

RISICO-ONDERZOEK VERKEERSDEELNEMERS IN NEDERLAND (ROVIN) III

Eindrapport over het subonderzoek naar de mate van nauwkeurigheid van opgegeven afstanden, behorende bij het deelonderzoek Validiteit van het CBS-onderzoek Verplaatsingsgedrag

R-83-48

F.J. de Bruin & J.M.J. Bos

Leidschendam, 1983

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV

SAMENVATTING

Inzicht in kwaliteit en bruikbaarheid van mobiliteitsgegevens uit het CBS-onderzoek Verplaatsingsgedrag is van belang voor wetenschappelijk verkeersveiligheidsonderzoek. Het CBS was door omstandigheden niet in de gelegenheid dit inzicht te geven. Hoewel het niet direct tot haar taak behoort, is de SWOV de mogelijkheid geboden hiernaar onderzoek te doen. In het onderzoek Verplaatsingsgedrag (OVG) verzamelt het CBS door middel van een rittenboek gegevens over verplaatsingen van personen. Over de interpretatie van aldus verkregen gegevens bestaan vragen vanuit de toepassing bij het volgen van de ontwikkeling van verkeersonveiligheid, het evalueren van maatregelen en het opsporen van aanknopingspunten voor onderzoek en beleid. Deze vragen betreffen in het bijzonder de zuiverheid en nauwkeurigheid van de verkregen afstanden voor de verschillende wijzen van vervoer en problemen die zich bij de enquête in dit verband voordoen. Noch in zijn algemeenheid, noch toegespitst op het OVG is terzake voldoende bekend, vandaar het deelonderzoek Validiteit, waaronder het subonderzoek Nauwkeurigheid opgegeven afstanden.

Tot de gegevens uit het OVG behoren de afstanden die geënquêteerde personen opgeven voor door hen gemaakte verplaatsingen. Literatuur over het schatten van afgelegde afstanden is er weinig. Proefonderzoek van de SWOV in 1976 leert dat op grote verschillen met de werkelijke lengten van gemaakte verplaatsingen moet worden gerekend. Het is nu de vraag of het hierbij om toevallige of systematische verschillen gaat en of eventuele systematische verschillen samengaan met bepaalde kenmerken van de verplaatsingen of van de personen.

In het algemeen zullen in verkeersveiligheidsonderzoek niet de afstanden van afzonderlijke verplaatsingen, maar de totalen voor volgens bepaalde kenmerken gegroepeerde verplaatsingen nodig zijn. Omdat de vraagstelling specifiek de toepassing van OVG-cijfers betreft, moet het onderzoek echter nauw bij het OVG aansluiten. Noch experimenteel onderzoek naar het vermogen van mensen om de lengte van een gemaakte verplaatsing te schatten, noch bijvoorbeeld onderzoek naar de mate waarin zij totalen van afgelegde afstanden misschatten, zijn zodoende overwogen.

Er moet dus worden nagegaan welke omvang in het OVG de toevallige en

eventuele systematische schattingsfouten voor afgelegde afstanden hebben, welke factoren er op een systematische fout van invloed zouden zijn en of het OVG-resultaat voor dergelijke fouten zou zijn te corrigeren.

Nauwe aansluiting bij het OVG kan worden bereikt als per vervoerswijze van de in het OVG opgegeven afzonderlijke verplaatsingsafstanden ook de werkelijke lengten bekend geraken. Het subonderzoek Nauwkeurigheid opgegeven afstanden is er dan ook op gericht deze werkelijke afstanden te achterhalen. Deze afstanden zullen daartoe moeten worden gemeten. Het eenvoudigst gaat dat vanaf een plattegrond of topografische kaart, waarop de gevolgde route is ingetekend.

Een steekproef uit de OVG-respondenten is gevraagd naar de gevolgde routes van de opgegeven verplaatsingen, deze routes zijn ter plaatse op plattegronden en/of topografische kaarten door getrainde enquêteurs ingetekend en vervolgens later nagemeten. Om de respondenten met succes te kunnen benaderen diende het tijdsverloop tussen de OVG-enquête en de enquête ten behoeve van het subonderzoek zo kort mogelijk te zijn. Dit is bereikt door de laatstgenoemde enquête ter afsluiting van de OVG-enquête te laten plaatsvinden.

De OVG- huishoudingensteekproef omvat van de meewerkende huishoudingen alle personen vanaf 12 jaar en deze vullen over twee (of drie) dagen gegevens over hun verplaatsingen op rittenboeken in. Maandelijks worden ca 1200 huishoudingen, gespreid over heel Nederland, geënquêteerd. De steekproef voor het subonderzoek Nauwkeurigheid opgegeven afstanden was gelijk aan de huishoudingensteekproef van het OVG, echter met in principe willekeurig één persoon per huishouding, geënquêteerd over de routes van de op de laatste OVG-enquêtedag opgegeven gemaakte verplaatsingen, waarbij zondagse verplaatsingen vanwege de veldwerkprocedure buiten het onderzoek bleven. Meerdere personen per huishouding zou te veel druk opleveren. Om praktische redenen moest daarbij worden gekozen voor de maanden novemberdecember 1978, terwijl om organisatorische redenen het veldwerk, anders dan standaard, niet door CBS- maar door medewerkers van de N.V. v/h Nederlandse Stichting voor Statistiek moest worden verricht. Vanwege het intekenen van routes is met speciaal getrainde enquêteurs gewerkt.

Niet van alle gemeenten waren goede plattegronden beschikbaar, met name niet van plattelandsgemeenten. Daarom zijn waar mogelijk in vergelijk-

bare, vervangende gemeenten naar evenredigheid niet één, maar twee personen per huishouding gevraagd.

De nametingen van de ingetekende routes die met een digitaliseertafel van de Meetkundige Dienst van de Rijkswaterstaat zijn verricht, hebben door de vele soorten kaartmateriaal overstap- en schaalproblemen ondervonden, terwijl ook het nauwkeurig aangeven van vertrek- en aankomstpunten en het precies volgen van de routes hoge eisen stelden. Zowel korte als lange verplaatsingen gaven extra moeilijkheden, bijvoorbeeld door het ontbreken van aansluitend kaartmateriaal. De koppeling van de nameetresultaten aan de bijbehorende oorspronkelijke enquêtegegevens heeft daarna opnieuw problemen en enige uitval opgeleverd.

De opgegeven en nagemeten verplaatsingsafstanden zijn vervolgens per vervoerswijze lopen, fiets en personenauto in twee fasen geanalyseerd, voor andere vervoerswijzen waren er te weinig metingen.

In de eerste fase is gezocht naar kenmerken van de verplaatsingen die van invloed zouden zijn op de mate waarin er eventueel van een systematische misschatting van de afgelegde afstand sprake is. Dat is gedaan met behulp van variantie-analyse.

Daarbij dient afhankelijkheid tussen de waarnemingen zoveel mogelijk te worden vermeden. Zodoende zijn eerst de identieke verplaatsingen verwijderd, overigens zonder hiermee onafhankelijkheid geheel te hebben bereikt. Bovendien moet de toevallige fout overal in de nu gereduceerde steekproef ongeveer dezelfde kansverdeling hebben. De grootte van de toevallige fout neemt echter toe met de lengte van de opgegeven afstand, zodat een transformatie van de opgegeven en nagemeten afstanden nodig was die de variantie voldoende zou stabiliseren.

Tenslotte was een inperking van het aantal te analyseren kenmerken noodzakelijk. Deze is gerealiseerd met behulp van CANALS-analyses.

Uit zowel de CANALS- als de variantie-analyses bleek dat de toevallige fout in het misschatten van afgelegde afstanden groot is.

Op 1%-niveau is bij elk van de drie vervoerswijzen niettemin de opgegeven afstand significant van invloed op de grootte van de systematische schattingsfout. In het algemeen zijn kortere opgegeven afstanden onder- en langere overschattingen.

Bij de vervoerswijze fiets is er nog een significant verschil tussen weekeinde en werkdagen, hoewel er voor het weekeinde slechts weinig, mogelijk niet geheel aselechte waarnemingen zijn. De opgegeven lengten van weekeindeverplaatsingen zouden gemiddeld verder boven de werkelijk afgelegde afstanden liggen dan die van werkdagverplaatsingen. Bij personenauto's is het verschil tussen de regio's juist niet significant.

In de tweede fase van de analyse zijn de grootten van de systematische afwijkingen tussen opgegeven en nagemeten afstanden voor de eerder significant gebleken verplaatsingskenmerken over de oorspronkelijke steekproef berekend en de bijbehorende betrouwbaarheden geschat. Met de opgegeven lengten van loopverplaatsingen bleken de werkelijke lengten gemiddeld ca 30% te hoog te zijn geschat, bij de fiets gaat het om een gemiddelde overschatting van ca 25% en bij de personenauto om gemiddeld ca 15%. Deze totaal-gemiddelden zijn hier genoemd om een globale indruk te geven van de effecten. Zij gelden voor de onderzochte steekproef en zijn dus, gezien de invloed van de opgegeven afstand en in het bijzonder het grote gewicht daarbij van langere opgegeven afstanden voor de vervoerswijzen fiets en personenauto, afhankelijk van de verdeling van de waarnemingen hierover.

Op basis van de geconstateerde, significante afwijkingen blijkt in principe correctie van OVG-cijfers mogelijk, mits reeds gemelde en andere problemen omtrent steekproefselectie, onderzoekmethodologie en de invloed van de herweging die het CBS bij de verwerking van de enquêtegegevens toepast, in ogenschouw worden genomen.

Een belangrijke conclusie is ook dat de OVG-cijfers bij een uitsplitsing naar meer verplaatsingskenmerken, als gevolg van de grote toevallige schattingsfouten pas een nauwkeurig beeld van de afgelegde verplaatsingsafstanden kunnen geven, als zij zijn gebaseerd op relatief grote aantallen waarnemingen.

