Onderzoekstechnieken

Deel I. Theorie

1 Onderzoekproces

1.1 Wetenschappelijke methode

1.2 Basisconcepten in onderzoek

1.2.1 Meetniveaus

> Definitie 1.2.1 (Variabele): Algemene eigenschap van een object waardoor we objecten van elkaar kunnen onderscheiden

> Definitie 1.2.2 (Waarde): Specifieke eigenschap, invulling voor die variabele

Kwalitatief

Nominaal beperkt aantal categorieën, geen volgorde

Ordinaal categorieën met logische volgorde

Kwantitatief

Interval metingen (getal + meeteenheid), geen nulpunt

Ratio intervalniveau met nulpunt

1.2.2 Onderzoekproces

1. Formuleren van de probleemstelling: wat is de onderzoeksvraag
2. Exacte informatiebehoefte definiëren: welke specifieke vragen moeten we stellen

3. Uitvoeren van het onderzoek: enquêtes, simulaties, ...
4. Verwerken van de gegevens: statistische software
5. Analyseren van de gegevens: uitvoeren van de statistische methodes
6. Conclusies schrijven: schrijven van onderzoeksverslag

Definitie 1.2.3 (Oorzakelijk verband): Een variabele veroorzaakt een oorzakelijk verband wanneer een verandering in die variabele op een betrouwbare manier een geassocieerde verandering van een andere variabele tot gevolg heeft, op voorwaarde dat alle andere potentiële oorzaken geëlimineerd zijn.

2 Analyse op 1 variabele

2.1 Gemiddelde

Definitie 2.1.1 (Gemiddelde): Het gemiddelde (symbool μ) van een set waarden is de som van al deze waarden gedeeld door het aantal waarden. De formule staat beschreven in 2.1.1.

$$\mu = \frac{1}{n} \times \sum_{i=1}^{n} x_i \tag{2.1.1}$$

2.2 Mediaan

Definitie 2.2.1 (Mediaan): Indien we alle cijfers sorteren van klein naar groot, is de mediaan het middelste cijfer, of het gemiddelde van de twee middelste cijfers indien het aantal cijfers oneven is.

2.3 Modus

Definitie 2.3.1 (Modus): De modus is het cijfer dat het meest voorkomt in een set van cijfers.

2.4 Bereik

Definitie 2.4.1 (Bereik): Het bereik in een set van getallen is de absolute waarde van het verschil tussen het laagste en grootste getal.

2.5 Kwartielen & kwartielafstand

Definitie 2.5.1 (Kwartielen): De kwartielen zijn de waarden die de lijst van nummers in 4 gelijke delen deelt. Elk deel vormt dus een kwart van de dataset. Men spreekt van een eerste, tweede en derde kwartiel (Q_1, Q_2, Q_3).

Definitie 2.5.2 (Kwartielafstand): Kwartielafstand is het verschil tussen Q_3 en Q_1 (dus $Q_3 - Q_1$).

2.6 Variantie & standaardafwijking

Definitie 2.6.1 (Variantie): De **variantie** is een maat voor de spreiding van een reeks waarden, dat wil zeggen de mate waarin de waarden onderling verschillen. Hoe groter de variantie, hoe meer de afzonderlijke waarden onderling verschillen, en dus ook hoe meer de waarden van het "gemiddelde" afwijken.

$$\sigma^2 = \frac{1}{n} \times \sum_{i=1}^{n} (x_i - \mu)^2 \tag{2.6.1}$$

Definitie 2.6.2 (Standaardafwijking): De **standaardafwijking** of **standaarddeviatie** is gedefinieerd als de wortel uit de variantie, en daardoor vergelijkbaar met de waarden van de variabele zelf.

$$\sigma = \sqrt{\sigma^2} \tag{2.6.2}$$

2.7 Centrum- en spreidingsmaten

Analyse	**Nominaal**	**Ordinaal**	**Interval** of **Ratio**
Centrum	Modus Modale klasse	Mediaan Modus Modale klasse	Gemiddelde Mediaan Modale klasse
Spreiding		Range Interkwartielafstand	Range Interkwartielafstand Standaarddeviatie

Tab. 1: Meetniveaus en mogelijkheden op variabelen

2.8 Grafieken

2.8.1 Boxplot

De boxplot wordt gevormd door een rechthoek begrensd door de kwartielwaarden (25% en 75%). In deze rechthoek wordt ook de mediaan getekend. De stelen, die aan de rechthoek zitten, bevatten de rest van de waarnemingen op de uitschieters en extremen na.

