

D9 85.16

RAPPORT B-85-647

STATISCHE VERWERKING CEL- EN TRAXIAALPROEVEN

IBBC

BEZOEKADRES:
LANGE KLEIWEG 5
RIJSWIJK Z.H.

POSTBUS 49
2600 AA DELFT
TEL 015-138222
TELEX 38270

RAPPORT

NO. B-85-647/64.3.0924

BETR.: STATISCHE VERWERKING CEL- EN TRAXIAALPROEVEN

door: ir. A.C.W.M. Vrouwenfelder

December 1985

AAN Rijkswaterstaat COW



INHOUD

1. Inleiding
2. Regressie-analyse
3. Keuze van regressie-variabelen
4. Bepaling van veilige schattingen voor ϕ' en c'
5. Ruimtelijke spreiding
6. Regressie-analyse bij één waarneming per locatie
7. Samenvatting en conclusies

APPENDICES

1. Coëfficiënten regressie-analyse
2. Statische eigenschappen bij lineaire transformatie
3. Regressie-analyse met meetfouten in beide richtingen

1. INLEIDING

In de COW-leidraad "Cel- en triaxiaalproeven" die medio 1986 verschijnt is een hoofdstuk opgenomen over de statistische verwerking van de meetresultaten. Ter voorbereiding van dat betreffende hoofdstuk werd een verkennende studie uitgevoerd teneinde de verschillende mogelijkheden te onderzoeken en de invloed van bepaalde vereenvoudigingen na te gaan. Dit rapport geeft een overzicht van de belangrijkste resultaten van dat onderzoek. Voor wie uitsluitend een toelichting wil lezen op de leidraad kan zich beperken tot de hoofdstukken 2 t/m 4. De hoofdstukken 5 en 6 gaan op enkele punten verder.

Bij het samenstellen van de leidraad en ook bij het schrijven van dit rapport is als uitgangspunt gekozen dat grond exact gehoorzaamt aan het bezwijkcriterium van Mohr-Coulomb. Afwijkingen van deze rechte bezwijkfunctie worden geacht uitsluitend het gevolg te zijn van toevallige meetfouten ofwel van toevallige ruimtelijke fluctuaties in de grondeigenschappen. De leidraad beperkt zich zelfs helemaal tot meetfouten en laat de ruimtelijke variaties buiten beschouwing: aangenomen wordt dat de grondeigenschappen bepaald worden aan de hand van een meertrapsproef op een monster dan wel aan de hand van eentrapsproeven op drie of vier zeer dicht bij elkaar gestoken monsters.

Uiteraard is daarmee niet gezegd dat systematische afwijkingen niet zouden bestaan of dat het geen zin zou hebben onderzoek te verrichten naar een eventueel kromlijnige bezwijkcontour. Gemeend wordt echter dat dergelijke verfijningen op dit moment in de meeste gevallen niet van praktische betekenis zijn.

Tenslotte moet worden opgemerkt dat de statistische verwerking van de meetresultaten in dit rapport gebaseerd wordt op een Bayesiaanse interpretatie van het kansbegrip [1], [2]. Met name wordt bedoeld dat de te bepalen parameters van de bezwijkcontour als stochastische variabelen worden opgevat. Volgens het in de wereld van de statistici meer gebruikelijke frequentistische kansbegrip is dit niet mogelijk omdat deze parameters in wezen deterministisch zijn; wij kennen alleen de waarde niet precies. Slechts in enkele gevallen leidt het overigens tot merkbare verschillen.

2. REGRESSIE-ANALYSE

We beschouwen het regressie-probleem eerst in algemene termen. Gegeven zijn twee fysische grootheden x en y . Voor een aantal waarden x_i van de grootheid x zijn de bijbehorende waarden van y bepaald. Het resultaat kan er bijvoorbeeld uitzien als aangegeven in figuur 1.

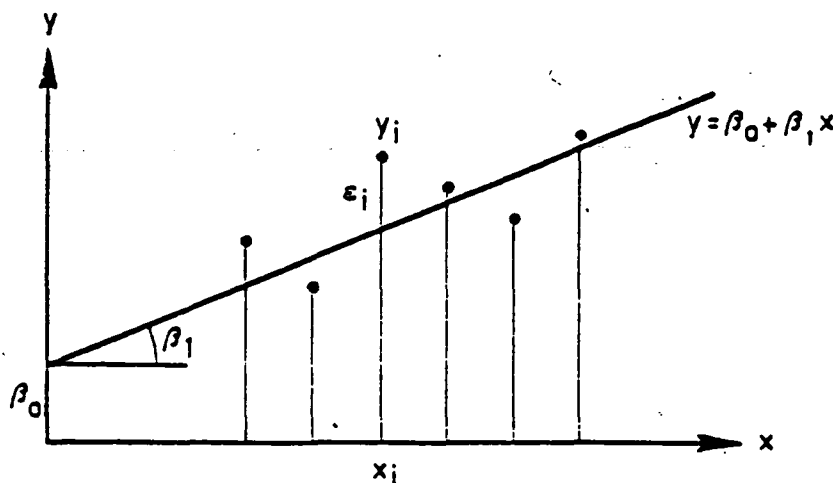
Stel dat men aanneemt dat tussen x en y in principe een rechtlijnig verband bestaat en dat het feit dat de gemeten punten niet keurig op een rechte lijn liggen het gevolg is van onnauwkeurigheden in de meting of andere toevallige factoren. Het probleem is dan om gegeven de meetresultaten van figuur 1 en zo goed mogelijke schatting van het verband tussen x en y te maken.

De gezochte rechte lijn kan worden geschreven als:

$$y = \beta_0 + \beta_1 x \quad (2.1)$$

Intuïtief is direct duidelijk dat de parameters β_0 en β_1 zodanig bepaald moeten worden dat een of andere fout minimaal wordt. Laten we per waarneming het verschil tussen gemeten waarde en de waarde volgens (2.1) aangeven met ε_i :

$$\varepsilon_i = y_i - \beta_0 - \beta_1 x_i \quad (2.2)$$



Figuur 1: Het principe van een regressie-analyse

Het is nu mogelijk om β_0 en β_1 zodanig te bepalen dat de som van deze verschillen ε_i in het kwadraat zo klein mogelijk is. Dit is de zogenaamde kleinste kwadratenmethode. In appendix 1 wordt aangetoond dat gegeven dit uitgangspunt de parameters β_0 en β_1 bepaald kunnen worden op:

$$\beta_0 = \bar{y} - \beta_1 \bar{x} \quad (2.3)$$

$$\beta_1 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2} \quad (2.4)$$

Hierin is $\bar{x}$ het gemiddelde van alle x_i -waarden en $\bar{y}$ het gemiddelde van de y_i . Het totale aantal waarnemingen bedraagt n en alle sommaties gaan van 1 tot n .

Als men de kleinste kwadratenmethode toepast in situaties met relatief weinig waarnemingen, kunnen toevallige factoren en grote invloed uitoefenen op de resulterende β_0 en β_1 . Met behulp van de waarschijnlijkheidsrekening is het mogelijk daar enig inzicht in te krijgen.

Laten we aannemen dat het werkelijke verband tussen x en y gegeven wordt door:

$$y = b_0 + b_1 x \quad (2.5)$$

De coëfficiënten b_0 en b_1 zijn deterministisch maar onbekend en moeten via metingen worden vastgesteld. Men kiest daartoe een aantal waarden x_i en bepaalt via proeven de bijbehorende y_i . Bij iedere proef wordt echter, zoals ook al in het begin gezegd, een meetfout ε_i gemaakt. Een veel gehanteerd model is om de meetfouten ε_i te beschouwen als onafhankelijke normaal verdeelde stochastische variabelen met gemiddelde nul en (nog onbekende) standaardafwijking σ . Het gevolg van de meetfout ε_i is dat het meetresultaat y_i gegeven wordt door:

$$y_i = b_0 + b_1 x_i + \varepsilon_i \quad (2.6)$$

De uitkomst van een meting y_1 is dus eveneens een stochastische variabele. Stel nu dat men een n -tal metingen doet en de resultaten verwerkt volgens de kleinste kwadratenmethode. Om te kunnen nagaan hoe de fouten ε_1 doorwerken in de parameters β_0 en β_1 vullen we (2.6) in in de uitdrukkingen (3) en (4):

$$\begin{aligned}\beta_0 &= \frac{1}{n} \sum (b_0 + b_1 x_1 + \varepsilon_1) - \beta_1 \bar{x} \\ \beta_1 &= \frac{\sum x_1 (b_0 + b_1 x_1 + \varepsilon_1) - \bar{x} \sum (b_0 + b_1 x_1 + \varepsilon_1)}{\sum x_1^2 - n \bar{x}^2}\end{aligned}\quad (2.7)$$

Na enige herleiding kan dit geschreven worden als:

$$\begin{aligned}\beta_0 &= b_0 + \sum \left(\frac{1}{n} - \bar{x} \alpha_1 \right) \cdot \varepsilon_1 \\ \beta_1 &= b_1 + \sum \alpha_1 \varepsilon_1\end{aligned}\quad (2.8)$$

met:

$$\alpha_1 = (x_1 - \bar{x}) / \sum (x_1 - \bar{x})^2 \quad (2.9)$$

Zoals te verwachten was kunnen we β_0 en β_1 dus niet zonder meer vereenzelvigen met b_0 en b_1 , maar zijn er stochastische afwijkingen als gevolg van de fouten ε_1 . In de hierboven ontwikkelde gedachtengang is er steeds vanuitgegaan dat b_0 en b_1 en de deterministische coëfficiënten zijn en de parameters β_0 en β_1 en de stochastische uitkomsten van het experiment. Dit is als het ware de situatie als het experiment nog moet beginnen. Ná het uitvoeren van het experiment komen we echter in een andere situatie. De parameters β_0 en β_1 zijn dan numeriek precies bekend terwijl t.a.v. b_0 en b_1 onzekerheden bestaan als gevolg van de meetfouten ε_1 . Zolang deze onzekerheden nog niet zijn geëlimineerd als gevolg van een zeer groot aantal metingen is het dus zinvol om b_0 en b_1 als stochastische grootheden te beschouwen. We brengen deze interpretatie volgens de zogenaamde "Bayesian Probability Theory" (zie [1] en [2]) tot uitdrukking door de vergelijkingen (2.8) als volgt te schrijven:

$$b_0 = \beta_0 - \sum \left(\frac{1}{n} - \alpha_1 \right) \varepsilon_1 \quad (2.10)$$

$$b_1 = \beta_1 - \sum \alpha_1 \varepsilon_1 \quad (2.11)$$

Opmerking: deze werkwijze leidt tot dezelfde resultaten als een formele Bayesiaanse analyse op basis van zogenaamde "vage apriori-veronderstellingen".

Uitgaande van (2.10) en (2.11) en de stochastische eigenschappen van de meetfouten ε_1 kunnen voor b_0 en b_1 gemiddelden, varianties en een covariatie worden bepaald. Gevonden wordt (zie appendix 2):

$$\mu(b_0) = \beta_0 \quad (2.12)$$

$$\mu(b_1) = \beta_1 \quad (2.13)$$

$$\sigma^2(b_0) = \sigma^2 \sum \left(\frac{1}{n} - \bar{x} \alpha_1 \right)^2 = \sigma^2 \left\{ \frac{1}{n} + \frac{\bar{x}^2}{\sum (x_1 - \bar{x})^2} \right\} \quad (2.14)$$

$$\sigma^2(b_1) = \sigma^2 \sum \alpha_1^2 = \sigma^2 \left\{ \frac{1}{\sum (x_1 - \bar{x})^2} \right\} \quad (2.15)$$

$$\text{cov}(b_0, b_1) = \sigma^2 \sum \alpha_1 \left(\frac{1}{n} - \bar{x} \alpha_1 \right) = \sigma^2 \left\{ \frac{-\bar{x}}{\sum (x_1 - \bar{x})^2} \right\} \quad (2.16)$$

Met behulp van deze resultaten zijn we in staat om voor b_0 en b_1 betrouwbare waarden aan te geven. De interesse ligt echter niet zo zeer in betrouwbare waarden voor b_0 en b_1 als wel in betrouwbare waarden voor y . We zullen daarom voor y ook een gemiddelde en standaardafwijking afleiden. Het verband tussen y enerzijds en b_0 en b_1 anderzijds wordt gegeven door (2.5):

$$y = b_0 + b_1 x \quad (2.5)$$

Voor het gemiddelde van y geldt:

$$\mu(y) = \mu(b_0) + \mu(b_1) x = \beta_0 + \beta_1 x \quad (2.17)$$

De kleinste kwadratenmethode levert dus een gemiddelde schatting op voor y . Voor de variantie van y geldt (zie appendix 2):

$$\sigma^2(y) = \sigma^2(b_0) + x^2 \sigma^2(b_1) + 2 x \text{cov}(b_0, b_1) \quad (2.18)$$

De eerste twee termen zullen zonder meer aanspreken; de derde term dient om eventuele afhankelijkheid tussen b_0 en b_1 in rekening te brengen. Invullen van (2.14) t/m (2.16) leidt tot:

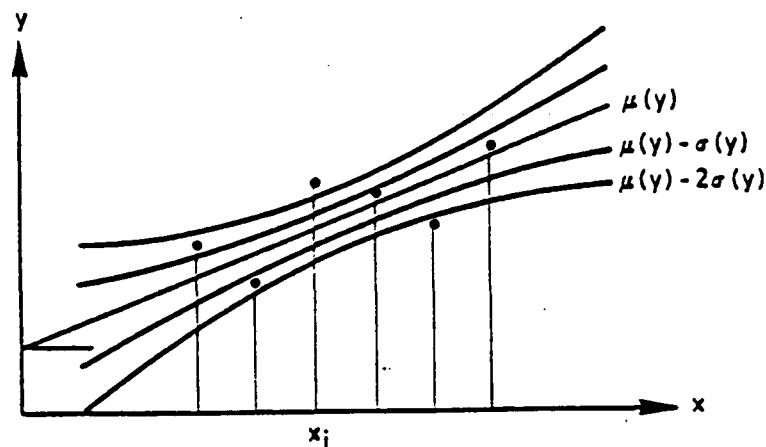
$$\sigma^2(y) = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2} + \frac{x^2}{\sum (x_i - \bar{x})^2} - \frac{2 x \bar{x}}{\sum (x_i - \bar{x})^2} \right]$$

Ofwel:

$$\sigma^2(y) = \frac{\sigma^2}{n} \left[1 + \frac{n (x - \bar{x})^2}{\sum (x_i - \bar{x})^2} \right] \quad (2.19)$$

Vergelijking (2.19) leert dat de spreiding in y het kleinst is voor $x = \bar{x}$, dat wil zeggen, in het zwaartepunt van het meetgebied. De standaardafwijking bedraagt daar $\sigma/\sqrt{n}$. Voor alle andere waarden van x is de spreiding groter. De lijnen $\mu(y) \pm \sigma(y)$, $\mu(y) \pm 2\sigma(y)$ enz. (zie figuur 2) hebben daarom een uitwaaiend karakter).