ABSTRACT: THE ACCURACY OF TRIP-DISTANCE ESTIMATIONS

The Netherlands National Traffic Survey (IRRD No. 606350) uses the diary method for recording the distances of trips made. No systematic research into the accuracy in which respondents can estimate the distances of the trips they make was found.

A subsample of the National Traffic Survey described the routes taken of their trips during one day, and these routes (ca. 3,500) were drawn on maps by specially-trained interviewers. The map distances were measured by means of a Digitalizer, and the measured distance of each trip was compared with the distance as estimated by the respondent and entered in the trip diary. The sample was only sufficiently large to permit analysis of trips by car, bicycle, and on foot. Trips by motorcycle, moped, bus, and tram had to be left out.

The random errors in estimation were much larger than the systematic errors. The systematic errors were mainly that of overestimation, dependant on reported trip length. Short trips (estimated less than 1 km) were generally underestimations but practically all others were overestimations.

For car trips there are also regional differences and for bicycle trips there is a difference between the weekend and other days. Only the distance is important for pedestrian trips. Influence of characteristics such as age, sex, and education were not found.

The average correction needed for overestimation is approx. 15% for car trips, ca. 25% for bicycle trips, and ca. 30% for pedestrian trips.

Weighting the results of the National Travel Survey is permissible under certain circumstances, but a large sample is more important.

INHOUD

Voorwoord

1. Inleiding

2. Beschrijving van het probleem

2.1. Algemeen

2.2. Registratie van afgelegde afstand en nauwkeurigheid

2.3. Omvang van het probleem

3. Doel van het subonderzoek Nauwkeurigheid opgegeven afstanden

4. Methode van onderzoek

4.1. Methoden voor het vaststellen van de werkelijk afgelegde afstand

4.1.1. Opnemen van de stand van de kilometerteller

4.1.2. Het bepalen van de meest logische route en bijbehorende afstand via een computergestuurd netwerk

4.1.3. Het vastleggen van afgelegde routes op plattegronden en het nameten van ingetekende routes

4.2. Keuze van de meetmethode

4.3. Opzet van het subonderzoek Nauwkeurigheid opgegeven afstanden

5. Uitvoering van het onderzoek

5.1. Werkprocedure CBS-onderzoek Verplaatsingsgedrag (OVG)

5.2. Uitvoering veldwerk subonderzoek Nauwkeurigheid opgegeven afstanden

5.3. De mate van nauwkeurigheid van de meetmethode in de praktijk

5.4. Meten van op plattegronden ingetekende routes

5.5. De analyse van de verzamelde afstandgegevens

5.6. De uitvoering van de analyses

6. Resultaten van het onderzoek

6.1. De steekproefomvang

6.2. De representativiteit van de substeekproef

6.3. De inspectie van de gegevens

6.4. De resultaten van de CANALS-analyses

6.5. De resultaten van de (co)variantie-analyses

- 6.6. Verschillen tussen analyse- en onderzoekbestand
- 6.7. Het schatten van de grootte van significant bevonden systematische effecten
- 6.8. De toevallige misschatting van alle beschikbare ritafstanden
- 6.9. Samenvatting van de resultaten

7. Conclusies en aanbevelingen

- 7.1. Conclusies
- 7.2. Aanbevelingen

Literatuur

Tabellen 1 t/m 13

Bijlage 1. Verplaatsingsformulieren SWOV 1976, idem OVG 1978 en idem OVG 1979.

Bijlage 2. Nauwkeurigheid opgegeven verplaatsingsafstanden; Technisch verslag van het subonderzoek behorende bij het deelonderzoek Validiteit van het CBS-onderzoek Verplaatsingsgedrag. J.M.J. Bos. SWOV, 1983.

VOORWOORD

Het Centraal Bureau voor de Statistiek (CBS) verzamelt sinds 1 januari 1978 verkeers- en vervoersprestatiegegevens door huishoudens in Nederland mondeling te enquêteren over o.a. hun persoonlijke omstandigheden en hun verplaatsingen in het verkeer. Daartoe dient men o.a. in een zgn. verplaatsingsboekje gedurende twee of drie aaneengesloten dagen de gemaakte verplaatsingen in te vullen. Het enquêteren gebeurt in het kader van het onderzoek Verplaatsingsgedrag (OVG), waarbij maandelijks een aselecte steekproef van circa 1100 tot 1300 huishoudens wordt benaderd.

Met name de Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV heeft uitgesproken dat alvorens de door het CBS verzamelde gegevens te kunnen gebruiken eerst de gebruikswaarde van de gegevens vast moest staan. Daarvoor was aanvullend onderzoek nodig. Dit werd door vele instanties onderschreven.

Het CBS heeft ambtshalve de taak om waarde en kwaliteit van het onderzoek Verplaatsingsgedrag (OVG) te onderzoeken en te bewaken. Derhalve diende het CBS het aanvullende onderzoek uit te voeren. Echter, door gebrek aan arbeidskrachten en het ontbreken van financiële middelen bij het CBS, moest gezocht worden naar een andere financieringsbron.

De Directie Verkeersveiligheid (DVV) van het Ministerie van Verkeer en Waterstaat is bereid gevonden het aanvullende onderzoek financieel mogelijk te maken. De SWOV heeft niet tot taak onderzoek te doen naar verkeers- en vervoersprestatiegegevens en de waarde en kwaliteit daarvan. Toch meende de SWOV op grond van het grote belang dat zij hecht aan het onderzoek Verplaatsingsgedrag (OVG) de opdracht tot het uitvoeren van aanvullend onderzoek op zich te moeten nemen. Een andere reden was dat de SWOV in de periode 1975 t/m 1978 een aantal zgn. pilot studies had laten verrichten omtrent verschillende kwaliteitsaspecten van onderzoek naar verkeers- en vervoersprestatiegegevens. De daarmee verworven kennis zou toegepast kunnen worden in de uitvoering van het aanvullende onderzoek. In september 1978 heeft de DVV aan de SWOV de opdracht gegeven tot het uitvoeren van een aanvullend onderzoek op het onderzoek Verplaatsingsgedrag (OVG) van het CBS.

Het aanvullend onderzoek bestaat uit twee deelonderzoeken: het deelonderzoek Non-response en het deelonderzoek Validiteit. Ter begeleiding van

het aanvullend onderzoek heeft de DVV de zgn. Begeleidingsgroep Non-response- en Validiteitsonderzoek (BNRV) ingesteld. De Begeleidingsgroep is in december 1978 in functie getreden. Haar taak bestond uit het goedkeuren van de door de SWOV in te brengen voorstellen ten aanzien van de onderzoekopzet, de uitvoering en de verslaglegging daarvan. De SWOV had de verplichting één keer per maand een voortgangsrapport ter goedkeuring voor te leggen. Zodoende kon de BNRV toezicht uitoefenen op de uitwerking van de door de DVV geformuleerde onderzoekopdracht.

De samenstelling van de BNRV was:

Voorzitter: de heer B.J. Quist - DVV

Secretaris: drs. W.H.R. Mante - DVV

(m.i.v. mei 1979 tevens voorzitter)

m.i.v. mei 1980 de heer J. Staal - DVV

Leden : dr. W. Horn - Studiefdeling, V & W

(m.i.v. mei 1980 tevens waarnemend voorzitter)

ir. J.A.M. v.d. Kooij,

m.i.v. januari 1980 opgevolgd door

drs. J. Termorshuizen - DVK, V & W

drs. F.W.M. Hendrikx - CBS, Afd. Verkeers- en Vervoersstatistieken

drs. M.J.G. van Widdershoven - CBS, Afd. Persoons- en Gezinsenquêtees

de heer J. Houben - CBS, Afd. Verkeers- en Vervoersstatistieken

de heer S. Harris, M.A. - SWOV

de heer F.J. de Bruin - SWOV

Het rapport "Risico-onderzoek verkeersdeelnemers in Nederland (ROVIN) I" (SWOV R-82-30) is het eindrapport over het deelonderzoek naar de non-response in het CBS-onderzoek Verplaatsingsgedrag.

Het deelonderzoek Validiteit bestaat uit twee afzonderlijke onderzoeken. Het ene onderzoek heeft betrekking op het opsporen van zogenaamde vergeten verplaatsingen. Hieronder worden verstaan verplaatsingen die in eerste instantie door de respondenten niet zijn ingevuld in een verplaatsingsboekje. Over dit onderzoek handelt het rapport "Risico-onderzoek verkeersdeelnemers in Nederland (ROVIN) II" (SWOV R-82-32).

Het andere onderzoek heeft betrekking op het vaststellen van de mate van nauwkeurigheid waarmee respondenten de door hen opgegeven verplaatsingsafstanden schatten. Het verslag van dit onderzoek is nu in dit rapport vastgelegd.

Wij danken in het bijzonder ir. C.W. Corbet, hoofd van de Meetkundige Dienst van Rijkswaterstaat in Delft, voor zijn inspanningen die het mogelijk hebben gemaakt dat de SWOV kon beschikken over meetapparatuur ten behoeve van het opmeten van ingetekende routes.

Dit rapport is samengesteld door F.J. de Bruin, Afdeling Onderzoekondersteuning in samenwerking met J.M.J. Bos, Afdeling Methoden en Technieken. De eindverantwoordelijkheid voor de inhoud van het rapport berust bij de SWOV.