Uitschieter Een uitschieter is een waarde die meer dan 1.5 keer de interkwartielafstand boven/onder het derde/eerste kwartiel ligt. Wordt aangeduid met een cirkeltje.

Extremum Een extremum is een waarde die meer dan 3 keer de interkwartielafstand boven/onder het derde/eerste kwartiel ligt. Wordt aangeduid met een sterretje.

3 Analyse op 2 variabelen

3.1 Kruistabellen

Definitie 3.1.1 (Kruistabel): In een kruistabel wordt de onafhankelijke variabele in functie van de afhankelijke variabele uitgezet.

3.2 χ^2est voor associatie

1. Stel de kruistabel op samen met marginale totalen.
2. Stel voor elke cel een schatter (expected) op voor de theoretische kans om in die cel te geraken.
$$e = (\frac{n_{rij}}{n} \times \frac{n_{kolom}}{n}) \times n = \frac{n_{rij} \times n_{kolom}}{n} \tag{3.2.1}$$
3. Bereken het verschil tussen geobserveerde (notatie a) en verwachte frequentie (e).
4. Bereken de maat van afwijking voor elke cel.
$$\frac{(a-e)^2}{e} \tag{3.2.2}$$
5. Deze gekwadrateerde deviaties gaan we dan optellen en vormt de χ^2 [1]
$$\chi^2 = \sum \frac{(a-e)^2}{e} \tag{3.2.3}$$

[1] Let op dat er afgerond wordt.

3.3 Cramér's V

Definitie 3.3.1 (Cramér's V):

$$V = \sqrt{\frac{\chi^2}{n(k-1)}} \tag{3.3.1}$$

waarbij:

χ^2 = berekende chi-kwadraatwaarde
n = aantal waarnemingen
k = kleinste waarde van het aantal kolommen of rijen van de tabel

Interpretatie

$$V \approx \begin{cases} 0 & \text{geen samenhang} \\ 0,1 & \text{zwakke samenhang} \\ 0,25 & \text{redelijk sterke samenhang} \\ 0,50 & \text{sterke samenhang} \\ 0,75 & \text{zeer sterke samenhang} \\ 1 & \text{volledige samenhang} \end{cases}$$

3.4 Regressie

Monotoon Een monotoon verband is een verband waarbij de onderzoeker de algemene richting van de samenhang tussen de twee variabelen kan aanduiden, hetzij stijgend, hetzij dalend. De richting van het verband verandert nooit.

Niet-monotoon Bij een niet-monotoon verband wordt de aanwezigheid (of afwezigheid) van de ene variabele systematisch gerelateerd aan de aanwezigheid (of afwezigheid) van een andere variabele. De richting van het verband kan echter niet aangeduid worden.

Stelling 3.4.1:

$$y = \beta_0 + \beta_1 x \tag{3.4.1}$$

waarbij:

y = de afhankelijke
x = de onafhankelijke

3.5 Correlatie

3.5.1 Pearsons product-momentcorrelatiecoëfficiënt

Definitie 3.5.1 (Pearsons product-momentcorrelatiecoëfficiënt): Pearsons product momentcorrelatiecoëfficiënt (R): een maat voor de sterkte van de lineaire samenhang tussen X en Y. De waarde kan variëren van -1 tot 1. Hoe dichter de correlatiecoëfficiënt bij 1 of -1, hoe beter de kwaliteit van het lineair model.

$R > 0$ = positief lineair verband
$R < 0$ = negatief lineair verband
$R = 0$ = geen lineaire samenhang

3.6 Determiniatiecoëfficiënt

Definitie 3.6.1: De determiniatiecoëfficiënt (R^2) is het kwadraat van de correlatiecoëfficiënt en verklaart het percentage van de variantie van de waargenomen waarden t.o.v. de regressierechte.

R^2 = de verklaarde variantie
$1 - R^2$ = de onverklaarde variantie

4 Steekproefonderzoek

4.1 Populatie & Steekproeven

Definitie 4.1.1 (Populatie): De verzameling van **alle** objecten of personen waar men in geïnteresseerd is en onderzoek naar wil doen noemt men de populatie. Een ander woord is ook wel onderzoeksgroep of doelgroep.

