproevenverdeling van D is bekend onder H_0. Een tabel met de kritieke D-waarden voor een tweezijdige toets is beschikbaar. Indien $D \geq$ kritieke waarde, dan wordt H_0 verworpen.

Voorbeeld. De volgende reeks is een steekproef van het aantal aankomsten per uur van klanten aan een bankfiliaal.

$$6, 6, 6, 3, 5, 6, 9, 12, 10, 10, 8, 6, 9, 4, 12, 5, 5, 6, 6, 11$$

Kan men stellen dat deze steekproef afkomstig is van een Poisson verdeling?

1. *Hypothesen*:

   H_0 : geen significant verschil tussen geobserveerde en Poisson verdeling
   H_1 : significant verschil tussen geobserveerde en Poisson verdeling

2. *Toetsstatistiek*: Kolmogorov-Smirnov D-statistiek:

$$D = \max |F_X(x) - S_X(x)|$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: als $D \geq 0.294$ dan H_0 verwerpen.

5. *Berekening*: voor steekproef van n =20: $\overline{x} = 7.25$.
   Berekening van $F_X(x)$ uit cumulatie van

$$f_X(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 3, 4, ..., 12. \qquad \lambda = 7.25$$

x	n_i	$s_X(x)$	$f_X(x)$	$S_X(x)$	$F_X(x)$
3	1	0.05	0.05	0.05	0.05
4	1	0.05	0.08	0.10	0.13
5	3	0.15	0.12	0.25	0.25
6	7	0.35	0.14	0.60	0.39 (*)
7	0	0.00	0.15	0.60	0.54
8	1	0.05	0.13	0.65	0.67
9	2	0.10	0.11	0.75	0.78
10	2	0.10	0.08	0.85	0.86
11	1	0.05	0.05	0.90	0.91
12	2	0.10	0.03	1.00	0.94

$$D = 0.21$$

6. *Beslissing*: $0.21 < 0.294 \rightarrow H_0$ niet verwerpen.

7. *Conclusie*: De verdeling van de steekproefgegevens is niet significant verschillend van een Poisson verdeling. $p > 0.20$.

3.4.4 Ordinaal: 1 steekproef Runs toets

Doel. Toetsen van een reeks van getallen op toeval.

Voorwaarden.

- Alle waarnemingen van de reeks getallen (steekproef) moeten in dezelfde omstandigheden verkregen zijn.
- De waarnemingen moeten geordend kunnen worden (ordinale meetschaal).

Methode. Een *run* kan gedefinieerd worden als een sekwentie van gelijke symbolen voorafgegaan en gevolgd door een ander of geen symbool. De toets vereist het gebruik van 2 verschillende symbolen (A en B, $+$ en $-$,...). Sommige steekproeven zijn reeds uitgedrukt in de 2 symbolen. Indien het getallen betreft, dan gaat men bv. een $+$ toekennen aan waarden boven het gemiddelde (of mediaan) en een $-$ aan de waarden daaronder. Aldus wordt de reeks getallen uitgedrukt in 2 symbolen. De hypothesen zijn:

H_0 : de 2 symbolen staan in een toevallige volgorde
H_1 : de 2 symbolen staan niet in een toevallige volgorde

Voor $n_1, n_2 \leq 20$ zijn tabellen beschikbaar met kritieke onder- en bovenwaarden voor het aantal runs R. Voor $n_1 > 20$ of $n_2 > 20$ is de steekproevenverdeling van R benaderend normaal verdeeld.

$$R \sim N(\mu_R, \sigma_R)$$

$$\mu_R = \frac{2n_1n_2}{n_1 + n_2} + 1$$

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}$$

De Z-statistiek is dan de toetsstatistiek:

$$Z = \frac{R - \mu_R}{\sigma_R}$$

Voorbeeld. De kwaliteitscontrole wil nagaan of de variaties in het over- en ondergewicht van een product dat normaal 17 gr. moet wegen al of niet het gevolg zijn van het toeval. Een steekproef van 20 achtereenvolgende geproduceerde stukken wordt genomen:

17.9, 17.5, 17.2, 17.3, 16.5, 16.8, 16.7, 17.2, 17.4, 17.6,

17.5, 17.8, 16.8, 16.5, 16.6, 17.7, 17.6, 17.7, 17.8, 17.2

1. *Hypothesen*:

   H_0 : volgorde van over- en ondergewicht is toevallig
   H_1 : volgorde van over- en ondergewicht is niet toevallig

2. *Toetsstatistiek*: aantal runs R

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: $n_1 = 6, n_2 = 14$; als $R \leq 5$ dan H_0 verwerpen.
5. *Berekening*: R=5
6. *Beslissing*: $5 \leq 5 \rightarrow H_0$ verwerpen.
7. *Conclusie*: geen toevallige volgorde van over- en ondergewicht.

3.4.5 Nominaal: χ^2 toets voor 1 steekproef

Doel. Toets op significant verschil tussen geobserveerde gegevens opgedeeld in k klassen en de theoretische verwachte frekwenties in de k klassen (toets voor verdelingsfunctie).

Voorwaarden.

- De geobserveerde en theoretische verdelingen bevatten een gelijk aantal elementen.
- Dezelfde opdeling in klassen geldt voor de geobserveerde en de verwachte verdeling.
- Minder dan 20% van de cellen hebben een verwachte frekwentie kleiner dan 5 en geen enkele cel heeft een verwachte frekwentie minder dan 1. In voorkomend geval moeten cellen gecombineerd worden.
- De geobserveerde frekwenties zijn verkregen geworden door aselecte steekproefname.

Methode. De aard van de gegevens is nominaal, t.t.z. elke klasse bevat tellingen van objecten met een bepaalde karakteristiek. Er moeten minstens 2 klassen zijn. De bedoeling is om te testen of er een significant verschil bestaat tussen het geobserveerd aantal objecten in elke klasse en het verwachte aantal onder H_0. De hypothesen zijn:

H_0 : Er is geen significant verschil tussen de geobserveerde en de verwachte frekwenties.
H_1 : Er is een significant verschil tussen de geobserveerde en de verwachte frekwenties.

De toetsstatistiek is de χ^2 statistiek:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Hierbij is O_i de geobserveerde en E_i de verwachte frekwentie voor klasse i. Indien de verschillen tussen de O_i's en E_i's klein zijn, dan zal de resulterende χ^2 klein zijn en zal H_0 niet verworpen worden. Indien grote verschillen resulteren in een χ^2 die groter of gelijk is aan de kritieke χ^2 waarde van de tabel, dan wordt H_0 verworpen.
De steekproevenverdeling van χ^2 onder H_0 is de χ^2 verdeling met $k-1$ vrijheidsgraden. Indien voor het opstellen van de theoretische verdeling er bijkomend nog m parameters uit de steekproefgegevens moeten worden geschat, dan daalt het aantal vrijheidsgraden tot $k-1-m$. Deze χ^2 geschikheidstoets is duidelijk minder krachtig dan de Kolmogorov-Smirnov geschiktheidstoets (cfr. § 3.4.3). De Kolmogorov-Smirnov toets gebruikt meer informatie die in de gegevens vervat zit, namelijk de volgorde, wat door de χ^2 niet gebruikt wordt.

Voorbeeld. Aan een steekproef van 60 personen wordt gevraagd naar het TV-station van hun voorkeur. De volgende resultaten werden opgetekend:

Station	1	2	3	4	5	6
Aantal	5	8	10	12	12	13

Zijn er TV-stations die significant meer geprefereerd worden?

1. *Hypothesen*:

   H_0 : De 6 TV-stations worden even graag bekeken (uniforme verdeling)
   H_1 : De 6 TV-stations worden niet even graag bekeken

2. *Toetsstatistiek*:
$$\chi^2 = \sum_{i=1}^{6} \frac{(O_i - E_i)^2}{E_i}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: 6-1=5 vrijheidsgraden, $\alpha = 0.05$; H_0 verwerpen als $\chi^2 \geq 11.07$

5. *Berekening*:
$$\chi^2 = \frac{(5-10)^2}{10} + \frac{(8-10)^2}{10} + \frac{(10-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(12-10)^2}{10} + \frac{(13-10)^2}{10} = 4.6$$

6. *Beslissing*: $4.6 < 11.07 \rightarrow H_0$ niet verworpen.

7. *Conclusie*: De preferenties zijn uniform verdeeld over de 6 TV-stations. $p > 0.05$.

Voorbeeld. De χ^2 geschikheidstoets kan uitgevoerd worden voor elke verdeling. Als voorbeeld wordt nagegaan of de jobtijd op basis van de volgende steekproef van 700 personen een benaderend normaal verdeling volgt.

Tijd	*Aantal werknemers*
< 10	38
10-19.99	51
20-29.99	62
30-39.99	74
40-49.99	83
50-59.99	91
60-69.99	81
70-79.99	72
80-89.99	61
90-99.99	52
≥ 100	35
Totaal:	700

1. *Hypothesen*:

   H_0 : De steekproefgegevens komen uit een normaal verdeelde populatie.

   H_1 : De steekproefgegevens komen niet uit een normaal verdeelde populatie.

2. *Toetsstatistiek*:
$$\chi^2 = \sum_{i=1}^{11} \frac{(O_i - E_i)^2}{E_i}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: vrijheidsgraden: $11 - 1 - 2 = 8$ (2 parameters schatten uit steekproefgegevens: $\overline{x} = 54.71$ en $s = 27.61$); $\alpha = 0.05$.
   H_0 verwerpen als $\chi^2 \geq 15.51$

5. *Berekening*:

Tijd	O_i	Z	*Kans*	E_i	$\frac{(O_i-E_i)^2}{E_i}$
< 10	38	-	0.0526	36.82	0.0378
10-19.99	51	-1.62	0.0512	35.84	6.4125
20-29.99	62	-1.26	0.0829	58.03	0.2716
30-39.99	74	-0.89	0.1114	77.98	0.2031
40-49.99	83	-0.53	0.1344	94.08	1.3049
50-59.99	91	-0.17	0.1428	99.96	0.8031
60-69.99	81	0.19	0.1335	93.45	1.6587
70-79.99	72	0.55	0.1124	78.68	0.5671
80-89.99	61	0.92	0.0785	54.95	0.6661
90-99.99	52	1.28	0.0498	34.86	8.4274
≥ 100	35	1.64	0.0505	35.35	0.0035
			1.000	700	20.3558

   De verwachte frekwenties E_i worden verkregen door de verwachte probabiliteit van klasse i te vermenigvuldigen met 700. De verwachte probabiliteit (kans) wordt verkregen door de probabiliteit onder de normaalcurve te bepalen uit de tabel, overeenkomend met de Z-waarde:
$$Z = \frac{X - \overline{X}}{S}$$
   Hierbij is X de ondergrens van de klasse. De eerste klasse heeft geen ondergrens en bijgevolg is er geen Z-waarde. De kans 0.0526 wordt verkregen door de kans te bepalen onder de standaardnormaal curve tussen $z = -\infty$ en $z = (10-54.71)/27.61 = -1.62$. Zo ook is de kans 0.0512 het verschil tussen kans 0.1038 ($z = (20 - 54.71)/27.61 = -1.26$) en 0.0526 ($z = (10 - 54.71)/27.61 = -1.62$). Aldus wordt $0.0512 \times 700 = 35.84$. Door dit te doen voor elke klasse wordt $\chi^2 = 20.3558$.

6. *Beslissing*: $20.3558 > 15.51 \rightarrow H_0$ verworpen.

7. *Conclusie*: De steekproefgegevens volgen geen normaal verdeling. $p < 0.01$.

3.4.6 Nominaal: Z toets voor proportie

Doel. Toetsen van significantie van het verschil tussen een hypothetische proportie π_0 en populatieproportie π op basis van een geobserveerde proportie P.

Voorwaarden.

- Populatie bestaat uit 2 klassen.
- $np > 5, n(1-p) > 5$ voor normale benardering van de binomiaal verdeling.

Methode. Voor zeer kleine steekproeven ($n \leq 20$) kan de binomiaal verdeling gebruikt worden. Uit de tabel van de binomiaal verdeling kan dan rechtstreeks de p-waarde worden afgelezen.
Vanaf $np > 5$ en $n(1-p) > 5$ kan de normaal benadering toegepast worden met de Z-toetsstatistiek:

$$Z = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

Indien $n/N > 0.05$ dient een eindige populatiecorrectie te worden toegepast.

Voorbeeld. Een elektrische hersteldienst beweert dat 10% van haar herstellingen van huishoudtoestellen het gevolg zijn van het niet correct aansluiten van een toestel. Op basis van een onderzoek van 200 herstellingen vond het bedrijf 15 herstellingen t.g.v. het niet correct aansluiten. Tonen deze resultaten aan dat de bewering van de hersteldienst correct is? $\alpha = 1\%$.

1. *Hypothesen*:

$$H_0 : \pi = 0.10$$
$$H_1 : \pi \neq 0.10$$

2. *Toetsstatistiek*:

$$Z = \frac{P - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

3. *Significantieniveau*: $\alpha = 0.01$.

4. *Beslissingsregel*: H_0 verwerpen als $Z \leq -2.58$ en $Z \geq 2.58$.

5. *Berekening*:

$$z = \frac{0.075 - 0.10}{\sqrt{\frac{0.10(1-0.10)}{200}}} = -1.18$$

6. *Beslissing*: $-2.57 \leq -1.18 \leq 2.57 \rightarrow H_0$ niet verwerpen.

7. *Conclusie*: De resultaten tonen niet aan dat de bewering incorrect is. $p = 0.238$.

3.5 Twee verwante steekproeven

Veel van de toetsen voor twee verwante steekproeven zijn eveneens toetsen voor het 1 steekproef geval, aangezien in vele toetsen de gepaarde steekproeven worden teruggebracht tot 1 steekproef van de verschillen.

3.5.1 Interval: t toets op gepaarde waarnemingen

Doel. Toetsen van de significantie van het verschil tussen 2 populatiegemiddelden μ_1 en μ_2 op basis van 2 steekproeven.

Voorwaarden.

- De waarnemingen van de 2 steekproeven zijn beschikbaar in paren.
- De beide oorspronkelijke populaties zijn normaal.
- Minstens interval gegevens.
- Er zijn geen veronderstellingen m.b.t. de varianties van beide populaties vereist.

Methode. Voor elk paar waarnemingen i wordt het verschil D berekend.

$$D_i = X_{1i} - X_{2i}$$

Aldus wordt er één steekproef met de verschillen D verkregen met gemiddelde en standaarddeviatie:

$$\overline{D} = \frac{\sum_i D_i}{n}$$

$$S_D = \sqrt{\frac{\sum(D_i - \overline{D})^2}{n-1}}$$

De toetsstatistiek wordt dan de t-statistiek met $n-1$ vrijheidsgraden:

$$t = \frac{\overline{D} - \mu_{\overline{D}_0}}{S_D/\sqrt{n}}$$

In het zeldzaam geval dat de variantie σ_D bekend zou zijn, is het gebruik van de Z-toetsstatistiek aangewezen:

$$Z = \frac{\overline{D} - \mu_{\overline{D}_0}}{\sigma_D/\sqrt{n}}$$

Voorbeeld. Negen personen worden onderworpen aan 2 opleidingscursussen met het volgende resultaat.

Persoon	1	2	3	4	5	6	7	8	9
Cursus 1	90	95	87	85	90	94	85	88	92
Cursus 2	85	88	87	86	82	82	70	72	80
Verschil	5	7	0	-1	8	12	15	16	12

Is de uitslag van cursus 1 significant beter dan die van cursus 2 ?

1. *Hypothesen*:

$$H_0 : \mu_D \leq 0$$
$$H_1 : \mu_D > 0$$

2. *Toetsstatistiek*:

$$t = \frac{\overline{D} - \mu_{\overline{D}_0}}{S_D/\sqrt{n}}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: $n - 1 = 9 - 1 = 8$ vrijheidsgraden,
H_0 verwerpen als $t \geq 1.860$

5. *Berekening*:

$$\overline{D} = \frac{5 + 7 + ... + 12}{9} = 8.2$$

$$S_D = \sqrt{\frac{(5 - 8.2)^2 + (7 - 8.2)^2 + ... + (12 - 8.2)^2}{8}} = 6.12$$

$$t = \frac{8.2 - 0}{6.12/\sqrt{9}} = 4.02$$

6. *Beslissing*: $4.02 > 1.860 \rightarrow H_0$ verworpen.

7. *Conclusie*: Uitslag cursus 1 is significant beter dan uitslag cursus 2. $p < 0.005$

3.5.2 Ordinaal: Teken toets

Doel. Toetsen van de significantie van het verschil tussen de medianen van 2 verdelingen. Deze toets kan eveneens gebruikt worden voor het toetsen van de significantie van het verschil tussen een populatiemediaan en een hypothetische mediaanwaarde M_0 i.g.v. 1 steekproef.

Voorwaarden.

- Waarnemingen van de 2 steekproeven moeten gepaard zijn.
- Minstens ordinale gegevens.