Leidschendam, december 1983

Prof. ir. E. Asmussen, directeur

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV

1. INLEIDING

In het rapport "Risico-onderzoek verkeersdeelnemers in Nederland (ROVIN) I" (SWOV, 1982) is de noodzaak tot het kunnen beschikken over gedetailleerde landelijke verkeers- en vervoersprestatiegegevens beschreven. Daarbij is vermeld dat het CBS-onderzoek Verplaatsingsgedrag in een aantal belangrijke opzichten voldoet aan de behoeften voor het verkeersveiligheidsonderzoek van de SWOV. Het betreft hier de wijze van verzamelen van gegevens en de mate van uitvoerigheid. De wijze van verzamelen heeft betrekking op aspecten als: aselecte steekproef, landelijke enquête waarbij vanaf 1 januari 1978 in principe over elke dag van het jaar verkeers- en vervoersprestatiegegevens zijn (worden) geregistreerd. En de mate van uitvoerigheid heeft betrekking op het feit dat per afzonderlijk afgelegde verplaatsing een aantal voor de SWOV essentiële gegevens worden geregistreerd als: wijze van verkeersdeelname, tijd(stip) van verplaatsing, afgelegde afstand en (indirect) het gebied van de verplaatsing. Daarnaast worden gegevens geregistreerd omtrent de persoon die de desbetreffende verplaatsing maakt, waaronder leeftijd, geslacht en woonplaats.

Alvorens de door het CBS verzamelde gegevens te kunnen gebruiken in het wetenschappelijk verkeersveiligheidsonderzoek is het noodzakelijk te weten in welke mate de gegevens representatief zijn. Daarvoor is kennis nodig over de populatie waarop de steekproef gebaseerd is. In het CBS-onderzoek worden geen verkeers- en vervoersprestatiegegevens verzameld over personen jonger dan 12 jaar, over tehuusbewoners en zgn. kleine groepen verkeersdeelnemers als bestuurders/passagiers van vrachtvoertuigen, mede vanwege de omvang van de steekproef. Daarnaast is kennis nodig over de non-response bij steekproeven en is het noodzakelijk te weten of de gegevens volledig zijn en welke kwaliteit de verzamelde gegevens hebben. De kwaliteit heeft betrekking op de exacte en juiste weergave van de gevraagde gegevens.

Met name over de kwaliteit van via enquêtes verzamelde verkeers- en vervoersprestatiegegevens is weinig bekend.

Er zijn echter aanwijzingen in de literatuur (zie o.a. Hoofdstuk 2) dat de gegevens, in een situatie waarbij respondenten moeten aangeven welke activiteiten zij hebben verricht, niet exact en juist worden ingevuld.

Dit leidt tot onnauwkeurigheden vooral bij het schatten van gegevens als tijdstip van vertrek en aankomst bij verplaatsingen en de afstanden die men aflegt. Van de twee genoemde kenmerken is de afgelegde afstand het meest relevant voor het wetenschappelijk verkeersveiligheidsonderzoek, aangezien dit gegeven gebruikt wordt om verkeersongevallengegevens aan te relateren.

2. BESCHRIJVING VAN HET PROBLEEM

2.1. Algemeen

Behoudens een door de SWOV in 1976 uitgevoerd onderzoek naar de relatie tussen opgegeven verplaatsingsafstanden en de in werkelijkheid afgelegde afstanden is er geen onderzoek op dit terrein bij de SWOV bekend. Wel zijn er in het verleden onderzoeken gedaan naar het vermogen van personen tot het schatten van afstanden. De resultaten van dergelijk (buitenlands) onderzoek van Harte (1975), House & Waller (1971) en White (1974) zijn een bevestiging van het vermoeden dat het schatten van afstanden voor mensen een niet onaanzienlijk probleem is.

2.2. Registratie van afgelegde afstand en nauwkeurigheid

In 1976 heeft de SWOV een proefonderzoek uitgevoerd waarbij de door de respondenten opgegeven afgelegde verplaatsingsafstanden werden vergeleken met de berekende werkelijke afstanden. Aan de respondenten werd gevraagd gedurende één dag hun eventuele verplaatsingen te noteren op een verplaatsingsformulier (zie Bijlage 1). Het verplaatsingsformulier en de bijbehorende vragenlijst werden twee dagen na het introductiegesprek opgehaald. Tijdens het ophaalbezoek werden de ingevulde formulieren nagekeken en informatie verzameld over de route die men had afgelegd per ingevulde verplaatsing. Deze informatie werd op stadsplattegronden en wegenkaarten vastgelegd. De ingetekende routes werden vervolgens opgemeten om zo de berekende werkelijke afstanden te verkrijgen.

De belangrijkste resultaten uit het proefonderzoek zijn:

1. Er is sprake van een grote toevallige fout bij het schatten van afgelegde afstanden.
2. Er is een redelijke correlatie, samenhangend met de actieradius, tussen de werkelijk afgelegde verplaatsingsafstanden en de opgegeven verplaatsingsafstanden bij de drie onderzochte vervoerswijzen lopen (0,69), fiets (0,89) en personenauto (0,97).
3. Korte opgegeven afstanden liggen gemiddeld ver onder de werkelijk afgelegde afstanden. Opgegeven afstanden, afgelegd te voet tot 500 meter, moeten met gemiddeld circa 43% worden opgehoogd, met de fiets tot 1000 meter met gemiddeld circa 25%, met de personenauto tot 5000 meter met gemiddeld circa 3%.

4. Naarmate de afstand als groter is opgegeven is er meer sprake van een overschatting. Opgegeven afstanden, afgelegd te voet vanaf 1000 meter, moeten met gemiddeld circa 24% worden verlaagd, met de fiets vanaf 5000 meter met gemiddeld circa 12%, met de personenauto vanaf 30.000 meter met gemiddeld circa 13%.

5. Gemiddeld worden verplaatsingsafstanden overschat. De gemiddelde overschattingen voor de verschillende vervoerswijzen waren in de onderzoeksteekproef als volgt: lopen circa 6%, fiets circa 8%, bromfiets circa 15%, personenauto circa 10% en openbaar vervoer circa 2%. Door de grote toevallige fout was met name het effect voor lopen niet significant, dat voor openbaar vervoer is niet getoetst.

6. Op de mate van het misschatten heeft naast de opgegeven afstand van de onderzochte variabelen ook de gemeentegrootte nog een belangrijke invloed.

2.3. Omvang van het probleem

Het enige onderzoek waaruit een indruk van de omvang van het probleem kan worden verkregen is weer het eerder genoemde proefonderzoek van de SWOV, waarin onder andere een selectie naar gemeentegrootte, leeftijd en geslacht, dag van de week en zo mogelijk vervoermiddelenbezit was toegepast. Dit proefonderzoek was bovendien beperkt in ruimte (beperkt geografisch gebied) en tijd (laatste twee weken november 1976), terwijl gebruik werd gemaakt van enquêteformulieren die sterk afwijken van die van het huidige OVG (zie Bijlage 1). Hierdoor is een vergelijking met het OVG niet goed mogelijk. Derhalve resteert de indruk dat de omvang van het probleem groot kan zijn.

3. DOEL VAN HET SUBONDERZOEK NAUWKEURIGHEID OPGEGEVEN AFSTANDEN

Een van de kwaliteitseisen waaraan het OVG dient te voldoen heeft betrekking op de nauwkeurigheid van de verzamelde gegevens en met name de afgelegde verplaatsingsafstanden. Behoudens de resultaten van het door de SWOV in 1976 uitgevoerde proefonderzoek is weinig bekend over de omvang van de onnauwkeurigheid van opgegeven verplaatsingsafstanden. Bij het verkeers- en vervoersprestatieonderzoek van het CBS bestaat de mogelijkheid dat de opgegeven verplaatsingsafstanden bepaalde systematische vertekeningen vertonen. Aanwijzingen daartoe zijn beschreven in par. 2.2. De verkeers- en vervoersprestatie zijn uitermate relevant door hun relatie met het vóórkomen van verkeersongevallen. Bij het beschrijven van de omvang en de ontwikkeling van de verkeersonveiligheid wordt gebruik gemaakt van onderverdelingen als vervoerswijze, leeftijd en geslacht van verkeersslachtoffers, tijdstip van de dag, dag van de week e.d. Het was tot voor kort niet mogelijk om de aldus verkregen meer gedetailleerde beschrijving van de verkeersonveiligheid te relateren aan en te corrigeren voor de verkeers- en vervoersprestatie. Het OVG heeft hierin verandering gebracht. In dit verband is het noodzakelijk inzicht te hebben in de mate van nauwkeurigheid waarmee OVG-gegevens worden geregistreerd. Dit geldt in het bijzonder voor de afgelegde afstanden. Uitgangspunt van het subonderzoek zijn de afzonderlijke verplaatsingsafstanden, zoals deze in het OVG worden verzameld. Door analyse van de mate waarin afzonderlijke verplaatsingsafstanden worden misgeschat kunnen de toevallige en mogelijk systematische afwijkingen worden onderscheiden. Tevens dient de vraag beantwoord te worden of er alsdan gecorrigeerd kan worden voor de systematische afwijkingen.

4. METHODE VAN ONDERZOEK

4.1. Methoden voor het vaststellen van de werkelijk afgelegde afstand

Er zijn verschillende methoden denkbaar om de werkelijk afgelegde afstand vast te stellen. De te hanteren methode is afhankelijk van de methode die gekozen wordt om de verkeers- en vervoersprestatie vast te stellen en van de wijze van vervoer waarvan men de gegevens wil verzamelen.

Wanneer gevraagd wordt naar de met een personenauto over een periode van bijvoorbeeld één jaar afgelegde afstand dan is de meest voor de hand liggende meetmethode het opnemen van de stand van de kilometerteller aan het begin en het eind van de waar te nemen periode. Deze methode is alleen mogelijk bij voertuigen waarbij een kilometerteller aanwezig is. In het geval dat gevraagd wordt naar de afgelegde afstand van een verplaatsing of op een dag dan zijn er andere methoden mogelijk. Aangezien in het onderzoek Verplaatsingsgedrag gewerkt wordt met de zgn. dagboekmethode waarbij per afgelegde verplaatsing o.a. de afgelegde afstand wordt geregistreerd, wordt uitgegaan van de laatstgenoemde situatie. De verschillende meetmethoden zijn:

1. Opnemen van de stand van de kilometerteller.
2. Het bepalen volgens een computergestuurd netwerk van wegen van de meest logische route en bijbehorende afstand.
3. Het vastleggen van afgelegde routes op aanwijzing van de respondent op plattegronden en het nameten van de routes.