Definitie 4.1.2 (Steekproef): Wanneer met een subgroep uit een populatie gaat onderzoeken, dan noemen we die groep een steekproef.

Definitie 4.1.3 (Steekproefkader): Een steekproefkader is een lijst van alle leden van een te onderzoeken populatie.

1. Definitie populatie

2. Bepalen van steekproefkader
3. Budget en Tijd

4.2 Steekproefmethode

Definitie 4.2.1 (Gestratificeerde steekproef): Een populatie kan worden ingedeeld in groepen, deze deelpopulaties worden 'strata' genoemd. Bij een gestratificeerde steekproef wordt per stratum een enkelvoudig aselecte steekproef getrokken. De verdeling naar omvang van de deelsteekproeven dient evenredig te zijn aan de verdeling naar omvang van de strata. Een gestratificeerde steekproef is proportioneel als het aandeel van de subpopulatie in de steekproef gelijk is aan het aandeel van de subpopulatie in de populatie als geheel.

Aselecte steekproef elk element uit de onderzoekspopulatie heeft een even grote kans om in de steekproef terecht te komen.

Selecte steekproef of een element in de steekproef terecht komt is afhankelijk van een persoonlijke beoordeling van een onderzoeker.

4.2.1 Fouten

Toevallige steekproeffouten Wanneer er puur door het toeval een verschil is in een waarde voor de populatie en de steekproef.

Systematische steekproeffouten Een procedure in de steekproef die een fout oplevert die een systematische oorzaak heeft en dus niet te wijten is aan toevallige effecten. Bijvoorbeeld door systematisch een bevoordeeld deel van de populatie te ondervragen. Als we onze superhelden zouden ondervragen via het internet, sluiten we alle superhelden uit die geen internetverbinding hebben.

Toevallige niet-steekproeffouten Hieronder vallen bijvoorbeeld verkeerd aangekruiste antwoorden of verschil in interpretatie van de vragen.

Systematische niet-steekproeffouten Wanneer bijvoorbeeld respondenten met een sterke band met het onderzoek eerder geneigd zijn om een vragenlijst in te vullen, ga je positievere antwoorden krijgen - terwijl ze niet representatief zijn voor de gehele populatie.

4.3 Kansverdeling

4.3.1 Stochatisch experiment

Definitie 4.3.1 (Universum of Uitkomstenruimte): Het universum of uitkomstenruimte van een experiment is de verzameling van alle mogelijke uitkomsten van dit experiment en wordt genoteerd met Ω.

Definitie 4.3.2 (Gebeurtenis): Een gebeurtenis is een deelverzameling van de uitkomstenruimte. Een enkelvoudige of elementaire gebeurtenis is een singleton; een samengestelde gebeurtenis heeft cardinaliteit groter dan 1.

Definitie 4.3.3 (Kansruimte): Het toekennen van kansen aan gebeurtenissen dient aan de volgende drie regels te voldoen.

1. Kansen zijn steeds positief: $P(A) \geq 0$ voor elke A.
2. De uitkomstenruimte heeft kans 1: $P(\Omega) = 1$.
3. Wanneer A en B *disjuncte* gebeurtenissen zijn dan is

$$P(A \cup B) = P(A) + P(B).$$

   Dit noemt men de somregel.

Wanneer de functie P aan de bovenstaande eigenschappen (axioma's) voldoet dan noemt men het drietal $(\Omega, \mathcal{P}(\Omega), P)$ een kansruimte (met $\mathcal{P}(\Omega)$ de *machtsverzameling* van Ω, d.w.z. de verzameling van alle deelverzamelingen van Ω).

4.4 Normale verdeling

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \tag{4.4.1}$$

Eigenschappen

- Normale verdeling is klokvormig
- De normale verdeling is symmetrisch
- Vanwege symmetrie is gemiddelde, mediaan en modus aan elkaar gelijk
- De totale oppervlakte onder de klokvormige figuur is 1

- In gebied σ onder μ en σ boven μ (het zogenoemde sigma gebied) ligt ongeveer 68% van de waarnemingen.
- In het gebied 2σ boven en onder μ ligt ongeveer 95% van alle waarnemingen.