Methode. De niet-parametrische tekentoets is een alternatief voor de parametrische t-toets voor gepaarde waarnemingen (cfr. § 3.5.1), indien de verdelingen van de 2 populaties verre van normaal zijn of indien de gegevens ordinaal geschaald zijn. Voor niet-symmetrische verdelingen is de mediaan een betere maat voor de centrale tendens dan het gemiddelde.
De tekentoets verleent zijn naam aan de "+" en "-" tekens die gebruikt worden om de verschillen tussen de paren aan te duiden. Indien een gepaard verschil gelijk aan 0 is, dan wordt deze niet mee beschouwd en dient de steekproefomvang hiervoor te worden aangepast.
De nulhypothese is dat de medianen van de beide populaties gelijk zijn, d.w.z.:

$$H_0 : P(X_i > Y_i) = P(X_i < Y_i) = \frac{1}{2}$$

Uitgedrukt in tekens geeft dit:

$$H_0 : P(+) = P(-) = \frac{1}{2}$$

De alternatieve hypothese is voor een tweezijdige toets is

$$H_1 : P(+) \neq P(-)$$

In voorkomend geval is de toetsstatistiek K het kleinste aantal tussen "+" en "-" tekens. De p-waarde wordt rechtstreeks gevonden in de Binomiaal tabel. Voor de tweezijdige toets dient de gevonden waarde verdubbeld te worden. Voor een eenzijdige toets zijn er 2 mogelijkheden. De eerste,

$$H_1 : P(+) < P(-)$$

heeft als toetsstatistiek K het aantal "+" tekens. De tweede,

$$H_1 : P(+) > P(-)$$

heeft als toetsstatistiek K het aantal "-" tekens.
Voor een effectieve steekproefomvang $n > 20$, kan de normaal benadering gebruikt worden. Volgens sommige auteurs is deze benadering zelfs toegestaan vanaf $n > 10$. De normaal verdeling heeft, met P=0.5 (onder H_0), als gemiddelde en standaarddeviatie:

$$\mu = nP = 0.5n$$

$$\sigma = \sqrt{nP(1-P)} = 0.5\sqrt{n}$$

en dus

$$Z = \frac{K - 0.5n}{0.5\sqrt{n}}$$

Aangezien het hier om een normale (continu) benadering van een binomiaal (discreet) verdeling gaat, is een continuiteitscorrectie aangewezen:

$$Z = \frac{(K \pm 0.5) - 0.5n}{0.5\sqrt{n}}$$

$$\begin{cases} K + 0.5 & \text{als } K < 0.5n \\ K - 0.5 & \text{als } K > 0.5n \end{cases}$$

Het is eveneens mogelijk deze toets uit te voeren om de significantie van het verschil tussen een populatiemediaan en een hypothetische mediaanwaarde M_0 te bepalen, d.w.z. het 1-steekproef geval. Hiervoor dient een "+" teken toegekend te worden indien de steekproefwaarneming groter is dan M_0 en een "-" teken anders.

Voorbeeld. 16 tennisspelers krijgen de opdracht om een nieuwe soort tennisbal te evalueren. Hiervoor wordt aan elke speler gevraagd om zijn voorkeur uit te spreken voor de oude of de nieuwe ballen. Het resultaat was: 1 onbeslist, 3 voor de oude en 12 voor de nieuwe ballen. Moet men overschakelen op de nieuwe ballen?

1. *Hypothesen*: voorkeur oude: "+"; voorkeur nieuwe "-"
$$H_0 : P(+) = P(-)$$
$$H_1 : P(+) < P(-)$$

2. *Toetsstatistiek*: K: aantal + tekens; of
$$Z = \frac{(K \pm 0.5) - 0.5n}{0.5\sqrt{n}}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $P[K \leq 3|n = 15; p = 0.5] \leq 0.05$ of H_0 verwerpen als $Z \leq -1.645$

5. *Berekening*:
$$P[K \leq 3|n = 15; p = 0.5] = 0.018$$
of met normaal benadering:
$$\mu = 0.5 \times 15 = 7.5$$
$$\sigma = 0.5\sqrt{15} = 1.94$$
$$Z = \frac{(3 + 0.5) - 7.5}{1.94} = -2.06$$

6. *Beslissing*: $0.018 < 0.05 \rightarrow H_0$ verworpen of $-2.06 < -1.645 \rightarrow H_0$ verworpen.

7. *Conclusie*: De nieuwe ballen zijn beduidend meer in trek. $p = 0.018$.

3.5.3 Ordinaal: Wilcoxon Rangtekentoets

Doel. Toetsen van de significantie van het verschil tussen de gemiddelden van 2 gelijkvormige verdelingen. Deze toets kan eveneens gebruikt worden voor het toetsen van de significantie van het verschil tussen een populatiegemiddelde μ en een hypothetische waarde μ_0.

Voorwaarden.

- De waarnemingen in de 2 steekproeven zijn gepaard.
- De steekproef van de verschillen is minstens interval geschaald.

Methode. Deze toets is een uitbreiding op de rangtekentoets, in zoverre dat niet enkel de richting van het verschil, maar ook de grootte-orde van het verschil wordt beschouwd.
Deze toets is een alternatief op de t-toets voor gepaarde waarnemingen (§ 3.5.1) indien aan een aantal veronderstellingen niet is voldaan (normaliteit, symmetrische verdeling,...). De hypothesen voor een eenzijdige en een tweezijdige toets zijn voor deze toets dezelfde als voor de voornoemde t-toets.

Door het gepaarde verschil tussen de beide steekproeven te nemen, worden de 2 steekproeven teruggebracht tot 1 steekproef van de verschillen:
$$D_i = X_{1i} - X_{2i}$$

De $|D_i|$'s worden vervolgens gerangschikt (zonder daarbij naar hun teken te kijken). Voor $|D_i|$'s met eenzelfde waarde worden gemiddelde rangscores toegekend. Daarna wordt het geschikte "+" of het "-" teken opnieuw toegekend aan de rangen. De toetsstatistiek T is de som van de "+" of "-" rangen, afhankelijk welk van de beide het kleinste is. De steekproefomvang n wordt aangepast door enkel de rangen voor verschillen verschillend van 0 te beschouwen.
Indien $n \leq 25$ kunnen de kritieke T-waarde afgelezen worden uit de tabel van de rangtekentoets. H_0 wordt verworpen als de berekende T-waarde kleiner of gelijk is aan de kritieke waarde. Voor $n > 25$ kan een normale benadering gebruikt worden met de Z-statistiek:

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Deze toets kan eveneens gebruikt worden voor het 1-steekproef geval, waarbij het een ordinaal alternatief vormt voor de t-toets voor een populatiegemiddelde (§ 3.4.1).

Voorbeeld. Een kinderpsycholoog wenst na te gaan of er een effect bestaat van de kinderopvang buitenshuis op het sociale gedrag van kinderen. Het sociaal gedrag wordt weergegeven als het resultaat van een aantal testen op 100. Hierbij kan men moeilijk stellen dat als een kind een score van 60 haalt, hij daarom tweemaal zo sociaal is als een kind met een score van 30. Wel kan gezegd worden dat het verschil tussen een score van 60 en 40 groter is dan het verschil tussen 40 en 30. Acht identieke tweelingen worden gebruikt als testsubject, waarvan er 1 thuis wordt opgevoed en 1 in de kinderopvang. De resultaten waren

Paar	*Score kinderopvang*	*Score thuis*	D	Rang van D
1	82	63	19	7
2	69	42	27	8
3	73	74	-1	-1
4	43	36	7	4.5
5	58	51	7	4.5
6	56	43	13	6
7	76	80	-4	-3
8	65	62	3	2

1. *Hypothesen*:

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

2. *Toetsstatistiek*: T_-

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $T_- \leq 4$

5. *Berekening*: T_-=4 (som van de "-"rangen; cfr. tabel Wilcoxon)

6. *Beslissing*: $4 \leq 4 \rightarrow H_0$ verworpen.

7. *Conclusie*: Er is een significante invloed van de kinderopvang op het sociaal gedrag. $0.02 < p < 0.05$

3.5.4 Nominaal: McNemar toets

Doel. Toetsen van significantie van de verschillen tussen 2 dichotome verdelingen.

Voorwaarden.

- De waarnemingen in de 2 steekproeven zijn gepaard.
- Nominale gegevens, meer specifiek dichotoom (0,1).

Methode. Deze toets wordt gebruikt om significante veranderingen vast te stellen van 0 naar 1 of van 1 naar 0. Dit is bijzonder nuttig in voor en na experimenten, waar men bijvoorbeeld wil vaststellen of een aantal testsubjecten van mening zijn veranderd na onderworpen te zijn geweest aan een behandeling. Voor deze toets kan een 2×2 contingentie tabel opgesteld worden.

	Na	
Voor	$Y_i = 1$	$Y_i = 0$
$X_i = 0$	A	B
$X_i = 1$	C	D

Met:

A : aantal subjecten met vooraf $X_i = 0$ en achteraf $Y_i = 1$;

B : aantal subjecten met vooraf $X_i = 0$ en achteraf $Y_i = 0$;

C : aantal subjecten met vooraf $X_i = 1$ en achteraf $Y_i = 1$;

D : aantal subjecten met vooraf $X_i = 1$ en achteraf $Y_i = 0$;

De hypothesen zijn dan:

$$H_0 : P(X_i = 0, Y_i = 1) = P(X_i = 1, Y_i = 0) \quad \forall i$$
$$H_1 : P(X_i = 0, Y_i = 1) \neq P(X_i = 1, Y_i = 0) \quad \forall i$$

De toetsstatistiek is de χ^2 toets met 1 vrijheidsgraad gedefinieerd als:

$$\chi^2 = \frac{(|A - D| - 1)^2}{A + D}$$

Voorbeeld. Twee kandidaten dingen naar het presidentschap. Voor en na een TV debat wordt aan 75 mensen naar de kandidaat van hun keuze gevraagd. De onderstaande tabel geeft het resultaat van die bevraging.

	Na	
Voor	Kandidaat 2	Kandidaat 1
Kandidaat 1	13	28
Kandidaat 2	27	7

Heeft het debat een significante invloed gehad op de keuze van de mensen?

1. *Hypothesen*:

$$H_0 : P(\text{Kandidaat 1} \rightarrow \text{Kandidaat 2}) = P(\text{Kandidaat 2} \rightarrow \text{Kandidaat 1})$$
$$H_1 : P(\text{Kandidaat 1} \rightarrow \text{Kandidaat 2}) \neq P(\text{Kandidaat 2} \rightarrow \text{Kandidaat 1})$$

2. *Toetsstatistiek*: χ^2 met 1 v.g.

$$\chi^2 = \frac{(|A - D| - 1)^2}{A + D}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $\chi^2 \geq 3.84$

5. *Berekening*:

$$\chi^2 = \frac{(|13 - 7| - 1)^2}{13 + 7} = 1.25$$

6. *Beslissing*: $1.25 < 3.84 \rightarrow H_0$ niet verworpen.

7. *Conclusie*: Er wordt geen significante verandering in de keuze van de mensen vastgesteld. $p > 0.05$.

3.6 Twee onafhankelijke steekproeven

3.6.1 Interval: Z of t toets op verschil van gemiddelden

Doel. Toetsen van de significantie van het verschil tussen 2 populatiegemiddelden, gebaseerd op 2 onafhankelijke steekproeven.

Voorwaarden.

- Normaliteit van de 2 oorspronkelijke populaties. In het geval van niet-normaal verdeelde populaties, is volgens de centrale limietstelling de steekproevenverdeling van $\overline{X_1} - \overline{X_2}$ eveneens normaal, i.g.v. voldoende grote steekproeven.
- Minstens interval gegevens.

Methode. De hypothesen voor een eenzijdige toets zijn:

$$H_0 : \mu_1 - \mu_2 \geq (\mu_1 - \mu_2)_0 \qquad H_0 : \mu_1 - \mu_2 \leq (\mu_1 - \mu_2)_0$$
$$H_1 : \mu_1 - \mu_2 < (\mu_1 - \mu_2)_0 \qquad H_1 : \mu_1 - \mu_2 > (\mu_1 - \mu_2)_0$$

en in geval van een tweezijdige toets:

$$H_0 : \mu_1 - \mu_2 = (\mu_1 - \mu_2)_0$$
$$H_1 : \mu_1 - \mu_2 \neq (\mu_1 - \mu_2)_0$$

In vele gevallen is $(\mu_1 - \mu_2)_0 = 0$.
De toetsstatistiek is afhankelijk van het geval:

1. *2 normaal verdeelde populaties, bekende populatievarianties.* Toetsstatistiek Z:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

2. *2 normaal verdeelde populaties, onbekende maar gelijke populatievarianties.* Toetsstatistiek t met $n_1 + n_2 - 2$ vrijheidsgraden:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

   met de gepoolde variantie:

$$s_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

3. *2 normaal verdeelde populaties, onbekende maar ongelijke populatievarianties.* Toetsstatistiek t'

$$t' = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

De kritieke t' waarde voor een tweezijdige toets wordt berekend uit:

$$t'_{1-\alpha/2} = \frac{\frac{s_1^2}{n_1} t_{(1-\alpha/2, n_1-1)} + \frac{s_2^2}{n_2} t_{(1-\alpha/2, n_2-1)}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Voor een eenzijdige toets wordt $t_{1-\alpha/2}$ hierin vervangen door $t_{1-\alpha}$.

4. *2 niet-normaal verdeelde populaties* (centrale limietstelling).
Toetsstatistiek Z:

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

of

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Voorbeeld. In een bedrijf is men geïnteresseerd in de tijd nodig om een bepaalde job uit te voeren. Hiervoor wordt een steekproef van 10 arbeiders van shift 1 en 8 arbeiders van shift 2 genomen met de volgende resultaten. De gemiddelde tijd voor de arbeiders van shift 1 is 26.1 sec. met een variantie van 144 sec.^{2}.De gemiddelde tijd voor de arbeiders van shift 2 is 17.6 sec. met een variantie van 110 sec.^{2}. Mag men hieruit besluiten dat de gemiddelde tijd om de job uit te voeren korter is voor een arbeider van shift 2 dan voor een arbeider van shift 1? De jobtijden worden verondersteld normaal verdeeld te zijn.

1. *Hypothesen*:

$$H_0 : \mu_1 - \mu_2 \leq 0$$
$$H_1 : \mu_1 - \mu_2 > 0$$

2. *Toetsstatistiek*:
Gelijke populatievarianties: t statistiek met $n_1 + n_2 - 2$ vrijheidsgraden:

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

Niet-gelijke populatievarianties:

$$t' = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

met kritieke t' waarde:

$$t'_{1-\alpha} = \frac{\frac{s_1^2}{n_1} t_{(1-\alpha, n_1-1)} + \frac{s_2^2}{n_2} t_{(1-\alpha, n_2-1)}}{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*:
   Gelijke populatievarianties: H_0 verwerpen als $t \geq 1.746$
   Niet-gelijke populatievarianties: H_0 verwerpen als $t' \geq 1.863$

$$t'_{1-\alpha} = \frac{\frac{144}{10}1.833 + \frac{110}{8}1.895}{\frac{144}{10} + \frac{110}{8}} = 1.863$$

5. *Berekening*:
   Gelijke populatievarianties:

$$t = \frac{(26.1 - 17.6) - 0}{\sqrt{\frac{129.13}{10} + \frac{129.13}{8}}} = 1.58$$

$$s_p^2 = \frac{(10-1)144 + (8-1)110}{10 + 8 - 2} = 129.13$$

   Niet-gelijke populatievarianties:

$$t' = \frac{(26.1 - 17.6) - 0}{\sqrt{\frac{144}{10} + \frac{110}{8}}} = 1.60$$

6. *Beslissing*:
   Gelijke populatievarianties: $1.58 < 1.746 \rightarrow H_0$ niet verworpen.
   Niet-gelijke populatievarianties:$1.60 < 1.863 \rightarrow H_0$ niet verworpen.

7. *Conclusie*: de gemiddelde tijd van de arbeiders van shift 2 is niet significant korter dan die van de arbeiders van shift 1. $0.05 < p < 0.10$.

Voorbeeld. Twee procedures kunnen gebruikt worden voor de fabricatie van draad. Een steekproef van 50 draden geproduceerd met procedure 1 heeft een gemiddelde treksterkte van 150 kg en een s.d. van 10 kg.; voor steekproef 2 met 100 draden gefabriceerd met procedure 2 is het gemiddelde 153 kg. en de s.d. 5 kg. Mag men stellen dat de treksterkte van de draad van procedure 1 meer dan 2 kg. lager is dan die vervaardigd met procedure 2? ($\alpha = 0.01$)

1. *Hypothesen*:

$$H_0 : \mu_2 - \mu_1 \leq 2$$
$$H_1 : \mu_2 - \mu_1 > 2$$

2. *Toetsstatistiek*:

$$Z = \frac{(\overline{X}_2 - \overline{X}_1) - (\mu_2 - \mu_1)_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

3. *Significantieniveau*: $\alpha = 0.01$.

4. *Beslissingsregel*: H_0 verwerpen als $Z \geq 2.33$

5. *Berekening*:

$$Z = \frac{(153 - 150) - 2}{\sqrt{\frac{10^2}{50} + \frac{5^2}{100}}} = 0.67$$

6. *Beslissing*: $0.67 < 2.33 \rightarrow H_0$ niet verworpen

7. *Conclusie*: Het verschil in treksterkte van beide draden is niet significant meer dan 2 kg. $p = 0.2514$

3.6.2 Interval: F toets voor verschil tussen varianties

Doel. Toetsen van de significantie van het verschil tussen 2 populatievarianties, gebaseerd op 2 onafhankelijke steekproeven.

Voorwaarden.

- Normaal verdeelde populaties.
- Minstens interval gegevens.

Methode. De hypothesen zijn

$$\begin{array}{ll} H_0 : \sigma_1^2 \geq \sigma_2^2 & H_0 : \sigma_1^2 \leq \sigma_2^2 \\ H_1 : \sigma_1^2 < \sigma_2^2 & H_1 : \sigma_1^2 > \sigma_2^2 \end{array}$$

voor een eenzijdige toets, en

$$\begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{array}$$

voor een tweezijdige toets.
De steekproefvarianties S_1^2 en S_2^2 worden berekend uit de steekproef.

$$S_1^2 = \frac{\sum(X_{1i} - \overline{X}_1)^2}{n_1 - 1} \qquad S_2^2 = \frac{\sum(X_{2i} - \overline{X}_2)^2}{n_2 - 1}$$

De toetsstatistiek is de F statistiek met $n_1 - 1$ vrijheidsgraden in de teller en $n_2 - 1$ vrijheidsgraden in de noemer:

$$F = \frac{S_1^2}{S_2^2}$$

Hierbij is vereist dat S_1^2 steeds de grootste van de 2 varianties is.

Voorbeeld. Een bedrijf heeft de keuze tussen 2 opleidingsprogramma's voor werknemers. 21 werknemers worden onderworpen aan het eerste, 16 aan het tweede opleidingsprogramma. Veronderstel dat de beide populaties normaal verdeeld zijn. Een t-toets wordt gebruikt om te bepalen of de gemiddelde score van een test afgenomen na de opleidingen significant verschillende resultaten opleverde. Kunnen we hiervoor een t toets met gelijke populatievarianties gebruiken, indien de varianties 225.5 en 70.3 bedragen?