Buiten beschouwing blijft om praktische en financiële redenen een methode waarbij personen gevolgd zouden worden tijdens hun verplaatsingen of waarbij zij hun verplaatsingen samen met een enquêteur zouden herhalen.

4.1.1. Opnemen van de stand van de kilometerteller

Deze meetmethode is alleen te gebruiken bij verplaatsingen volgens vervoerswijzen waarbij een goede kilometerteller kan worden toegepast. Slechts een gering percentage fietsen is met een kilometerteller uitgerust. Dit is te verhelpen door kilometertellers ter beschikking te stellen. Voor verplaatsingen te voet is een telinstrument beschikbaar in de vorm van een pedometer. Het nadeel kan zijn dat het uitreiken van een

kilometerteller een a-typisch verkeersgedrag oproept. Een beschikbaarstelling en het feitelijk werken ermee levert bovendien een aanzienlijk praktisch probleem op zoals blijkt uit het aandeel van de vervoerswijzen fietsen en lopen in het aantal afgelegde verplaatsingen dat resp. circa 11% en circa 45% van het totaal bedraagt (OVG, 1978).

Maar met de introductie van het gebruik van kilometertellers in het onderzoek ontstaan tegelijk vragen over de nauwkeurigheden van dergelijke instrumenten. Uit de in autotijdschriften gepubliceerde zgn. road tests is bekend dat de afwijkingen van kilometertellers in veel gevallen variëren tussen 0 en 2%. De kilometerteller geeft in de meeste gevallen meer aan dan de werkelijke afstand. Het is niet bekend wat de variatie in de praktijk als gevolg van bandspanning- en slijtage verschillen zal zijn. Verder geven de meeste bestaande kilometertellers alleen hele kilometers aan.

Ondanks dat het een éénduidige meetmethode lijkt, is gezien de hiervoor aangeduide problemen de conclusie dat de kilometerteller voor het beoogde niet geschikt is.

4.1.2. Het bepalen van de meest logische route en bijbehorende afstand via een computergestuurd netwerk

In de verkeers- en vervoerswetenschappen wordt veel gebruik gemaakt van modellen. Hiervoor worden gegevens over het wegennet in de vorm van x- en y-coördinaten in een bestand opgenomen. Gegevens over het verplaatsingsgedrag, verkregen door middel van enquêtes, worden ingevoerd via het vastleggen van de vertrek- en aankomstadressen in eerdergenoemde x- en y-coördinaten. Met behulp van een computerprogramma kan de meest logische route worden bepaald en de bijbehorende afstand berekend. De gehanteerde modellen worden regelmatig geijkt aan de hand van gegevens afkomstig uit enquêtes waarbij routegegevens worden gevraagd.

Het is vooralsnog niet voldoende duidelijk of met dergelijke modellen de afstanden met een grote mate van betrouwbaarheid nauwkeurig berekend kunnen worden. Dit laatste zou in ieder geval niet kunnen plaatsvinden voor routes die in de directe woonomgeving worden afgelegd, aangezien de bekende netwerken voornamelijk de doorgaande verkeerswegen omvatten. Met deze methode kan niet worden voldaan aan de belangrijke eis dat om tot onderling vergelijkbare resultaten te kunnen komen, alle afstanden met

één meetmethode dienen te worden gemeten. Bovendien roept de keuze voor een dergelijke indirecte meetmethode gezien het karakter van het beoogde onderzoek zelf opnieuw calibratieproblemen op.

4.1.3. Het vastleggen van afgelegde routes op plattegronden en het nameten van ingetekende routes

Aan deze methode zijn verschillende randvoorwaarden verbonden. De eerste randvoorwaarde is dat de respondent in staat is de afgelegde route juist weer te geven. Als tweede randvoorwaarde geldt dat de medewerkers in staat moeten zijn kaarten te lezen en routebeschrijvingen op kaarten in te tekenen. De derde randvoorwaarde is dat men kan beschikken over voor het doel geschikte kaarten. Daarnaast is de vierde en laatste randvoorwaarde dat een methode beschikbaar is waarmee met een grote mate van nauwkeurigheid de vastgelegde route kan worden opgemeten.

Het vragen aan een persoon naar de routebeschrijving van een bepaalde verplaatsing kan aanleiding zijn voor het ontstaan van fouten. Het kan zijn dat de ondervraagde persoon bijvoorbeeld de door hem/haar gevolgde route niet meer precies kan beschrijven. In het algemeen is er dus geen zekerheid of de gegeven routebeschrijving in overeenstemming is met de werkelijkheid.

In de voorbereidingsfase van het proefonderzoek van de SWOV in 1976 is gebleken dat niet iedere interviewer in staat is vlot kaarten te lezen en nauwkeurig routes in te tekenen. Derhalve dienden interviewers geselecteerd te worden op de hiervoor genoemde vaardigheden en getraind te worden met als afsluiting een proeve van bekwaamheid. Aan dit aspect moest veel aandacht worden besteed in de voorbereiding van het subonderzoek. Ten aanzien van het beschikbaar zijn van geschikt kaartmateriaal moet eerst worden bepaald wat het gebruiksdoel is. In dit geval is dat het op kaarten en plattegronden kunnen intekenen van een route. De routebeschrijvingen moeten geleverd worden door de personen die deel uitmaken van de substeekproef van het Validiteitsonderzoek. Vragen met betrekking tot geschikt kaartmateriaal zijn: staan alle wegen op de plattegrond, staat op de plattegrond het gehele bebouwde gebied van een gemeente, heeft de plattegrond een formaat waarmee te werken is, is de plattegrond op één schaal getekend etc? Het grootste probleem doet zich voor bij de kleine plattelandsgemeenten. In veel gevallen is geen plattegrond be-

schikbaar en als er een plattegrond beschikbaar is dan voldoet deze veelal niet aan de hiervoor genoemde criteria. Dit heeft tot gevolg dat er vooraf een selectie moet plaatsvinden van gemeenten waarbinnen een steekproef getrokken moet worden van adressen of personen ten behoeve van het onderzoek naar de mate van nauwkeurigheid van opgegeven afstanden. Een selectie zoals hiervoor aangegeven, hoeft nauwelijks gevolgen te hebben voor de landelijke of regionale representativiteit van de steekproef. Een ondervertegenwoordiging van kleine gemeenten kan door ophoging van resultaten worden rechtgetrokken mits die op te hogen resultaten representatief mogen worden verondersteld ook voor de weggevallen kleine gemeenten. Daarbij heeft elke onderzoekopzet zijn eigen implicaties voor de betrouwbaarheid van de onderzoekuitkomsten.

De vierde randvoorwaarde heeft betrekking op de wijze waarop op plattegronden ingetekende routes nauwkeurig opgemeten kunnen worden. Bij het opmeten kunnen drie verschillende methoden worden toegepast.

De eerste methode houdt in dat met een curvimeter de ingetekende route wordt opgemeten. Deze methode leidt in de praktijk tot grote schommelingen in de uitkomsten. Afwijkingen tussen twee identieke metingen van 30% tot 40% zijn geen uitzondering. De afwijkingen zijn het gevolg van de onmogelijkheid om er nauwkeurig en geconcentreerd mee te werken bij het opmeten. Daarnaast zijn er twijfels inzake de kwaliteit van de in de handel verkrijgbare curvimeters.

Bij de tweede methode worden de ingetekende routes opgemeten met behulp van een digitaliseertafel met een zgn. cursor. Een cursor is een apparaat voorzien van een stuk glas met daarin twee kruisdraden voor het bepalen van coördinaten. Verder heeft het apparaat enkele bedieningsknoppen. Het geheel staat in verbinding met meetapparatuur waaronder een tafelcomputer en een daaraan gekoppelde digitaliseertafel. De digitaliseertafel bestaat uit een tekentafel die voorzien is van een fijnmazig netwerk van draden. Met de apparatuur kan heel nauwkeurig worden gewerkt. In de tafelcomputer is een programma opgeslagen voor het opmeten van routes. Het opmeten zelf vindt plaats met de cursor die met de hand over de ingetekende routes wordt bewogen. Het beginpunt van een route wordt in x- en y-coördinaten vastgelegd. Vervolgens wordt de cursor over de ingetekende route bewogen waarna ook het eindpunt in x- en y-coördinaten wordt vastgelegd. Door vervolgens de schaal van de desbetreffende plattegrond in de tafelcomputer in te voeren verkrijgt men de afgelegde verplaatsingsafstand. Ook bij

deze methode blijft het mogelijk dat er fouten worden gemaakt, zo kan de cursor onzorgvuldig over de ingetekende route worden bewogen en een verkeerde schaal in de tafelcomputer worden ingevoerd.

Uit proefnemingen in 1976 door de SWOV uitgevoerd is gebleken dat het gebruik van een digitaliseertafel het aantal grote afwijkingen tussen twee meetresultaten sterk reduceert ten opzichte van het gebruik van een curvimeter. Het betreft hier meetresultaten die betrekking hebben op één ingetekende route. Bij het gebruik van een curvimeter heeft slechts 20% van de meetresultaten een onderlinge afwijking van minder dan 2%. Bij het gebruik van een digitaliseertafel heeft ruim 50% van de meetresultaten een onderlinge afwijking van minder dan 2%. Derhalve moet aan het resultaat van de proefneming de conclusie worden verbonden dat het gebruik van een digitaliseertafel te prefereren is boven het gebruik van een curvimeter.