De oorzaak van het uitwaaieren ligt in het stochastische karakter van de parameter b_1 . Als alleen b_0 stochastisch zou zijn dan waren de betrouwbaarheidslijnen recht en evenwijdig. De spreiding in de richtingscoëfficiënt b_1 zorgt echter voor een grote afwijking in y op verder weg gelegen punt.



Figuur 2: Betrouwbaarheidsintervallen bij een lineaire regressie-analyse.

Om uitspraken te kunnen doen over de overschrijdingskansen van de betrouwbaarheidskrommen moet de verdeling van y bekend zijn. In principe heeft y een normale verdeling omdat de meetfouten ε_1 normaal waren aangenomen. Aangezien de standaardafwijking σ van de meetfouten echter niet bekend is maar uit de gevonden waarden moet worden berekend, gaat de normale verdeling van y over in een student-verdeling met $n-2$ graden van vrijheid (zie [1]). Er geldt dus:

$$y = \mu(y) + t_{n-2} \cdot \sigma(y) \quad (2.20)$$

In tabel 1 zijn waarden voor t_n opgenomen voor verschillende betrouwbaarheidsniveau's. Het gemiddelde $\mu(y)$ en de standaardafwijking $\sigma(y)$ worden respectievelijk gegeven door (2.17) en (2.19). De onbekende standaardafwijking σ kan worden geschat via:

$$s^2 = \frac{\sum \varepsilon_1^2}{n-2} = \frac{\sum (y_1 - \beta_0 - \beta_1 x_1)^2}{n-2} \quad (2.21)$$

Het aantal vrijheidsgraden bedraagt ook hier $(n - 2)$ omdat van het totale aantal vrijheidsgraden n er reeds twee gebruikt worden om de regressielijn vast te leggen.

v	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$
1	3.08	6.31	12.71
2	1.89	2.92	4.30
3	1.64	2.35	3.18
4	1.53	2.13	2.78
5	1.47	2.02	2.57
6	1.44	1.94	2.45
7	1.42	1.90	2.36
8	1.40	1.86	2.31
9	1.38	1.83	2.26
10	1.37	1.81	2.23
15	1.34	1.75	2.13
20	1.32	1.72	2.09
25	1.32	1.71	2.06
30	1.31	1.70	2.04
∞	1.28	1.64	1.96

Tabel 2.1 : Waarden van t_v voor verschillende onderschrijdingskans α .

Voorbeeld

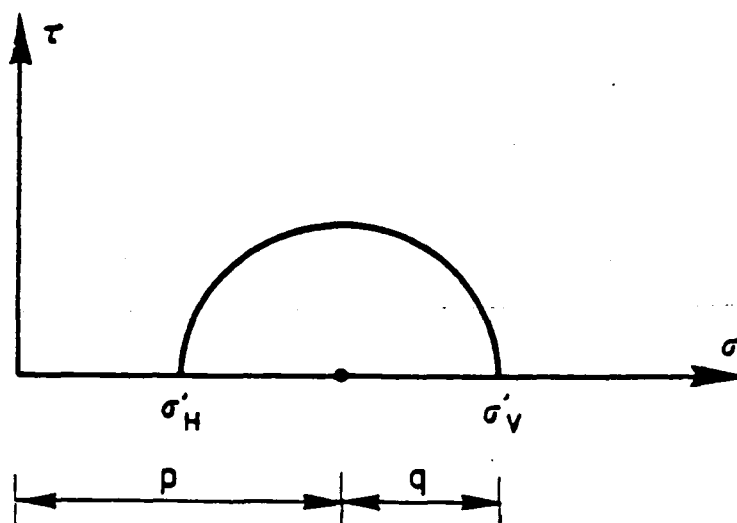
Neem aan dat een viertraps-celproef (CL-4-proef) op een kleimonsters de volgende resultaten heeft opgeleverd:

p	q
17,1 kN/m ²	8,4 kN/m ²
41,4	13,6
73,1	21,6
106,1	27,4

Onder verwijzing naar de discussie hierover in hoofdstuk 3 kiezen we p als onafhankelijke regressie-variabele en q als afhankelijke regressie-variabele. De grootheden p en q zijn gedefinieerd als (zie figuur 3):

$$p = \frac{1}{2} (\sigma'_v + \sigma'_h) \quad q = \frac{1}{2} (\sigma'_v - \sigma'_h)$$

Hierbij is σ'_v de verticale korrelspanning en σ'_h de horizontale steundruk.



Figuur 3: Definitie van p en q in een cirkel van Mohr.

Met behulp van de formules (2.3) en (2.4) kunnen de β -parameters worden uitgerekend. De parameter β_0 correspondeert fysisch met $c' \cos \phi'$ en de parameter β_1 met $\sin \phi'$, waarbij c' de cohesie en ϕ' de hoek van inwendige wrijving voorstelt. Gevonden wordt:

$$\beta_0 = 4,84 \text{ kN/m}^2 \quad \beta_1 = 0,217$$

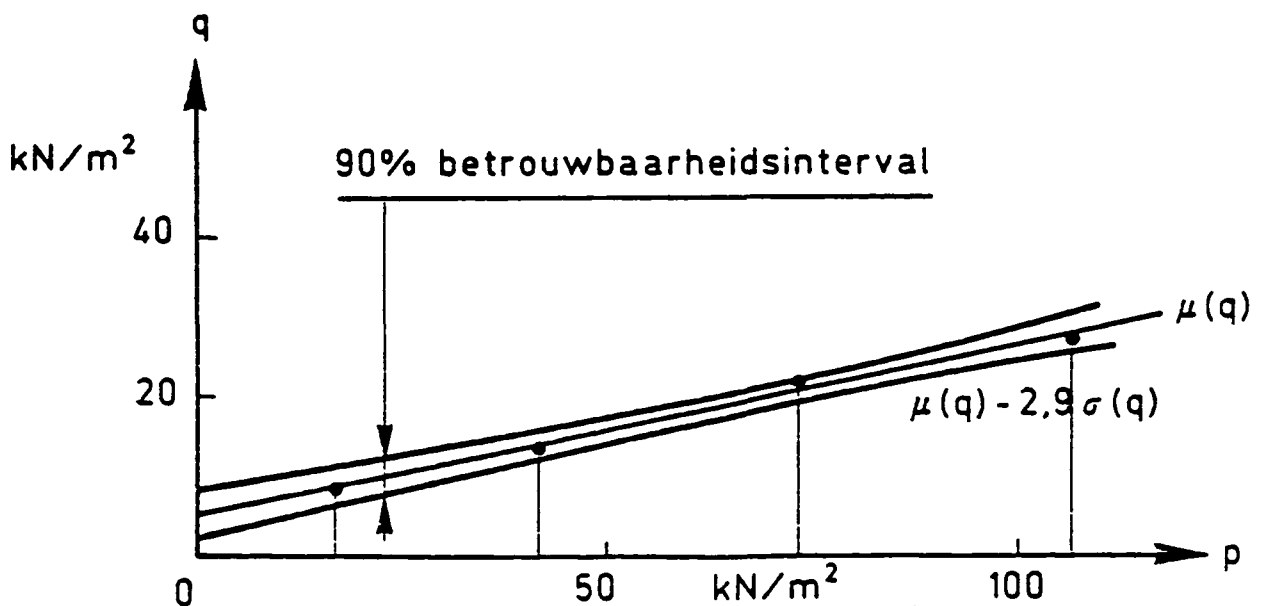
Dit correspondeert dus met:

$$c' = 4,95 \text{ kN/m}^2 \quad \phi' = 12,5^\circ$$

De meetwaarden en de resulterende regressielijn zijn uitgezet in figuur 4. We zien dat de punten redelijk op een rechte lijn liggen. Dit blijkt ook uit de standaardafwijking van de fouten ϵ_1 die met behulp van (2.21) bepaald kan worden op:

$$s(\epsilon_1) = 0,74 \text{ kN/m}^2$$

Ten opzichte van de gemiddelde q -waarde van $17,75 \text{ kN/m}^2$ betekent dit een variatiecoëfficiënt van $s/\bar{q} = 4\%$.



Figuur 4: Gemiddelde regressielijn en 90% tweezijdig betrouwbaarheidsinterval voor een CL-4-proef op klei.

In figuur 4 is verder nog het 90% betrouwbaarheidsinterval aangegeven, zoals kan worden berekend op basis van de formules (2.17), (2.19) en (2.20). De waarde van t_{n-2} voor $n = 4$ en een eenzijdige overschrijdingskans van 5% bedraagt 2,92 (zie tabel 2.1). Voor $p = \bar{p} = 59,43 \text{ kN/m}^2$ vinden we bijvoorbeeld:

$$\mu(q) = 17,73 \text{ kN/m}^2$$

$$\sigma(q) = s/\sqrt{n} = 0,37 \text{ kN/m}^2$$

$$q_{\text{betrouwbaar}} = \mu(q) - t_{n-2} \sigma(q) = 17,73 - 2,92 \cdot 0,37 = 16,65 \text{ kN/m}^2$$

Een andere interessante waarde is natuurlijk $p = 0$; daarvoor kan berekend worden:

$$\mu(q) = 4,84 \text{ kN/m}^2$$

$$\sigma(q) = \frac{s}{\sqrt{n}} \cdot (2,04) = 0,37 \cdot 2,04 = 0,75 \text{ kN/m}^2$$

$$q_{\text{betr}} = 4,84 - (2,92) (0,75) = 2,64 \text{ kN/m}^2$$

De q -waarde voor $p = 0$ correspondeert vrijwel met de schuifweerstand c' . We moeten dus concluderen dat deze parameter met de gegeven CL-4-proef niet erg nauwkeurig bepaald is, ondanks het feit dat de meetpunten een vrij goede fit met de rechte lijn vertoonden.

Voor de volledigheid merken we op dat de grote spreiding voor lage p -waarden niets te maken heeft met een grote spreiding in de proefresultaten aldaar. De spreiding $\sigma(\epsilon_1)$ is immers voor alle waarden van p even groot verondersteld. De grote relatieve spreiding voor de gevonden c' -waarde is het gevolg van de lage absolute waarde van c' in combinatie met de situering buiten het meetgebied.

Gevoelsmatig is het gevonden resultaat niet helemaal bevredigend. Een mogelijke verbetering hierin zou kunnen worden bereikt door uit te gaan van een model waarbij $\sigma(\epsilon_1)$ toeneemt met p .

3. KEUZE VAN DE REGRESSIE-VARIABELEN

Bij de toepassing van de standaard regressie-analyse bij de verwerking van de resultaten van cel- en triaxiaalproeven staat men allereerst voor de keuze welke variabele men als x en welke variabele men als y moet aanmerken. In beginsel moet men voor x een grootte nemen die men precies kent en er voor zorgen dat y alle meetonnauwkeurigheden bevat.

Er dienen zich, nog afgezien van de x - y -verwisseling, drie mogelijkheden aan:

- | | |
|-------------------|----------------|
| 1) $x = p$ | $y = q$ |
| 2) $x = \sigma_v$ | $y = \sigma_h$ |
| 3) $x = \sigma_a$ | $y = \tau_a$ |

Hierin is σ_v de verticale spanning, σ_h de horizontale spanning, p het gemiddelde van de horizontale en verticale spanning ($p = (\sigma_h + \sigma_v)/2$), q de maximale schuifspanning ($q = (\sigma_v - \sigma_h)/2$), σ_a de normale spanning in het vlak van afschuiven en τ_a de schuifspanning in het vlak van afschuiven.

Van de drie genoemde mogelijkheden levert alleen de derde direct de verlangde parameters ϕ en c' op. In de andere gevallen bepaalt men eerste de parameters voor de regressielijn in het σ_h - σ_v of p - q -vlak. Daarna moeten deze parameters en de onzekerheden daarin worden omgerekend naar ϕ en c' . Een nadeel van de regressie-analyse met $x = \sigma_a$ en $y = \tau_a$ is echter dat σ_a en τ_a alleen bepaald kunnen worden als de afschuifrichting in de proef bekend is, hetgeen in de praktijk nagenoeg onmogelijk is. Een andere weg is om de afschuifrichting te berekenen uit de helling van de regressielijn. We komen dan echter op een iteratieve procedure terecht, hetgeen onaantrekkelijk is. De mogelijkheid ($x = \sigma_a$, $y = \tau_a$) blijft daarom verder buiten beschouwing.

Ten aanzien van de andere twee mogelijkheden geldt dat geen van beide volledig voldoet aan de gestelde uitgangspunten. Beschouw eerst de celproef. Er geldt:

$$\sigma'_v = F/A \quad (3.1)$$

$$\sigma'_h = \sigma_{cel} \quad (3.2)$$

De bovenbelasting F ligt bij de celproef goed vast, maar voor de bepaling van σ'_v moet ook het oppervlak A in vervormde toestand worden bepaald, hetgeen enkele onzekerheden met zich meebrengt. De celspanning σ_{cel} , en daarmee σ'_h , is wel direct en goed meetbaar. Daaruit zou dus volgen: $x = \sigma'_h$ en $y = \sigma'_v$. Een probleem is echter dat moeilijk interpreteerbaar is welke waarde van σ_{cel} tijdens de proef precies met bezwijken in verband gebracht moet worden.