4.4.1 Standaardnormale verdeling

Definitie 4.4.1 (Z-Score): Deze score geeft dus aan hoe extreem een waarneming is of anders gezegd, hoeveel standaarddeviaties is de waarneming x van het gemiddelde μ verwijderd.

$$z = \frac{x - \mu}{\sigma} \tag{4.4.2}$$

Voor deze z-scores heeft men tabellen opgesteld met de kansen dat een waarde kleiner dan z getrokken wordt uit Z, de zgn. linkerstaartkans[a]:

$$P(Z < z) \tag{4.4.3}$$

[a] Er bestaan ook tabellen met de rechterstaartkans

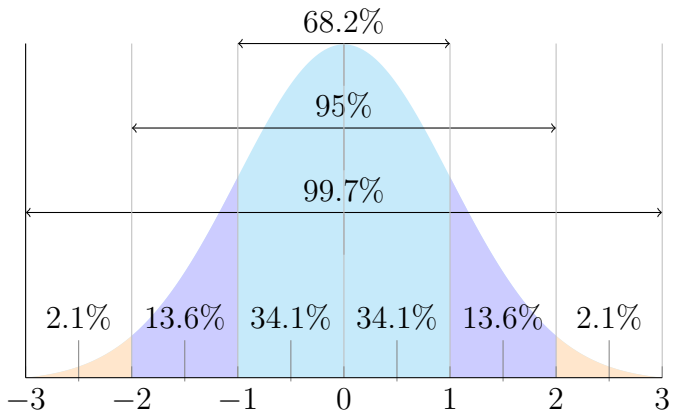

Fig. 1: De standaardnormale verdeling met opdeling in zones

Berekenen kansen met de normale verdeling

1. Bepaal de kansvariabele met de bijbehorende normale verdeling
2. Bereken de z-score bij de bijhorende x-waarde.
3. Schets de plaats van de gevraagde kans

4. Herleid de gevraagde kans met behulp van de schets tot een linkerstaartkans en gebruik de z-tabel van de standaardnormale verdeling om deze te bepalen. Gebruik indien nodig de symmetrieregel en de regel van 100% kans.

4.4.2 Testen op normaliteit

- Construeer een histogram voor de gegevens en bekijk de vorm van de grafiek. Als de gegevens bij benadering een normale verdeling hebben, zal de vorm van het histogram een klokcurve vormen.

- Bereken de intervallen $\overline{x} \pm s$, $\overline{x} \pm 2s$, $\overline{x} \pm 3s$ en bepaal het percentage meetwaarden dat binnen elk van deze intervallen valt. Als de gegevens ongeveer normaal verdeeld zijn, zullen de percentages ongeveer gelijk zijn aan respectievelijk 68%, 95% en 99,7%.

- Construeer een QQ-plot (normaliteitsplot, zie Definitie 4.4.2) voor de gegevens. Als de gegevens ongeveer normaal verdeeld zijn, zullen de punten ongeveer op een rechte lijn liggen.

- Bereken de *kurtosis* ("welving" of "platheid"): duidt aan hoe scherp de "piek" van de verdeling is.
  - Een normale verdeling heeft een kurtosis = 0
  - Een vlakke distributie heeft een negatieve kurtosis
  - Een eerder piekvormige distributie heeft een positieve kurtosis
  - Let op: bij de originele definitie van kurtosis (zoek die eens op!) heeft de normale verdeling een kurtosis van 3. Wij gebruiken hier een alternatieve definitie, meestal de "excess kurtosis" genoemd, waar men 3 aftrekt van de originele waarde, zodat je op 0 uitkomt.

- Berken de *Skewness* (scheefheid): duidt aan hoe symmetrisch de data is.
  - Een symmetrische distributie heeft een skewness = 0
  - Bijgevolg: een normale verdeling heeft een skewness = 0.
  - Een distributie met een lange linkerstaart heeft een negatieve skewness
  - Een distributie met een lange rechterstaart heeft een positieve skewness

– Vuistregel: absolute waarde van skewness > 1, geen symmetrische distributie.

Definitie 4.4.2 (QQ-plot of normaliteitsplot): Een normaliteitsplot of QQ-plot[a] voor een gegevensverzameling is een spreidingsdiagram met de gesorteerde gegevenswaarden op de ene as en de bijbehorende verwachte z-waarden van een standaardnormale verdeling op de andere as.