De derde methode houdt in dat gewerkt wordt met een optisch leesapparaat dat uitgerust is met een omvangrijk computergeheugen. De zgn. optische lezer maakt gebruik van speciale inleesformulieren. Elke op een plattegrond en/of kaart ingetekende route dient op een afzonderlijk inleesformulier te worden overgebracht. Dit betekent extra werk en een nieuwe mogelijkheid tot onnauwkeurigheid. Op elk inleesformulier wordt tevens de schaalgrootte vermeld. Met behulp van een computerprogramma kunnen de routes op de inleesformulieren worden ingelezen en met behulp van de vermelde schaalgrootte omgezet worden in nagemeten afstanden. In tegenstelling tot de hiervoor beschreven methode vindt het opmeten van de route machinaal plaats. Het voordeel daarvan is dat éénduidig wordt opgemeten en dat meetfouten niet of nauwelijks optreden. De snelheid waarmee nagemeten kan worden is beduidend hoger dan bij de twee reeds besproken methoden. Echter de beschikbaarheid van de apparatuur is een groot probleem.

4.2. Keuze van de meetmethode

Bij de keuze van de meetmethode dient naast het geschikt zijn van een meetmethode ook de beschikbaarheid van de noodzakelijke apparatuur en de kosten van het toepassen een belangrijke rol te spelen. Dit geldt met name voor de keuze tussen de digitaliseertafel en de zgn. optische lezer. Het gebruik van een optische lezer vereist ruime ervaring en een uitge-

breide opleiding. De kosten per gemeten route zijn hoog. Ten aanzien van de beschikbaarheid van de apparatuur kan vermeld worden dat in Nederland in 1979 twee apparaten waren geïnstalleerd en dat de kans om een apparaat te huren gering was. Hiermee verviel deze mogelijkheid. De mogelijkheden om een digitaliseertafel te huren waren redelijk groot. Terwijl de kosten per gemeten route lager zijn dan bij een optische lezer. Derhalve is gekozen voor een digitaliseertafel als meetinstrument.

4.3. Opzet van het subonderzoek Nauwkeurigheid opgegeven afstanden

In het onderzoek Verplaatsingsgedrag van het CBS wordt aan ieder persoon van 12 jaar en ouder die op de aselechte steekproef van woonadressen woont gevraagd een dagboek bij te houden. De desbetreffende personen wordt verzocht gedurende een periode van twee dagen (of drie dagen als het introductiegesprek op een vrijdag valt) direct volgend op het introductiegesprek de door hen afgelegde verplaatsingen bij te houden.

Per afgelegde verplaatsing wordt een aantal gegevens gevraagd als reis-motief, vervoerswijze, tijdstip van vertrek en aankomst, aantal meereizende personen (niet bij verplaatsingen met openbaar vervoer) en de afgelegde afstand op honderd meter nauwkeurig. Op de dag na de laatste invuldag worden de ingevulde formulieren opgehaald en ter plaatse gecontroleerd op volledigheid en juiste invulling. Het enige moment waarop informatie kan worden verzameld inzake de door de desbetreffende persoon afgelegde route is het ophaalgesprek.

Het verzamelen van informatie omtrent afgelegde routes betekent een relatief sterke uitbreiding van de tijd die de interviewer en respondent moeten besteden aan het nalopen van de ingevulde verplaatsingsformulieren. Ten einde de hoeveelheid tijd die de interviewers nodig hebben om de verlangde extra informatie te verzamelen te beperken, is besloten in het Validiteitsonderzoek te volstaan met een substeekproef. Deze substeekproef moest aselekt getrokken worden uit de streekproef van het OVG. De deelname aan het Validiteitsonderzoek zou ook moeten plaatsvinden op basis van vrijwilligheid. In de substeekproef van adressen zou gemiddeld één persoon per adres voor het Validiteitsonderzoek worden gevraagd. Om een aselechte substeekproef te verkrijgen is een methode ontwikkeld die uit twee fasen is opgebouwd. De eerste fase werd in hoofdzaak bepaald door de beperking dat niet alle gemeenten in Nederland de beschikking

hebben over recent kaartmateriaal en/of kaartmateriaal van voldoende kwaliteit ten behoeve van het intekenen en/of opmeten van afgelegde routes (zie par. 4.1.3.). Vanwege dit gegeven moest in de voorbereiding van het Validiteitsonderzoek een selectie worden gemaakt van geschikte gemeenten. Het resultaat van de selectie was dat 30 gemeenten op een totaal van 257 gemeenten buiten het Validiteitsonderzoek bleven. Van de 30 gemeenten horen er 14 thuis in de categorie plattelandsgemeenten volgens de CBS-indeling gebaseerd op de volkstelling in 1971. De steekproef van het OVG van de maanden november en december 1978 bestond uit onder andere 42 plattelandsgemeenten.

In de tweede fase moest voor de afgevallen gemeenten naar vervanging worden gezocht, waarbij per steekproefadres niet één, maar twee personen tot de substeekproef behoren. Daarbij diende met twee aspecten rekening te worden gehouden:

1. De substeekproef moest een zo groot mogelijke geografische spreiding hebben.
 2. De verdeling van de substeekproef naar gemeentegrootteklasse moest zo veel mogelijk identiek zijn aan die van de basissteekproef.
- Dit hield in dat als vervangende gemeenten alleen de gemeenten uit de steekproef van het OVG in aanmerking kwamen die geografisch dicht in de buurt van de te vervangen gemeenten lagen en bovendien tot dezelfde urbanisatieklasse behoorden. De omvang van de totale substeekproef moest gelijk zijn aan één persoon per steekproefadres.

Tijdens het zgn. ophaalbezoek dienen de aldus geselecteerde personen bij de controle van de ingevulde verplaatsingen door de interviewer informatie te verschaffen over de routes die zij gebruiken om zich te verplaatsen. De interviewer heeft tot taak de te verkrijgen informatie in te tekenen op stadsplattegronden en/of wegenkaarten. Vervolgens dienen de ingetekende routes te worden opgemeten. Hierdoor heeft men de beschikking over twee waarden die betrekking hebben op één en dezelfde verplaatsing. De ene waarde is een zo nauwkeurig mogelijke berekening van de werkelijke afstand en de andere waarde is een door een persoon opgegeven geschatte afstand. Het subonderzoek spitst zich toe op het onderling vergelijken van de twee waarden behorende bij één en dezelfde verplaatsing. Door middel van het analyseren van de te constateren verschillen moet worden getracht eventueel aanwezige systematische verschillen en de aanwezige

toevallige verschillen te kwantificeren. Tevens moet onderzocht worden of en in welke mate er variabelen zijn die invloed hebben op het eventueel systematisch onder- of overschatten van afstanden. Inzicht in deze problematiek zou kunnen leiden tot het berekenen van correctiefactoren die gebruikt kunnen worden om de verkeersprestatie te corrigeren voor mis-schattingen.

5. UITVOERING VAN HET ONDERZOEK

5.1. Werkprocedure CBS-onderzoek Verplaatsingsgedrag (OVG)

Het CBS distribueert de steekproefadressen die in een periode van vier weken bezocht moeten worden onder zijn interviewers. De interviewer stuurt een brief aan de bewoners van het te bezoeken adres. In de brief wordt het doel van het onderzoek uiteengezet en wordt een voorstel voor een bezoek met vermelding van datum en tijdstip gedaan. De bewoners van het te bezoeken adres kunnen contact opnemen met de interviewer voor het maken van een andere afspraak of het inwinnen van aanvullende informatie. Indien niet wordt gereageerd, bezoekt de interviewer het adres op de voorgestelde datum en tijd. In het eerste gesprek wordt na een introductie van het onderzoek gevraagd om medewerking. Wanneer deze wordt verleend, vult de interviewer ter plaatse het huishoudformulier (het zgn. omslagformulier) in.

Vervolgens wordt in het introductiegesprek uitgelegd hoe de bewoners van het adres de persoonsvragenlijst en het daaraan gekoppelde verplaatsingsboekje (B-formulier) moeten invullen. De in het CBS-onderzoek Verplaatsingsgedrag gehanteerde formulieren zijn te vinden in het rapport "Risico-onderzoek verkeersdeelnemers in Nederland (ROVIN) I" (SWOV, 1982). Aan het einde van het eerste gesprek wordt door de interviewer een afspraak gemaakt voor het ophaalbezoek. Dit kan drie tot vier dagen na het eerste gesprek zijn, afhankelijk van het aantal dagen dat men verplaatsingen moet invullen (twee of drie dagen). Tijdens het ophaalbezoek worden de ingevulde vragenlijsten en verplaatsingsboekjes globaal doorgenomen. Het CBS werkt uitsluitend met interviewsters.

In het geval dat de bewoners op de afgesproken dag en tijd niet thuis zijn, wordt het adres maximaal tweemaal herbezoekt. Heeft dan nog steeds geen gesprek plaatsgevonden, dan wordt het adres beschouwd als een non-response-adres. Ook adressen waar alle bewoners weigeren aan het onderzoek mee te doen, worden beschouwd als non-response-adressen. Wanneer op een adres een persoon weigert aan het onderzoek deel te nemen of tussentijds ophoudt met het invullen van het verplaatsingsboekje dan wordt de verkregen informatie niet verwerkt.

Het CBS-onderzoek behelst een mondelinge enquête op continue basis onder

een via een aselechte steekproef gevormde groep huishoudens. De continue basis houdt in dat regelmatig nieuwe steekproefadressen worden getrokken ten behoeve van een eenmalige deelname in het onderzoek.

Lezers die meer willen weten over het CBS-onderzoek Verplaatsingsgedrag, kunnen zich wenden tot de hoofdafdeling Verkeer en vervoer van het CBS.