Voor de triaxiaalproef geldt:

$$\sigma'_v = \sigma_{cel} - u + F/A$$

$$\sigma'_h = \sigma_{cel} - u$$

Bij de triaxiaalproef ligt in principe de celdruk goed vast en kan goed gemeten worden. Een bron van onnauwkeurigheden is natuurlijk de meting van de waterspanning u . In principe kan die worden opgevangen door te transformeren naar het p-q-systeem:

$$p = \frac{\sigma'_v + \sigma'_h}{2} = \sigma_{cel} - u + \frac{1}{2} \frac{F}{A} \quad (3.3)$$

$$q = \frac{\sigma'_v - \sigma'_h}{2} = \frac{1}{2} \frac{F}{A} \quad (3.4)$$

Alle onzekerheid afkomstig van u is nu in p ondergebracht. Wat echter blijft in q is de onzekerheid afkomstig van A (net als bij de celproef) en, de vraag welke waarde van F deze laatste met bezwijken in verband gebracht moet worden. (Merk op dat bij de celproef onzekerheid wordt toegekend aan de horizontale steunspanning en bij de triaxiaalproef aan de verticale spanning. Dit verschil is terug te voeren tot de wijze waarop de proef wordt uitgevoerd).

De conclusie is dat geen de voorgestelde werkwijzen tot een theoretisch volkomen bevredigende oplossing leidt. Strikt genomen is dan de standaard regressie-analyse niet toepasbaar en moet naar andere oplosmethode worden uitgezien. In verband daarmee is een literatuurstudie uitgevoerd naar de mogelijkheden voor het uitvoeren van een regressie-analyse waarbij in beide variabelen fouten aanwezig worden geacht. Gebleken is overigens dat de literatuur op dit punt uiterst dun gezaaid is. De meeste boeken behandelen uitsluitend de standaardmethode en maken zich er vervolgens met de opmerking vanaf dat in de praktijk meestal wel duidelijk zal zijn welke variabele als afhankelijk en welke als onafhankelijk dient te worden gekozen. Een uitzondering hierop is het boek van Seber [3], waarin enkele paragrafen aan dit onderwerp zijn gewijd. Ook in de vaktijdschriften komt men niet veel over het onderwerp tegen. Gevonden zijn de in de literatuurlijst als [4] tot en met [9] opgenomen referenties. Hiervan zijn Madansky [6] en Sprent [7] het meest informatief. Het blijkt dat de kleinste kwadraten-methode minimaal de variantie moet kennen van de fout in de ene of de andere richting, dan wel hun onderlinge verhouding. Als een dergelijk gegeven ontbreekt moet men omzien naar een geheel andere schattingsprocedure voor de bepaling van de parameters β_0 en β_1 . Deze methoden zijn echter weinig efficient, tamelijk ingewikkeld en daardoor niet erg geschikt voor het hier beoogde doel. Nadere informatie over de uitgevoerde studie is te vinden in bijlage 3.

We concluderen dat, ondanks theoretische bezwaren, teruggevallen moet worden op de standaard-methode. Op het eerst gezicht lijkt de regressierichting dan te moeten corresponderen met de variabele waarin de onnauwkeurigheid maximaal is. Dit is echter niet juist omdat ook de richting van de gezochte regressielijn van invloed is: een fout in die richting van de lijn heeft minder invloed dan een fout er loodrecht op. (Opmerking: men stelt om die reden wel voor de kleinste kwadratenmethode loodrecht op de regressielijn uit te voeren; er is dan echter geen standaard geval omdat de richting zelf nog onbekend is; formules met betrekking tot de afschatting van de fout worden daardoor gecompliceerd).

Een belangrijke vraag is hoeveel verschil in het eindresultaat optreedt, afhankelijk van de keuze van de regressiemethode. Voor de viertrapscel-

proef en voor een set van triaxiaalproeven op dicht bij elkaar gelegen monsters is het verschil niet erg groot. Als voorbeeld is in tabel 3.1 een overzicht gegeven van de resultaten van de vier verschillende keuzemogelijkheden voor hetzelfde geval als in het voorbeeld van hoofdstuk 3. Voor de gemiddelde schattingen van c' en ϕ' maakt het heel weinig uit; voor de 5% onderschrijdingswaarde van bijvoorbeeld de cohesie zijn de verschillen relatief iets groter, maar nog altijd zeer klein.

De keuze van de regressierichting heeft maar weinig invloed omdat de spreiding klein is. Indien bijvoorbeeld proeven gedaan zouden worden op verder uit elkaar gelegen monsters, en ook de ruimtelijke fluctuaties van grondeigenschappen van invloed zijn, komt het probleem anders te liggen (zie tabel 3.2). In de hoofdstukken 5 en 6 wordt daarop terug gekomen. Voor de standaardprocedure waarbij uitsluitend toevallige meetfouten optreden is de keuze van de regressievariabelen een duidelijk ondergeschikt probleem.

Gegeven het feit dat het in veel gevallen gebruikelijk is om p als onafhankelijk en q als afhankelijke waarde te gebruiken ligt het gezien bovenstaande overwegingen voor de hand deze praltijk onveranderd te laten. Voor voordeel van deze keuze is in ieder geval dat voor minder goed geconditioneerde proeven bij klei de regressielijn vrij flauw loopt en eventuele fouten in p t.g.v. onnauwkeurigheden in de spanningsmeting minder belangrijk zijn.

Methode	β_0	β_1	$c' \text{ [kN/m}^2\text{]}$	c'_{veilig}	$\text{tg } \phi'$
1a) $x=p$ $y=q$	4,84	0,217	$-\beta_0 / \sqrt{1-\beta_1^2} = 4,96$	2,74	$\beta_1 / \sqrt{1-\beta_1^2} = 0,222$
1b) $x=q$ $y=p$	-21,86	4,58	$-\beta_0 / \beta_1^2 - 1 = 4,89$	2,66	$1 / \beta_1^2 - 1 = 0,224$
2a) $x=\sigma'_v$ $y=\sigma'_h$	-7,96	0,643	$-\beta_0 / 2\beta_1 = 4,96$	2,74	$(1-\beta_1) / 2\beta_1 = 0,223$
2b) $x=\sigma'_h$ $y=\sigma'^1_v$	12,44	1,553	$\beta_0 / 2\beta_1 = 4,99$	2,77	$(\beta_1 - 1) / 2\beta_1 = 0,222$

Tabel 3.1: Vier mogelijke berekeningsresultaten uitgaande van een vier-trapscelproef op een enkel grondmonster (voor gegevens zie voorbeeld in hoofdstuk 3).

Methode	β_0	β_1	$c' [\text{kN/m}^2]$	c'_{veilig}	$\text{tg } \phi'$
1a) $x=p$ $y=q$	3,39	0,25	3,50	-1,67	0,258
1b) $x=q$ $y=p$	-12,03	3,91	3,18	-2,11	0,264
2a) $x=\sigma'_v$ $y=\sigma'_h$	- 5,30	0,60	3,42	-1,78	0,259
2b) $x=\sigma'_h$ $y=\sigma'_v$	9,26	1,66	3,59	-1,59	0,256

Tabel 3.2: Resultaten volgens 4 verschillende berekeningsresultaten bij 4 eentrapscelproeven op monsters van verschillende locaties.

Genomen is uit tabel 5.1:

Monster 1, eerste punt: $p = 17,1$ en $q = 8,4 \text{ kN/m}^2$

monster 2, tweede punt: $p = 41,5$ en $q = 13,6$

monster 3, derde punt : $p = 74,1$ en $q = 20,1$

monster 4, vierde punt: $p = 102,0$ en $q = 30,2$

Merk op dat de veilige schatting van c' zelfs negatief uitkomt. In hoofdstuk 6 wordt een methode behandeld waarbij die niet het geval is.

4. BEPALING VAN VEILIGE SCHATTINGEN VOOR ϕ' en c'

Indien, overeenkomstig de conclusie van het voorgaande hoofdstuk, de regressie-analyse wordt uitgevoerd met p en q, dan resulteren daaruit waarden voor β_0 en β_1 , ofwel verwachtingswaarden voor b_0 en b_1 . We zijn echter niet geïnteresseerd in b_0 en b_1 , maar in de cohesie c' en de hoek van inwendige vrijwing ϕ' . Tussen deze grootheden bestaat, zoals ook reeds eerder aangegeven het volgende verband:

$$c' = \frac{b_0}{\sqrt{1-b_1^2}} \quad \text{tg } \phi' = \frac{b_1}{\sqrt{1-b_1^2}} \quad (4.1)$$

Met behulp van deze formule kunnen de gemiddelden of verwachtingswaarden voor c' en $\text{tg } \phi'$ direct worden uitgerekend. Voor de goede orde zij opgemerkt dat een dergelijk procedure niet helemaal exact is omdat (4.1) geen lineaire betrekking is.

De benadering is overigens geheel binnen de grenzen van de zogenaamde niveau II methode, zoals die de laatste jaren veel binnen de civiele techniek worden toegepast [2]. Met behulp van deze methode zullen we ook uitdrukkingen afleiden voor de varianties en covarianties van c' en $\text{tg } \phi'$. Voor de variantie van c' geldt:

$$\sigma^2(c') = \left\{ \frac{\partial c'}{\partial b_0} \sigma(b_0) \right\}^2 + \left\{ \frac{\partial c'}{\partial b_1} \sigma(b_1) \right\}^2 + 2 \left\{ \frac{\partial c'}{\partial b_0} \frac{\partial c'}{\partial b_1} \right\} \text{cov}(b_0, b_1) \quad (4.2)$$

Voor het lineaire geval is deze uitdrukking exact zoals volgt via vergelijking met de overeenkomstige uitdrukking voor $\sigma^2(u)$ in appendix 2. Voor niet-lineaire functies is sprake van een benadering. De partiele afgeleiden worden ontwikkeld voor de gemiddelde waarden van b_0 en b_1 . Dit staat bekend als de mean-value-variant. De fout is klein zolang de spreiding klein is en daaraan wordt bij deze toepassing voldaan. Uitwerking van (4.2) geeft:

$$\sigma^2(c') = \frac{\sigma^2(b_0)}{1-\mu^2(b_1)} + \frac{\mu^2(b_0)\mu^2(b_1)\sigma^2(b_1)}{\{1-\mu^2(b_1)\}^3} + \frac{2\mu(b_0)\mu(b_1)\text{cov}(b_0, b_1)}{\{1-\mu^2(b_1)\}^2} \quad (4.3)$$

We werken deze formule numeriek uit voor het voorbeeld van hoofdstuk 2. Berekend was daar $\sigma(\varepsilon) = 0,74 \text{ kN/m}^2$. Met behulp van (2.14), (2.15) en

(2.16) volgt dan:

$$\sigma^2(b_0) = (0,74)^2 \left\{ \frac{1}{4} + 59^2/4482 \right\} = 0,567 \quad (\text{kN/m}^2)^2$$

$$\sigma^2(b_1) = (0,74)^2/4482 = 0,000122$$

$$\text{cov}(b_0 b_1) = (0,74)^2 * (-59/4482) = - 0,0072 \quad (\text{kN/m}^2)^2$$

Ter informatie: daaruit volgt $\sigma(b_0) = 0,75 \text{ kN/m}^2$ en $\sigma(b_1) = 0,011$, ofwel variatiecoëfficiënten van 15 % en 5 %; de correlatiecoëfficiënt $\rho(b_0 b_1) = \text{cov}(b_0 b_1)/\sigma(b_0) \sigma(b_1)$ is gelijk aan - 0,87 duidend op een vrij grote negatieve correlatie.

Met behulp van deze gegevens volgt voor de variantie van de cohesie:

$$\sigma^2(c') = 0,595 + 0,0001 - 0,008 = 0,587$$

We concluderen dat de eerste term verreweg het meest dominant is; dit is een reden geweest om in de leidraad alleen deze ene term als benadering op te nemen:

$$\sigma^2(c') = \frac{\sigma^2(b_0)}{1 - \mu^2(b_1)} \quad (4.4)$$

Analoog volgt als benadering voor de variantie van de wrijvingshoek:

$$\sigma^2(\text{tg}\phi') = \frac{\sigma^2(b_1)}{1 - \mu^2(b_1)} \quad (4.5)$$

Tenslotte wordt, in overeenstemming met de hierboven toegepaste benaderingen, de correlatiecoëfficiënt van c' en $\text{tg}\phi'$ gelijk gesteld aan die van b_0 en b_1 :

$$\rho(c', \text{tg}\phi') = \rho(b_0 b_1) \quad (4.6)$$

Als laatste stap moeten voor τ als functie van σ de statistische eigenschappen worden bepaald. Zoals bekend wordt het verband tussen σ en τ gegeven door:

$$\tau = c' + \sigma \operatorname{tg} \phi' \quad (4.7)$$

Voor de gemiddelde waarde van τ bij gegeven σ volgt dan:

$$\mu(\tau) = \mu(c') + \sigma \mu(\operatorname{tg} \phi') \quad (4.8)$$

De variantie van τ wordt, gebruik makend van de formules van appendix 2, gegeven door:

$$\sigma^2(\tau) = \sigma^2(c') + \sigma^2 \sigma^2(\operatorname{tg} \phi') + 2\sigma\rho(c', \operatorname{tg} \phi')\sigma(c')\sigma(\operatorname{tg} \phi') \quad (4.9)$$

Opmerking: in deze formule staat σ voor 'normaalspanning' en $\sigma(\dots)$ voor standaardafwijking.