[a] Q staat hier voor *quantile*, kwantiel

4.5 Centrale limietstelling

Definitie 4.5.1 (Lineaire combinatie van onafhankelijke, normaal verdeelde stochasten): Formeel: Een lineaire combinatie van onafhankelijke, normaal verdeelde stochasten is steeds normaal verdeeld.

$$X_i \sim Nor(\mu_i, \sigma_i) \Rightarrow Y = \sum_i \alpha_i X_i \quad \text{ook normaal verdeeld}$$

Bijgevolg zal ook het steekproefgemiddelde van een steekproef uit een populatie met een willekeurige verdeling, nagenoeg normaal verdeeld zijn voor een voldoende grote n.

Definitie 4.5.2 (Centrale limietstelling): Beschouw een aselecte steekproef van n waarnemingen die uit een populatie met verwachtingswaarde μ en standaardafwijking σ wordt genomen. Als n groot genoeg is zal de kansverdeling van het steekproefgemiddelde $\overline{x}$ een normale verdeling benaderen met verwachting $\mu_{\overline{x}} = \mu$ en standaardafwijking $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$. Hoe groter de steekproef is, des te beter zal de kansverdeling van $\overline{x}$ de verwachtingswaarde van de populatie benaderen.

4.5.1 Concreet

4.5.2 Schatten van een parameter

Definitie 4.5.3 (Puntschatter): Een puntschatter voor een populatieparameter is een regel of een formule die ons zegt hoe we uit de steekproef een getal moeten berekenen om de populatieparameter te schatten. Een puntschatter is dus een steekproefgrootheid.

4.5.3 Betrouwbaarheidsinterval

Definitie 4.5.4 (Betrouwbaarheidsinterval): Een betrouwbaarheidsinterval is een regel of een formule die ons zegt hoe we uit de steekproef een interval moeten berekenen dat de waarde van de parameter met een bepaalde hoge waarschijnlijkheid bevat.

Definitie 4.5.5 (Betrouwbaarheidsinterval kleine steekproef): Om een betrouwbaarheidsinterval voor het gemiddelde te bepalen op basis van een klein steekproef bepalen we:

$$\overline{x} \pm t_{\frac{\alpha}{2}}(\frac{s}{\sqrt{n}})$$

waarbij $t_{\frac{\alpha}{2}}$ gebaseerd is op $(n-1)$ vrijheidsgraden. We veronderstellen wel dat we een aselecte steekproef genomen hebben uit een populatie die bij benadering normaal verdeeld is.

Definitie 4.5.6 (Betrouwbaarheidsinterval voor p gebaseerd op grote steekproef):

$$\overline{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\overline{pq}}{n}} \tag{4.5.1}$$

waarbij:

$\overline{p} = \frac{x}{n}$
$\overline{q} = 1 - \overline{p}$

5 Toetsingsprocedures

5.1 Theorie

Definitie 5.1.1 (Hypothese): Een *hypothese* of *onderzoeksvraag* is in de empirische wetenschap een stelling die (nog) niet bewezen is en dient als uitgangspunt voor een experiment.

Definitie 5.1.2 (Statistische hypothese): Een statische hypothese is een uitspraak over de numerieke waarde van een populatieparameter.

5.2 Elementen hypothesetoets

Definitie 5.2.1 (Nulhypothese H_0): Deze hypothese proberen we te ontkrachten door een redenering in het ongerijmde. We gaan deze hypothese ac-

cepteren, tenzij de gegevens overtuigend wijzen op het tegendeel.

Definitie 5.2.2 (Alternatieve hypothese H_1): De hypothese die meestal gesteund wordt door de onderzoeker. Deze hypothese zal alleen worden geaccepteerd als de gegevens overtuigend wijzen op zijn juistheid.