1. *Hypothesen*:

$$\begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{array}$$

2. *Toetsstatistiek*: F statistiek met $n_1 - 1 = 20$ vrijheidsgraden in de teller en $n_2 - 1 = 15$ vrijheidsgraden in de noemer:
$$F = \frac{S_1^2}{S_2^2}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $F \leq 0.39$ ($F_{(20,15),0.025} = 1/F_{(15,20),0.975}$) of $F \geq 2.76$ ($F_{(20,15),0.975}$)

5. *Berekening*:
$$F = \frac{225.5}{70.3} = 3.21$$

6. *Beslissing*: $3.21 > 2.76 \rightarrow H_0$ verworpen.

7. *Conclusie*: De beide varianties zijn significant verschillend. $p < 0.05$.

3.6.3 Ordinaal: Mann-Whitney toets

Doel. Toetsen of twee onafhankelijke steekproeven afkomstig zijn van een zelfde populatie.

Voorwaarden.

- De variabelen zijn continu.
- De twee steekproeven zijn onafhankelijk.
- De verdelingen van de 2 populaties verschillen enkel m.b.t. hun locatie.
- Minstens ordinale gegevens.

Methode. De Mann-Whitney toets is het niet-parametrisch alternatief met het grootste onderscheidingsvermogen voor de t-toets voor het verschil van 2 gemiddelden indien aan een aantal veronderstellingen van de t-toets niet is voldaan (normaliteit, interval schaal gegevens, symmetrische verdeling,...).
Toetsen of twee verdelingsfuncties overeenstemmen komt in vele gevallen neer op het toetsen van de nulhypothesen dat de kans 0.5 is dat de waarnemingen van de variabele X (populatie 1) groter zijn dan die van de variabele Y (populatie 2). De alternatieve hypothese geeft aan dat deze kans verschillend is van 0.5, hetgeen wijst op een significant verschil in een plaatsparameter (gemiddelde voor een symmetrische verdeling, mediaan voor een niet-symmetrische verdeling,...)

$$H_0 : P(X < Y) = \tfrac{1}{2}$$
$$H_1 : P(X < Y) \neq \tfrac{1}{2}$$

voor een tweezijdige toets, en

$$H_0 : P(X < Y) \leq \tfrac{1}{2} \quad H_0 : P(X < Y) \geq \frac{1}{2}$$
$$H_1 : P(X < Y) > \tfrac{1}{2} \quad H_1 : P(X < Y) < \frac{1}{2}$$

voor de eenzijdige toetsen.
De toetsprocedure vereist eerst de bepaling van de steekproefomvang n_2 van de steekproef van X en steekproefomvang n_1 van de steekproef van Y, met $(n_2 < n_1)$. Vervolgens worden beide steekproeven als 1 grote steekproef beschouwd en worden $n_1 + n_2$ rangen toegekend aan de hele steekproef. Gemiddelde rangen worden toegekend aan gelijke getallen.
De toetsstatistiek W_x is de som van de rangen van X, de kleinste van de 2 steekproeven. Daarnaast wordt ook W'_x berekend als

$$W'_x = n_2(n_1 + n_2 + 1) - W_x$$

Indien $n_2 \leq 16$ en $n_1 \leq 19$ dan wordt kritieke waarde van de Mann-Whitney tabel afgelezen. Indien W_x of W'_x kleiner is dan de kritieke waarde, dan wordt H_0 verworpen.
Voor grotere steekproeven $n_2 > 16$ en $n_1 > 19$ kan de normaal verdeling gebruikt worden:

$$Z = \frac{W_x \pm 0.5 - n_2(n_2 + n_1 + 1)/2}{\sqrt{n_2 n_1 (n_2 + n_1 + 1)/12}}$$

De waarde $+0.5$ wordt toegepast indien een probabiliteit aan de linkerkant van de verdeling wordt gezocht. Voor een probabiliteit aan de rechterkant wordt -0.5 gebruikt.

Voorbeeld. Een test wordt opgelost door een groep van 5 mannen en 4 vrouwen uitgevoerd met de volgende resultaten.

Vrouw (X)	110	70	53	51	
rang	9	5	3	2	
Man (Y)	78	64	75	45	82
rang	7	4	6	1	8

Kan op basis van deze resultaten besloten worden dat de mediane score van de vrouwen significant beter is dan die van de mannen? De testresultaten kunnen in geen van beide groepen normaal verdeeld verondersteld worden.

1. *Hypothesen*:

$$H_0 : P(X < Y) \geq \tfrac{1}{2}$$
$$H_1 : P(X < Y) < \tfrac{1}{2}$$

2. *Toetsstatistiek*: W_x: som van de rangen van X.

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verworpen als $W_x < 12$ of $W'_x < 12$ $(n_2 = 4, n_1 = 5)$.

5. *Berekening*:

$$W_x = 9 + 5 + 3 + 2 = 19$$
$$W'_x = 4(4 + 5 + 1) - 19 = 21$$

6. *Beslissing*: $19 > 12,\ 21 > 12 \rightarrow H_0$ niet verworpen.

7. *Conclusie*: De resultaten van de vrouwen zijn niet significant beter dan die van de mannen.

3.6.4 Nominaal: χ^2 toets voor 2 onafhankelijke steekproeven

Doel. Toetsen van de significantie van de verschillen tussen 2 verdelingen, gebaseerd op 2 steekproeven, gespreid over r klassen.

Voorwaarden.

- Nominale gegevens.
- Minder dan 20% van de cellen hebben een verwachte frekwentie kleiner dan 5 en geen enkele cel heeft een verwachte frekwentie minder dan 1. In voorkomend geval moeten rijen en cellen gecombineerd worden.

Methode. De hypothesen luiden:

H_0 : De verschillen tussen de beide steekproeven zijn niet significant

H_1 : De verschillen tussen de beide steekproeven zijn significant

De geobserveerde frekwenties O_{ij} worden in een $(r \times 2)$ tabel gebracht.

	Steekproef 1	Steekproef 2	Totaal
Klasse 1	O_{11}	O_{12}	$n_{1.}$
...	...	...	...
Klasse r	O_{r1}	O_{r2}	$n_{r.}$
Totaal	$n_{.1}$	$n_{.2}$	n

Hierbij stelt 0_{ij} de geobserveerde frekwentie van klasse i in steekproef j voor. De omvang van steekproef 1 is $n_{.1}$, die van steekproef 2 $n_{.2}$.
De verwachte frekwentie E_{ij} wordt berekend middels:

$$E_{ij} = \frac{n_{i.}n_{.j}}{n}$$

De toetsstatistiek is de χ^2 met $(r-1)$ vrijheidsgraden:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Indien de χ^2 waarde de kritieke waarde overstijgt, dan wordt H_0 verworpen.

Voorbeeld. Een onderzoek probeert het verband na te gaan tussen de lichaamslengte van een persoon en zijn leiderscapaciteit. Een steekgroep met kleine mensen en één met grote mensen wordt hiervoor beschouwd. Leiderschap wordt vertaald in "Volger", "Niet-classificeerbaar" en "Leider". De volgende resultaten werden opgetekend:

	Grote mensen	Kleine mensen	Totaal
Volger	22 (16.3)	14 (19.7)	36
Niet-classificeerbaar	9 (6.8)	6 (8.2)	15
Leider	12 (19.9)	32 (24.1)	44
Totaal	43	52	95

1. *Hypothesen*:

   H_0 : De verschillen in leiderscapaciteit tussen grote en kleine mensen zijn niet significant

   H_1 : De verschillen in leiderscapaciteit tussen grote en kleine mensen zijn significant

2. *Toetsstatistiek*:
$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $\chi^2 \geq 5.99$ ($r - 1 = 2$ vrijheidsgraden).

5. *Berekening*: De E_{ij}'s worden tussen haakjes vermeld voor elke cel.
$$\chi^2 = \frac{(22 - 16.3)^2}{16.3} + \frac{(9 - 6.8)^2}{6.8} + ... + \frac{(32 - 24.1)^2}{24.1} = 10.67$$

6. *Beslissing*: $10.67 > 5.99 \rightarrow H_0$ verworpen.

7. *Conclusie*: De verschillen in leiderscapaciteit tussen grote en kleine mensen zijn significant. $p < 0.01$.

3.6.5 Nominaal: Z toets op verschil tussen 2 proporties

Doel. Toetsen van de significantie tussen 2 populatieproporties, gebaseerd op 2 steekproeven.

Voorwaarden.

- $n_1p_1 > 5, n_1(1 - p_1) > 5, n_2p_2 > 5, n_2(1 - p_2) > 5$ voor de normale benadering van de binomiaal verdeling.
- Twee aselecte onafhankelijke steekproeven.
- Twee dichotome populaties.

Methode. In geval van de normale benadering, is de toetsstatistiek de Z-statistiek:
$$Z = \frac{(P_1 - P_2) - (\pi_1 - \pi_2)_0}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}}$$

Gezien π_1 en π_2 meestal niet gekend zijn, moeten ze geschat worden door respectievelijk P_1 en P_2.
Indien de nulhypothese toetst of beide proporties gelijk zijn (enkel tweezijdige toets), is dit een argument om beide proporties te combineren tot een *gepoolde schatting* van de hypothetische gemeenschappelijke proportie
$$\overline{P} = \frac{P_1n_1 + P_2n_2}{n_1 + n_2}$$

met standaardfout:
$$S_{P_1-P_2} = \sqrt{\frac{\overline{P}(1 - \overline{P})}{n_1} + \frac{\overline{P}(1 - \overline{P})}{n_2}}$$

De toetsstatistiek is dan:

$$Z = \frac{(P_1 - P_2) - 0}{S_{P_1 - P_2}}$$

Voorbeeld. Een enquête toonde aan dat 54 van de 225 klanten komende van buiten de stad en 52 van de 175 klanten wonende in de stad enkel inkopen deden in grootwarenhuizen. Kan men op basis van deze 2 steekproeven besluiten of er een significant verschil bestaat tussen de beide groepen klanten?

1. *Hypothesen*:

$$H_0 : \pi_1 = \pi_2$$
$$H_1 : \pi_1 \neq \pi_2$$

2. *Toetsstatistiek*:

$$Z = \frac{(P_1 - P_2) - 0}{S_{P_1 - P_2}}$$

met

$$S_{P_1 - P_2} = \sqrt{\frac{\overline{P}(1 - \overline{P})}{n_1} + \frac{\overline{P}(1 - \overline{P})}{n_2}}$$
$$\overline{P} = \frac{P_1 n_1 + P_2 n_2}{n_1 + n_2}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $Z \leq -1.96$ of $Z \geq 1.96$

5. *Berekening*:

$$P_1 = \frac{54}{225} = 0.240 \qquad P_2 = \frac{52}{175} = 0.297$$
$$Z = \frac{(0.240 - 0.297) - 0}{0.044} = -1.28$$

met

$$S_{P_1 - P_2} = \sqrt{\frac{0.265(1 - 0.265)}{225} + \frac{0.265(1 - 0.265)}{175}} = 0.044$$
$$\overline{P} = \frac{0.240(225) + 0.297(175)}{225 + 175} = 0.265$$

6. *Beslissing*: $-1.96 < -1.28 < 1.96 \rightarrow H_0$ niet verworpen.

7. *Conclusie*: Er is geen significant verschil tussen de 2 proporties. $p = 2(0.1003) = 0.2006$.

3.7 k verwante steekproeven

3.7.1 Interval: F toets voor 2-weg variantie-analyse

Doel. Toetsen of k gerelateerde steekproeven afkomstig zijn van een zelfde populatie.

Voorwaarden.

- Alle k populaties zijn normaal verdeeld.
- Alle k populaties hebben gelijke variantie.
- Minstens interval gegevens.

Methode. De 2-weg variantie-analyse onderzoekt het effekt van 2 behandelingen. De tussen-variantie ten gevolge van elk van de 2 behandelingen afzonderlijk wordt vergeleken met de residuele variantie middels 2 separate F statistieken.

X_{11}	...	X_{1k}	$\overline{X_{1.}}$
...	...	...	...
X_{n1}	...	X_{nk}	$\overline{X_{n.}}$
$\overline{X_{.1}}$	...	$\overline{X_{.k}}$	$\overline{\overline{X}}$

De k steekproeven zijn net even groot en hebben omvang n. De twee-weg variantie-analyse wordt gebruikt om het verschil tussen de k gerelateerd steekproeven (het zogenaamde *behandelingseffect*) op significantie te toetsen. Bovendien kan ook het verschil tussen de n observaties over de k steekproeven (het zogenaamde *blokeffect*) op significantie getoetst worden. Bv. zowel het verschil tussen k machines als het verschil tussen n operatoren van de machines kan op significantie getoetst worden. Deze verschillen worden weergegeven door een tussenvariantie.
De *tussenvariantie* S_k^2 is de variabiliteit tussen het globaal gemiddelde $\overline{\overline{X}}$ en de k gemiddelden $\overline{X}_{.j}$ van elke steekproef $j = 1, ..., k$.

$$S_k^2 = \frac{n \sum_{j=1}^{k} (\overline{X}_{.j} - \overline{\overline{X}})^2}{k-1}$$

De tussenvariantie S_n^2 geeft de variabiliteit aan tussen het globaal gemiddelde $\overline{\overline{X}}$ en elk blokgemiddelde $\overline{X}_{i.}$ voor $i = 1, ..., n$.

$$S_n^2 = \frac{k \sum_{i=1}^{n} (\overline{X}_{i.} - \overline{\overline{X}})^2}{n-1}$$

Van de andere kant is er *binnenvariantie* S_e^2, die variabiliteit ten gevolge van het toeval weergeeft. De binnenvariantie kan geïnterpreteerd als het onverklaarde deel van de totale variantie.

$$S_e^2 = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n} (X_{ij} - \overline{X}_{.j} - \overline{X}_{i.} + \overline{\overline{X}})^2}{(n-1)(k-1)}$$

De hypothesen zijn tweeledig. Enerzijds wordt er getoetst of er significante verschillen bestaan tussen de k steekproeven. Daarvoor wordt de $F_{k-1,(n-1)(k-1)}$ statistiek gebruikt met $k-1$ vrijheidsgraden in de teller en $(n-1)(k-1)$ vrijheidsgraden in de teller.

$$F = \frac{S_k^2}{S_e^2}$$

Indien de steekproefgemiddelden van de k steekproeven sterk verschillen van het globaal gemiddelde, dan wordt een hoge waarde voor de tussenvariantie in verhouding tot de binnenvariantie verkregen. Dit geeft een hoge F-waarde. Indien deze waarde minstens even groot is als de kritieke waarde, dan wordt de nulhypothese verworpen en mag men stellen dat er significante verschillen tussen de steekproeven waargenomen kunnen worden.

Een gelijkaardige toets kan uitgevoerd worden om de significante verschillen tussen de n blokken vast te stellen. Hiervoor wordt de $F_{n-1,(n-1)(k-1)}$ statistiek gebruikt met $n-1$ vrijheidsgraden in de teller en $(n-1)(k-1)$ vrijheidsgraden in de teller.

$$F = \frac{S_n^2}{S_e^2}$$

Voorbeeld. Het aantal geproduceerde eenheden per uur van 3 machines wordt vergeleken. De test wordt uitgevoerd met een steekproef van 5 uren en 5 operatoren, die elk 1 uur aan de 3 machines hebben gestaan. Kan men op basis van de onderstaande gegevens stellen dat er een significant verschil bestaat tussen de machines. Is er een effect van de operatoren?

Machine	1	2	3	$\overline{X_{i.}}$
Operator 1	53	61	51	55
Operator 2	47	55	51	51
Operator 3	46	52	49	49
Operator 4	50	58	54	54
Operator 5	49	54	51	51
$\overline{X_{.j}}$	49	56	51	$\overline{\overline{X}} = 52$

1. *Hypothesen*:

H_0 : $\mu_1 = \mu_2 = \mu_3$

H_1 : Er zijn significante verschillen tussen de 3 machines

en

H_0 : Er zijn geen significante verschillen tussen de 5 operatoren

H_1 : Er zijn significante verschillen tussen de 5 operatoren

2. *Toetsstatistiek*:
Machines: F statistiek $k-1 = 2$ vrijheidsgraden in de teller en $(n-1)(k-1) = 8$ vrijheidsgraden in de noemer:

$$F = \frac{S_k^2}{S_e^2}$$

met

$$S_k^2 = \frac{n\sum_{j=1}^{k}(\overline{X}_{.j} - \overline{\overline{X}})^2}{k-1}$$

$$S_e^2 = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij} - \overline{X}_{.j} - \overline{X}_{i.} + \overline{\overline{X}})^2}{(n-1)(k-1)}$$

Operatoren: F statistiek $n-1 = 4$ vrijheidsgraden in de teller en $(n-1)(k-1) = 8$ vrijheidsgraden in de noemer:

$$F = \frac{S_n^2}{S_e^2}$$

met

$$S_n^2 = \frac{k\sum_{i=1}^{n}(\overline{X}_{i.} - \overline{\overline{X}})^2}{n-1}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*:
   Machines: H_0 verwerpen als $F \geq 4.46$
   Operatoren: H_0 verwerpen als $F \geq 3.84$

5. *Berekening*:

$$S_k^2 = \frac{130}{2} = 65$$

$$S_n^2 = \frac{72}{4} = 18$$

$$S_e^2 = \frac{22}{8} = 2.75$$

$$F = \frac{S_k^2}{S_e^2} = 23.6$$

$$F = \frac{S_n^2}{S_e^2} = 6.5$$

6. *Beslissing*:
   Machines: $23.6 > 4.46 \rightarrow H_0$ verworpen.
   Operatoren: $6.5 > 3.84 \rightarrow H_0$ verworpen.

7. *Conclusie*:
   Er zijn significante outputverschillen tussen de machines.$p < 0.025$
   Er zijn significante verschillen tussen de operatoren.$p < 0.025$

3.7.2 Ordinaal: Friedmann 2-weg variantie-analyse toets

Doel. Toetsen of k gerelateerde steekproeven afkomstig zijn van een zelfde populatie.

Voorwaarden.

- Minstens ordinale gegevens.

Methode. De scores worden in een tabel met k kolommen en n rijen gezet, waarbij n de steekproefomvang van elke steekproef voorstelt. Binnen elke rij worden de scores gerangschikt van 1 tot k. Vervolgens wordt voor elke kolom $j = 1, ..., k$ de rangsom berekend R_j. De toetsstatistiek is de χ^2 statistiek met $k - 1$ vrijheidsgraden:

$$\chi^2 = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1)$$

Voor kleine waarden van n en k bestaan er specifieke tabellen voor de kritieke waarden van de χ^2. De gewone tabellen van de χ^2 verdeling bieden echter een voldoende goede benadering. Indien de berekende χ^2 waarde minstens even groot is dan de kritieke waarde, dan wordt de nulhypothese verworpen, wat wil zeggen dat er significante verschillen bestaan tussen de k steekproeven. Bij verwerping van de nulhypothese bestaat de mogelijkheid om via paarsgewijze vergelijkingen de significant van elkaar verschillende steekproeven te detecteren. Daarop wordt hier niet verder ingegaan.