Tenslotte kan, overeenkomstig formule (2.20) een veilige schatting gegeven worden voor de opneembare schuifspanning τ gegeven normaalspanning σ :

$$\tau_{\text{veilig}} = \mu(\tau) - t_{n-2} \sigma(\tau) \quad (4.10)$$

We continueren nu ons rekenvoorbeeld. Met behulp van (4.4) t/m (4.6) volgt:

$$\sigma(c') = \frac{0,75}{\sqrt{1 - (0,217)^2}} = 0,77$$

$$\sigma(\operatorname{tg} \phi') = \frac{0,011}{\sqrt{1 - (0,217)^2}} = 0,011$$

$$\rho(c' \operatorname{tg} \phi') = -0,87$$

Het resultaat van (4.8) t/m (4.10) is voor $t_{n-2} = t_2 = 2,92$ voor $\alpha = 0.05$ weergegeven in figuur 5. Een nadeel van dit resultaat is dat in de glijvakberekeningen niet met één waarde voor de cohesie en een waarde voor de hoek van inwendige wrijving gewerkt kan worden. Om dit toch mogelijk te maken zijn in de leidraad 2 alternatieven opgenomen. Het eerste alternatief is om te werken met een veilige schatting voor c' en $\operatorname{tg} \phi'$:

$$c'_{\text{veilig}} = \mu(c') - t_{n-2} \sigma(c') \quad (4.11)$$

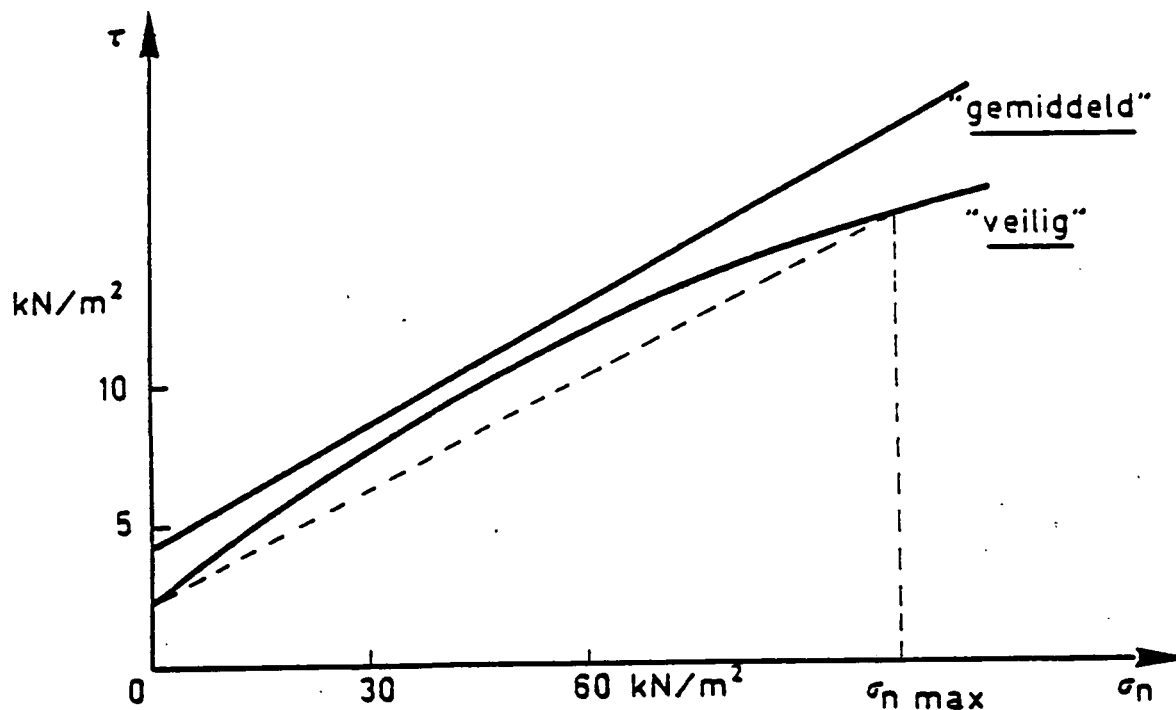
$$tg\phi'_{\text{veilig}} = \mu(tg\phi') - t_{n-2} \sigma(tg\phi') \quad (4.12)$$

De resulterende bezwijklijn voor het criterium van Mohr-Coulomb is eveneens weergegeven in figuur 5. Deze benadering blijkt erg aan de conservatieve kant. Een tweede voorgestelde procedure is om op curve (4.10) het punt aan te wijzen met de maximaal optredende normaalspanning. Een lineaire interpolatie tussen dit punt en het punt $(0, c'_{\text{veilig}})$ is een veilige benadering, die duidelijk minder conservatief is dan de benadering op grond van (4.11) en (4.12).

Indien men uit een bepaalde geologische formatie over monsters beschikt afkomstig van meerdere lokaties, maar wel zodanig dat per lokatie een c' en ϕ' -waarde bepaald kan worden, dan zal men in het algemeen verschillende schattingen vinden voor c' en ϕ' . De oorzaak hiervan is onder meer dat de grondeigenschappen, ook in een als homogeen beschouwde geologische laag, ruimtelijke fluctuaties vertonen. In hoofdstuk 5 wordt daar nader op ingegaan. Wel heeft het nog zin te wijzen op het volgende: verondersteld mag worden dat de meetfout ϵ voor de beschouwde formatie in alle lokaties dezelfde onzekerheden bevat en in ongeveer gelijke mate. Het voordeel daarvan is dat $\sigma(\epsilon)$ geschat kan worden op grond van alle lokaties tezamen. Indien c' en ϕ' op m plaatsen bepaald zijn, volgt daarmee $\sigma^2(\epsilon)$ uit:

$$\sigma^2(\epsilon) = \frac{1}{m(n-2)} \sum_{i=1}^m \epsilon_i^2 \quad (4.13)$$

Overeenkomstig kan in alle formules t_{n-2} vervangen worden door $t_{m(n-2)}$. Bij grote spreiding in de $\sigma(\epsilon)$ -waarden van de individuele lokaties kan de geldigheid van deze pooling eventueel nog getest worden met een F-test.



Figuur 5: De gemiddelde lijn komt overeen met formule (4.8):

$$\mu(\tau) = \mu(c') + \sigma E(\operatorname{tg}\phi');$$

De veilige lijn is formule (4.10):

$$\tau_{\text{veilig}} = \mu(\tau) - t_{n-2} \sigma(\tau) \text{ met } \mu(\tau) \text{ volgens (4.8) en } \sigma(\tau) \text{ volgens (4.9).}$$

De stippellijn geeft een praktische benadering aan als $\sigma_{n \max}$ de maximale normaalspanning in een bepaalde berekening voorstelt.

5. RUIMTELIJKE SPREIDING

Ten behoeve van een glijvlakberekening van een dijklichaam worden in beginsel een aantal monsters gestoken en beproefd. Indien voor ieder monster een analyse wordt uitgevoerd als aangegeven in de voorgaande hoofdstukken, dan zal men in het algemeen per monster een ander resultaat vinden. Een deel van de gevonden verschillen kan worden toegeschreven aan meetonauwkeurigheden, maar een ander deel is het gevolg van de ruimtelijke spreiding van de grondeigenschappen. Terwille van de eenvoud zullen in dit hoofdstuk de meetonauwkeurigheden buiten beschouwing worden gelaten. We nemen dus aan dat de spreiding in gevonden c' en ϕ' waarden geheel en al afkomstig is van ruimtelijke fluctuaties in de grond.

Als illustratie voor de verdere beschouwingen kiezen we de resultaten van een zestal CL-4-proeven als aangegeven in tabel 5.1. Deze resultaten zijn ontleend aan het rapport "Vergelijkend onderzoek Cel- en Triaxiaalproeven" van het LGM (lit. [10]). Het rekenvoorbeeld van de voorgaande hoofdstukken had betrekking op het eerste monster van deze serie. Na omrekening van de β_0 β_1 waarden naar c' en ϕ' vinden we de volgende resultaten:

monster	c' kN/m^2	$\text{tg } \phi'$
11	4,96	0,222
33	3,70	0,243
39	2,42	0,257
68	3,37	0,275
95	2,45	0,252
106	2,32	0,241

Om een indruk te kunnen opbouwen van de ruimtelijke spreiding in de grondeigenschappen bepalen we van bovengenoemde resultaten de

mon- ster nr.	σ'_v kN/m ²	$\sigma'_H(M)$ kN/m ²	ϵ_x	ϵ_{vol}	P kN/m ²	q kN/m ²	Opmerkingen
11	25,5 55,0 94,7 133,4	8,7 27,8 51,5 78,7	0,027 0,052 0,076 0,095		17,1 41,4 73,1 106,1	8,4 13,6 21,6 27,4	voorgeconsolideerd bij 30 kN/m ²
33	25,3 55,1 95,1 135,0	9,8 28,0 53,3 77,5	0,018 0,041 0,067 0,086		17,6 41,5 74,2 106,2	7,7 13,6 20,9 28,8	" "
39	25,0 54,4 94,2 133,9	10,7 29,4 54,0 76,0	0,016 0,038 0,067 0,090		17,8 41,7 74,1 104,9	7,2 12,5 20,1 29,0	" "
68	24,8 54,1 93,2 132,1	9,8 25,6 49,1 71,8	0,017 0,047 0,079 0,102		17,3 39,8 71,2 102,0	7,5 14,3 22,0 30,2	" "
95	25,3 55,0 95,6 135,7	10,7 30,3 54,9 77,9	0,022 0,046 0,066 0,097		18,0 42,7 75,2 106,8	7,3 12,3 20,3 28,9	" "
102	24,8 53,6 93,1 132,1	10,9 30,0 53,9 74,8	0,023 0,044 0,069 0,099		17,7 41,8 73,5 99,2	6,9 11,8 19,6 33,4	" "
106	25,3 55,1 95,7 136,1	11,2 31,6 55,9 80,3	0,020 0,043 0,069 0,098		18,3 43,3 75,8 108,2	7,1 11,7 19,9 27,9	" "

Opmerking: Proef 102 is niet meegenomen omdat daarbij vermoedelijk een scheur in de celwand is ontstaan.

Tabel 5.1: Resultaten Cl-4-proeven (kopie van pagina 34 van LGM rapport 230640-I, lit [10]). Proef 102 is niet meegenomen omdat daarbij een scheur in de celwand zou zijn ontstaan.

gemiddelden, varianties en covariantie. De standaardformules daarvoor luiden:

$$m(c') = \frac{1}{n} \sum c'_i \quad (5.1)$$

$$m(\operatorname{tg} \phi') = \frac{1}{n} \sum \operatorname{tg} \phi'_i \quad (5.2)$$

$$s^2(c') = \frac{1}{n-1} \sum [c'_i - m(c')]^2 \quad (5.3)$$

$$s^2(\operatorname{tg} \phi') = \frac{1}{n-1} \sum [\operatorname{tg} \phi'_i - m(\operatorname{tg} \phi')]^2 \quad (5.4)$$

$$\operatorname{cov}(c' \operatorname{tg} \phi) = \frac{1}{n-1} \sum [c'_i - m(c')][\operatorname{tg} \phi'_i - m(\operatorname{tg} \phi')] \quad (5.5)$$

Voor de gemiddelden volgt:

$$m(c') = 3,20 \text{ kN/m}^2 \quad m(\operatorname{tg} \phi') = 0,248$$

De varianties en covarianties worden gegeven door:

$$s^2(c') = 1,06 \text{ (kN/m}^2\text{)}^2$$

$$s^2(\operatorname{tg} \phi') = 0,00592$$

$$\operatorname{cov}(c' \operatorname{tg} \phi) = -0,007107 \text{ kN/m}^2$$

We bepalen hieruit eenvoudig de standaardafwijkingen voor c' en ϕ' en de correlatiecoëfficiënt $\rho(c', \operatorname{tg} \phi')$:

$$s(c') = 1,03 \text{ kN/m}^2 \quad s(\operatorname{tg} \phi') = 0,023$$

$$r(c' \operatorname{tg} \phi') = \frac{\operatorname{cov}(c' \operatorname{tg} \phi')}{s(c') s(\operatorname{tg} \phi')} = -0,30$$

Niet onverwachts blijkt de spreiding in de cohesie c' relatief groter te zijn dan de spreiding in de tangens van ϕ' . De negatieve waarde van r duidt erop dat men bij grote waarden van ϕ' meestal kleine waarden van c' aantreft en omgekeerd. Ter voorkoming van mogelijke misverstanden nog het volgende: de correlatiecoëfficiënt van c' en $\operatorname{tg} \phi'$ zoals hier

bepaald, moet niet worden verward met de correlatie-coëfficiënt zoals die wordt berekend bij een regressie-analyse. De correlatie-coëfficiënt bij de regressie-analyse geeft niet de mate van afhankelijkheid weer van de regressie-coëfficiënten maar van de regressie-variabelen x en y (of σ' en τ').

Bij glijvlakberekeningen is men in principe geïnteresseerd in de maximaal mogelijke schuifspanningswaarde τ bij een gegeven normaalspanning σ .

Het verband tussen deze grootheden wordt gegeven door:

$$\tau = c' + \sigma \operatorname{tg} \phi' \quad (5.6)$$

De ruimtelijke spreiding van c' en $\operatorname{tg} \phi'$ maakt dat ook de maximaal opneembare schuifspanning van punt tot punt in de grond zal verschillen. Het gemiddelde en de standaardafwijking volgen daarbij uit (analoog aan (4-8 en (4-9)):

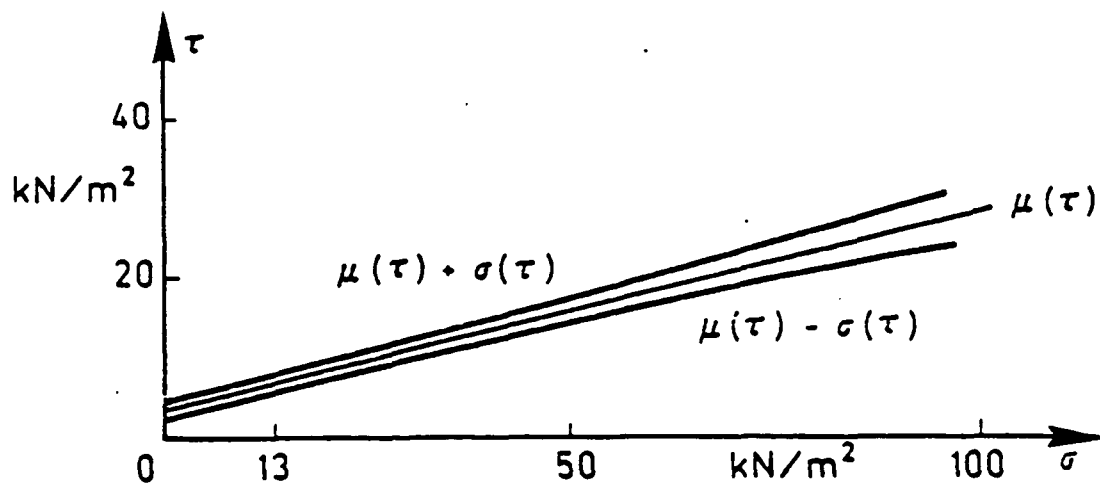
$$\mu(\tau) = \mu(c') + \sigma \mu(\operatorname{tg} \phi') \quad (5.7)$$

$$\sigma^2(\tau) = \sigma^2(c') + \sigma^2 \cdot \sigma^2(\operatorname{tg} \phi') + 2 \cdot \sigma \cdot \operatorname{cov}(c' \operatorname{tg} \phi') \quad (5.8)$$

Figuur 6 geeft de lijnen $\mu(\tau)$ en $\mu(\tau) \pm \sigma(\tau)$ als functie van σ . De lijnen hebben een vergelijkbaar verloop als de lijn voor τ_{veilig} in figuur 5. Het is daarom goed de verschillen in aard en interpretatie nog eens duidelijk te laten uitkomen.