Definitie 5.2.3 (Teststatistiek): De veranderlijke die berekend wordt uit de steekproef

Definitie 5.2.4 (Aanvaardingsgebied): Het gebied van waarden die de nulhypothese H_0 ondersteunt

Definitie 5.2.5 (Verwerpingsgbied): Het gebied dat waarden bevat die de nulhypothese verwerpen (ook kritiek gebied genoemd)

5.3 Stappenplan

1. Bepalen hypotheses
2. Vastleggen significantieniveau α en steekproefomvang n
3. Toetsingsgrootheid en de waarde hiervan in de steekproef

$$M \sim Nor(\mu, \frac{\sigma}{\sqrt{n}})$$

4. Bereken grens en teken het kritiek gebied of bepaal de p-waarde

$$g = \mu \pm z \times \frac{\sigma}{\sqrt{n}} \tag{5.3.1}$$

5.4 Kritiek gebied, aanvaardingsgebied & kritieke grenswaarde

5.5 Eenzijdige of tweezijdige toetsen

5.5.1 Doel

Test op gemiddelde waarde μ van de populatie aan de hand van een steekproef van n onafhankelijke steekproefwaarden.

5.5.2 Voorwaarde

De populatie is willekeurig verdeeld, n voldoende groot.

5.5.3 Type test

	Tweezijdig	Eenzijdig links	Eenzijdig rechts
H_0	$\mu = \mu_0$	$\mu = \mu_0$	$\mu = \mu_0$
H_1	$\mu \neq \mu_0$	$\mu < \mu_0$	$\mu > \mu_0$
Verwerpingsgebied	$\lvert z \rvert > g$	$z < -g$	$z > g$

5.5.4 Teststatistiek

$$z = \frac{\overline{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

5.6 Fouten in hypothesetoets

	Werkelijke stand van zaken	
	H_0 correct	H_1 correct
H_0 geaccepteerd	Juist	Fout type II
H_0 verworpen	Fout type I	Juist

6 χ^2 toets

Definitie 6.0.1:

$$e = n \times \pi \tag{6.0.1}$$

waarbij:

$e =$ expected

Definitie 6.0.2:

$$\chi^2 = \sum_{i=1}^{n} \frac{(o_i - e_i)^2}{e_i} \tag{6.0.2}$$

waarbij:

$o =$ observed

Definitie 6.0.3 (Gestandaardizeerde residuen): De gestandaardiseerde residuen duiden aan welke klassen de grootste bijdrage leveren aan de waarde van de grootheid.

$$r_i = \frac{O_i - n\pi_i}{\sqrt{n\pi_i(1-\pi_i)}} \tag{6.0.3}$$

7 Indexcijfers

Definitie 7.0.1 (Indexcijfer): Een indexcijfer is een getal dat de verandering van een variabele in de loop van de tijd meet in verhouding tot de waarde van de variabele tijdens een bepaalde basisperiode.

7.1 Notatie

l = indexcijfer
p = prijs
q = hoeveelheid
w = waarde
b = basisperiode
v = verslagperiode
n = periode / tijdstip
$l_{p,n}$ = prijs-indexcijfer
$l_{q,n}$ = hoeveelheid-indexcijfer
$l_{w,n}$ = waarde-indexcijfer
P_b = prijs in de basisperiode
P_v = prijs in de verslagperiode
Q_v = hoeveelheid in de basisperiode
Q_b = hoeveelheid in de verslagperiode

7.2 Enkelvoudig

Definitie 7.2.1 (Enkelvoudige indexcijfers): Enkelvoudige indexcijfers hebben betrekking op slechts één object of artikel. Voorbeelden hiervan zijn hieronder beschreven.

Prijs-indexcijfer

$$I_p = \frac{P_v}{P_b} \times 100 \tag{7.2.1}$$

Hoeveelheid-indexcijfer

$$I_q = \frac{Q_v}{Q_b} \times 100 \tag{7.2.2}$$

Waarde indexcijfer

$$I_w = \frac{P_v \times Q_v}{P_b \times Q_b} \times 100 = \frac{I_p \times I_q}{100} \tag{7.2.3}$$

7.3 Samengesteld

Definitie 7.3.1 (Samengestelde index): Een samengesteld indexcijfer is het rekenkundig gemiddelde van een aantal enkelvoudige indexcijfers.