Voorbeeld. Vijf softwarepakketten (A,B,C,D en E) voor routeplanning worden geëvalueerd door ze elk vijf afstanden tussen 2 steden te laten berekenen. De resultaten zijn in grootte-orde niet vergelijkbaar met elkaar (t.t.z. afstand Antwerpen - Mechelen is kleiner dan Oostende - Luik). De vergelijking gebeurt bijgevolg door voor elk probleem de pakketten te rangschikken m.b.t. hun oplossing. De volgende resultaten werden hierbij verkregen.

	A	*B*	*C*	*D*	*E*
Probleem 1	5	1	3	4	2
Probleem 2	5	3	2	4	1
Probleem 3	4	2	3	5	1
Probleem 4	5	3	4	2	1
Probleem 5	4	5	3	2	1
Totaal	23	14	15	17	6

Zijn er significante verschillen tussen de resultaten van de softwarepakketten?

1. *Hypothesen*:

   H_0 : De verschillen tussen de 5 pakketten zijn niet significant

   H_1 : De verschillen tussen de 5 pakketten zijn significant

2. *Toetsstatistiek*: χ^2 met $k - 1 = 5 - 1 = 4$ vrijheidsgraden

$$\chi^2 = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3n(k+1)$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $\chi^2 \geq 9.488$.

5. *Berekening*:

$$\chi^2 = \frac{12}{5(5)(5+1)} 23^2 + 14^2 + 15^2 + 17^2 + 6^2 - 3(5)(5+1) = 12$$

6. *Beslissing*: $12 > 9.488 \rightarrow H_0$ verworpen.

7. *Conclusie*: Er zijn significante verschillen tussen de verschillende softwarepakketten. $0.01 < p < 0.025$.

3.7.3 Nominaal: Cochran Q toets

Doel. Toetsen van de significantie van het verschil tussen k sets van frekwenties of proporties.

Voorwaarden.

- De k sets van frekwenties hebben betrekking op dezelfde n elementen.
- De observaties zijn nominaal in 2 klassen (dichotoom 0,1).

Methode. De toets onderzoekt of er significante verschillen bestaan tussen k gerelateerde steekproeven van omvang n en bestaande enkel uit 0 en 1. Hiervoor wordt een tabel opgesteld met k kolommen en n rijen. De toetsstatistiek is de Q statistiek die χ^2 verdeeld is met $k-1$ vrijheidsgraden:

$$Q = \frac{(k-1)\left[k\sum_{j=1}^{k} C_j^2 - (\sum_{j=1}^{k} C_j)^2\right]}{k\sum_{i=1}^{n} L_i - \sum_{i=1}^{n} L_i^2}$$

- C_j: kolomtotaal van j-de kolom;
- L_i: rijtotaal van i-de rij;

Indien Q minstens even groot is dan de kritieke χ^2 waarde, dan wordt de nulhypothese verworpen, wat betekent dat er significante verschillen tussen de k dichotome steekproeven zijn.

Voorbeeld. Drie statistische hypothesetoetsen werden toegepast op acht gegevenssets. Een "0" wijst op de verwerping van H_0, een "1" op het niet verwerpen ervan.

	A	*B*	*C*	*Totaal*
Set 1	1	1	1	3
Set 2	1	1	0	2
Set 3	1	0	1	2
Set 4	0	0	1	1
Set 5	0	0	1	1
Set 6	0	1	0	1
Set 7	1	0	0	1
Set 8	0	0	0	0
Totaal	4	3	4	11

Is er verschil tussen de drie hypothesetoetsen m.b.t. hun onderscheidingsvermogen?

1. *Hypothesen*:

$$H_0 : \text{De verschillen tussen de 3 toetsen zijn niet significant}$$
$$H_1 : \text{De verschillen tussen de 3 toetsen zijn significant}$$

2. *Toetsstatistiek*: Q statistiek die χ^2 verdeeld is met $k-1$ vrijheidsgraden:

$$Q = \frac{(k-1)\left[k\sum_{j=1}^{k} C_j^2 - (\sum_{j=1}^{k} C_j)^2\right]}{k\sum_{i=1}^{n} L_i - \sum_{i=1}^{n} L_i^2}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $Q \geq 5.99$.

5. *Berekening*:

$$Q = \frac{(3-1)[3(4^2+3^2+4^2) - (11)^2}{3(11) - (3^2+2^2+...+1^2+0^2)} = 0.33$$

6. *Beslissing*: $0.33 < 5.99 \rightarrow H_0$ niet verworpen.

7. *Conclusie*: Er zijn geen significante verschillen tussen de 3 hypothesetoetsen m.b.t. hun onderscheidingsvermogen. $p > 0.05$.

3.8 k onafhankelijke steekproeven

3.8.1 Interval: F toets voor 1-weg variantie-analyse

Doel. Toetsen of k onafhankelijke steekproeven afkomstig zijn uit k populaties met eenzelfde gemiddelde.

Voorwaarden.

- Alle k populaties zijn normaal verdeeld.
- Alle k populaties hebben gelijke variantie.
- Minstens interval gegevens.

Methode. De hypothesen zijn:

$$H_0 : \quad \mu_1 = \mu_2 = ... = \mu_k$$
$$H_1 : \text{ Er zijn significante verschillen tussen de } k \text{ gemiddelden}$$

De toetsstatistiek is de $F_{k-1,n-k}$ statistiek met $k-1$ vrijheidsgraden in de teller en $n-k$ vrijheidsgraden in de noemer, waarbij n het totaal aantal waarnemingen in de k steekproeven voorstelt.

$$F = \frac{S_k^2}{S_e^2}$$

Hierbij wordt S_1 beschouwd als de variantie binnen elke steekproef, de zogenaamde *binnenvariantie*, die enkel en alleen het gevolg is van het toeval. De binnenvariantie wordt gewoon berekend door de variantie binnen elke steekproef j te berekenen en deze op te tellen voor de k steekproeven:

$$S_e^2 = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)^2}{n-k} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} X_{ij}^2 - \sum_{j=1}^{k} n_j \overline{X}_j^2}{n-k}$$

Hierbij is $\overline{X}_j = (\sum_{i=1}^{n_j} X_{ij})/n_j$ het steekproefgemiddelde van steekproef j met n_j observaties. De term S_k^2 is de zogenaamde *tussenvariantie*, de variantie tussen de steekproefgemiddelden en het totale gemiddelde. Deze variantie vertegenwoordigt de afwijking van elk steekproefgemiddelde ten opzichte van het totaal gemiddelde, niet ten gevolge van het toeval, maar ten gevolge van een systematisch effekt.

$$S_k^2 = \frac{\sum_{j=1}^{k} n_j(\overline{X}_j - \overline{\overline{X}})^2}{k-1} = \frac{\sum_{j=1}^{k} n_j \overline{X}_j^2 - n\overline{\overline{X}}^2}{k-1}$$

Hierbij is $\overline{\overline{X}} = \sum_{j=1}^{k}\sum_{i=1}^{n_j} X_{ij}/n$ het globaal gemiddelde.
Indien de verschillen tussen de k steekproefgemiddelden groot is, dan is de tussenvariantie groot t.o.v. de binnenvariantie, wat leidt tot een hoge F-waarde. Indien deze F-waarde de kritieke $F_{k-1,n-k}$ waarde overstijgt, dan wordt de nulhypothese verworpen en kan men besluiten dat er significante verschillen tussen de k gemiddelden zijn. Met paarsgewijze vergelijkingstoetsen kan dan verder gezocht worden naar de significante verschillen tussen de k steekproeven. Op dit laatste wordt hier echter niet verder ingegaan.

Voorbeeld. Drie insecticides worden vergeleken. Het criterium is het aantal dode insecten op een bepaalde oppervlakte met gewassen behandeld met de insecticides. Kan men op basis van deze resultaten besluiten dat er een significant verschil bestaat in effectiviteit tussen de 3 insecticides?

Insecticide	n_j	X_{ij}	$\overline{X}_j$
1	4	11, 9, 13, 11	11
2	6	25, 28, 31, 27, 30, 33	29
3	5	19, 23, 19, 21, 20	20.4

1. *Hypothesen*:

$$H_0: \quad \mu_1 = \mu_2 = \mu_3$$
$$H_1: \quad \text{Er zijn significante verschillen tussen de 3 insecticides}$$

2. *Toetsstatistiek*: F statistiek met $k-1 = 3-1 = 2$ vrijheidsgraden in de teller en $n-k = 15-3 = 12$ vrijheidsgraden in de noemer:

$$F = \frac{S_k^2}{S_e^2}$$

met

$$S_k^2 = \frac{\sum_{j=1}^{k} n_j(\overline{X}_j - \overline{\overline{X}})^2}{k-1} = \frac{\sum_{j=1}^{k} n_j \overline{X}_j^2 - n\overline{\overline{X}}^2}{k-1}$$

$$S_e^2 = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(X_{ij} - \overline{X}_j)^2}{n-k} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} X_{ij}^2 - \sum_{j=1}^{k} n_j \overline{X}_j^2}{n-k}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $F \geq 3.88$

5. *Berekening*:

$$S_e^2 = \frac{8 + 42 + 11.2}{15-3} = 5.1$$
$$S_k^2 = \frac{7610.8 - 6826.66}{3-1} = 392.07$$
$$F = \frac{392.07}{5.1} = 76.87$$

6. *Beslissing*: $76.87 > 3.88 \rightarrow H_0$ verworpen

7. *Conclusie*: Er zijn significante verschillen tussen de 3 insecticiden. $p < 0.025$.

3.8.2 Ordinaal: Kruskal-Wallis 1-weg variantie-analyse

Doel. Toetsen of k onafhankelijke steekproeven afkomstig zijn uit verschillende populaties.

Voorwaarden.

- De steekproefomvang van elke steekproef moet minstens 5 zijn ($n \geq 5$).
- Minstens ordinale gegevens.

Methode. Dit is de ordinale versie van de 1-weg variantie-analyse, die hiervoor beschreven staat. De Kruskal-Wallis toets kan gebruikt worden indien aan de assumpties van de F-toets (normaliteit, gelijke varianties) niet volledig voldaan is.
Alle waarnemingen van alle steekproeven worden in één reeks gerangschikt met rang 1 tot n (som van alle steekproefgroottes). Bepaal vervolgens voor elke steekproef j afzonderlijk R_j (de som der rangen). De toetsstatistiek KW volgt een χ^2 verdeling met $k-1$ vrijheidsgraden:

$$KW = \frac{12}{n(n+1)} \left(\sum_{j=1}^{k} \frac{R_j^2}{n_j} \right) - 3(n+1)$$

Indien KW minstens even groot is dan de kritieke χ^2 waarde, dan wordt de nulhypothese verworpen, wat betekent dat de k steekproeven niet afkomstig zijn uit dezelfde populatie.

Voorbeeld. Drie types van pijnstillers worden op hun effectiviteit getest door ze toe te dienen aan 3 groepen van in het totaal 22 mensen. De gemiddelde duurtijd van hoofdpijn geobserveerd bij de 22 mensen gedurende 2 weken wordt weergegeven in onderstaande tabel. Aangezien het 22 verschillende mensen betreft, zijn de tijden enkel benaderend vergelijkbaar. Zijn de drie types van pijnstillers even effectief op een 0.01 significantieniveau?

type A	*type B*	*type C*
5.3 (12.5)	6.3 (15)	2.4 (3)
4.2 (9)	8.4 (20)	3.1 (5)
3.7 (6.5)	9.3 (21)	3.7 (6.5)
7.2 (17)	6.5 (16)	4.1 (8)
6.0 (14)	7.7 (18)	2.5 (4)
4.8 (11)	8.2 (19)	1.7 (2)
	9.5 (22)	5.3 (12.5)
		4.5 (10)
		1.3 (1)
R_A=70	R_B=131	R_C=52

1. *Hypothesen:*

   H_0 : De verschillen tussen de 3 pijnstillers zijn niet significant

   H_1 : De verschillen tussen de 3 pijnstillers zijn significant

2. *Toetsstatistiek*: KW is χ^2 verdeeld met $k-1=2$ vrijheidsgraden.

$$KW = \frac{12}{n(n+1)}\left(\sum_{j=1}^{k}\frac{R_j^2}{n_j}\right) - 3(n+1)$$

3. *Significantieniveau*: $\alpha = 0.01$.

4. *Beslissingsregel*: H_0 verwerpen als $KW \geq 9.21$

5. *Berekening*:

$$KW = \frac{12}{22(22+1)}(\frac{70^2}{6} + \frac{131^2}{7} + \frac{52^2}{9}) - 3(22+1) = 15.63$$

6. *Beslissing*: $15.63 > 9.21 \rightarrow H_0$ verworpen.

7. *Conclusie*: Er is verschil in de effectiviteit van de pijnstillers. $p < 0.01$.

3.8.3 Nominaal: χ^2 toets voor k onafhankelijke steekproeven

Doel. Toetsen van de significantie van de verschillen tussen k verdelingen, gebaseerd op k steekproeven, gespreid over r klassen.

Voorwaarden.

- Nominale gegevens.
- Minder dan 20% van de cellen hebben een verwachte frekwentie kleiner dan 5 en geen enkele cel heeft een verwachte frekwentie minder dan 1. In voorkomend geval moeten rijen en cellen gecombineerd worden.

Methode. Deze toets is een uitbreiding van χ^2 toets voor 2 onafhankelijke steekproeven. De hypothesen luiden:

H_0 : De verschillen tussen de k steekproeven zijn niet significant
H_1 : De verschillen tussen de k steekproeven zijn significant

De geobserveerde frekwenties O_{ij} worden in een $(r \times k)$ contingentietabel gebracht, met r rijen en k kolommen.

	Steekproef 1	...	Steekproef k	Totaal
Klasse 1	O_{11}	...	O_{1k}	$n_{1.}$
...	...	...	...	...
Klasse r	O_{r1}	...	O_{rk}	$n_{r.}$
Totaal	$n_{.1}$	...	$n_{.k}$	n

De verwachte frekwentie E_{ij} wordt berekend middels:

$$E_{ij} = \frac{n_{i.}n_{.j}}{n}$$

De toetsstatistiek is de χ^2 met $(r-1)(k-1)$ vrijheidsgraden:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Indien de χ^2 waarde de kritieke waarde overstijgt, dan wordt H_0 verworpen. Bij verwerping van de H_0 is het echter niet mogelijk de cellen te lokaliseren die aanleiding geven tot de significante verschillen.

Voorbeeld. Er wordt beweerd dat vrouwen die kinderen hebben slechter slapen dan in de periode voor hun kinderen. Om dit te staven werden 60 vrouwen ondervraagd, opgedeeld in 3 groepen volgens het aantal kinderen. Wat kan er uit deze gegevens afgeleid worden?

Aantal kinderen	1	2	≥ 3	Totaal
Slaapt slechter	25	10	5	40
Slaapt even goed	5	4	7	16
Slaapt beter	0	1	3	4
Totaal	30	15	15	60

Bij de berekening van de verwachte frekwenties hebben meerdere cellen van de klassen "Slaapt even goed" en "Slaapt beter" frekwenties kleiner dan 5. Een samenvoeging van beide klassen dringt zich bijgevolg op.

Aantal kinderen	1	2	≥ 3	Totaal
Slaapt slechter	25 (20)	10 (10)	5 (10)	40
Slaapt minstens even goed	5 (10)	5 (5)	10(5)	20
Totaal	30	15	15	60

1. *Hypothesen*:

   H_0 : De verschillen tussen de 3 steekproeven zijn niet significant

   H_1 : De verschillen tussen de 3 steekproeven zijn significant

2. *Toetsstatistiek*: χ^2 statistiek met $(r-1)(k-1) = (2-1)(3-1) = 2$ vrijheidsgraden.

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $\chi^2 \geq 5.99$.

5. *Berekening*:

$$\chi^2 = \frac{(25-20)^2}{20} + \frac{(10-10)^2}{10} + \frac{(5-10)^2}{10} + \frac{(5-10)^2}{10} + \frac{(5-5)^2}{5} + \frac{(10-5)^2}{5} = 11.25$$

6. *Beslissing*: $11.25 > 5.99 \rightarrow H_0$ verwerpen.

7. *Conclusie*: Er zijn significante verschillen in slaapgedrag tussen de 3 groepen van vrouwen. $p < 0.01$.

3.9 Samenvatting

1. Een hypothesetoets is een statistische procedure die toelaat om te kiezen tussen 2 tegenstrijdige hypothesen.
2. Een hypothesetoets bestaat uit 7 stappen.
3. De nulhypothese is de hypothese die getest wordt, de hypothese van gelijkheid.
4. Het verwerpen van de nulhypothese is een sterker besluit dan het aanvaarden ervan.
5. De toetsstatistiek is de beslissingsnemer.
6. De 2 fouttypes I en II zijn complementair; enkel een grotere steekproef kan hun probabiliteit van voorkomen doen dalen.
7. Het significantieniveau α bepaalt de kans op een type I fout.
8. Het onderscheidingsvermogen van een toets geeft aan hoe duidelijk een toets de beslissing tussen aanvaarden en verwerpen van de nulhypothese kan nemen.
9. De p-waarde geeft het minimale significantieniveau weer.
10. De hypothesetoetsen worden geclassificeerd op basis van de meetschaal (nominaal, ordinaal, interval, (ratio)) en op basis van het aantal steekproeven (1, 2 onafhankelijke, 2 verwante, k onafhankelijke en k verwante).
11. Er bestaan hypothesetoetsen voor plaatsparameters (gemiddelde, mediaan,...) en spreidingsparameters (variantie,...), maar ook om het toevallige karakter van een steekproef te onderzoeken, om te toetsen op verdelingen,....
12. Een aantal belangrijke vertalingen:

toetsen van hypothesen	*hypothesis testing*
onderscheidingsvermogen	*power*
toetsstatistiek	*test statistic*
meetschaal	*measurement scale*
verwante steekproeven	*related samples*
onafhankelijke steekproeven	*independent samples*

3.10 Oefeningen

1. Een kopietoestel werkt aan een gemiddelde snelheid van 45 kopies per minuut. Door een kleine wijziging aan te brengen hoopt de fabrikant de snelheid verder op te drijven. Drie testen leveren de volgende snelheden op: 46, 47 en 48 kopies per minuut. Is de snelheidstoename significant te noemen? Veronderstel dat de snelheid normaal verdeeld is.