In figuur 5 is het stochastische karakter van τ het gevolg van het feit dat een beperkt aantal onnauwkeurige meetresultaten een exacte vaststelling onmogelijk maakt. Bij opvoering van het aantal metingen zal $\sigma(\tau)$ voor alle waarden van σ naar nul gaan, vooropgesteld natuurlijk dat het gekozen model juist is.

De vorm van $\sigma(\tau)$ als functie van σ is geheel en al bepaald door de gekozen meetpunten. De spreiding is het kleinste in het midden van het meetgebied en groeit naar buiten toe. Dit ondanks de veronderstelde gelijke waarde van $\sigma(\epsilon)_1$ voor alle meetpunten.



Figuur 6: Gemiddelde $\pm$ standaardafwijking voor de schuifspanning τ als functie van σ , berekend op basis van een zestal monsterresultaten.

Bij de spreiding $\sigma(\tau)$ in figuur 6 ligt de zaak anders. Hierbij valt van te voeren niet uit te maken waar de spreiding minimaal is. Dit wordt gedicteerd door de natuur. In een ander geval juist bij een grotere waarde. De grootte van $\sigma(\tau)$ heeft ook niets te maken met grotere of kleinere meetnauwkeurigheid. Tenslotte zal bij toenemend aantal waarnemingen ook geen daling van $\sigma(\tau)$ optreden.

In het onderhavige voorbeeld blijkt $\sigma(\tau)$ het kleinste te zijn bij een normaalspanning $\sigma = 13 \text{ kN/m}^2$. De waarde van $\sigma(\tau)$ is daar $0,98 \text{ kN/m}^2$, vrijwel even groot dus als $\sigma(c')$. Bij $\sigma = 60 \text{ kN/m}^2$ is de spreiding in de schuifspanning opgelopen tot $1,45 \text{ kN/m}^2$. We zullen voor deze middelgrote spanning de ruimtelijke spreiding eens vergelijken met de onzekerheid t.g.v. de meetfouten:

ruimtelijke spreiding: $\sigma = 60 \text{ kN/m}^2$

$$\frac{\sigma(\tau)}{\mu(\tau)} = \frac{1,45}{18,00} = 8\%$$

meetnauwkeurigheid : $\sigma = 60 \text{ kN/m}^2$

$$\frac{\sigma(\tau)}{\mu(\tau)} = \frac{0,37}{17,75} = 2\%$$

Dit resultaat kan gezien worden als een rechtvaardiging achteraf van de verwaarlozing van de meetfouten in dit voorbeeld. In het middengebied blijkt de variatiecoëfficiënt van de meetfouten inderdaad een orde kleiner. Voor kleine waarden van σ kan de interactie overigens nog best aanzienlijk zijn.

Een volgend belangrijk punt is welke waarde van τ in een berekening gehanteerd moet worden. Een veel gebruikte waarde is die van het zogenaamde veilig geschatte gemiddelde gegeven door:

$$\tau_{\text{gem.veilig}} = \mu(\tau) - t_{n-1} \frac{\sigma(\tau)}{\sqrt{n}}$$

Hierbij is n de steekproefomvang, in dit geval dus 6; voor 5% onderschrijding en $n = 6$ volgt verder $t_{n-1} = 2,02$ zodat:

$$\tau_{\text{gem.veilig}} = \mu(\tau) - 0.83 \sigma(\tau)$$

Als de steekproefomvang n toeneemt wordt het veilig geschatte gemiddelde tenslotte gelijk aan het werkelijke gemiddelde. Het is overigens nog maar de vraag of men met deze gemiddelde waarde moet rekenen. Per slot van rekening vertoont de grond een ruimtelijke spreiding en zijn er dus sterke en zwakke plaatsen. Het zou dus wel eens de voorkeur kunnen verdienen om niet met het gemiddelde maar met een karakteristieke waarde te werken, bijvoorbeeld gegeven door:

$$\tau_{\text{karakteristiek}} = \mu(\tau) - 1,64 \sigma(\tau)$$

Dit is de karakteristieke waarde horende bij een 5% onderschrijdingskans. Uiteraard moet men bij een beperkte steekproefomvang, net als bij het gemiddelde op deze waarde ook nog een "veilige schattingscorrectie" aanbrengen. De grootte van deze correctie wordt dan (zie [11]):

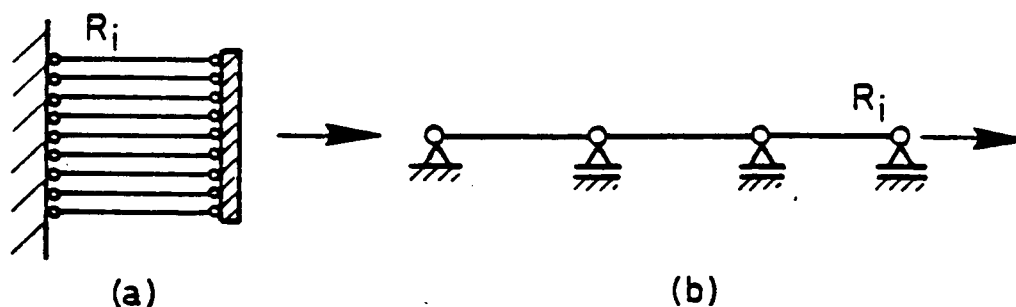
$$\tau_{\text{kar.veilig}} = \mu(\tau) - \sigma(\tau) \sqrt{(2.04)^2 + (0.83)^2} = \mu(\tau) - 2.2 \sigma(\tau)$$

De vraag waarmee nu in een glijvlakberekening gewerkt moet worden, een gemiddelde of een karakteristieke waarde, is overigens niet zo gemakkelijk te beantwoorden. Laten we om de problematiek duidelijk te maken

eerst kijken naar de beide elementaire draagkrachtsystemen zoals weergegeven in figuur 7. In termen van betrouwbaarheidsanalyse staat het eerste systeem bekend als een parallelsysteem en het tweede als een seriesysteem [2]. De sterkte van de afzonderlijke elementen duiden we aan met R_i . Bij het parallelsysteem is de sterkte van het systeem dan gelijk aan de som van alle R_i -waarden (bij plastisch materiaalgedrag).

De sterkte van het seriesysteem is daarentegen gelijk aan de kleinste van alle R_i . Zonder op details in te gaan zal het duidelijk zijn dat men bij het parallelsysteem redelijkerwijs mag werken met de gemiddelde sterkte, terwijl men bij het seriesysteem beter een lage waarde kan kiezen.

Bij de stabiliteitsberekening van een dijk is een afzonderlijk glijvlak te vergelijken met een parallelsysteem. Het rekenen met gemiddelde waarden ligt daar dan ook voor de hand. In werkelijkheid zijn er echter een groot aantal mogelijke glijvlakken. Hierdoor ontstaat een serie-



Figuur 7: Twee elementaire systemen uit de betrouwbaarheidsanalyse:
(a) parallelsysteem; (b) seriesysteem

systeem-effect en het werken met gemiddelde waarden zou daarom best eens iets te optimistisch kunnen zijn. Met name bij een lange dijk moet rekening worden gehouden met een sterkte-reductie door serie-werking.

Een belangrijk aspect dat in bovenstaande problemen nog een rol speelt is de correlatie. Twee monsters kunnen in principe niet als statistisch onafhankelijk worden beschouwd maar bezitten een afhankelijkheid als functie van hun onderlinge afstand. Deze correlatie werkt nivellerend op de beide genoemde volume-effecten. Het is de bedoeling dit verschijnsel te onderzoeken binnen het project "Risicoanalyse voor stabiliteitsberekeningen" [12]. De benodigde correlatiegegevens daarvoor zullen overigens nog verzameld moeten worden. In principe staan hiervoor twee wegen open:

1. Door het beproeven van monsters met verschillende afmetingen.
2. Via een gerichte analyse van gewone monsters met bekende ruimtelijke coördinaten.

Momenteel wordt de voorkeur gegeven aan methode 2; een onderzoek in die richting zal in 1985 worden uitgevoerd. Voor een voorlopige uitwerking van de veiligheidsfilosofie met betrekking tot de serie- en parallelwerking wordt verwezen naar [13].

Indien men bij het ontwerpen aandacht wilt geven aan het probleem van de ruimtelijke spreiding, dan zou men kunnen overwegen per locatie slechts één waarneming te doen. De waarden van c' en ϕ' stelt men dan dus vast via een regressie-analyse op proefresultaten van ver uit elkaar gelegen monsters. Men name voor de triaxiaal proef zou een dergelijke handelwijze best voor de hand liggen. Hoofdstuk 6 nader op deze procedure in.

6. REGRESSIE-ANALYSE BIJ EEN WAARNEMING PER LOCATIE

In dit hoofdstuk beschouwen we de situatie dat per locatie niet meer dan één punt in het p-q-diagram wordt gemeten. Laten we aannemen dat de metingen geen fouten bevatten. Het feit dat de verschillende meetpunten niet op één rechte lijn liggen, is dus uitsluitend het gevolg van ruimtelijke spreiding van c' en $\text{tg } \phi'$, ofwel van b_0 en b_1 in het p-q-assenstelsel. De regressie-coëfficiënten b_0 en b_1 hebben dus ruimtelijke gemiddelden $\mu(b_0)$ en $\mu(b_1)$, standaardafwijkingen $\sigma(b_0)$ en $\sigma(b_1)$ en een correlatie-coëfficiënt $\rho(b_0, b_1)$. Als voor de ruimtelijke spreiding van b_0 en b_1 een normale verdeling wordt verondersteld, kunnen we de regressie-coëfficiënten van het i-de grondmonster schrijven als:

$$b_{0i} = \mu(b_0) + u_i \sigma(b_0) \quad (6.1)$$

$$b_{1i} = \mu(b_1) + v_i \sigma(b_1) \quad (6.2)$$

Hierin zijn u_i en v_i normaal verdeelde variabelen met gemiddelden nul en standaardafwijking 1; de correlatie-coëfficiënt $\rho(u_i, v_i)$ van u_i en v_i is gelijk aan $\rho(b_0, b_1)$. Van correlaties tussen de monsters wordt in dit rapport afgezien.

Indien nu een meting wordt verricht aan het i-de grondmonster dan geldt voor het meetpunt (p_i, q_i) dat exact wordt voldaan aan:

$$q_i = b_{0i} + b_{1i}p_i \quad (6.3)$$

De meetfouten zijn immers nul verondersteld. Via (6.1) en (6.2) volgt:

$$q_i = \mu(b_0) + u_i \sigma(b_0) + p_i \mu(b_1) + p_i v_i \sigma(b_1)$$

Ofwel anders gerangschikt:

$$\mu(b_0) + \mu(b_1)p_i = q_i - u_i \sigma(b_0) - p_i v_i \sigma(b_1) \quad (6.4)$$

We kunnen dit ook opvatten als:

$$\mu(b_0) + \mu(b_1) p_1 = q_1 + \varepsilon_1 \quad (6.5)$$

Hierbij is ε_1 een fout met een normale verdeling, een gemiddelde nul en een variantie σ_ε^2 volgens:

$$\sigma_\varepsilon^2 = \sigma^2(b_0) + 2p_1\sigma(b_0)\sigma(b_1)\rho(b_0b_1) + p_1^2\sigma^2(b_1) \quad (6.6)$$

We concluderen dat dit geval leidt tot een regressiesituatie met de gemiddelden $\mu(b_0)$ en $\mu(b_1)$ als regressie-coëfficiënten en fouten uitsluitend in de q -richting (y -richting). Het verschil met de standaardregressie-situatie is dat σ_ε niet constant is maar een kwadratische functie van p_1 . Voor de schatting van de gemiddelden $\mu(b_0)$ en $\mu(b_1)$ geeft dat geen problemen. We kiezen hiervoor de gewone kleinste kwadratenschatters $m(b_0)$ en $m(b_1)$ volgens:

$$m(b_0) = \bar{q} - m(b_1) \cdot \bar{p} \quad (6.7)$$

$$m(b_1) = \frac{\sum(q_1 p_1) - n\bar{p}\bar{q}}{\sum(p_1^2) - n\bar{p}^2} \quad (6.8)$$

Invullen van q_1 volgens (6-3) leidt op de gebruikelijke manier (zie ook hoofdstuk 2) tot:

$$m(b_0) = \mu(b_0) + \sum\left(\frac{1}{n} - \alpha_1 \bar{p}\right)\varepsilon_1 \quad (6.9)$$

$$m(b_1) = \mu(b_1) + \sum \alpha_1 \varepsilon_1 \quad (6.10)$$

$$\text{met } \alpha_1 = (p_1 - \bar{p}) / \sum (p_1 - \bar{p})^2 \quad (6.11)$$

De schatters $m(b_0)$ en $m(b_1)$ zijn derhalve zuivere schatters in de zin van de klassieke mathematische statistiek. We vervolgen overigens vanaf dit punt, net als in hoofdstuk 2, met een Bayesiaanse interpretatie en vatten verder $\mu(b_0)$ en $\mu(b_1)$ op als stochastische variabelen waarvan de spreiding het gevolg is van de beperkte omvang van de steekproef. De

verwachtingswaarden van $\mu(b_0)$ en $\mu(b_1)$ zijn respectievelijk gelijk aan $m(b_0)$ en $m(b_1)$ zoals die via (6.7) en (6.8) uit de steekproef kunnen worden bepaald. De varianties en covariantie van $\mu(b_0)$ en $\mu(b_1)$ worden (zie hoofdstuk 2 gegeven door:

$$\sigma^2(\mu(b_0)) = \Sigma \left(\frac{1}{n} - \alpha_1 \bar{p} \right)^2 \sigma^2(\epsilon_1) \quad (6.12)$$

$$\sigma^2(\mu(b_1)) = \Sigma \alpha_1^2 \sigma^2(\epsilon_1) \quad (6.13)$$

$$\text{cov}(\mu(b_0)\mu(b_1)) = \Sigma \alpha_1 \left(\frac{1}{n} - \alpha_1 p_1 \right) \sigma^2(\epsilon_1) \quad (6.14)$$