Definitie 7.3.2 (Ongewogen samengesteld indexcijfer): Aan elk element (partiële indexcijfer) wordt hetzelfde belang gehecht.

$$\overline{I} = \frac{\sum_{i=0}^{n} I_{i,0}}{n} \tag{7.3.1}$$

waarbij:

$\overline{I}$ = ongewogen samengesteld indexcijfer
$I_{i,0}$ = partiële indexcijfers op tijdstop t met basis jaar 0
n = aantal partiële indexcijfers

Definitie 7.3.3 (Gewogen samengesteld indexcijfer): Het belang van de onderscheiden partiële indexcijfers is ongelijk: men geeft aan elk partieel indexcijfer een wegingscoëfficiënt.

$$\overline{I} = \frac{\sum_{i=0}^{n} g_i \times I_i}{\sum_{i}^{n} g_i} \tag{7.3.2}$$

waarbij:

$\overline{I}$ = gewogen samengesteld indexcijfer
g_i = gewicht of wegingscoëfficiënt van het i^{de} partiële cijfer
I_i = partiële indexcijfers

Deze coëfficiënten kunnen bepaald worden in functie van hoeveelheden, volumes, uitgaves, proefondervindelijk ....

Definitie 7.3.4 (Indexcijfer van Laspeyres): Het gewogen samengesteld indexcijfer van **Laspeyres** is het gewogen gemiddelde van partiële indexcijfers waarbij de hoeveelheden in de **basisperiode** constant gehouden zijn m.a.w. indexcijfer met vaste gewichten of berekening volgens basisjaarmethode. Zie vergelijking 7.3.3.

$$I_p^l = \frac{\sum P_v \times Q_b}{\sum P_b \times Q_b} \tag{7.3.3}$$

Definitie 7.3.5 (Indexcijfer van Paasche): Het gewogen samengesteld indexcijfer van **Paasche** neemt de hoeveelheden in de **verslagperiode** als wegingcoëfficiënt m.a.w. indexcijfer met veranderlijke gewichten. Zie vergelijking 7.3.4.

$$I_p^P = \frac{\sum P_v \times Q_v}{\sum P_b \times Q_v} \tag{7.3.4}$$

Definitie 7.3.6 (Indexcijfer van Fischer): Het indexcijfer van Fischer is het meetkundig gemiddelde van Laspeyres en Paasches indexcijfer. De vierkantswortel uit het product van de indexcijfers van Laspeyres en Paasche. Zie vergelijking 7.3.5.

$$I_p^f = \sqrt{I_p^L \times I_p^P} \tag{7.3.5}$$

8 Tijdreeksen

Definitie 8.0.1 (Tijdsreeks): Een tijdreeks is een opeenvolging van observaties van een willekeurige variabele in functie van de tijd.

8.1 Tijdreeksmodellen

8.2 Schatten parameters

8.2.1 Moving average

Definitie 8.2.1 (Moving average): Algemeen is het moving average het gemiddelde van de m laatste observaties.

$$\widehat{b} = \sum_{i=k}^{t} \frac{x_i}{m} \tag{8.2.1}$$

met $k = t - m + 1$. m is de time range en is de parameter van de methode.

8.2.2 Meten nauwkeurigheid voorspellingen

Definitie 8.2.2 (Gemiddelde van de deviaties):

$$MAD = \frac{1}{n} \sum_{1}^{n} |e_i| \tag{8.2.2}$$

Definitie 8.2.3 (Gemiddelde absolute procentuele afwijking):

$$MAPE = \frac{1}{n} \sum_{1}^{n} \left| \frac{e_i}{X_i} \right| \tag{8.2.3}$$

Definitie 8.2.4 (Gemiddelde variantie):

$$s_e^2 = \frac{1}{m} \sum_{1}^{n} (e_i - \overline{e})^2 \tag{8.2.4}$$

Definitie 8.2.5 (Gemiddelde kwadratische afwijking):

$$RMSE_e = \sqrt{\frac{1}{m} \sum_{1}^{n} (e_i)^2} \tag{8.2.5}$$

8.3 Exponentiële smoothing

8.3.1 Enkelvoudig

Definitie 8.3.1 (Exponentiële smoothing):

$$X_T = \alpha x_{t-1} + (1 - \alpha) X_{t-1}, 0 \leq \alpha \leq 1, t \geq 3 \tag{8.3.1}$$

waarbij:

$\alpha =$ smoothing constante

8.3.2 Dubbel

Definitie 8.3.2 (Holt-voorspelling of dubbele exponentiële voorspelling):

$$\begin{aligned} X_t &= \alpha x_t + (1-\alpha)(X_{t-1} + b_{t-1}) & 0 \leq \alpha \leq 1 \\ b_t &= \gamma(X_t - X_{t-1}) + (1-\gamma)b_{t-1} & 0 \leq \gamma \leq 1 \end{aligned}$$