2. De gemiddelde levensduur van een batterij is 22 u. Veronderstel dat die levensduur normaal verdeeld is. Voor batterijen is het bekend dat de standaarddeviatie op die levensduur 3 u. bedraagt. Indien een steekproef van 9 batterijen een gemiddelde levensduur van 20 u. oplevert, mogen we dan concluderen dat de werkelijke levensduur minder is dan 22 u.?

3. Veronderstel een discrete uniforme verdeling met 5 waarden: 1, 2, 3, 4, 5. Elke waarde heeft een kans 1/5. Kan de volgende aselecte steekproef van 10 waarden 1,1,1,2,2,2,3,3,3,3 beschouwd worden als zijnde afkomstig van de voornoemde uniforme verdeling. Toets dit aan de hand van een Kolmogorov-Smirnov toets.

4. Een drankenfabrikant beweert dat zijn flesjes een inhoud hebben van 32 cl. Op basis van een steekproef van 50 flesjes wordt een gemiddelde inhoud van 31 cl. met een standaarddeviatie van 0.75 cl. gevonden. Kan men met een significantieniveau van 1% besluiten dat de fabrikant zijn klanten bedriegd?

5. Een bedrijf spendeert 1500 BEF per man per jaar aan veiligheidsopleiding. De directie vermoedt dat meer dan 75% van de vergelijkbare bedrijven een bedrag uitgeven dat niet hoger is. Op basis van een steekproef van 60 bedrijven, waren er 50 die het bedrag niet overstegen hadden. Wordt hiermee de bewering van de directie bevestigd?

6. De wachtij voor een kassa bevat 50 mensen, mannen en vrouwen in deze volgorde:
MVMVMMMVVMVMVMVMMMMVMVMVM
MVVVMVMVMVMMVMMVMMMMVMVMM.
Is de volgorde van mannen en vrouwen toevallig?

7. Een kantoorhouder bestudeert het aankomstpatroon van zijn klanten gedurende 90 uren. Het resultaat was als volgt:

Aantal aankomsten	0	1	2	3	4	5	6	7	8	≥ 9
Frekwentie	9	17	25	15	11	7	2	2	2	0

Mag de kantoorhouder hierbij besluiten dat het aantal aankomsten per uur een toevallige verdeling (Poisson verdeling) volgt?

8. Voor de productie van staalplaat mag de variantie niet hoger zijn dan 0.016 kg^2. Een toevallige steekproef van 26 platen geeft een variantie van 0.025 kg^2. Kan men besluiten dat er niet aan de specificaties wordt voldaan?

9. Aan 27 personen wordt gevraagd om 2 dezelfde chocoladerepen te eten, weliswaar ingepakt in 2 verschillende verpakkingen A en B. Elke persoon moet zeggen welk van de beide repen hij het lekkerste vond.

Persoon	Lekkerste reep	Persoon	Lekkerste reep
1	A	15	-
2	A	16	A
3	A	17	A
4	B	18	A
5	B	19	A
6	-	20	A
7	A	21	B
8	A	22	A
9	A	23	A
10	A	24	A
11	A	25	A
12	B	26	A
13	A	27	B
14	A		

Is de reep met verpakking A significant lekkerder dan die met verpakking B?

10. Twee leveranciers van een metalen component worden geëvalueerd. Aangezien de kost van hun product gelijk is, zal de kwaliteit doorslaggevend zijn. Op basis van een steekproef van 31 stukken bij de beide leveranciers, worden de volgende varianties verkregen: 0.0058 en 0.0041. Zal men een beslissing kunnen nemen op basis van een significant verschil in kwaliteit?

11. Acht personen moeten 2 verschillende reclamefilmpjes van hetzelfde product evalueren. Toets de bewering al zou de score van het eerste filmpje beter zijn dan dat van het tweede. Veronderstel normaliteit van de scores. $\alpha = 0.05$.

Persoon	1	2	3	4	5	6	7	8
Score filmpje 1	26.2	20.3	25.4	19.6	21.5	28.3	23.7	24.0
Score filmpje 2	24.1	21.3	23.7	18.0	20.1	25.8	22.4	21.4

12. Een groep van 40 mensen wordt voor en na een reclamespot gevraagd naar hun keuze tussen 2 producten. De reclamespot betreft 1 van die 2 producten. Men stelt vast dat van de 25 mensen die vooraf een preferentie voor product 1 vertonen er 7 overstappen naar product 2. Van de 15 mensen die vooraf voor product 2 hadden gekozen zijn er 6 na de spot overgestapt naar product 1. Is er een significant effect van de reclamespot?

13. Twee machines vervaardigen stalen ringen. Een aselecte steekproef van 10 ringen van machine 1 heeft een gemiddelde diameter van 1.051 cm. met een variantie van 0.000441 cm.^{2}. Een steekproef van 15 ringen van machine 2 levert een gemiddelde van 1.036 cm met een variantie van 0.000225 cm.^{2} op. Mag men hieruit besluiten dat de ringen van machine 1 een grotere diameter hebben dan die van machine 2? Veronderstel hierbij dat de diameters van de ringen een normaal verdeling volgen en dat de populatievarianties zowel gelijk als ongelijk kunnen zijn. ($\alpha = 0.01$).

14. IQ testen bij personen voor en na een assertiviteitstraining, leverden de volgende resultaten.

Persoon	*IQ voor*	*IQ na*
1	86	88
2	71	77
3	77	76
4	68	64
5	91	96
6	72	72
7	77	65
8	91	90
9	70	65
10	71	80
11	88	81
12	87	72

Is er een significante invloed van de assertiviteitstraining? Hierbij mag men niet veronderstellen dat iemand met een score van 90 driemaal slimmer is dan iemand met een score van 30.

15. Een steekproef van 50 bedienden in een bedrijf in een derde-wereld land leert dat het gemiddeld dagloon 160 BEF bedraagt met een standaardfout van 1.44 BEF. Een steekproef van 30 arbeiders in datzelfde bedrijf geeft een gemiddelde van 155 BEF en een standaardfout van 1.50 BEF. Is er een significant verschil tussen het dagloon van de arbeider en dat van de bediende?

16. Een studentenvereniging doet een referendum bij haar leden m.b.t. de invoering van een nieuw evaluatiesysteem. Hieruit bleek dat 221 van de 325 mannelijke studenten en 120 van de 200 vrouwelijke studenten voor waren. Duiden deze resultaten op een significant verschil in mening tussen mannelijke en vrouwelijke studenten?

17. De resultaten voor een examen statistiek worden vergeleken tussen een steekproef van 9 TEW studenten en een steekproef van 10 Geneeskunde studenten.

TEW	50.5	37.5	49.8	56.0	42.0	56.0	50.0	54.0	48.0	
Geneeskunde	57.0	52.0	51.0	44.2	55.0	62.0	59.0	45.2	53.5	44.4

De verdeling van de examenresultaten is in geen van beide groepen normaal verdeeld. Kan men op basis van deze gegevens besluiten dat er een significant verschil is tussen de examenresultaten van de studenten TEW en Geneeskunde?

18. Een steekproef van studenten uit school 1 en een steekproef van studenten uit school 2 worden onderworpen aan een aantal PMS testen met de volgende resultaten.

	Test score				
	0-275	276-350	351-425	426-500	Totaal
School 1	6	14	17	9	46
School 2	30	32	17	3	82
Totaal	36	46	34	12	128

Zijn de testscores significant verschillend tussen de beide scholen?

19. Een bedrijf werkt in 3 shiften met 4 machines (A, B, C en D). Over een periode van 1 jaar werd de frekwentie van machine-uitval opgetekend per machine per shift. Zijn er significant verschillen in de frekwentie van pannes tussen de 3 shiften?

Machines	A	B	C	D
Shift 1	10	6	13	13
Shift 2	10	12	19	21
Shift 3	13	10	13	18

20. Voor een schoonheidswedstrijd moeten 12 juryleden 4 kandidaten rangschikken.

Jurylid	*Kand. 1*	*Kand. 2*	*Kand. 3*	*Kand. 4*
1	4	3	2	1
2	4	2	3	1
3	3	1.5	1.5	4
4	3	1	2	4
5	4	2	1	3
6	2	2	2	4
7	1	3	2	4
8	2	4	1	3
9	3.5	1	2	3.5
10	4	1	3	2
11	4	2	3	1
12	3.5	1	2	3.5

Is er een significant verschil tussen de 4 kandidaten?

21. Drie weermannen (A,B en C) worden gevraagd om weersvoorspellingen voor 12 verschillende dagen te maken. Een juiste voorspelling werd als een 1 gekwoteerd, een foute als een 0. Kan men op basis van de volgende resultaten besluiten dat er significante verschillen tussen de 3 weermannen zijn?

	A	*B*	*C*	*Totaal*
Dag 1	1	1	1	3
Dag 2	1	1	1	3
Dag 3	0	1	0	1
Dag 4	1	1	0	2
Dag 5	0	0	0	0
Dag 6	1	1	1	3
Dag 7	1	1	1	3
Dag 8	1	1	0	2
Dag 9	0	0	1	1
Dag 10	0	1	0	1
Dag 11	1	1	1	3
Dag 12	1	1	1	3

22. Een bedrijf is op zoek naar een nieuwe verpakking voor een bepaalde soort van haar koekjes. 4 verpakkingen A, B, C en D worden hiervoor geëvalueerd door telkens 5 mensen. De mensen werden gevraagd naar de prijs die ze voor de koekjes met de gepaste verpakking zouden bereid zijn te betalen. Wijzen deze resultaten op een significant verschil tussen de vier verpakkingen? Er kan niet uitgegaan worden van normaliteitsveronderstellingen voor de oorspronkelijke populaties.

Verpakking A	60	40	40	60	50
Verpakking B	70	70	60	70	75
Verpakking C	55	70	60	55	70
Verpakking D	55	45	35	40	50

23. Een bedrijf wenst over te gaan tot de aankoop van nieuwe bestelwagens. Drie types worden beschouwd (A, B en C) met betrekking tot hun verbruik van benzine per 100 km. Van elk type wordt een beperkte steekproef genomen. Kan men op basis van deze gegevens aantonen dat het branstofverbruik significant verschilt tussen de 3 types?. Men kan er van uitgaan dat het brandstofverbruik voor elk type normaal verdeeld is.

A	B	C
23	27	24
23	29	26
20	25	24
21	23	
	24	

24. Om het effect van een reclamecampagne voor een lokale autogarage te bepalen worden 3 reclamecampagnes over 3 verschillende mediamixes geprobeerd. Is er een significant verschil tussen de 3 reclamecampagnes m.b.t. de verkopen 1 week nadien? Geven de drie mediamixes aanleiding tot significante verschillen tussen de verkopen?

	Campagne 1	Campagne 2	Campagne 3
Mix 1	3	5	4
Mix 2	11	10	12
Mix 3	16	21	17

Hoofdstuk 4

Associatiematen

4.1 Inleiding

De term *associatie* duidt op samenhang, relatie, verband tussen fenomenen. Meestal wordt de associatie voorgesteld door een *associatiemaat.* Een associatiemaat is principieel een getal, gestandaardiseerd tussen 0 en 1, waarbij 0 wijst op afwezigheid van enige samenhang en 1 op een maximale sterkte van de samenhang. Sommige associatiematen zijn gedefinieerd tussen -1 en 1 en geven bijgevolg ook de richting van het verband aan.

Zoals steeds in de statistische inferentie, kan een associatiemaat berekend op basis van de steekproef toevallig een slechte schatting zijn voor de werkelijke waarde van de associatie in de populatie. Daarom is het aangewezen in plaats van enkel de puntschatting te gebruiken eveneens een *betrouwbaarheidsinterval* te berekenen. Het betrouwbaarheidsinterval geeft aan of een associatiemaat significant van 0 verschilt, indien 0 niet tot het interval behoort. In het geval van associatiematen is het uitvoeren van een *significantietoets* overwegend eenvoudiger dan het opstellen van een betrouwbaarheidsinterval.
Zowel voor de berekening van een betrouwbaarheidsinterval als voor het uitvoeren van een hypothesetoets om de significantie van een associatiemaat te bepalen, zijn de belangrijkste elementen: de steekproefomvang n, het significantieniveau α en de toetsstatistiek.

Het dient opgemerkt te worden dat associatie *niet* noodzakelijk een beïnvloeding, een *causaal verband* weerspiegelt. Bv. uit een analyse blijkt dat er een significante associatie bestaat tussen "al of niet alleenwonend zijn" en het "al of niet eten van kant-en-klare maaltijden". Alleenwonenden eten significant meer kant-en-klare maaltijden. Maar dit betekent nog niet dat het alleenwonen het eten van kant-en-klare maaltijden tot gevolg heeft...

4.2 Classificatie van associatiematen

Net zoals voor de hypothesetoetsen, kan de keuze van een geschikte associatiemaat gebaseerd worden op verschillende keuzecriteria: meetniveau van de variabelen (nominaal, ordinaal of metrisch), op de vorm van de analyse (symmetrisch, asymmetrisch), op het achterliggend interpretatiemodel (statistische afhankelijkheid, optimale predictie,...), op het onderscheidingsvermogen,...

Meetschaal	Associatiemaat
Nominaal	Cramer C en ϕ coëfficient (4.5.1)
Ordinaal	Spearman rangcorrelatie R_s (4.4.1) Kendall τ (4.4.2)
Interval	Pearson R (4.3.1)

Tabel 4.1: Classificatie van associatiematen.

In deze cursus opteren we voor een classificatie op basis van het meetniveau van de variabelen. Voor een beschrijving van de vier meetniveau's wordt verwezen naar § 3.3.2. Tabel 4.1 bevat de associatiematen op de verschillende meetniveau's die in dit hoofdstuk aan bod zullen komen. Net zoals voor de hypothesetoetsen, is dit overzicht niet exhaustief, maar bevat wel de meest representatieve associatiematen.
De onderscheiden associatiematen worden besproken in functie van de hypothesetoets die vereist is om hun significantie te bepalen.

4.3 Interval meetniveau

4.3.1 Pearson R correlatie

Doel. Toetsen van de lineaire correlatie tussen 2 metrische variabelen.

Voorwaarden.

- Minstens interval schaal gegevens.
- Bij elke waarde van X behoort een normaal verdeelde subpopulatie van Y waarden.
- Bij elke waarde van Y behoort een normaal verdeelde subpopulatie van X waarden.
- De gezamelijk verdeling van X en Y is bivariaat normaal.

Methode. De correlatiecoëfficient van Pearson wordt ook wel eens product-moment correlatiecoëfficient genoemd en wordt gedefinieerd als:

$$R = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2 \sum(Y_i - \overline{Y})^2}}$$

De teller stelt de covariantie voor en geeft een idee over het samen varieren van X en Y t.o.v. hun gemiddelde. Indien X en Y in eenzelfde zin afwijken t.o.v. hun gemiddelde is de covariantie en dus ook R positief. Wijken X en Y af in tegengestelde zin van hun gemiddelde, dan zijn de covariantie en de R negatief. Indien er geen dergelijke afwijkingspatronen zijn tussen beide variabelen, dan is $R \approx 0$ en is er geen *lineair* verband. De correlatiecoëfficient is m.a.w. een gestandaardiseerde maat ($R \in [-1, 1]$) voor het samen variëren van de X en Y.
Teneinde significantietoetsen uit te voeren voor de correlatiecoëfficient R van Pearson, wordt de toetsstatistiek

$$t = R\sqrt{\frac{n-2}{1-R^2}}$$

gebruikt, die een t verdeling volgt met $n-2$ vrijheidsgraden.
De hypothesen zijn

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

voor een tweezijdige toets, en

$$H_0 : \rho \geq 0 \quad H_0 : \rho \leq 0$$
$$H_1 : \rho < 0 \quad H_1 : \rho > 0$$

voor een eenzijdige toets.
Indien op basis van de steekproefcorrelatie R de nulhypothese verworpen wordt, dan kan men stellen dat de populatiecorrelatiecoefficient ρ significant verschillend is van 0.

Voor het toetsen van de hypothesen

$$H_0 : \rho = \rho_0$$
$$H_1 : \rho \neq \rho_0$$

en

$$H_0 : \rho \geq \rho_0 \quad H_0 : \rho \leq \rho_0$$
$$H_1 : \rho < \rho_0 \quad H_1 : \rho > \rho_0$$

met ρ_0 een hypothetische correlatiewaarde verschillend van 0, is de Fisher transformatie van R in Z_R aangewezen:

$$Z_R = \frac{1}{2} \ln \frac{1+R}{1-R}$$

De steekproefgrootheid Z_R is benaderend normaal verdeeld met gemiddelde en standaardfout:

$$Z_\rho = \frac{1}{2} \ln \frac{1+\rho}{1-\rho}$$

$$S_{Z_R} = \frac{1}{\sqrt{n-3}}$$

De toetsstatistiek is dan de Z statistiek

$$Z = \frac{Z_R - Z_{\rho_0}}{S_{Z_R}}$$

Voorbeeld. Een studie wordt uitgevoerd om het verband tussen verkoopsvolumes Y (in miljoen BEF) en de oppervlakte van warenhuizen X (in $1000m^2$) te bepalen. Een steekproef van 10 warenhuizen levert het volgende resultaat op. Is er een significante correlatie tussen de beide variabelen?

1. *Hypothesen*:

$$H_0 : \rho = 0$$
$$H_1 : \rho \neq 0$$

X	Y
40	3.5
600	25.0
60	4.8
72	3.5
400	30.0
90	5.0
200	12.0
70	4.5
80	5.0
84	6.0

2. *Toetsstatistiek*: t statistiek met $n-2=10-2=8$ vrijheidsgraden

$$t = R\sqrt{\frac{n-2}{1-R^2}}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $t \leq -2.306$ of $t \geq 2.306$.