Het verschil met de formules (2.14) t/m (2.16) is dat $\sigma^2(\epsilon_1)$ niet voor het sommatie-teken mag worden gebracht omdat $\sigma^2(\epsilon_1)$ een functie is van p_1 (zie 6.6). Als deze $\sigma^2(\epsilon_1)$ als functie van p_1 helemaal bekend zouden zijn, dat wil dus zeggen als $\sigma(b_0)$, $\sigma(b_1)$ en $\rho(b_0 b_1)$ bekend zouden zijn, dan was het uitwerken van (6.12) enz. geen probleem. We zouden vervolgens dan kunnen overgaan tot het bepalen van "veilige" gemiddelde waarden voor b_0 en b_1 (ofwel c en $\text{tg } \phi$), dan wel "veilige" karakteristieke totale sterkten of "veilige" karakteristieke glijvlaksterkten (zie [11]). In principe zijn de spreidingseigenschappen van b_0 en b_1 echter niet bekend en moeten deze ook uit de steekproefresultaten worden bepaald. In beginsel is dit mogelijk al blijkt het moeilijk efficiënte schatters te vinden. In een hierna uit te voeren rekenvoorbeeld zullen de volgende statistics worden gebruikt:

$$D_1 = \frac{1}{n-2} \Sigma (\Delta q_1)^2 \quad (6.15)$$

$$D_2 = \frac{1}{n-2} \Sigma (\Delta q_1)^2 / p_1 \quad (6.16)$$

$$D_3 = \frac{1}{n-2} \Sigma (\Delta q_1)^2 / p_1^2 \quad (6.17)$$

Hierin zijn Δq_1 de residuen, d.w.z. de verschillen tussen de meetpunten en de gevonden regressielijn. Op grond van (6.4) geldt voor Δq_1 :

$$\Delta q_1 = u_1 \sigma(b_0) + p_1 v_1 \sigma(b_1) \quad (6.18)$$

De normale standaardvariabelen u_1 en v_1 zijn daarbij onderling gecorreleerd met correlatiecoëfficiënt $\rho(b_0 b_1)$. We bepalen de verwachtingswaarden van D_1 t/m D_3 :

$$E(D_1) = \frac{n}{n-2} [\sigma^2(b_0) + 2 \bar{p} \text{cov}(b_0 b_1) + \bar{p}^2 \sigma^2(b_1)] \quad (6.19)$$

$$E(D_2) = \frac{n}{n-2} [(\bar{1/p}) \sigma^2(b_0) + 2 \text{cov}(b_0 b_1) + \bar{p} \sigma^2(b_1)] \quad (6.20)$$

$$E(D_3) = \frac{n}{n-2} [(\bar{1/p^2}) \sigma^2(b_0) + 2 (\bar{1/p}) \text{cov}(b_0 b_1) + \sigma^2(b_1)] \quad (6.21)$$

De horizontale balken geven aan dat de gemiddelde waarde genomen moet worden; gebruik is gemaakt van $\text{cov}(b_0 b_1) = \rho(b_0 b_1) \sigma(b_0) \sigma(b_1)$. We concluderen dat schattingen voor $\sigma^2(b_0)$, $\sigma^2(b_1)$ en $\text{cov}(b_0 b_1)$ gevonden kunnen worden door de volgende drie vergelijkingen met drie onbekenden op te lossen:

$$\begin{bmatrix} 1 & 2 & \bar{p}^2 \\ (\bar{1/p}) & 2 & \bar{p} \\ (\bar{1/p^2}) & 2(\bar{1/p}) & 1 \end{bmatrix} \begin{bmatrix} \sigma^2(b_0) \\ \text{cov}(b_0 b_1) \\ \sigma^2(b_1) \end{bmatrix} = \frac{1}{n-2} \begin{bmatrix} \Sigma (\Delta q_1)^2 \\ \Sigma (\Delta q_1)^2 / p_1 \\ \Sigma (\Delta q_1)^2 / p_1^2 \end{bmatrix} \quad (6.22)$$

De factor $(n/n-2)$ in (6.19) t/m (6.21) wordt verwaarloosd omdat bij het bepalen van de regressielijn reeds twee onafhankelijke vrijheidsgraden zijn gebruikt. Het probleem met formule (6.22) is dat de matrix niet sterk geconditioneerd is. Met andere woorden, de drie vergelijkingen zijn bijna afhankelijk en daardoor is de oplossing erg gevoelig voor kleine veranderingen in het rechterlid. Mede onder invloed daarvan kan het voorkomen dat voor één van de varianties een negatieve waarde resulteert of dat de covariantie groter of kleiner wordt dan het produkt van de standaardafwijkingen. Men dient hierop bedacht te zijn en de oplossing op deze punten te controleren. Indien gewenst kan in een vervolgonderzoek nog gezocht worden naar betere schattingsprocedures, al zullen problemen zoals hierboven geschetst nooit helemaal vermeden kunnen worden.

Voorbeeld

We beschouwen nogmaals het voorbeeld van tabel 3.2. De meetresultaten daarbij waren:

$p_1 = 17,1 \text{ kN/m}^2$	$q_1 = 8,4 \text{ kN/m}^2$
$p_2 = 41,5$	13,6
$p_3 = 74,1$	20,1
$p_4 = 102,0$	30,2

We gaan ervan uit dat deze meetresultaten exact kloppen en dat de punten niet op een rechte lijn liggen, omdat b_0 en b_1 (ofwel c en $\tan \phi$) ruimtelijke spreiding vertonen. De grondparameters b_0 en b_1 hebben dus gemiddelden $\mu(b_0)$ en $\mu(b_1)$, standaardafwijkingen $\sigma(b_0)$ en $\sigma(b_1)$ en een correlatiecoëfficiënt $\rho(b_0 b_1)$. Al deze statistische eigenschappen zijn echter onbekend en dienen uit de steekproef bepaald te worden. Aangezien de steekproef beperkt maakt met een bepaalde mate van onzekerheid. In principe zou dus voor alle vijf de statistische parameters een gemiddelde en een standaardafwijking moeten worden bepaald. We zullen dit echter alleen doen voor $\mu(b_0)$ en $\mu(b_1)$; voor $\sigma(b_0)$, en $\rho(b_0 b_1)$ volstaan we met alleen het gemiddelde.

Het gemiddelde voor $\mu(b_0)$ en $\mu(b_1)$ volgen via (6.7) en (6.8) met als resultaat:

$$\mu \{ \mu(b_0) \} = m(b_0) = 3,39 \text{ kN/m}^2$$

$$\mu \{ \mu(b_1) \} = m(b_1) = 0,25$$

Deze getalwaarden waren reeds gevonden in hoofdstuk 2.

We bepalen vervolgens de schattingen voor $\sigma(b_0)$, $\sigma(b_1)$ en $\rho(b_0 b_1)$ m.b.v. (6.22). Het resultaat is:

$$\sigma(b_0) = 0,32 \text{ kN/m}^2$$

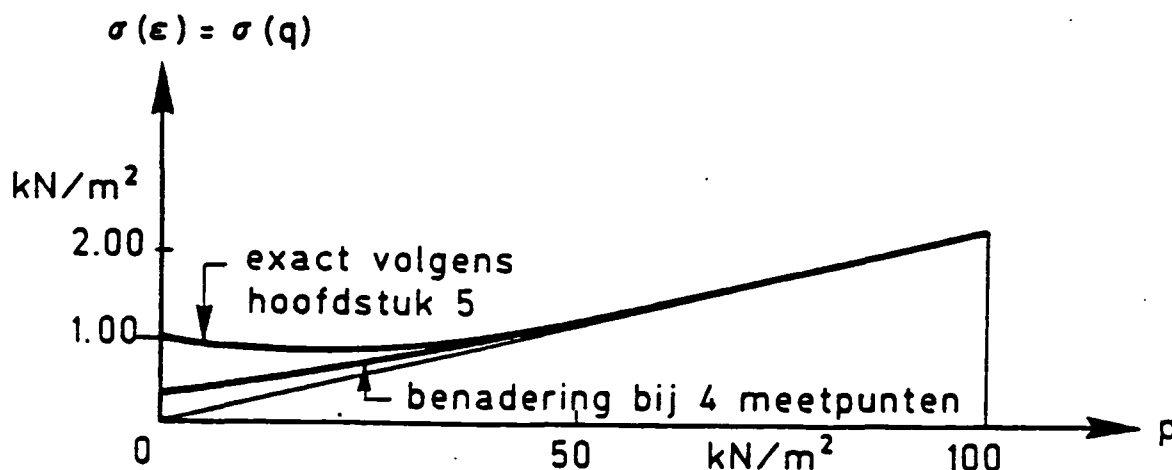
$$\sigma(b_1) = 0,0236$$

$$\rho(b_0 b_1) = + 0,20$$

De hier berekende standaardvariatiën en correlatie zijn ook berekend in hoofdstuk 5, maar dan op basis van alle gegevens, d.w.z. de 24 meetpunten van de 6 viertrapsceelproeven. Het aldaar berekende resultaat kunnen we gemakshalve beschouwen als de exacte oplossing. Gevonden was:

$$\sigma(c') = 1,03 \text{ kN/m}^2; \sigma(\text{tg } \phi') = 0,023 \text{ en } \rho(b_0 b_1) = - 0,30.$$

We constateren dat met de huidige berekening een goed resultaat is behaald voor de dominante spreiding horend bij b_1 ($\text{tg } \phi'$) maar dat voor de andere grootheden de resultaten vrij slecht zijn. De waarde van $\sigma(b_0)$ is een factor 3 telaar en in plaats van een zwak negatieve wordt een zwak positieve correlatie berekend. In figuur 8 is aangegeven wat de consequenties zijn voor $\sigma(\epsilon)$ ofwel $\sigma(q)$. Bij de berekening van $\sigma(\epsilon)$ is uitgegaan van formule (6.6). Voor $p = 0$ is $\sigma(\epsilon) = \sigma(b_0)$ voor $p > 0$ gaat $\sigma(\epsilon)$ langzaam naar $\rho(b_1)$ toe waardoor het huidige resultaat bij grotere p steeds nauwkeuriger wordt.



figuur 8 Vergelijking van $\sigma(\epsilon)$ op basis van 24 en 4 waarnemingen; voor grote waarden van p is de benadering goed maar niet nabij $p = 0$.

Nu $\sigma(b_0)$, $\sigma(b_1)$ en $\rho(b_0, b_1)$ bekend zijn kunnen we van $\mu(b_0)$ en $\mu(b_1)$ de varianties en covarianties bepalen m.b.v. (5-12) en (5-13). We geven het resultaat hier in de vorm van hun standaardafwijkingen en de correlatiecoëfficiënten:

$$\sigma(\mu(b_0)) = 1,16 \text{ kN/m}^2$$

$$\sigma(\mu(b_1)) = 0,028$$

$$\rho(\mu(b_0), \mu(b_1)) = -0,88$$

We zullen het rekenvoorbeeld afronden door rekenwaarden voor q af te leiden bij gegeven p . (Globaal kan men dat opvatten als rekenwaarden voor τ bij gegeven σ). Als eerste mogelijke rekenwaarde bepalen we de 95% betrouwbaarheidsgrens voor $\mu(q)$. Daartoe moet $\sigma(\mu(q))$ worden uitgerekend. Op grond van $q = b_0 + b_1 p$ geldt: $\mu(q) = \mu(b_0) + \mu(b_1)p$ en derhalve volgt voor de variantie van $\mu(q)$:

$$\sigma^2\{\mu(q)\} = \sigma^2\{\mu(b_0)\} + 2p \text{ cov}\{\mu(b_1)\} + p^2 \sigma^2\{\mu(b_1)\} \quad (6.23)$$

Op basis hiervan kan het veilige gemiddelde worden bepaald via:

$$\mu(q)_{\text{veilig}} = \mu(\mu(q)) - t_{n-2} \cdot \sigma(\mu(q)) \quad (6.24)$$

Aangezien $n = 4$ en 95% betrouwbaarheid wordt gewenst is $t_{n-2} = 2,92$. De resulterende lijn is weergegeven in figuur 9.