8.3.3 Driedubbel

Definitie 8.3.3 (Holt-Winters methode):

$$\begin{aligned} X_t &= \alpha\frac{x_t}{c_{t-L}} + (1-\alpha)(X_{t-1} + b_{t-1}) && \text{Smoothing} \\ b_t &= \gamma(X_t - X_{t-1}) + (1-\gamma)b_{t-1} && \text{Trend smoothing} \\ c_t &= \beta\frac{x_t}{X_t} + (1-\beta)c_{t-L} && \text{Seasonal smoothing} \\ F_{t+m} &= (X_t + mb_t)c_{t-L+m} && \text{Voorspelling} \end{aligned}$$

waarbij:

x_t = observatie op tijdstip t
X_t = smoothed observatie op tijdstip t
b_t = trendfactor op tijdstip t
c_t = seizoensindex op tijdstip t
F_t = voorspelling op tijdstip t
L = periode (bv. van de seizoenen)

Deel II. Formules

Gemiddelde

$$\mu = \frac{1}{n} \times \sum_{i=1}^{n} x_i \tag{8.3.2}$$

Variantie

$$\sigma^2 = \frac{1}{n} \times \sum_{i=1}^{n} (x_i - \mu)^2 \tag{8.3.3}$$

Standaardafwijking

$$\sigma = \sqrt{\sigma^2} \tag{8.3.4}$$

Cramér's V

$$V = \sqrt{\frac{\chi^2}{n(k-1)}} \tag{8.3.5}$$

Normale verdeling

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}} \tag{8.3.6}$$

Z-Score

$$Z = \frac{X - \mu}{\sigma} \tag{8.3.7}$$

Prijs-indexcijfer

$$I_p = \frac{P_v}{P_b} \times 100 \tag{8.3.8}$$

Hoeveelheid-indexcijfer

$$I_q = \frac{Q_v}{Q_b} \times 100 \tag{8.3.9}$$

Waarde indexcijfer

$$I_w = \frac{P_v \times Q_v}{P_b \times Q_b} \times 100 = \frac{I_p \times I_q}{100} \tag{8.3.10}$$

Ongewogen samengesteld indexcijfer

$$\overline{I} = \frac{\sum_{i=0}^{n} I_{i,0}}{n} \tag{8.3.11}$$

Gewogen samengesteld indexcijfer

$$\overline{I} = \frac{\sum_{i=0}^{n} g_i \times I_i}{\sum_{i}^{n} g_i} \tag{8.3.12}$$

Indexcijfer van Laspeyres

$$I_p^l = \frac{\sum P_v \times Q_b}{\sum P_b \times Q_b} \tag{8.3.13}$$

Indexcijfer van Paasche

$$I_p^P = \frac{\sum P_v \times Q_v}{\sum P_b \times Q_v} \tag{8.3.14}$$

Indexcijfer van Fischer

$$I_p^f = \sqrt{I_p^L \times I_p^P} \tag{8.3.15}$$

Exponentiële smoothing

$$X_T = \alpha x_{t-1} + (1-\alpha)X_{t-1}, 0 \leq \alpha \leq 1, t \geq 3 \tag{8.3.16}$$

Holt-voorspelling / dubbele exponentiële voorspelling

$$\begin{aligned} X_t &= \alpha x_t + (1-\alpha)(X_{t-1} + b_{t-1}) && 0 \leq \alpha \leq 1 \\ b_t &= \gamma(X_t - X_{t-1}) + (1-\gamma)b_{t-1} && 0 \leq \gamma \leq 1 \end{aligned}$$

Holt-Winters methode / driedubbele exponentiële voorspelling

$$\begin{aligned} X_t &= \alpha\frac{x_t}{c_{t-L}} + (1-\alpha)(X_{t-1} + b_{t-1}) && \text{Smoothing} \\ b_t &= \gamma(X_t - X_{t-1}) + (1-\gamma)b_{t-1} && \text{Trend smoothing} \\ c_t &= \beta\frac{x_t}{X_t} + (1-\beta)c_{t-L} && \text{Seasonal smoothing} \\ F_{t+m} &= (X_t + mb_t)c_{t-L+m} && \text{Voorspelling} \end{aligned}$$