5. *Berekening*:

$$t = 0.92\sqrt{\frac{8}{1-(0.92)^2}} = 6.64$$

met

$$R = \frac{\sum(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum(X_i - \overline{X})^2 \sum(Y_i - \overline{Y})^2}} = \frac{14871.8}{\sqrt{(309198.4)(836.74)}} = 0.92$$

$$\overline{X} = 1696/10 = 169.6 \qquad \overline{Y} = 99.3/10 = 9.93$$

6. *Beslissing*: $6.64 > 2.306 \rightarrow H_0$ verwerpen

7. *Conclusie*: de verkopen en de omvang van warenhuizen zijn lineair gecorreleerd. $p < 0.01$.

Voorbeeld. Veronderstel dat men in het vorige voorbeeld wenst na te gaan of de correlatie verschillend is van 0.95.

1. *Hypothesen*:

$$H_0 : \rho = 0.95$$
$$H_1 : \rho \neq 0.95$$

2. *Toetsstatistiek*:

$$Z = \frac{Z_R - Z_{\rho_0}}{S_{Z_R}}$$

met

$$Z_R = \frac{1}{2}\ln\frac{1+R}{1-R}$$

$$Z_{\rho_0} = \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0}$$

$$S_{Z_R} = \frac{1}{\sqrt{n-3}}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $Z \leq -1.96$ of $Z \geq 1.96$

5. *Berekening*:

$$Z = \frac{1.59 - 1.83}{0.38} = -0.63$$

met

$$Z_R = \frac{1}{2} \ln \frac{1+0.92}{1-0.92} = 1.59$$

$$Z_{\rho_0} = \frac{1}{2} \ln \frac{1+0.95}{1-0.95} = 1.83$$

$$S_{Z_R} = \frac{1}{\sqrt{10-3}} = 0.38$$

6. *Beslissing*: $-1.96 < -0.63 < 1.96 \rightarrow H_0$ niet verworpen

7. *Conclusie*: De correlatie tussen X en Y is niet significant verschillend van 0.95. $p = 2 \times 0.3228 = 0.6456$.

4.4 Ordinaal meetniveau

4.4.1 Spearman R_s rangcorrelatie

Doel. Toetsen van de correlatie tussen 2 ordinale variabelen.

Voorwaarden.

- Minstens ordinaal schaal gegevens.

Methode. Voor de berekening van de Spearman rangcorrelatiecoëfficient R_s worden de waarnemingen van elk van de beide variabelen X en Y afzonderlijk gerangschikt van 1 tot n. Indien twee of meer waarnemingen eenzelfde waarde hebben, dan wordt een gemiddelde rang toegekend.
Vervolgens wordt voor elk object $i = 1, ..., n$ het verschil in rangen, D_i berekend om de rangcorrelatiecoëfficient te kunnen berekenen:

$$R_s = 1 - \frac{6 \sum_{i=1}^{n} D_i^2}{n^3 - n}$$

met $R_s \in [-1, 1]$.
De hypothesen zijn

$$H_0 : \rho_s = 0$$
$$H_1 : \rho_s \neq 0$$

voor een tweezijdige toets, en

$$H_0 : \rho_s \geq 0 \quad H_0 : \rho_s \leq 0$$
$$H_1 : \rho_s < 0 \quad H_1 : \rho_s > 0$$

Voor $n \leq 30$ bestaat een tabel met de kritieke R_s waarden voor verschillende waarden van α. Voor $n > 30$ is de toetsstatistiek de Z statistiek:

$$Z = R_S\sqrt{n-1}$$

Voorbeeld. Voor 10 studenten wordt het aantal uren gestudeerd (X) en het overeenkomend aantal punten op een test (Y) weergegeven. Bestaat er een verband tussen de beide variabelen?

X	Y	rang X	rang Y	D	D^2
5	6	4	3	1	1
9	16	6	6	0	0
17	18	7	7	0	0
1	1	1	1	0	0
2	3	2	2	0	0
21	21	8	9	1	1
3	7	3	4	1	1
29	20	9	8	1	1
7	15	5	5	0	0
100	22	10	10	0	0

1. *Hypothesen*:
$$H_0 : \rho_s = 0$$
$$H_1 : \rho_s \neq 0$$

2. *Toetsstatistiek*:
$$R_s = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n^3 - n}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verworpen als $R_S \geq 0.6364$ of $R_S \leq -0.6364$ (tabel)

5. *Berekening*:
$$R_s = 1 - \frac{6(4)}{10^3 - 10} = 0.976$$

6. *Beslissing*: $0.976 > 0.6364 \rightarrow H_0$ verworpen.

7. *Conclusie*: Er is een positieve correlatie tussen het aantal uren studeren en het aantal punten op een test. $p < 2 \times 0.001$

4.4.2 Kendall τ rangcorrelatie

Doel. Toetsen van de correlatie tussen 2 ordinale variabelen.

Voorwaarden.

- Minstens ordinaal schaal gegevens.

Methode. De Kendall τ rangcorrelatiecoëfficient is toepasbaar op eenzelfde soort gegevens waarvoor de Spearman R_s rangcorrelatie eveneens geschikt is, t.t.z. twee ordinale variabelen. Een voordeel van het gebruik van de Kendall τ is dat er een mogelijkheid bestaat om de τ te veralgemenen tot een patiële correlatiecoëfficient waarmee het effect kan onderzocht worden van een mogelijke derde variabele Z die uiteindelijk de correlatie tussen de variabelen X en Y veroorzaakt.

Vooreerst dienen aan de n observaties van X en Y rangen te worden toegekend. Het volgend voorbeeld, waarbij de correlatie tussen de rangschikkingen van 2 juryleden wordt bepaald, verduidelijkt een en ander:

subject	a	b	c	d
Jurylid 1	3	4	2	1
Jurylid 2	3	1	4	2

De n waarnemingen van X worden vervolgens gerangschikt in hun natuurlijke volgorde van 1 tot n. Gemiddelde rangen worden toegekend aan gelijke waarnemingen. De rangen van Y passen zich aan aan de natuurlijke volgorde van X.

subject	d	c	a	b
Jurylid 1	1	2	3	4
Jurylid 2	2	4	3	1

Om de graad van overeenkomst tussen de rangschikkingen te bepalen, is het voldoende om enkel de rangschikking van Y te bekijken en na te gaan in hoeverre die afwijkt van de natuurlijke rangschikking. Hiervoor wordt voor elke rang van Y nagegaan of de daaropvolgende rangen in natuurlijke volgorde staan of niet. In het voorbeeld wordt begonnen met rang 2 links in de tabel; rang 4 komt na rang 2 en staat bijgevolg in natuurlijke volgorde waardoor een +1 wordt toegekend; rang 3 komt eveneens na rang 2 en dus wordt ook hieraan een +1 toegekend; rang 1 komt voor rang 2 en dus is deze volgorde niet natuurlijk en wordt een -1 toegekend. Dus, voor rang 2 geeft dit:

$$(+1) + (+1) + (-1) = +1$$

Voor de 2e rang, rang 4 wordt volgens hetzelfde principe de volgende score verkregen:

$$(-1) + (-1) = -2$$

Voor de 3e rang, rang 3, is het resultaat -1 (rang 1 na rang 3). De som van deze drie resultaten wordt voorgesteld door S:

$$S = +1 - 2 - 1 = -2$$

De maximaal mogelijke waarde voor S bedraagt:

$$\frac{n(n-1)}{2}$$

Door de waarde van S te relateren tot de maximaal mogelijke waarde wordt de correlatiecoëfficient τ verkregen:

$$\tau = \frac{2S}{n(n-1)}$$

Voor het beschouwde voorbeeld is $\tau = -0.33$. Toetsen op significantie van de correlatiecoëfficient kan zowel éénzijdig als tweezijdig.
Voor $n \leq 10$, is τ de toetsstatistiek en kan de p-waarde rechtstreeks uit de specifieke Kendall τ tabel afgelezen worden. Voor het beschouwde voorbeeld vindt men $p = 0.375$, hetgeen wijst op het niet verwerpen van H_0 voor $\alpha = 0.05$.
Indien $n > 10$ is de toetsstatistiek de Z statistiek:

$$Z = \frac{\tau - 0}{\sqrt{\frac{2(2n+5)}{9n(n-1)}}}$$

Voorbeeld. Zie voorbeeld van verband tussen het aantal uren gestudeerd en het overeenkomend aantal punten op een test (cfr. § 4.4.1).

Student	X	Y	rang X	rang Y
4	1	1	1	1
5	2	3	2	2
7	3	7	3	4
1	5	6	4	3
9	7	15	5	5
2	9	16	6	6
3	17	18	7	7
6	21	21	8	9
8	29	20	9	8
10	100	22	10	10

1. *Hypothesen*:

   H_0 : er is geen verband tussen X en Y
   H_1 : er is een verband tussen X en Y

2. *Toetsstatistiek*: $n \leq 10$:

   $$\tau = \frac{2S}{n(n-1)}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $p \leq 0.05$

5. *Berekening*:

$$\tau = \frac{2(41)}{10(10-1)} = 0.91$$

$$S = (9-0)+(8-0)+(6-1)+(6-0)+(5-0)+(4-0)+(3-0)+(1-1)+(1-0) = 41$$

6. *Beslissing*: $p = 0.00 < 0.05 \rightarrow H_0$ verworpen.

7. *Conclusie*: Er is een positieve correlatie tussen het aantal uren studeren en het aantal punten op een test. $p = 0.00$.

4.5 Nominaal meetniveau

4.5.1 Cramer C en ϕ coëfficient

Doel. Toetsen van de samenhang tussen 2 nominale variabelen.

Voorwaarden.

- Twee gegevenssets van ongeordende variabelen. De ene set bevat k klassen, de andere r klassen.
- Minder dan 20% van de cellen hebben een verwachte frekwentie kleiner dan 5 en geen enkele cel heeft een verwachte frekwentie minder dan 1. In voorkomend geval moeten rijen en cellen gecombineerd worden.

Methode. De Cramer C coëfficient wordt berekend op basis van de χ^2 statistiek. De hypothesen luiden:

$$H_0 : \text{Er is geen samenhang tussen de 2 variabelen}$$
$$H_1 : \text{Er is samenhang tussen de 2 variabelen}$$

De geobserveerde frekwenties O_{ij} worden in een $(r \times k)$ contingentietabel gebracht, met r rijen (attribuut of variabele **B**) en k kolommen (attribuut of variabele **A**).

	A_1	...	A_k	Totaal
B_1	O_{11}	...	O_{1k}	$n_{1.}$
...	...	...	...	...
B_r	O_{r1}	...	O_{rk}	$n_{r.}$
Totaal	$n_{.1}$	...	$n_{.k}$	n

De verwachte frekwentie E_{ij} wordt berekend door het produkt van de marginale totalen te delen door het totale aantal:

$$E_{ij} = \frac{n_{i.}n_{.j}}{n}$$

De toetsstatistiek is de χ^2 met $(r-1)(k-1)$ vrijheidsgraden:

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Indien de χ^2 waarde de kritieke waarde overstijgt, dan wordt H_0 verworpen.

De eigenlijke associatiemaat is de Cramer C coëfficient:

$$C = \sqrt{\frac{\chi^2}{n(L-1)}}$$

met

$$L = \min(r, k)$$

De waarde van C is begrepen in $[0, 1]$ en kan nooit negatief zijn, aangezien het nominale gegevens betreft waartussen geen orde-relatie bestaat. Een $C = 0$ geeft aan dat de variabelen onafhankelijk van elkaar zijn en dus geen samenhang vertonen. Een waarde $C = 1$ wijst echter niet op een perfecte correlatie; bovendien kan de relatie slechts éénzijdig zijn, d.i. een zogenaamde *asymmetrische* relatie.

Indien de beide gegevenssets slechts twee klassen bevatten (een 2×2 contingentietabel), en dus dichotoom zijn, dan kan de zogenaamde ϕ coëfficient gebruikt worden.

	$Y_i = 0$	$Y_i = 1$
$X_i = 0$	A	B
$X_i = 1$	C	D

$$\phi = \frac{|AD - BC|}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

De waarde van de ϕ coëfficient is begrepen in $[0, 1]$, met 0 wijzend op geen samenhang. De toetsstatistiek is de χ^2 met 1 vrijheidsgraad:

$$\chi^2 = \frac{n(|AD - BC| - n/2)^2}{(A+B)(C+D)(A+C)(B+D)}$$

Dit type van associatiemaat is niet zondermeer vergelijkbaar met de correlatiecoëfficienten R en R_s. Het voordeel van deze associatiemaat is echter dat de omvang van beide gegevensets niet noodzakelijk gelijk hoeft te zijn.

Voorbeeld. Voor een groep van 815 veroordeelden uit de Benelux wordt de aard van het delict in verband gebracht met de nationaliteit. Bestaat er een significante samenhang tussen de beiden?

	Zedendelict	Vermogensdelict	Geweldsdelict	Totaal
Belg	138 (84.19)	49(74.23)	16(44.58)	203
Nederlander	153(151.37)	150(133.46)	62(80.17)	365
Luxemburger	47(102.44)	99(90.31)	101(54.25)	247
Totaal	338	398	179	815

1. *Hypothesen:*

H_0 : geen samenhang tussen de aard van het delict en de nationaliteit van de veroordeelden

H_1 : samenhang tussen de aard van het delict en de nationaliteit van de veroordeelden

2. *Toetsstatistiek*: χ^2 statistiek met $(r-1)(k-1) = (3-1)(3-1) = 4$

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$$C = \sqrt{\frac{\chi^2}{n(L-1)}}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $\chi^2 \geq 9.49$

5. *Berekening*:

$$\chi^2 = \frac{(138 - 84.19)^2)}{84.19} + \frac{(49 - 74.23)^2)}{74.23} + ... + \frac{(101 - 54.25)^2)}{54.25} = 138.60$$

$$C = \sqrt{\frac{138.60}{815(3-1)}} = 0.29$$

6. *Beslissing*: $138.60 > 9.49 \rightarrow H_0$ verworpen.

7. *Conclusie*: Er is een significante samenhang ten belope van 0.29 tussen de aard van het delict en de nationaliteit van de veroordeelde. $p < 0.01$.

Voorbeeld. Een interview geeft het volgende verband tussen het geslacht en het al of niet lezen van een bepaald tijdschrift.

	Niet-lezer	Lezer	Totaal
Man	21	37	58
Vrouw	26	14	40
Totaal	47	51	98

1. *Hypothesen*:

H_0 : Er is geen samenhang tussen het geslacht en het al of niet lezen van het tijdschrift
H_1 : Er is een samenhang tussen het geslacht en het al of niet lezen van het tijdschrift

2. *Toetsstatistiek*: χ^2 statistiek met 1 vrijheidsgraad

$$\chi^2 = \frac{n(|AD - BC| - n/2)^2}{(A+B)(C+D)(A+C)(B+D)}$$

$$\phi = \frac{|AD - BC|}{\sqrt{(A+B)(C+D)(A+C)(B+D)}}$$

3. *Significantieniveau*: $\alpha = 0.05$.

4. *Beslissingsregel*: H_0 verwerpen als $\chi^2 \geq 3.84$

5. *Berekening*:

$$\chi^2 = \frac{98(|(21)(14) - (37)(26)| - 98/2)^2}{(21+37)(26+14)(21+26)(37+14)} = 6.75$$

$$\phi = \frac{|(21)(14) - (37)(26)|}{\sqrt{(21+37)(26+14)(21+26)(37+14)}} = 0.28$$

6. *Beslissing*: $6.75 > 3.84 \rightarrow H_0$ verworpen.

7. *Conclusie*: Er is een significante samenhang ten belope van 0.28 tussen het geslacht en het al of niet lezen van het tijdschrift.$p < 0.01$

4.6 Samenvatting

1. Een associatiemaat is een gestandaardiseerd getal dat de samenhang tussen fenomenen (hier 2) uitdrukt.

2. Voor metrische en ordinale gegevens zijn associatiematen correlatiecoëfficienten, gedefinieerd tussen -1 en 1 (inbegrepen). De correlatiecoëfficienten geven eveneens de richting van het verband aan.

3. Voor nominale gegevens is door het ontbreken van een volgorde relatie de associatiemaat gedefinieerd tussen 0 en 1 (inbegrepen). Bijgevolg kan enkel de sterkte, maar niet de richting van het verband hiermee aangegeven worden.

4.7 Oefeningen

1. Hiervolgend wordt voor 10 transportbedrijven het bedrag aan verzekeringen en veiligheid gespendeerd per jaar in 10000 BEF (Y) vergeleken met de verkopen in 1000 ton (X). Is er een significante correlatie tussen de beide variabelen? Is de correlatie significant verschillend van 0.95?

X	Y
13	10
18	16
14	12
18	18
23	17
21	17
14	9
25	19
23	17
14	11

2. Bepaal het beduidende karakter van de associatie tussen de reactie op een geneesmiddel en het geslacht.

	Reactie	Geen reactie	Totaal
Man	26	24	50
Vrouw	24	26	50
Totaal	50	50	100

3. Een panel experten rangschikt 15 succesvolle ondernemingen op basis van job-tevredenheidsscore van de werknemers en het groeipotentieel van het bedrijf. Is er directe relatie tussen beiden op basis van de onderstaande rangschikking? Los dit op 2 manieren op. ($\alpha = 0.01$).

Job tevredenheid	Groeipotentieel
1	6
2	3
3	1
4	2
5	4
6	5
7	11
8	10
9	7
10	8
11	12
12	9
13	14
14	13
15	15

4. Bepaal of er een significant verband bestaat tussen de reactie van de huid op ultra-violet licht en de kleur van ogen.

	Hevige reactie	Matige reactie	Geen reactie	Totaal
Blauwe ogen	19	27	4	50
Grijze ogen	7	8	5	20
Bruine ogen	1	13	16	30
Totaal	27	48	25	100

5. 10 professoren van een faculteit worden door de decaan en de faculteitssecretaris onafhankelijk van elkaar gekwoteerd op basis van hun lesgeverskwaliteiten. Is er enige vorm van overeenstemming tussen de 2 rangschikkingen? Toon dit aan op 2 manieren.