Indien men niet wil rekenen met een (veilige) gemiddelde waarde maar met een veilige karakteristieke waarde dan kan deze berekend worden via [11]:

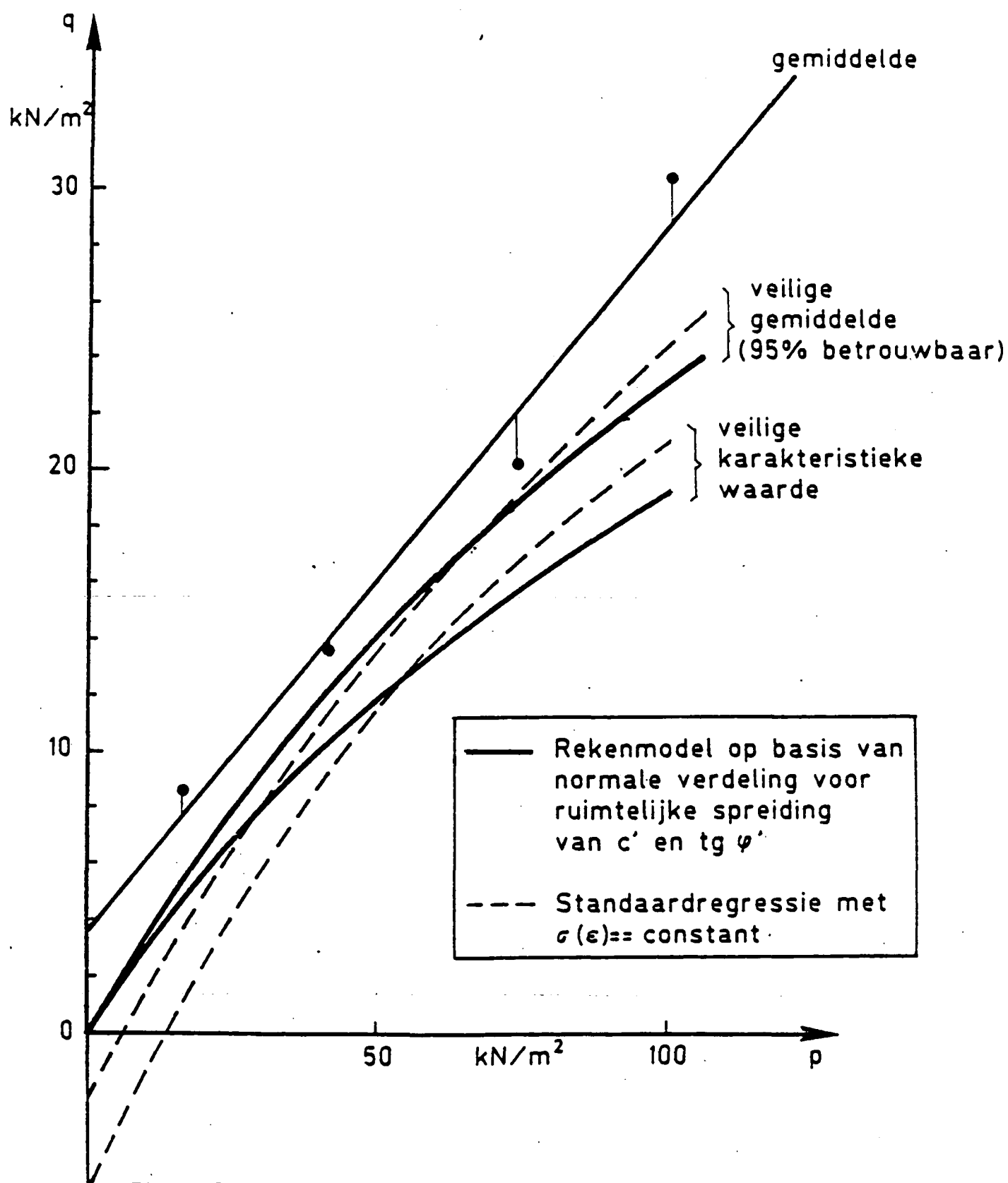
$$q_{\text{kar. veilig}} = \mu(\mu(q)) - t_{n-2} \sqrt{\sigma^2\{\mu(q)\} + \sigma^2(q)} \quad (6.25)$$

Voor $\sigma^2(q)$: zie figuur 9. Het resultaat van formule (6.25) is eveneens opgenomen in figuur 9. We zien dat in het gebied van lage p -waarden de variantie van $\mu(q)$ domineert, pas bij grotere waarden van p gaat de variantie van q zelf een rol spelen.

Uiteraard hadden we de gegeven meetpunten ook op de gewone regressie-manier kunnen behandelen, d.w.z. met $\sigma^2(\epsilon) = \text{constant}$. De resultaten van een dergelijke analyse zijn eveneens uitgezet in figuur 9 (gestip-pelde lijnen). Ook hierbij zijn berekend een veilig gemiddelde en een veilige karakteristieke waarde. Duidelijk valt te constateren dat deze waarden voor lage spanningen veel te pessimistisch uitvallen en dat een ingewikkelde analyse als hier voorgesteld nodig is.

Een belangrijke vraag tenslotte is nog deze: waar blijft in de hier gegeven analyse het keuze aspect van de afhankelijke variabele. Als we de afleidingen globaal langs lopen dan lijkt het alsof de keuze van de kleinste-kwadratenmethode in q-richting onvermijdelijk is. Dit is toch echter niet het geval.

De basis voor deze onvermijdelijkheid wordt in feite gelegd bij de arbitraire aanname dat b_0 en b_1 (ofwel c' en $\text{tg } \phi'$) normaal verdeelde variabelen zijn. Indien men daarentegen bijvoorbeeld zou aannemen dat $(1/\text{tg } \phi')$ en $(c'/\text{tg } \phi')$ normaal verdeeld zijn, dan zou men naderhand uitkomen op een kleinste-kwadratenmethode in p-richting. Uiteraard zou op dit punt nader onderzoek kunnen plaatsvinden. Voorlopig lijkt het om redenen van mathematische verwerkbaarheid het beste om de huidige keuzes van een normale verdeling voor $\text{tg } \phi'$ en c' aan te houden.



Figuur 9 Gemiddelde, veilig gemiddelde en veilige karakteristieke waarde op basis van 4 meetpunten afkomstig van 4 verschillende monsters.

7. SAMENVATTING EN CONCLUSIES

Dit rapport geeft een overzicht van de studie die is uitgevoerd naar de statistische verwerking van de cel- en triaxiaalproeven m.b.v. regressie-analyse. Een deel van de resultaten is terecht gekomen in de COW-leidraad "Cel- en Triaxiaalproeven" van 1986. De belangrijkste informatie met betrekking tot deze leidraad is opgenomen in de hoofdstukken 2 tot en met 4 van dit rapport.

Hoofdstuk 2 geeft de basisformules voor de bepaling van de regressie-coëfficiënten en het betrouwbaarheidsinterval.

In hoofdstuk 3 wordt het probleem aan de orde gesteld welke variabele als afhankelijk en welke als onafhankelijke variabele gekozen zou moeten worden. Volgens de standaardprocedure moet men voor de onafhankelijke x-variabele een parameter nemen die men precies kent en er voor zorgen dat de afhankelijke y-variabele alle onnauwkeurigheid bevat. Het blijkt echter dat een dergelijke keuze niet te maken is en daarom is een literatuurstudie uitgevoerd naar een methode waarbij beide variabelen onnauwkeurig mogen zijn. De gevonden methoden bleken echter weinig geschikt en uiteindelijk is besloten dat de meest gangbare praktijk van het ogenblik (gemiddelde spanning p als x-variabele en halve verschilspanning q als y variabele) goed verdedigbaar is. Bij een viertrapsproef of bij meerdere eentrapspoeven op dicht bij elkaar gelegen monsters heeft de keuze gelukkig ook niet veel invloed. Wel kan men enig verschil van betekenis vinden indien men meerdere eentrapspoeven doet op verder van elkaar gelegen monsters. In de huidige praktijk is dat echter niet gangbaar.

Hoofdstuk 4 geeft aan hoe men op basis van een uitgevoerde regressie-analyse komt tot waarden voor cohesie en wrijvingshoek waarmee operationeel in een glijkvlakberekening gewerkt kan worden. Het blijkt uiteindelijk mogelijk via redelijk eenvoudige formules (4.4 t/m 4.6) een bruikbare benadering op te stellen (figuur 5).

In de hoofdstukken 5 en 6 tenslotte komen nog enige onderwerpen aan de orde die uiteindelijk buiten de leidraad zijn gehouden. Dit betreft op

de eerste plaats de ruimtelijke spreiding en de wijze waarop daar in berekeningen rekening mee gehouden zou kunnen worden. In afwachting van verdere theoretisch en proefondervindelijk onderzoek is dit onderdeel kort gehouden. In hoofdstuk 6 wordt teruggekomen op de mogelijkheid cohesie en wrijvingshoek te bepalen als men slechts één waarneming per locatie heeft. De meetonnauwkeurigheden worden hier verwaarloosd en alleen ruimtelijke fluctuaties worden meegenomen. De procedure blijkt uitvoerbaar, doch is ingewikkeld en niet altijd goed geconditioneerd. Voor de praktijk is deze methode niet anders dan in "noodgevallen" aan te bevelen.

Literatuur

- [1] Benjamin, J.R., Cornell, C.A.
Probability, Statistics and Decision for Civil Engineers.
McGraw Hill Book Company, New York, 1970.

- [2] Vrouwenvelder, A., Vrijling, J.
Probabilistische Ontwerpen (Collegedictaat 63)
TH Delft, Afdeling Civiele Techniek, december 1984.

- [3] Seber, G.A.F.
Linear Regression Analysis.
John Wiley and Sons, New York, 1977.

- [4] Wald, A.
The fitting of straight lines if both variables are subject to error.
Annals of Mathematical Statistics, 1940, 11, page 284-300.

- [5] Berkson, J.
Are there two regressions.
Journal American Statistical Association, 1950, vol. 45, p. 164-168.

- [6] Madansky, A.
The fitting of straight lines when both variables are subject to error.
Journal American Statistical Association, 1959, vol. 54, p. 173-205.

- [7] Sprent, P.
Models in Regression.
Methuen and Co. Ltd, London, 1969.

- [8] Richardson, D., Wu, D.
Least squares and Grouping Method Estimators in the Errors in Variable Model.
Journal American Statistical Association, 1970, Vol. 65, p. 724-748.

-
- [9] Davies, R., Hutton, B.
The effect of errors in the independent variables in linear regression.
Biometrika, 1975, page 383-391.
- [10] Heijnen, W.J.
Vergelijkend onderzoek cel- en triaxiaalproeven.
Laboratorium voor Grondmechanica, Rapport 230640-I, Delft, 1978
(MD-78-6).
- [11] Vrouwenvelder, A.
Rekenwaarden voor grondeigenschappen.
TNO-rapport no. B-80-161/62.6.0406, Rijswijk, maart 1980, (MD-80-14),
- [12] Calle, E., Vrouwenvelder, A.
Notitie over probabilistische analyse van de stabiliteit van grond-
massieven.
TNO-IBBC / LGM, Delft, 1979, (MD-79-15).
- [13] Calle, E.O.F.
Probabilistic Analysis of Stability of Earth Slopes
XI ICSMFE, San Fransisco, 1985.

APPENDIX 1 COEFFICIENTEN REGRESSIEANALYSE

Gegeven zijn een aantal punten (x_i, y_i) . Bij benadering liggen deze punten op de rechte $y = \beta_0 + \beta_1 x$. Gevraagd wordt de coëfficiënten β_0 en β_1 zodanig te bepalen dat de som van de afwijkingen ϵ_i in het kwadraat, met $\epsilon_i = y_i - y$, zo klein mogelijk is. De vergelijking voor deze som is:

$$S = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

De waarden van β_0 en β_1 waarvoor S een minimum bereikt kunnen worden gevonden uit:

$$\frac{\partial S}{\partial \beta_0} = \sum_{i=1}^n 2 (y_i - \beta_0 - \beta_1 x_i) (-1) = 0$$

$$\frac{\partial S}{\partial \beta_1} = \sum_{i=1}^n 2 (y_i - \beta_0 - \beta_1 x_i) (-x_i) = 0$$

Via enige herleiding volgt:

$$\sum y_i - n\beta_0 - \beta_1 \sum x_i = 0 \quad \sum x_i y_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 = 0$$

Na deling door n en introductie van $\bar{x} = \sum x_i / n$ en $\bar{y} = \sum y_i / n$ kan de eerste vergelijking geschreven worden als:

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Dit komt precies overeen met vergelijking (2.3) uit hoofdstuk 2.

Invullen van β_0 in de onderste vergelijking levert vervolgens:

$$\sum x_i y_i - \bar{y} \sum x_i + \beta_1 \bar{x} \sum x_i - \beta_1 \sum x_i^2 = 0$$

$$\text{ofwel: } \beta_1 = \frac{\sum x_i y_i - n \bar{y} \bar{x}}{\sum x_i^2 - n \bar{x}^2}$$

Dit is vergelijking (2.4).

APPENDIX 2 STATISTISCHE EIGENSCHAPPEN BIJ LINEAIRE TRANSFORMATIE

Gegeven zijn de stochastische variabelen x en y met bekende gemiddelden $\mu(x)$ en $\mu(y)$, varianties $\sigma^2(x)$ en $\sigma^2(y)$ en covariantie $\text{cov}(xy)$. Alvorens de precieze definitie van deze grootheden te geven introduceren we eerst het begrip verwachtingswaarden van de functie $g(x,y)$. Deze verwachtingswaarde is gedefinieerd als:

$$E(g(x,y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(\xi, \eta) f_{xy}(\xi, \eta) d\xi d\eta$$

Hierin is $f_{xy}(\xi, \eta)$ de tweedimensionale kansdichtheidsfunctie van x en y . Eenvoudig te bewijzen eigenschappen van de verwachtingswaarde zijn (zie [1], [2]):

$$E(a) = a \quad (I)$$

$$E(ag) = a E(g) \quad (II)$$

$$E(g+h) = E(g) + E(h) \quad (III)$$

en als x en y onafhankelijk zijn:

$$E\{g(x)h(y)\} = E(g(x))E(h(y)) \quad (IV)$$

Het gemiddelde kan nu gedefinieerd worden als de verwachtingswaarden van de stochast zelf:

$$\mu(x) = E(x)$$

De variantie is de verwachtingswaarde van $(x-\mu(x))^2$:

$$\sigma^2(x) = E\{(x-\mu(x))^2\}$$

Analoog is de covariantie van x en y gedefinieerd als de gemengde kwadratische verwachtingswaarde

$$\text{cov}(xy) = E\{(x-\mu(x))(y-\mu(y))\}$$

Als x en y onafhankelijk zijn is de covariantie nul (zie eigenschap IV).
In plaats van met de covariantie werkt men ook vaak met de
correlatiecoëfficiënt ρ_{xy} :

$$\rho_{xy} = \frac{\text{cov}(xy)}{\sigma(x)\sigma(y)}$$

Bewezen kan worden dat $-1 \leq \rho_{xy} \leq +1$

Stel nu dat de variabelen (x,y) getransformeerd worden naar (u,v)
volgens:

$$u = a + bx + cy$$

$$v = d + ex + fy$$

Gevraagd gemiddelden, varianties en covarianties van u en v .