5.2. Uitvoering veldwerk subonderzoek Nauwkeurigheid opgegeven afstanden

De opdracht tot het uitvoeren van het Validiteitsonderzoek werd verleend aan het marktonderzoekbureau de NV v/h Ned. Stichting voor de Statistiek (NSS) te 's-Gravenhage. Dit bureau heeft ook het deelonderzoek Non-response uitgevoerd. In het verleden heeft de SWOV reeds verschillende verplaatsingsonderzoeken door het genoemde bureau laten uitvoeren. Deze verplaatsingsonderzoeken waren proefonderzoeken in het kader van het ontwikkelen van een onderzoek naar het verplaatsingsgedrag. De aspecten "vergeten verplaatsingen" en "nauwkeurigheid opgegeven afstanden" kwamen in enkele proefonderzoeken uitvoerig aan de orde. Zodoende bestond reeds enige ervaring op het gebied van het valideren van verplaatsingsonderzoeken. Dit gold ook voor de bij het marktonderzoekbureau in dienst zijnde interviewsters en interviewers. Op grond van deze feiten was een selectie van ervaren en geschikte interviewers betrekkelijk eenvoudig te realiseren. Een belangrijk onderdeel van de instructie van de geselecteerde interviewers was het kaartlezen en het trainen op het nauwkeurig intekenen van routes op plattegronden en/of wegenkaarten. Hierbij werden een aantal voorbeelden gebruikt die varieerden van eenvoudig tot gecompliceerd. Daarnaast werd de opzet en werkwijze van het CBS-onderzoek Verplaatsingsgedrag uitvoerig toegelicht en de methode voor het opsporen en registreren van eventueel vergeten verplaatsingen uitgelegd.

De voorbereiding van het Validiteitsonderzoek moest binnen twee maanden worden voltooid. Alle aandacht was daarbij gericht op het probleemloos overnemen van het veldwerk van het OVG door de NSS van het CBS. De continuïteit van het OVG stond centraal.

Ten behoeve van het subonderzoek Nauwkeurigheid opgegeven afstanden is het CBS-onderzoek Verplaatsingsgedrag enigzins uitgebreid. Het CBS zond de steekproefadressen naar de NSS die deze onder zijn interviewers verdeelde. Het uitvoerende onderzoeksbureau NSS had tot taak de uitvoering

van het veldwerk voor het tijdelijk uitgebreide OVG in zijn geheel van het CBS over te nemen gedurende de periode 5 november 1978 (start veldwerkperiode 12) tot en met 31 december 1978 (einde veldwerkperiode 13). Vooraf waren de in de steekproef van het onderzoek te betrekken gemeenten aangewezen op basis van de aanwezigheid van geschikt kaartmateriaal. Uiteindelijk bleek dat vooral bij de kleine, agrarische gemeenten onvoldoende vervangende gemeenten gevonden konden worden die voldeden aan de in par. 4.4. aangegeven criteria. Dit had tot gevolg dat de kleine, agrarische gemeenten in beperkte mate ondervertegenwoordigd zijn in de substeekproef van adressen. Voor de bepaling van de mate van misschatten van afgelegde afstanden lijkt deze selectiviteit van belang te kunnen zijn, omdat de gemeentegrootte in het proefonderzoek van 1976 enige invloed op het misschatten bleek te hebben.

De omvang van de totale substeekproef moest gelijk zijn aan één persoon per steekproefadres van het OVG. Door de beperking van het aantal steekproefgemeenten diende op steekproefadressen uit de vervangende gemeenten meer dan één persoon per steekproefadres geselecteerd te worden. Het maximum aantal personen per steekproefadres werd op twee gesteld om de vrijwillig meewerkende huishoudens zo min mogelijk te belasten.

Alle steekproefadressen werden vooraf voorzien van een code. De code 0 betekende dat het desbetreffende adres geen deel uitmaakte van de substeekproef. Een code 1 betekende dat één persoon op het desbetreffende steekproefadres geselecteerd moest worden. En met de code 2 werd aangegeven dat op het steekproefadres twee personen voor de substeekproef geselecteerd moesten worden.

De tweede stap van de selectie had betrekking op de keuze van personen binnen een huishouden dat behoorde tot de substeekproef. Hierbij werd gebruik gemaakt van de zgn. dobbelsteenmethode. De interviewer diende met behulp van een dobbelsteen één of meer personen voor het Validiteitsonderzoek te selecteren. Dit selecteren moest de interviewer thuis doen op een tijdstip gelegen tussen het introductiegesprek, waarbij onder andere de samenstelling van het desbetreffende huishouden wordt vastgelegd, en het ophaalbezoek. Tussen het CBS en de SWOV was overeengekomen dat de dobbelsteenmethode niet thuis bij de respondenten mocht plaatsvinden om eventuele irritatie bij de respondenten te voorkomen.

Het aselechte selecteren van de respondent door het gooien van een dobbelsteen op een moment voordat het ophaalbezoek plaatsvond is in de praktijk door drie factoren niet haalbaar gebleken.

Door het feit dat het meewerken aan het onderzoek op basis van vrijwilligheid geschiedde kon de situatie ontstaan dat bij het weigeren van de beoogde respondent een andere persoon binnen het huishouden werd gevraagd. Wanneer de beoogde respondent geen verplaatsingen had gemaakt op de dag waarover de aanvullende informatie (afgelegde routes) moest worden verstrekt, dan moest een andere respondent binnen het huishouden worden gevonden die zich wel verplaatst had. Dit was noodzakelijk om over voldoende aantallen ingetekende routes te kunnen beschikken.

Teneinde de extra inspanning die verricht moest worden om de verlangde route-informatie te verstrekken beperkt te houden, was besloten alleen gegevens te vragen over de laatste dag van de invulperiode. Aan het besluit lag tevens ten grondslag de opvatting dat het herinneringsvermogen van personen beperkt is en dat informatie over de eerste invuldag minder nauwkeurig is dan over de laatste invuldag. Hierdoor ontstond het nadeel dat van verplaatsingen die op een zondag waren afgelegd geen routebeschrijvingen konden worden verzameld. De zondag valt in de vooraf vastgestelde aaneengesloten invulperiode van zaterdag tot en met maandag. Deze selectiviteit naar dagsoort heeft uiteraard invloed op de reikwijdte van het onderzoekresultaat.

Daarnaast kon de situatie bestaan dat degene die met behulp van de dobbelsteen was aangewezen op het moment dat de informatie inzake afgelegde routes verzameld moest worden niet beschikbaar of aanwezig was. Hiermee is aangegeven dat de zogenaamde dobbelsteenmethode in feite niet toe te passen is. De interviewers zullen daarom in de praktijk van de methode zijn afgeweken of deze in het geheel niet hebben toegepast. Dit betekent dat de interviewers in veel gevallen de ten tijde van het ophaalbezoek aanwezige personen die verplaatsingen hadden afgelegd, benaderd hebben voor een aanvullend gesprek inzake de afgelegde routes. Of deze gang van zaken tot selectiviteit heeft geleid is de vraag.

Selectiviteit is er wel om andere redenen. Hij wordt o.a. veroorzaakt door het feit dat door de selectie de trefkans van personen om tot deze substeekproef te behoren, afhankelijk is van de grootte van het huishouden. Personen in kleine huishoudens of alleenstaanden hebben een grotere trefkans dan personen in grote huishoudens. Voor dit aspect kan in principe achteraf worden gecorrigeerd. De correctie is hier niet uitgevoerd. Overigens is in het proefonderzoek van de SWOV uit 1976 de invloed van de grootte van de huishoudens op de mate van het mis-schatten van afgelegde afstanden niet onderzocht.

Tijdens het zgn. ophaalgesprek moest de interviewer de door de ondervraagde personen geleverde informatie omzetten in getekende routes. De routes zijn op stadsplattegronden en provinciale- en/of landelijke wegenkaarten ingetekend. Voor de stadsplattegronden werd gebruik gemaakt van kaarten die worden uitgegeven door gemeenten, plaatselijke VVV's en boekhandels annex uitgevers en door een landelijk reclamebureau RIJNLAND. Voor de provinciale wegenkaarten zijn de ANWB-kaarten gebruikt en voor de landelijke autokaart de SHELL-autokaart van Nederland. Per plattegrond of wegenkaart mochten maximaal vijf routes worden ingetekend. Dit maximum werd gehanteerd teneinde de leesbaarheid en nauwkeurigheid van het intekenen enigzins te garanderen. Elke route werd met een aparte kleur of in een aparte vorm ingetekend en van een volgnummer voorzien. Dit volgnummer moest corresponderen met het volgnummer in het verplaatsingsboekje. De interviewers die belast waren met het intekenen van routes hadden ieder de beschikking over twee stadsplattegronden, een provinciale kaart en een landelijke autokaart per persoon behorende tot de substeekproef van het Validiteitsonderzoek. In een aantal gevallen was het beschikbare kaartmateriaal niet voldoende. Het is voorgekomen dat de interviewer bij een bepaald bezoek meer dan twee stadsplattegronden nodig had of dat de desbetreffende persoon zich had verplaatst in een andere gemeente dan zijn/haar woongemeente. In het laatste geval kon slechts een gedeelte van de afgelegde route op kaarten worden ingetekend. Bij het meten van de ingetekende routes zijn laatstgenoemde routes niet gemeten. Het is onduidelijk of dit verder van veel betekenis zou zijn. Met de hiervoor genoemde situaties kon vooraf geen rekening worden gehouden. Indien een interviewer voorzag dat hij/zij stadsplattegronden tekort kwam dan werd de desbetreffende interviewer aangespoord op eigen initiatief plattegronden te kopen.