Prof.	Decaan	Secretaris
A	5	6
B	4	4
C	3	5
D	1	2
E	2	3
F	6	1
G	7	9
H	10	8
I	9	7
J	8	10

Hoofdstuk 5

Aanvullende oefeningen

1. Drie verzekeringsagenten verkopen elk drie verzekeringspolissen. Kan men op basis van de onderstaande verkoopsgegevens besluiten dat er een significant verband bestaat tussen de verzekeringsagenten en de polissen, t.t.z. dat de variabelen "Polis" en "Agent" niet onafhankelijk zijn?

	Agent 1	Agent 2	Agent 3	Totaal
Polis 1	10	6	14	30
Polis 2	5	2	2	9
Polis 3	5	12	4	21
Totaal	20	20	20	60

2. 1240 sportwinkels verkopen een bepaald merk ping-pong balletjes. De aanbevolen verkoopprijs is 35 BEF. Het bedrijf dat de balletjes fabriceert vermoedt dat de balletjes aan meer dan 35 BEF. worden verkocht. Een steekproef van 100 winkels geeft een gemiddelde verkoopprijs van 38.90 BEF. en een standaarddeviatie van 15 BEF. Is het vermoeden van het bedrijf gerechtvaardigd?

3. De fabricant van batterij A wil aantonen dat zijn batterij minstens 45 minuten langer werkt dan die van zijn concurrent. Hiervoor worden 2 onafhankelijke steekproeven genomen van 100 batterijen van elk merk. De gemiddelde levensduur van batterij A is 308 min. met een standaarddeviatie van 84 min. De gemiddelde levensduur van de batterij van de concurrent is 254 min. met een standaarddeviatie van 67 min. Is de bewering van de fabricant van batterij A correct op basis van de steekproefgegevens?

4. Een manager beweert dat de fractie mensen die meer dan 50% witte producten kopen meer dan 10% hoger ligt in gebieden met hoge werkloosheid dan in die met lage werkloosheid. Om dit aan te tonen wordt een steekproef van 300 kopers genomen in een gebied met hoge werkloosheid en 700 kopers in een gebied met lage werkloosheid. Respectievelijk 120 en 140 kopers daarvan kopen meer dan 50% witte producten. Is de bewering van de manager correct?

5. Een bedrijf wenst een tekstverwerker aan te kopen. Drie paketten (A, B, C) worden getest door drie groepen van telkens 6 werknemers. Elke werknemer werd gevraagd om de voor hem benodigde leertijd te noteren. Zijn er significante verschillen in leertijden tussen de 3 paketten? ($\alpha = 0.01$).

A	B	C
45	30	22
38	40	19
56	28	15
60	44	31
47	25	27
65	42	17

6. Bestaat er een significant positief verband tussen het verloop van de volgende twee beursindices? Het verloop van een index is nooit normaal.

Datum	Index 1	Index 2
1/05	220	151
8/05	218	150
15/05	216	148
21/05	217	149
28/05	215	147
04/06	213	146
11/06	219	152
18/06	236	165
25/06	237	162
02/07	235	161

7. Een steekproef van 9 zout-watervissen geeft een gemiddelde lengte van 8.5 cm en een variantie van 7.2 cm.^{2}. Een andere steekproef van 7 zoet-watervissen levert een gemiddelde van 9 cm. en een variantie van 3.6 cm.^{2}. Is de variabiliteit in de lengte van de zoet- en zoutwatervissen significant verschillend? Toon dit aan met een 0.90 betrouwbaarheidsinterval. Veronderstel dat de beide populaties normaal verdeeld zijn.

8. Een politieke partij wenst te weten hoeveel inwoners van een gemeente voorstander zijn van een gemeentelijk referendum m.b.t. het bouwen van een containerpark naast een woonwijk. De partij wenst hiertoe een enquête te doen bij de inwoners. Indien de enquête een fractie moet opleveren dat met 95% betrouwbaarheid binnen 0.025 van de werkelijke fractie voorstanders moet liggen, hoe groot moet dan de steekproef zijn? Een gelijkaardige enquête in een naburige gemeente leverde een fractie van 60% voorstanders op.

9. De leeftijden van 29 werknemers van een KMO zijn als volgt:
Mannen: 44, 30, 34, 47, 35, 46, 35, 47, 48, 34, 32, 42, 43, 49, 46, 47
Vrouwen: 26, 25, 38, 33, 42, 40, 44, 26, 25, 43, 35, 48, 37
Is er een significant verschil in leeftijd tussen de mannelijke en de vrouwelijke werknemers. De leeftijdsverdelingen zijn scheef (niet symmetrisch), wat betekent dat het gemiddelde geen goede maat is voor de centrale tendens.

10. Een leraar fysica wil de beste methode bepalen waarmee hij de relativiteitstheorie kan uitleggen. Hij doet het op 3 manieren (A, B, C) en vraagt achteraf aan elk van de 18 studenten van zijn klas de door hun verkozen manier. Kan de leraar op basis van deze gegevens iets besluiten?

Student	Voorkeur	Student	Voorkeur
1	A,B,C	10	-
2	B	11	-
3	B	12	B
4	B	13	B
5	B,C	14	B,C
6	-	15	B
7	B	16	A,B
8	A,B	17	A,B,C
9	A,B,C	18	B

11. Een drinkwedstrijd tussen drie studentenclubs wordt afgesloten met de volgende resultaten in liters per deelnemer.
Studentenclub 1: 1, 3, 4.
Studentenclub 2: 4, 5, 6, 6, 7, 8, 9.
Studentenclub 3: 2, 3, 3, 4.
Zijn er significante verschillen tussen de drie studentenclubs m.b.t. hun drankprestaties? Veronderstel normaliteit van de populaties.

12. Een studie beweert dat het aantal nachtelijke auto-ongevallen op kruispunten kan teruggedrongen worden door een betere verlichting van de kruispunten. Voor 12 kruispunten werd een meting gedaan van het aantal auto-ongelukken in het jaar voor de installatie van de verlichting en één in het jaar daarna. Is de stelling van de studie correct? Het aantal auto-ongevallen aan alle vergelijkbare kruispunten kan als normaal verdeeld worden beschouwd.

Locatie	Voor	Na
1	8	5
2	12	3
3	5	2
4	4	1
5	6	4
6	3	2
7	4	2
8	3	4
9	2	3
10	6	5
11	6	4
12	9	3

13. Een lampenbedrijf produceert lampen met een gemiddelde levensduur die normaal verdeeld is met gemiddelde 800 uren en standaardfout 40 uren. Indien men een steekproef van 100 lampen neemt, wat is dan de kans dat de levensduur tussen 778 en 834 uren zal liggen?

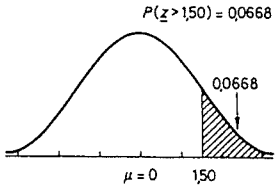
Bibliografie

1. Aczel, A.D. (1993), *Complete Business Statistics*, Irwin, Homewood.
2. Buijs, A. (1994), *Statistiek om mee te werken*, Stenfert Kroese, Leiden.
3. Barrow, M. (1996), *Statistics for Economics, Accounting and Business Studies*, Longman, London.
4. Conover, W.J. (1980), *Practical Nonparametric Statistics*, J. Wiley & Sons, New York.
5. Daniel, W.W. & Terrell, J.C. (1995), *Business Statistics for Management and Economics*, Houghton Mifflin, Boston.
6. Kanji, G.K. (1995), *100 Statistical Tests* , Sage Publications, London.
7. Moors, J.J.A. (1991), *Statistiek in de economie II: Steekproeftheorie en Analyserende Statistiek*, Academic Service Economie en Bedrijfskunde, Schoonhoven.
8. Siegel, S., Castellan, N.J. (1989), *Nonparametric Statistics for the Behavioral Sciences*, McGraw-Hill, Singapore.
9. Triola,M.F., Franklin, L.A. (1994), *Business Statistics*, Addison Wesley, Reading.
10. Wonacott & Wonacott (1990), *Introductory Statistics for Business and Economics*, J. Wiley, Singapore.
11. *Schaum's outline series of theory and problems of Statistics*, Spiegel.
12. *The Statistics Problem Solver*, REA, New York.

Tabellen

de standaardnormale verdeling

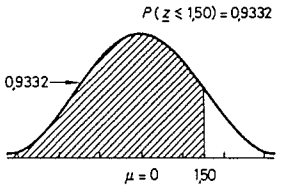
Aangegeven is $P(\underline{z}>z)$.



	Tweede decimaal van z									
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0,0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0,1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0,2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0,3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0,4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0,5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0,6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0,7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0,8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0,9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1,0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1,1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1,2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1,3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1,4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
1,5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1,6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1,7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1,8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1,9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2,0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2,1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2,2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2,3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2,4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2,5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2,6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2,7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2,8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2,9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3,0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010

de cumulatieve standaardnormale verdeling

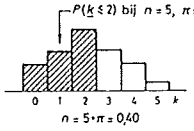
Aangegeven is $F(z)=P(\underline{z}\leq z)$.



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

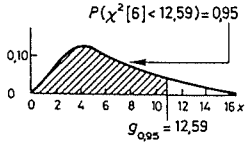
de cumulatieve binomiale verdeling

Cumulatieve kansen bij de binomiale verdeling $F(k)=P(\underline{k}\leq k)$ voor $n=2, \ldots, 10, 12, 15$ en 20



n	k	π = 0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9928	0.9720	0.9392	0.8960	0.8438	0.7840	0.7182	0.6480	0.5748	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.1160	0.0778	0.0503	0.0312
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.4284	0.3370	0.2562	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.7648	0.6826	0.5931	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9460	0.9130	0.8688	0.8125
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9947	0.9898	0.9815	0.9688
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1176	0.0754	0.0647	0.0277	0.0156
	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.3191	0.2333	0.1636	0.1094
	2	0.9978	0.9842	0.9527	0.9011	0.8306	0.7443	0.6471	0.5443	0.4415	0.3438
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8826	0.8208	0.7447	0.6562
	4	1.0000	0.9999	0.9996	0.9984	0.9954	0.9891	0.9777	0.9590	0.9308	0.8906
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9982	0.9959	0.9917	0.9844
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0490	0.0280	0.0152	0.0078
	1	0.9556	0.8503	0.7166	0.5767	0.4449	0.3294	0.2338	0.1586	0.1024	0.0625
	2	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.5323	0.4199	0.3164	0.2266
	3	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.8002	0.7102	0.6083	0.5000
	4	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9444	0.9037	0.8471	0.7734
	5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9910	0.9812	0.9643	0.9375
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9994	0.9984	0.9963	0.9922
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0319	0.0168	0.0084	0.0039
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1691	0.1064	0.0632	0.0352
	2	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.4278	0.3154	0.2201	0.1445
	3	0.9996	0.9950	0.9786	0.9437	0.8862	0.8059	0.7064	0.5941	0.4770	0.3633
	4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8939	0.8263	0.7396	0.6367
	5	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9747	0.9502	0.9115	0.8555
	6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9964	0.9915	0.9819	0.9648
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9983	0.9961
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0404	0.0207	0.0101	0.0046	0.0020
	1	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.1211	0.0705	0.0385	0.0195
	2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.3373	0.2318	0.1495	0.0898
	3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.6089	0.4826	0.3614	0.2539
	4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.8283	0.7334	0.6214	0.5000
	5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9464	0.9006	0.8342	0.7461
	6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9888	0.9750	0.9502	0.9102
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9986	0.9962	0.9909	0.9805
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9992	0.9980
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0282	0.0135	0.0060	0.0025	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0860	0.0464	0.0232	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.2616	0.1673	0.0996	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.5138	0.3823	0.2660	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.7515	0.6331	0.5044	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9527	0.9051	0.8338	0.7384	0.6230
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9740	0.9452	0.8980	0.8281
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9952	0.9877	0.9726	0.9453
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9983	0.9955	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990
12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0057	0.0022	0.0008	0.0002
	1	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0424	0.0196	0.0083	0.0032
	2	0.9804	0.8891	0.7358	0.5583	0.3907	0.2528	0.1513	0.0834	0.0421	0.0193
	3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.3467	0.2253	0.1345	0.0730
	4	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.5833	0.4382	0.3044	0.1938
	5	1.0000	0.9995	0.9954	0.9806	0.9456	0.8822	0.7873	0.6652	0.5269	0.3872
	6	1.0000	0.9999	0.9993	0.9961	0.9857	0.9614	0.9154	0.8418	0.7393	0.6128
	7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9745	0.9427	0.8883	0.8062
	8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9944	0.9847	0.9644	0.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9992	0.9972	0.9921	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998
15	0	0.4633	0.2059	0.0874	0.0352	0.0134	0.0047	0.0016	0.0005	0.0001	0.0000
	1	0.8290	0.5490	0.3186	0.1671	0.0802	0.0353	0.0142	0.0052	0.0017	0.0005
	2	0.9638	0.8159	0.6042	0.3980	0.2361	0.1268	0.0617	0.0271	0.0107	0.0037
	3	0.9945	0.9444	0.8227	0.6482	0.4613	0.2969	0.1727	0.0905	0.0424	0.0176
	4	0.9994	0.9873	0.9383	0.8358	0.6865	0.5155	0.3519	0.2173	0.1204	0.0592
	5	0.9999	0.9978	0.9832	0.9389	0.8516	0.7216	0.5643	0.4032	0.2608	0.1509
	6	1.0000	0.9997	0.9964	0.9819	0.9434	0.8689	0.7548	0.6098	0.4522	0.3036
	7	1.0000	1.0000	0.9996	0.9958	0.9827	0.9500	0.8868	0.7869	0.6535	0.5000
	8	1.0000	1.0000	0.9999	0.9992	0.9958	0.9848	0.9578	0.9050	0.8182	0.6964
	9	1.0000	1.0000	1.0000	0.9999	0.9992	0.9963	0.9876	0.9662	0.9231	0.8491
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9972	0.9907	0.9745	0.9408
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995	0.9981	0.9937	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9989	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.3585	0.1216	0.0388	0.0115	0.0032	0.0008	0.0002	0.0000	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119
	10	1.0000	1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

enkele kritieke grenzen voor de χ^2-verdeling

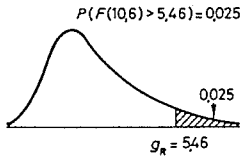


Aangegeven zijn: grenswaarden g_α met $P(\chi^2 < g_\alpha) = \alpha$ v is het aantal vrijheidsgraden

v	1	2	3	4	5	6	7	8	9	10
$g_{0,01}$	–	0,02	0,11	0,30	0,55	0,87	1,24	1,65	2,09	2,56
$g_{0,025}$	–	0,05	0,22	0,48	0,83	1,24	1,69	2,18	2,70	3,25
$g_{0,05}$	–	0,10	0,35	0,71	1,15	1,64	2,17	2,73	3,33	3,94
$g_{0,95}$	3,84	5,99	7,81	9,49	11,07	12,59	14,07	15,51	16,92	18,31
$g_{0,975}$	5,02	7,38	9,35	11,41	12,83	14,45	16,01	17,53	19,02	20,48
$g_{0,99}$	6,63	9,21	11,34	13,28	15,09	16,81	18,48	20,09	21,67	23,21

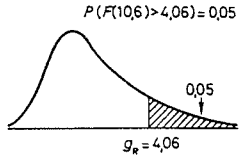
v	12	14	16	18	20	25	30	50	100
$g_{0,01}$	3,57	4,66	5,81	7,01	8,26	11,52	14,95	29,71	70,06
$g_{0,025}$	4,40	5,63	6,91	8,23	9,59	13,12	16,79	32,36	74,22
$g_{0,05}$	5,23	6,57	7,96	9,39	10,85	14,61	18,49	34,76	77,93
$g_{0,95}$	21,03	23,68	26,30	28,87	31,41	37,65	43,77	67,50	124,34
$g_{0,975}$	23,34	26,12	28,85	31,53	34,17	40,65	46,98	71,42	129,56
$g_{0,99}$	26,22	29,14	32,00	34,81	37,57	44,31	50,89	76,15	135,81

de F-verdeling



Overschrijdingskans 0,025 Dus $P(F(v_1, v_2) > g_R) = 0{,}025$

v_1 / v_2	1	2	3	4	5	6	8	10	15	20	30
1	647	799	864	900	922	937	957	969	985	993	1001
2	38,5	39,0	39,2	39,2	39,3	39,3	39,4	39,4	39,4	39,5	39,5
3	17,4	16,0	15,4	15,1	14,9	14,7	14,5	14,4	14,2	14,2	14,1
4	12,2	10,65	9,98	9,60	9,36	9,20	8,98	8,84	8,66	8,56	8,46
5	10,0	8,43	7,76	7,39	7,15	6,98	6,76	6,62	6,43	6,33	6,23
6	8,81	7,26	6,60	6,23	5,99	5,82	5,60	5,46	5,27	5,17	5,07
7	8,07	6,54	5,89	5,52	5,29	5,21	4,90	4,76	4,57	4,47	4,36
8	7,57	6,60	5,42	5,05	4,82	4,65	4,43	4,30	4,10	4,00	3,89
9	7,21	5,71	5,08	4,72	4,48	4,32	4,10	3,96	3,77	3,67	3,56
10	6,94	5,46	4,83	4,47	4,24	4,07	3,85	3,72	3,52	3,42	3,31
15	6,20	4,77	4,15	3,80	3,58	3,41	3,20	3,06	2,86	2,76	2,64
20	5,87	4,46	3,86	3,51	3,29	3,13	2,91	2,77	2,57	2,46	2,35
30	5,57	4,18	3,59	3,25	3,03	3,87	2,65	2,51	2,31	2,20	2,07



Overschrijdingskans 0,05 Dus $P(F(v_1, v_2) > g_R) = 0{,}05$

v_1 / v_2	1	2	3	4	5	6	8	10	12	16	20
1	161	200	216	225	230	234	239	242	244	246	248
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37	19,39	19,41	19,43	19,44
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84	8,78	8,74	8,69	8,66
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04	5,96	5,91	5,84	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82	4,74	4,68	4,60	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15	4,06	4,00	3,92	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73	3,64	3,57	3,49	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44	3,34	3,28	3,20	3,15
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07	2,97	2,91	2,82	2,77
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85	2,76	2,69	2,60	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59	2,49	2,42	2,33	2,28
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45	2,35	2,28	2,18	2,12
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34	2,24	2,16	2,06	2,00
30	4,17	3,32	2,92	2,69	2,53	2,42	s,27	2,16	2,09	1,99	1,93
50	4,03	3,18	2,79	2,56	2,40	2,29	2,13	2,02	1,95	1,85	1,78