Het gemiddelde van u kan met behulp van de eigenschappen I, II en III
als volgt worden afgeleid:

$$\begin{aligned}\mu(u) &= E(u) = \\ &= E\{a + bx + cy\} = \\ &= a + b E(x) + c E(y) = \\ &= a + b \mu(x) + c \mu(y)\end{aligned}$$

Analoog kan voor de variantie van u worden afgeleid:

$$\begin{aligned}\sigma^2(u) &= E\{(u - \mu(u))^2\} = \\ &= E\{(a + bx + cy - a - b\mu(x) - c\mu(y))^2\} = \\ &= E\{[b(x - \mu(x)) + c(y - \mu(y))]^2\}\end{aligned}$$

$$= b^2 E\{ (x - \mu(x))^2 \} + 2bc E\{ (x - \mu(x))(y - \mu(y)) \} + \\ c^2 E\{ (y - \mu(y))^2 \} = b^2 \sigma^2(x) + 2bc \operatorname{cov}(xy) + c^2 \sigma^2(y)$$

Als x en y onafhankelijk zijn geldt:

$$\sigma^2(u) = b^2 \sigma^2(x) + c^2 \sigma^2(y)$$

Voor $\mu(v)$ en $\sigma^2(v)$ gelden uiteraard overeenkomstige formules. We leiden tot slot nog de uitdrukking voor de covariantie van u en v af:

$$\begin{aligned} \operatorname{cov}(uv) &= E\{ (u - \mu(u))(v - \mu(v)) \} = \\ &= E\{ [b(x - \mu(x)) + c(y - \mu(y))] [e(x - \mu(x)) + f(y - \mu(y))] \} \\ &= be E\{ (x - \mu(x))^2 \} + (bf + ce) E\{ (x - \mu(x))(y - \mu(y)) \} + \\ &\quad cf E\{ (y - \mu(y))^2 \} = \\ &= be \sigma^2(x) + (bf + ce) \operatorname{cov}(xy) + cf \sigma^2(y) \end{aligned}$$

Als x en y onafhankelijk zijn geldt:

$$\operatorname{cov}(uv) = be \sigma^2(x) + cf \sigma^2(y)$$

In het rapport komen soms transformaties voor over meer dan 2 variabelen, in verband daarmee volgen hier zonder afleiding nog enkele algemene formules:

$$y_i = a_{i0} + \sum_{k=1}^n a_{ik} x_k$$

Voor het gemiddelde en de standaardafwijking van y_i geldt:

$$\mu(y_i) = a_{i0} + \sum_{k=1}^n a_{ik} \mu(x_k)$$

$$\sigma^2(y_i) = \sum_{k=1}^n \sum_{l=1}^n a_{ik} a_{il} \text{cov}(x_k x_l)$$

Hierbij is $\sigma^2(x_k)$ genoteerd als $\text{cov}(x_k x_k)$

De covariantie van y_i en y_j wordt gegeven door:

$$\text{cov}(y_i y_j) = \sum_{k=1}^n \sum_{l=1}^n a_{ik} a_{jl} \text{cov}(x_k x_l)$$

Als de x_i onafhankelijk zijn geldt:

$$\sigma^2(y_i) = \sum_{k=1}^n a_{ik} \sigma^2(x_k) \quad \text{cov}(y_i y_j) = \sum_{k=1}^n a_{ik} a_{jk} \sigma^2(x_k)$$

APPENDIX 3 REGRESSIE ANALYSE MET MEETFOUTEN IN BEIDE VARIABELEN

We beschouwen het geval dat tussen twee deterministische variabelen X en Y een lineair verband bestaat volgens:

$$Y = bX \quad \text{III-1}$$

We nemen in deze appendix ter vereenvoudiging aan dat de lijn door de oorsprong gaat. In veel opzichten verschilt dit geval niet van het algemene geval waarbij de lijn niet door de oorsprong gaat. Daar waar het verschil wel van belang is zal hierop afzonderlijk worden ingegaan.

Stel dat de parameter b onbekend is en aan de hand van n meetpunten (x_1, y_1) moet worden geschat. De gemeten waarden (x_1, y_1) worden daarbij verondersteld meetfouten (δ_1, ϵ_1) te bevatten t.o.v. de exacte waarden (X_1, Y_1) :

$$x_1 = Y_1 + \delta_1 \quad \text{III-2}$$

$$y_1 = Y_1 + \epsilon_1 \quad \text{III-3}$$

De fouten δ_1 en ϵ_1 worden aangenomen als onafhankelijke normaal verdeelde stochastische variabelen met gemiddelden nul en standaardafwijkingen σ_δ en σ_ϵ .

Laten we om te beginnen analyseren wat er resulteert als dit probleem wordt behandeld volgens de kleinste-kwadratenmethode. De waarde van b wordt dan geschat met behulp van de schatter $= \beta$ volgens:

$$\beta = \frac{\sum x_1 y_1}{\sum x_1^2}$$

We onderzoeken nu de aard van deze schatter binnen het raamwerk van de klassieke mathematische statistiek. (Opmerking: bij een Bayesiaanse uitwerking komt men hier op bepaalde punten tot tegenovergestelde conclusies). Daartoe bepalen we allereerst de verwachtingswaarde van β :

$$\begin{aligned}
 E(\beta) &= E \left\{ \frac{\sum x_1 y_1}{\sum x_1^2} \right\} = \\
 &= E \left\{ \frac{\sum (X_1 + \delta_1) (bX_1 + \epsilon_1)}{\sum (X_1 + \delta_1)^2} \right\} \\
 &= E \left\{ \frac{b \sum X_1^2 + b \sum X_1 \delta_1 + \sum X_1 \epsilon_1 + \sum \delta_1 \epsilon_1}{\sum X_1^2 + \sum \delta_1^2 + 2 \sum X_1 \delta_1} \right\}
 \end{aligned}$$

Hierbij is gebruik gemaakt van (III-1) t/m (III-3). Een exacte uitwerking van $E(\beta)$ is niet goed mogelijk. We passen daarom de volgende benadering toe:

$$E(\beta) \approx \frac{E \{ b \sum X_1^2 + b \sum X_1 \delta_1 + \sum X_1 \epsilon_1 + \sum \delta_1 \epsilon_1 \}}{E \{ \sum X_1^2 + \sum \delta_1^2 + 2 \sum X_1 \delta_1 \}}$$

In [9] wordt aangetoond dat deze benadering asymptotisch juist is voor grote n . Via $A = \frac{1}{n} \sum X_1^2$ kan $E(\beta)$ verder worden uitgewerkt tot:

$$E(\beta) = \frac{nbA + 0 + 0 + 0}{nA + 0 + \sigma_\delta^2} = b \left\{ \frac{A}{A + \sigma_\delta^2} \right\} \quad (\text{III-4})$$

Uit deze formule, die vaak in de literatuur voorkomt, blijkt dat β geen zuivere schatter is van b . Aangetoond kan worden dat de variantie van β met kans 1 convergeert naar een waarde ongelijk b . Het verschil tussen $E(\beta)$ en b noemt men de "bias". De bias is volgens (III-5) niet groot als de meetpunten ver uit elkaar liggen (A groot) en de variantie van de meetfouten σ_δ^2 vrij klein is.

Het werken met een "biased estimator" heeft natuurlijk geen enkel bezwaar als men de grootte van de bias precies kent, bijvoorbeeld doordat σ_δ^2 bij voorbaat bekend is. Als σ_δ^2 niet bij voorbaat bekend is moet deze uit de meetgegevens worden afgeleid. Hier doet zich echter de kern van het probleem voor: als zowel de x -metingen als de y -metingen

fouten bevatten met onbekende standaardafwijkingen dan is het niet mogelijk om deze op consistente wijze uit de meetgegevens te destilleren (zie Sprent [7]). Minimaal moet men of σ_δ^2 , of σ_ϵ^2 , of hun onderlinge verhouding kennen. Als men over geen van deze informatie beschikt is er sprake van zogenaamde "onidentificeerbaarheid".

Met bovenstaande is overigens geenszins gezegd dat het probleem om een goede schatting voor b te vinden onoplosbaar zou zijn. We zullen alleen naar een andere schatter moeten omzien dan de kleinste-kwadrantenmethode. Men zou bijvoorbeeld kunnen denken aan de methode van de maximum likelihood. Deze methode leidt tot $\beta = \Sigma y_1^2 / \Sigma x_1^2$ (zie [6], [7]) maar ook dit is bij nadere beschouwing een onzuivere schatter. Een andere schatter die wel zuiver blijkt te zijn is de volgende:

$$\beta = \frac{\Sigma y_1}{\Sigma x_1} \quad (\text{III-6})$$

De uitwerking van $E(\beta)$ gebeurt op dezelfde wijze als bij de kleinste kwadratenschatter, d.w.z. met een benadering die allen asymptotisch juist is voor grote n . De schatter (III-6) is derhalve ook niet zonder meer zuiver maar slechts asymptotisch zuiver. Voor eindige n bevat ook deze β een bias die niet kwantificeerbaar is, hooguit kan men hiervoor een bovengrens aangeven door aan te nemen dat alle spreiding die zich in x -richting bevindt. In principe kan voor β ook een standaardafwijking worden bepaald. Daartoe wordt β als volgt uitgeschreven:

$$\beta = \frac{\Sigma (Y_1 + \delta_1)}{\Sigma X_1 + \epsilon_1} = \frac{b\bar{X} + \frac{1}{n} \Sigma \delta_1}{\bar{X} + \frac{1}{n} \Sigma \epsilon_1} = b + \frac{1}{n\bar{X}} \Sigma \delta_1 - \frac{b}{n\bar{X}} \Sigma \epsilon_1$$

Hierin is $\bar{X} = (1/n) \Sigma X_1$; voor $\sigma^2(\beta)$ volgt nu:

$$\sigma^2(\beta) = \left\{ \frac{1}{n\bar{X}} \right\}^2 [\sigma_\delta^2 + n b^2 \sigma_\epsilon^2] = \frac{\sigma_\delta^2 + b^2 \sigma_\epsilon^2}{n\bar{X}^2}$$

Beschouw nu de residuen

$$\Delta y_i = y_i - bx_i = Y_i + \epsilon_i - bX_i - b\delta_i = \epsilon_i - b\delta_i$$

Er geldt:

$$E(\Delta y) = 0$$

$$\sigma^2(\Delta y) = \sigma_{\epsilon}^2 + b^2 \sigma_{\delta}^2$$

Derhalve kan $\sigma_{\epsilon}^2 + b^2 \sigma_{\delta}^2$ geschat worden uit $1/(n-2) \sum \Delta^2 y_i$; indien verder $\bar{x}$ benaderd wordt met $\bar{x}$ ligt de schatting voor $\sigma(\beta)$ vast.

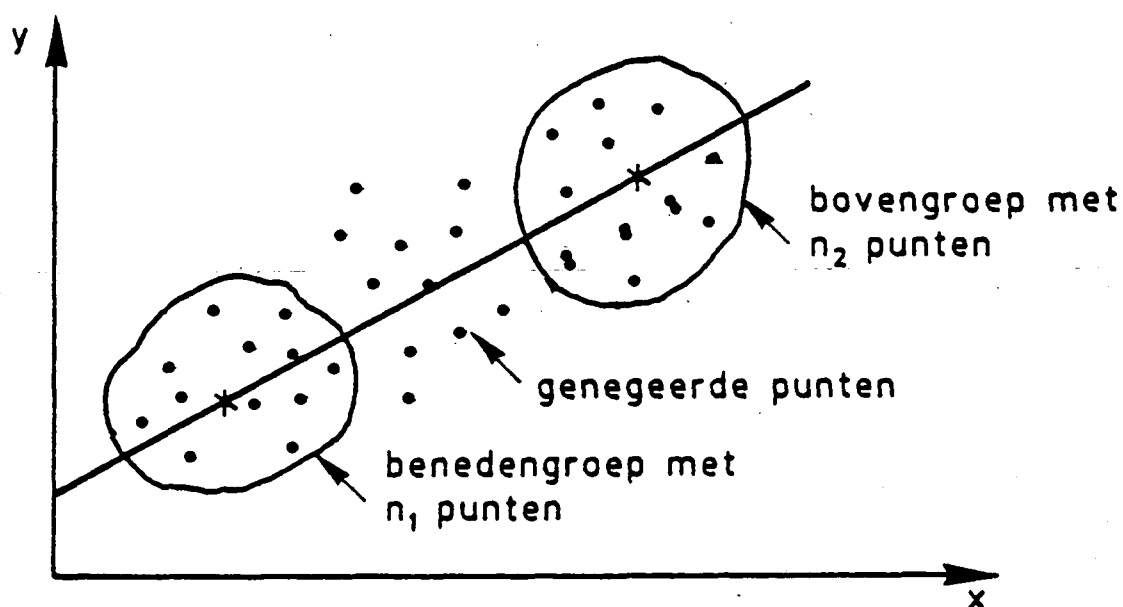
Schatter (III-6) kan natuurlijk ook worden toegepast in de normale regressie-situatie, d.w.z. als er alleen fouten in y-richting zijn. De bias blijkt dan verdwenen te zijn. Vergelijken we de standaardafwijking van (III-6) met die van de kleinste kwadraten-methode dan blijkt de kleinste-kwadratenmethode een kleinere standaardafwijking te hebben. Dit is niet verwonderlijk aangezien de kleinste-kwadratenmethode juist volgt door de variantie van de regressiecoëfficiënten te minimaliseren. We zouden kunnen zeggen dat de ruimere toepasbaarheid van (III-6) betaald moet worden met een geringere mate van efficiency.

Naast deze geringere efficiency heeft (III-6) nog een ander nadeel, namelijk dat de procedure veel gecompliceerder wordt als wordt overgestapt naar het algemene geval waarbij de regressielijn niet door de oorsprong gaat.

We moeten dan zogenaamde "grouping" toepassen (Madansky [6], Richardson, Wu [8]) wat neerkomt op zwaartepuntsbepalingen van twee groepjes punten, waar doorheen dan een recht lijn getrokken kan worden. De schatting van β volgt dan bijvoorbeeld uit:

$$\beta = \frac{\left(\frac{1}{n_1}\right) \sum y_{i1} - \left(\frac{1}{n_2}\right) \sum y_{ij}}{\left(\frac{1}{n_1}\right) \sum x_{i1} - \left(\frac{1}{n_2}\right) \sum x_{ij}} \quad (\text{III-7})$$

In figuur III-1 is deze formule nader toegelicht. Het probleem bij deze methode is dat de meetfouten niet zo groot mogen zijn dat een punt als het ware in de verkeerde groep terecht komt. Men zou namelijk de punten moeten selecteren op hun werkelijke X-Y-waarden maar men weet natuurlijk alleen de x-y-meetresultaten. Hierdoor kan een punt dat eigenlijk bij de bovengroep thuishoort in de ondergroep terecht komen en omgekeerd. In principe is het mogelijk de kansen op dit soort verstoringen af te schatten en de invloed ervan op β na te gaan. Deze analyses worden dan echter gauw erg ingewikkeld. Samenvattend dient te worden geconstateerd dat deze methoden weinig aantrekkelijk zijn voor praktische toepassing.



Figuur III-1 Groeperingsmethode ter bepaling van een regressielijn.