5.3. De mate van nauwkeurigheid van de meetmethode in de praktijk

Bij het omzetten van routebeschrijvingen in getekende routes doet zich een aantal problemen voor. De vier belangrijkste problemen die tevens kunnen leiden tot significante afwijkingen worden in het kort besproken. De ondervraagde persoon wordt op de dag na afloop van de invulperiode tijdens het zgn. ophaalgesprek ondervraagd over de op de laatste invuldag afgelegde routes. Hierbij kan het voorkomen dat de desbetreffende persoon

zich niet meer precies kan herinneren door welke straten hij zich verplaatst heeft. Daarnaast kan de betrokkene uit onwil niet willen vertellen welke route hij/zij heeft genomen en bewust een andere route opgeven. Aangezien de controle op wat de desbetreffende persoon in werkelijkheid heeft gedaan ontbreekt, zijn eventuele afwijkingen niet vast te stellen. Hierdoor is er sprake van mogelijke fouten die op geen enkele wijze zijn te kwantificeren.

Ongeveer de helft van de in het subonderzoek gebruikte ruim 210 stadsplattegronden heeft een schaal van of in de buurt van 1:5.000 (enkele uitschieters beneden 1:1.000) en de anderen hebben schalen van of in de buurt van 1:10.000, 1:15.000 en 1:25.000. De provinciale wegenkaarten hebben een schaal van 1:100.000 en de landelijke wegenkaart heeft een schaal van 1:250.000.

De stadsplattegronden zijn voordat met het nameten van ingetekende routes werd gestart met behulp van kaarten 1:10.000 van de Topografische Dienst te Delft geijkt.

In het meest extreme geval zijn zowel het vertrek- als het aankomstadres niet goed aangegeven en wel met een afwijking van twee keer 1 cm. Dit betekent op een plattegrond met schaal 1:5.000 een maximale meetfout van 2 x 50 meter, hetgeen bij verplaatsingen met een geringe afstand een duidelijke vertekening van de meetuitkomst oplevert. Bij loopverplaatsingen kan dit een afwijking van 25% betekenen bij een afstand van circa 400 meter (meer dan de helft van de loopverplaatsingen heeft een afstand van minder dan 400 meter).

Bij verplaatsingen met personenauto's kunnen afwijkingen optreden van maximaal 10% zowel voor wat betreft de intergemeentelijke verplaatsingen als wat betreft verplaatsingen over grote afstanden (meer dan 25 km). Dit gegeven leidt ertoe dat de vertrek- en aankomstadressen met een grote mate van nauwkeurigheid moeten worden aangegeven. In de praktijk is dat een moeilijke, zo niet onmogelijke opgave gebleken.

Een ander probleem vormt het intekenen van de route op een plattegrond. Het is regelmatig voorgekomen dat de route niet precies de op de plattegronden aangegeven straten volgden als gevolg van bochtafsnijdingen en het over gebouwen heen tekenen. In een aantal gevallen is dat ook niet goed te realiseren vanwege bijvoorbeeld het feit dat een straat als een dunne lijn is getekend. Hierdoor is het noodzakelijk gebleken de ingetekende routes voordat zij werden opgemeten na te lopen op mogelijk onzorgvuldig intekenen.

Een ander probleem houdt verband met de mate van onnauwkeurigheid die kan ontstaan bij het tekenen van de overgang van de ene kaart naar de andere (bijvoorbeeld van een stadsplattegrond naar een provinciale- of landelijke wegenkaart). In de voorbereidingsfase van het meten van ingetekende routes is gecontroleerd in hoeverre de overgang goed is ingetekend.

Wanneer bleek dat de overgang onnauwkeurig of verkeerd was ingetekend dan werd dat achteraf zo goed mogelijk gecorrigeerd.

Geconstateerd is uit de controlemetingen dat de effecten van een niet op elkaar aansluitende overgang van de ene kaart op een andere kaart in een groot aantal gevallen uitmiddelen tegen andere afwijkingen bij het volgen van routes. Daarmee is niet vastgesteld dat de nameting geen systematische verschillen met de werkelijke afstand kan opleveren. Deze worden echter te verwaarlozen klein geacht. Per afzonderlijke verplaatsing kan een maximale afwijking van drie- tot vierhonderd meter zijn opgetreden. Aangezien dit verschijnsel in de meeste gevallen optreedt bij verplaatsingen met een relatief grote afstand (groter dan tien kilometer) zal het effect in een beperkt aantal gevallen ten hoogste vijf percent kunnen zijn.

Tenslotte moet opgemerkt worden dat de kaarten kunnen rekken en krimpen als gevolg van veranderingen in vochtigheid. De afwijkingen die hierdoor optreden kan op enkele percenten worden geschat.

5.4. Meten van op plattegronden ingetekende routes

Bij het meten van de ingetekende routes is dankbaar gebruik gemaakt van de diensten van de Meetkundige Dienst van Rijkswaterstaat te Delft. Na enig speurwerk bleek deze te beschikken over een digitaliseertafel met bijbehorende apparatuur die geschikt was voor het meten van ingetekende routes. Een nadeel was wel dat deze apparatuur voor onze doeleinden slechts in beperkte mate beschikbaar was. De Meetkundige Dienst heeft ten behoeve van de meetwerkzaamheden een besturingsprogramma geschreven. Ter voorbereiding van de meetwerkzaamheden zijn door de NSS de volgende activiteiten verricht:

- het kopiëren van de ingevulde verplaatsingsformulieren waarvan ook een ingetekende route beschikbaar was, e.e.a. met inachtneming van eisen m.b.t. bescherming van de persoonlijke levenssfeer en in overeenstemming met eisen van het CBS;

- het op een hanteerbaar formaat knippen van de plattegronden en wegenkaarten;
- het nagaan of in het geval van twee kaarten per route het eindpunt op de ene kaart identiek was aan het beginpunt op de andere kaart en eventuele afwijkingen corrigeren;
- het berekenen van de exacte schalen van de gebruikte stadsplattegronden met behulp van kaarten van de Topografische Dienst;
- het controleren of de route nauwkeurig is getekend;
- het aanleggen van een enveloppe inhoudende gekopieërde verplaatsingsformulieren en bijbehorende ingetekende routes met vermelding van enkele gegevens als enquêtenummer, leeftijd en geslacht van de desbetreffende persoon, woongemeente en de schalen van de gebruikte plattegronden.

Alvorens een ingetekende route gemeten werd diende men een aantal gegevens in de computer in te voeren. De gegevens waren: enquêtenummer, verplaatsingsnummer, type plattegrond, en vervoerswijze. Vervolgens moest het beginpunt van de ingetekende route worden gedigitaliseerd (in coördinaten vastgelegd). Daarna werd de cursor over de ingetekende route bewogen tot aan het eindpunt van de ingetekende route. De serie handelingen werd afgesloten met het vastleggen van de coördinaten van het eindpunt van de ingetekende route. Wanneer een route op twee verschillende kaarten was ingetekend, dan moesten bij het einde van de route op een kaart ook drie referentiepunten in coördinaten worden vastgelegd. Deze handeling moest op de tweede kaart herhaald worden. Door middel van een in het besturingsprogramma opgenomen zgn. driehoeksmetmethode en het intoetsen van de twee bij de gehanteerde kaarten behorende schalen werd uitgerekend het verschil in de overgang van de ene kaart naar de andere kaart. Als het verschil in verhouding tot de totaal te berekenen afstand binnen een marge van 2 à 3% bleef, werd het meten van de ingetekende route op de tweede kaart vervolgd. Was het verschil groter dan 2 à 3% dan moest de zgn. driehoeksmeting worden overgedaan. In de nacontrole werd veel aandacht besteed aan de uitgevoerde zgn. driehoeksmetingen. Teneinde de kwaliteit van de meetwerkzaamheden te controleren was vastgelegd dat elke tiende meting op een andere dag door andere personen werd overgedaan. Hierbij werd geconstateerd dat er grote onderlinge meetverschillen voorkwamen. Afwijkingen van tien percent of meer bleken echter gering in aantal te zijn. In deze gevallen werd een meting voor een derde keer

uitgevoerd en één van de twee meetwaarden die het dichtst bij elkaar in de buurt lagen als definitieve meetwaarde gebruikt. Echter in veruit de meeste gevallen bedroeg het onderlinge verschil in afstand minder dan drie percent. De standaard meetfout bleek berekend over alle controlemetingen in de buurt van de drie percent te liggen. Het meten van ingetekende routes volgens de hiervoor beschreven werkwijze kan leiden tot verschillende onnauwkeurigheden waaronder het niet nauwkeurig digitaliseren van het vertrek- en of aankomstadres. Dit zal niet tot grote afwijkingen hebben geleid, behoudens in de gevallen waarbij de te meten afstand minder dan tweehonderdvijftig meter bedroeg. Andere mogelijke onnauwkeurigheden zijn in par. 5.3. besproken.

Wanneer de mogelijke foutenbronnen (rek en krimp van het kaartmateriaal, niet goed intekenen van afgelegde route, meetfouten) allemaal aanwezig zouden zijn en in de zelfde richting zouden werken, zou de maximale fout in een enkel geval ongeveer 10% kunnen bedragen.

5.5. De analyse van de verzamelde afstandgegevens

Volgens de definitie van het CBS wordt onder een verplaatsing verstaan de afstand afgelegd tussen een herkomstadres en een bestemmingsadres. Hierbij speelt geen rol of bij de verplaatsing van meer dan één vervoerswijze gebruik is gemaakt.

De ingetekende routes zijn daar waar van verschillende vervoerswijzen gebruik is gemaakt bij het nameten opgesplitst naar vervoerswijzen. Hierdoor kan één ingetekende route verschillende nagemeten afstanden opleveren. Elk afzonderlijk deel wordt rit genoemd.