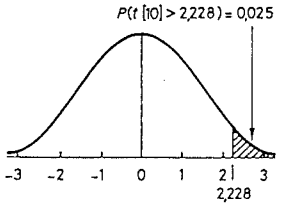
de cumulatieve Poisson-verdeling[1]

Aangegeven is telkens $P(k \leq k)$

$k=$	$\mu=4{,}0$	$\mu=4{,}5$	$\mu=5{,}0$	$\mu=6{,}0$	$\mu=7{,}0$	$\mu=8{,}0$	$\mu=9{,}0$	$\mu=10{,}0$
0	.0183	.0111	.0067	.0025	.0009	.0003	.0001	.0000
1	.0916	.0611	.0404	.0174	.0073	.0030	.0012	.0005
2	.2381	.1736	.1247	.0620	.0296	.0138	.0062	.0028
3	.4335	.3423	.2650	.1512	.0818	.0424	.0212	.0103
4	.6288	.5321	.4405	.2851	.1730	.0996	.0550	.0293
5	.7851	.7029	.6160	.4457	.3007	.1912	.1157	.0671
6	.8893	.8311	.7622	.6063	.4497	.3134	.2068	.1301
7	.9489	.9134	.8666	.7440	.5987	.4530	.3239	.2202
8	.9786	.9597	.9319	.8472	.7291	.5926	.4557	.3328
9	.9919	.9829	.9682	.9161	.8305	.7166	.5874	.4579
10	.9972	.9933	.9863	.9574	.9015	.8159	.7060	.5830
11	.9991	.9976	.9945	.9799	.9467	.8881	.8030	.6968
12	.9997	.9992	.9980	.9912	.9730	.9362	.8758	.7916
13	.9999	.9997	.9993	.9964	.9872	.9658	.9261	.8645
14	1.0000	.9999	.9998	.9986	.9943	.9827	.9585	.9165
15		1.0000	.9999	.9995	.9976	.9918	.9780	.9513
16			1.0000	.9998	.9990	.9963	.9889	.9730
17				.9999	.9996	.9984	.9947	.9857
18				1.0000	.9999	.9993	.9976	.9928
19					1.0000	.9997	.9989	.9965
20						.9999	.9996	.9984
21						1.0000	.9998	.9993
22							.9999	.9997
23							1.0000	.9999
24								1.0000

1. Als gevolg van afrondingsverschillen kunnen de cumulatieve kansen die in deze tabel worden afgelezen in geringe mate afwijken van cumulatieve kansen die door optelling zijn berekend met de gewone tabel van de Poisson-verdeling.

de *t*-verdeling



Aantal vrijheidsgraden v	rechteroverschrijdingskans				
	.1	.05	.025	.01	.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Critical values of r in the runs test*

Given in the tables are various critical values of r for values of m and n less than or equal to 20. For the one-sample runs test, any observed value of r which is less than or equal to the smaller value, or is greater than or equal to the larger value in a pair is significant at the $\alpha = .05$ level.

m \ n	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2											2 –	2 –	2 –	2 –	2 –	2 –	2 –	2 –	2 –
3					2 –	2 –	2 –	2 –	2 –	2 –	2 –	2 –	2 –	3 –	3 –	3 –	3 –	3 –	3 –
4				2 9	2 9	2 –	3 –	3 –	3 –	3 –	3 –	3 –	3 –	3 –	4 –	4 –	4 –	4 –	4 –
5			2 9	2 10	3 10	3 11	3 11	3 –	3 –	4 –	4 –	4 –	4 –	4 –	4 –	4 –	5 –	5 –	5 –
6		2 –	2 9	3 10	3 11	3 12	3 12	4 13	4 13	4 13	4 13	5 –	5 –	5 –	5 –	5 –	5 –	6 –	6 –
7		2 –	2 –	3 11	3 12	3 13	4 13	4 14	5 14	5 14	5 14	5 15	5 15	6 15	6 –	6 –	6 –	6 –	6 –
8		2 –	3 –	3 11	3 12	4 13	4 14	5 14	5 15	5 15	6 16	6 16	6 16	6 16	6 17	7 17	7 17	7 17	7 17
9		2 –	3 –	3 –	4 13	4 14	5 14	5 15	5 16	6 16	6 16	6 17	7 17	7 18	7 18	7 18	8 18	8 18	8 18
10		2 –	3 –	3 –	4 13	5 14	5 15	5 16	6 16	6 17	7 17	7 18	7 18	7 18	8 19	8 19	8 19	8 20	9 20
11		2 –	3 –	4 –	4 13	5 14	5 15	6 16	6 17	7 17	7 18	7 19	8 19	8 19	8 20	9 20	9 20	9 21	9 21
12	2 –	2 –	3 –	4 –	4 13	5 14	6 16	6 16	7 17	7 18	7 19	8 19	8 20	8 20	9 21	9 21	9 21	10 22	10 22
13	2 –	2 –	3 –	4 –	5 –	5 15	6 16	6 17	7 18	7 19	8 19	8 20	9 20	9 21	9 21	10 22	10 22	10 23	10 23
14	2 –	2 –	3 –	4 –	5 –	5 15	6 16	7 17	7 18	8 19	8 20	9 20	9 21	9 22	10 22	10 23	10 23	11 23	11 24
15	2 –	3 –	3 –	4 –	5 –	6 15	6 16	7 18	7 18	8 19	8 20	9 21	9 22	10 22	10 23	11 23	11 24	11 24	12 25
16	2 –	3 –	4 –	4 –	5 –	6 –	6 17	7 18	8 19	8 20	9 21	9 21	10 22	10 23	11 23	11 24	11 25	12 25	12 25
17	2 –	3 –	4 –	4 –	5 –	6 –	7 17	7 18	8 19	9 20	9 21	10 22	10 23	11 23	11 24	11 25	12 25	12 26	13 26
18	2 –	3 –	4 –	5 –	5 –	6 –	7 17	8 18	8 19	9 20	9 21	10 22	10 23	11 24	11 25	12 25	12 26	13 26	13 27
19	2 –	3 –	4 –	5 –	6 –	6 –	7 17	8 18	8 20	9 21	10 22	10 23	11 23	11 24	12 25	12 26	13 26	13 27	13 27
20	2 –	3 –	4 –	5 –	6 –	6 –	7 17	8 18	9 20	9 21	10 22	10 23	11 24	12 25	12 25	13 26	13 27	13 27	14 28

* Adapted from Swed, and Eisenhart, C. (1943). Tables for testing randomness of grouping in a sequence of alternatives. *Annals of Mathematical Statistics*, 14, 83–86, with the kind permission of the authors and publisher.

Critical values of D in the Kolmogorov-Smirnov one-sample test*

Sample size (N)	Level of significance for D = maximum $\lvert F_0(X) - S_N(X)\rvert$				
	.20	.15	.10	.05	.01
1	.900	.925	.950	.975	.995
2	.684	.726	.776	.842	.929
3	.565	.597	.642	.708	.828
4	.494	.525	.564	.624	.733
5	.446	.474	.510	.565	.669
6	.410	.436	.470	.521	.618
7	.381	.405	.438	.486	.577
8	.358	.381	.411	.457	.543
9	.339	.360	.388	.432	.514
10	.322	.342	.368	.410	.490
11	.307	.326	.352	.391	.468
12	.295	.313	.338	.375	.450
13	.284	.302	.325	.361	.433
14	.274	.292	.314	.349	.418
15	.266	.283	.304	.338	.404
16	.258	.274	.295	.328	.392
17	.250	.266	.286	.318	.381
18	.244	.259	.278	.309	.371
19	.237	.252	.272	.301	.363
20	.231	.246	.264	.294	.356
25	.21	.22	.24	.27	.32
30	.19	.20	.22	.24	.29
35	.18	.19	.21	.23	.27
Over 35	$\frac{1.07}{\sqrt{N}}$	$\frac{1.14}{\sqrt{N}}$	$\frac{1.22}{\sqrt{N}}$	$\frac{1.36}{\sqrt{N}}$	$\frac{1.63}{\sqrt{N}}$

* Adapted from Massey, F. J., Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, 46, 70, with the kind permission of the author and publisher.

Table of Critical Values of T in the Wilcoxon Matched-pairs Signed-ranks Test*

N	Level of significance for one-tailed test		
	.025	.01	.005
	Level of significance for two-tailed test		
	.05	.02	.01
6	0	—	—
7	2	0	—
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

* Adapted from Table I of Wilcoxon, F. 1949. *Some rapid approximate statistical procedures.* New York: American Cyanamid Company, p. 13, with the kind permission of the author and publisher.

Critical values of the smallest rank sum for the Wilcoxon–Mann–Whitney test

n_1 = number of elements in the largest sample;
n_2 = number of elements in the smallest sample.

		Level of significance α						Level of significance α			
Two-sided		0.20	0.10	0.05	0.01	Two-sided	0.20	0.10	0.05	0.01
One-sided		0.10	0.05	0.025	0.005	One-sided	0.10	0.05	0.025	0.005
n_1	n_2					n_1	n_2			
3	2	3	–	–	–	10	6	38	35	32	27
3	3	7	6	–	–	10	7	49	45	42	37
4	2	3	–	–	–	10	8	60	56	53	47
4	3	7	6	–	–	10	9	73	69	65	58
4	4	13	11	10	–	10	10	87	82	78	71
5	2	4	3	–	–	11	1	1	–	–	–
5	3	8	7	6	–	11	2	6	4	3	–
5	4	14	12	11	–	11	3	13	11	9	6
5	5	20	19	17	15	11	4	21	18	16	12
6	2	4	3	–	–	11	5	30	27	24	20
6	3	9	8	7	–	11	6	40	37	34	28
6	4	15	13	12	10	11	7	51	47	44	38
6	5	22	20	18	16	11	8	63	59	55	49
6	6	30	28	26	13	11	9	76	72	68	61
7	2	4	3	–	–	11	10	91	86	81	73
7	3	10	8	7	–	11	11	106	100	96	87
7	4	16	14	13	10	12	1	1	–	–	–
7	5	23	21	20	16	12	2	7	5	4	–
7	6	32	29	27	24	12	3	14	11	10	7
7	7	41	39	36	32	12	4	22	19	17	13
8	2	5	4	3	–	12	5	32	28	26	21
8	3	11	9	8	–	12	6	42	38	35	30
8	4	17	15	14	11	12	7	54	49	46	40
8	5	25	23	21	17	12	8	66	62	58	51
8	6	34	31	29	25	12	9	80	75	71	63
8	7	44	41	38	34	12	10	94	89	84	76
8	8	55	51	49	43	12	11	110	104	99	90
9	1	1	–	–	–	12	12	127	120	115	105
9	2	5	4	3	–	13	1	–	–	–	–
9	3	11	9	8	6	13	2	7	5	4	–
9	4	19	16	14	11	13	3	15	12	10	7
9	5	27	24	22	18	13	4	23	20	18	14
9	6	36	33	31	26	13	5	33	30	27	22
9	7	46	43	40	35	13	6	44	40	37	31
9	8	58	54	51	45	13	7	56	52	48	44
9	9	70	66	62	56	13	8	69	64	60	53
10	1	1	–	–	–	13	9	83	78	73	65
10	2	6	4	3	–	13	10	98	92	88	79
10	3	12	10	9	6	13	11	114	108	103	93
10	4	20	17	15	12	13	12	131	125	119	109
10	5	28	26	23	19	13	13	149	142	136	125

		Level of significance α						Level of significance α			
Two-sided		0.20	0.10	0.05	0.01	Two-sided	0.20	0.10	0.05	0.01
One-sided		0.10	0.05	0.025	0.005	One-sided	0.10	0.05	0.025	0.005
n_1	n_2					n_1	n_2			
14	1	1	–	–	–	17	4	28	25	21	16
14	2	7	5	4	–	17	5	40	35	32	25
14	3	16	13	11	7	17	6	52	47	43	36
14	4	25	21	19	14	17	7	66	61	56	47
14	5	35	31	28	22	17	8	81	75	70	60
14	6	46	42	38	32	17	9	97	90	84	74
14	7	59	54	50	43	17	10	113	106	100	89
14	8	72	67	62	54	17	11	131	123	117	105
14	9	86	81	76	67	17	12	150	142	135	122
14	10	102	96	91	81	17	13	170	161	154	140
14	11	118	112	106	96	17	14	190	182	174	159
14	12	136	129	123	112	17	15	212	203	195	180
14	13	154	147	141	129	17	16	235	225	217	201
14	14	174	166	160	147	17	17	259	249	240	223
15	1	1	–	–	–	18	1	1	–	–	–
15	2	8	6	4	–	18	2	9	7	5	–
15	3	16	13	11	8	18	3	19	15	13	8
15	4	26	22	20	15	18	4	30	26	22	16
15	5	37	33	29	23	18	5	42	37	33	26
15	6	48	44	40	33	18	6	55	49	45	37
15	7	61	56	52	44	18	7	69	63	58	49
15	8	75	69	65	56	18	8	84	77	72	62
15	9	90	84	79	69	18	9	100	93	87	76
15	10	106	99	94	84	18	10	117	110	103	92
15	11	123	116	110	99	18	11	135	127	121	108
15	12	141	133	127	115	18	12	155	146	139	125
15	13	159	152	145	133	18	13	175	166	158	144
15	14	179	171	164	151	18	14	196	187	179	163
15	15	200	192	184	171	18	15	218	208	200	184
16	1	1	–	–	–	18	16	242	231	222	206
16	2	8	6	4	–	18	17	266	255	246	228
16	3	17	14	12	8	18	18	291	280	270	252
16	4	27	24	21	15	19	1	2	1	–	–
16	5	38	34	30	24	19	2	10	7	5	3
16	6	50	46	42	34	19	3	20	16	13	9
16	7	64	58	54	46	19	4	31	27	23	17
16	8	78	72	67	58	19	5	43	38	34	27
16	9	93	87	82	72	19	6	57	51	46	38
16	10	109	103	97	86	19	7	71	65	60	50
16	11	127	120	113	102	19	8	87	80	74	64
16	12	145	138	131	119	19	9	103	96	90	78
16	13	165	156	150	130	19	10	121	113	107	94
16	14	185	176	169	155	19	11	139	131	124	111
16	15	206	197	190	175	19	12	159	150	143	129
16	16	229	219	211	196	19	13	180	171	163	147
17	1	1	–	–	–	19	14	202	192	182	168
17	2	9	6	5	–	19	15	224	214	205	189
17	3	18	15	12	8	19	16	248	237	228	210

Critical values of the Spearman test statistic

Approximate upper-tail critical values, r_s^*, where $P(r_s > r_s^*) \leq \alpha$, $n = 4(1)30$

	Significance Level, α					
n	0.001	0.005	0.010	0.025	0.050	0.100
4	——	——	——	——	.8000	.8000
5	——	——	.9000	.9000	.8000	.7000
6	——	.9429	.8857	.8286	.7714	.6000
7	.9643	.8929	.8571	.7450	.6786	.5357
8	.9286	.8571	.8095	.7143	.6190	.5000
9	.9000	.8167	.7667	.6833	.5833	.4667
10	.8667	.7818	.7333	.6364	.5515	.4424
11	.8364	.7545	.7000	.6091	.5273	.4182
12	.8182	.7273	.6713	.5804	.4965	.3986
13	.7912	.6978	.6429	.5549	.4780	.3791
14	.7670	.6747	.6220	.5341	.4593	.3626
15	.7464	.6536	.6000	.5179	.4429	.3500
16	.7265	.6324	.5824	.5000	.4265	.3382
17	.7083	.6152	.5637	.4853	.4118	.3260
18	.6904	.5975	.5480	.4716	.3994	.3148
19	.6737	.5825	.5333	.4579	.3895	.3070
20	.6586	.5684	.5203	.4451	.3789	.2977
21	.6455	.5545	.5078	.4351	.3688	.2909
22	.6318	.5426	.4963	.4241	.3597	.2829
23	.6186	.5306	.4852	.4150	.3518	.2767
24	.6070	.5200	.4748	.4061	.3435	.2704
25	.5962	.5100	.4654	.3977	.3362	.2646
26	.5856	.5002	.4564	.3894	.3299	.2588
27	.5757	.4915	.4481	.3822	.3236	.2540
28	.5660	.4828	.4401	.3749	.3175	.2490
29	.5567	.4744	.4320	.3685	.3113	.2443
30	.5479	.4665	.4251	.3620	.3059	.2400

Note: The corresponding lower-tail critical value for r_s is $-r_s^*$.

Upper-tail probabilities for T, the Kendall rank-order correlation coefficient ($N \leq 10$)*

Entries are $p = P[T \geq$ tabled value$]$.

N	T	p	N	T	p	N	T	p	N	T	p
4	.000	.625	7	.048	.500	9	.000	.540	10	.022	.500
	.333	.375		.143	.386		.056	.460		.067	.431
	.667	.167		.238	.281		.111	.381		.111	.364
	1.000	.042		.333	.191		.167	.306		.156	.300
				.429	.119		.222	.238		.200	.242
5	.000	.592		.524	.068		.278	.179		.244	.190
	.200	.408		.619	.035		.333	.130		.289	.146
	.400	.242		.714	.015		.389	.090		.333	.108
	.600	.117		.810	.005		.444	.060		.378	.078
	.800	.042		.905	.001		.500	.038		.422	.054
	1.000	.008		1.000	.000		.556	.022		.467	.036
							.611	.012		.511	.023
6	.067	.500	8	.000	.548		.667	.006		.556	.014
	.200	.360		.071	.452		.722	.003		.600	.008
	.333	.235		.143	.360		.778	.001		.644	.005
	.467	.136		.214	.274		.833	.000		.689	.002
	.600	.068		.286	.199		.889	.000		.733	.001
	.733	.028		.357	.138		.944	.000		.778	.000
	.867	.008		.429	.089		1.000	.000		.822	.000
	1.000	.001		.500	.054					.867	.000
				.571	.031					.911	.000
				.643	.016					.956	.000
				.714	.007					1.000	.000
				.786	.003						
				.857	.001						
				.929	.000						
				1.000	.000						

* Adapted and reproduced by permission of the publishers Charles Griffin & Co. Ltd., 16 Pembridge Road, London W11 3HL, from Appendix Table 5 of Kendall, M. G. (1970). *Rank correlation methods* (fourth edition).