De opgemeten afstanden van op plattegronden ingetekende routes zijn te zamen met een aantal kenmerken, die als identificatie van de desbetreffende verplaatsingen dienden, in een te koppelen bestand opgenomen. Het bestand is vervolgens gekoppeld aan het bestand van het CBS-onderzoek Verplaatsingsgedrag. Hiermee werd een directe relatie gelegd tussen een bepaalde opgegeven verplaatsingsafstand en bijbehorende kenmerken en de nagemeten verplaatsingsafstand.

Gezien het doel van het onderzoek (zie Hoofdstuk 3) is de analyse erop

gericht te onderscheiden tussen eventuele systematische invloeden op het misschatten van afstanden en toevallige schattingsfouten. Dit kan worden bereikt door te onderzoeken of de eventueel te vinden systematische invloeden zodanig groot zijn in verhouding tot de toevallige schattingsfouten, dat er sprake is van een significant niveau. Daartoe dient van de in het onderzoek te betrekken variabelen te worden berekend wat hun bijdragen zijn aan de verschillen tussen opgegeven en nagemeten afstanden, wat met behulp van variantie-analyse kan worden gedaan.

Aan het gebruik van variantie-analyse zijn in dit geval enige problemen verbonden. De techniek veronderstelt dat de toevallige misschattingen normaal verdeeld en onderling onafhankelijk zijn, met een over de gehele populatie gelijke variantie.

In dit onderzoek zijn er sterke afhankelijkheden door identieke ritten. Deze kan men ondervangen door van de identieke ritten slechts één rit in het analysebestand op te nemen.

Voor andere mogelijk aanwezige afhankelijkheden, als het feit dat dezelfde personen meer verplaatsingen maakten en dus ook vaker schattingen gaven, is correctie vooraf niet goed mogelijk binnen het kader van het huidige onderzoek.

In het proefonderzoek uit 1976 is geconstateerd dat de spreiding in de verschillen tussen de opgegeven en nagemeten afstanden toeneemt met de opgegeven afstand. Hiervoor zijn theoretisch verklaringen denkbaar, zodat een worteltransformatie van de afstand ervoor corrigeert. De worteltransformatie bleek in het genoemde proefonderzoek redelijk te voldoen en was ook voor de gegevens van het huidige onderzoek afdoende. De Annex II bij het als Bijlage 2 opgenomen zgn. Technische verslag toont overigens aan dat verschillen tussen de varianties, hoewel wezenlijk verminderd, toch niet geheel lijken te verdwijnen, al kan dit door het steekproefkarakter worden veroorzaakt.

Bij Bijlage 2 is als Annex I een theoretische beschouwing opgenomen over deze variantiestabilisatie zoals deze is nagestreefd ten behoeve van het kunnen toepassen van variantie-analyse.

De eigenlijke kwantificering van significant bevonden effecten is vervolgens geschied op basis van de totale nameetsteekproef, dus inclusief de identieke verplaatsingen. Ondanks alle representativiteitsproblemen is deze steekproef immers toch de aannemelijkste representant van het OVG in

dit onderzoek. De spreiding in de grootten van deze effecten is echter weer bepaald uit het analysebestand, omdat de afhankelijkheden anders tot moeilijkheden zouden kunnen leiden.

5.6. De uitvoering van de analyses

De eerste activiteit in de voorbereidingsfase van de analyse bestond uit het selecteren van de gegevens uit het OVG-bestand. De selectie vond plaats op aanwezigheid van verplaatsingen met een nagemeten afstand. Uit de geselecteerde gegevens werd een aantal vooraf vastgestelde variabelen geconstrueerd (zie par. 6.4.). Daarna hebben de opgegeven en nagemeten afstanden een worteltransformatie ondergaan en zijn de analysegegevens gehercodeerd en in klassen ondergebracht. Het aldus verkregen bestand omvat de totale nameetsteekproef. Hieruit is het analysebestand gevormd door er de identieke verplaatsingen uit weg te laten.

De eigenlijke analyse is voorafgegaan door een algemene inspectie van gegevens. Daarna zijn CANALS-analyses uitgevoerd om de belangrijkste variabelen op te sporen. Op deze variabelen zijn (co)variantie-analyses gedaan en van de significante effecten zijn de spreidingen berekend. Deze analyses hebben plaatsgevonden op het analysebestand. Vervolgens zijn de grootten van significant bevonden systematische effecten geschat uit het totale nameetbestand, nadat dit op enkele aspecten nog was vergeleken met het analysebestand.

Vanwege verschillen in actieradius moeten per vervoerswijze aparte afstandsklassen worden gehanteerd. Ook omdat het aannemelijk is dat het misschatten van afstanden afhankelijk is van de vervoerswijze en dat eventuele correctie voor misschatten per vervoerswijze plaats zal vinden, is besloten de gegevens per vervoerswijze afzonderlijk te analyseren. In de eerste fase van de analyse zijn de opgegeven afstanden uit het analysebestand tegen de nagemeten afstanden geplot waaruit een indruk van de totale schattingsfout wordt verkregen. Verder is een lineaire-regressie-analyse uitgevoerd ten einde een indruk te krijgen van de systematische schattingsfout. Ten behoeve van een overzichtelijke presentatie van de plots in één vorm en schaalgrootte zijn enkele extreme opgegeven of nagemeten afstand buiten beschouwing gebleven. Vervolgens zijn de verschillen tussen de wortelgetransformeerde opgegeven en nagemeten afstan-

den geplot tegen de nagemeten afstand. Ook over de aldus verkregen datapunten is een lineaire-regressie-analyse uitgevoerd.

Voor de uitvoering van de tweede fase van de analyse bleek het aantal bromfietsverplaatsingen (44) te gering te zijn. Hierdoor bleven de vervoerswijzen: lopen, fiets en personenauto over. Op de drie genoemde datasets zijn CANALS-analyses uitgevoerd.

Bij CANALS-analyses worden de klassen van de variabelen geschaald zodat het verschil tussen de wortelgetransformeerde opgegeven en nagemeten afstanden optimaal wordt beschreven met behulp van de andere aanwezige variabelen. Het resultaat van de CANALS-analyses is dat zij bepalen welke variabelen de grootste bijdrage leveren bij het verklaren van de verschillen tussen de wortelgetransformeerde opgegeven en nagemeten afstanden.

Aangezien bij de toegepaste variantie-analyses slechts een beperkt aantal variabelen tegelijk meegenomen kunnen worden, moest de keuze voor nader te analyseren variabelen bepaald worden door de CANALS-analyses die kunnen worden beschouwd als zoekstrategieën bij de aanwezigheid van veel kenmerken. De reden om naast CANALS-analyses ook variantie-analyses uit te voeren is dat de variantie-analyse betere toetsingsmogelijkheden biedt. CANALS-analyse-resultaten zijn vooral beschrijvend. Betrouwbaarheidsonderzoek is hierbij relatief duur.

In de derde fase van de analyse zijn over de belangrijkste gebleken variabelen (co)variantie-analyses uitgevoerd. Daarmee zijn de variabelen die op de mate van het misschatten een significante invloed hebben bekend.

In de laatste fase van de analyse moesten vervolgens de grootten van de effecten van deze variabelen nog uit het totale nameetbestand worden bepaald aan de hand van de sommen van opgegeven en nagemeten afstanden. De spreidingen van de grootten van deze effecten zijn tenslotte, zoals reeds vermeld, uit het analysebestand geschat.

6. RESULTATEN VAN HET ONDERZOEK

6.1. De steekproefomvang

Bij de voorbereiding van het onderzoek was uitgegaan van een substeekproef van circa 2000 personen (gemiddeld één persoon per steekproefadres). Door een hogere non-response dan vooraf berekend (840 adressen in plaats van 700 adressen) bleef uiteindelijk een substeekproef van 1384 personen over. Deze betreft personen die op de dag waarover gegevens over afgelegde routes moesten worden verzameld verplaatsingen hebben gemaakt. Naast een prognose voor het aantal te verwachten personen in de substeekproef is ook een raming gemaakt van het aantal te verwachten verplaatsingen volgens de definitie van het CBS. Uitgaande van een gemiddeld aantal verplaatsingen per persoon per dag van circa drie werd verondersteld circa 6000 verplaatsingen met ingetekende routes aan te treffen. Uiteindelijk bleek dat van 4795 verplaatsingen de volledige route was ingetekend op kaarten. Deze 4795 verplaatsingen zijn nagemeten. Hiervan waren 192 verplaatsingen gemaakt in 1979, toen nieuwe verplaatsingsformulieren zijn gebruikt. Dit kan geleid hebben tot veranderingen in de registratie van afgelegde verplaatsingen. Om te voorkomen dat de 192 verplaatsingen verstoring zouden werken op de analyse-uitkomsten zijn die verplaatsingen uiteindelijk niet in het analysebestand opgenomen. Het koppelen van het OVG-bestand aan het bestand met nagemeten afstanden heeft problemen opgeleverd. Dit heeft uiteindelijk geleid tot een verlies van 1036 niet te koppelen verplaatsingen. In par. 6.2. wordt besproken wat de consequenties zijn van het grote aantal uitgevallen verplaatsingen voor de interpretatie van de analyse-uitkomsten.

Ten behoeve van het zoveel mogelijk vermijden van correlaties tussen de verschillende gegevens, zijn in het analysebestand alleen de unieke ritten opgenomen. Dit heeft geleid tot een reductie van het rittenbestand. Daarnaast zijn ritten weggelaten waarbij niet alle gegevens bekend waren of het aantal per vervoerswijze te gering van omvang voor de analyse.

In Tabel 1 is aangegeven hoe de definitieve omvang van het analysebestand bestaande uit 2009 ritten (zie voor de definitie par. 5.4) ontstaan is. De ritten gemaakt met het openbare vervoer e.d. (124) waren te gering in aantal om te worden geanalyseerd.

6.2. De representativiteit van de substeekproef

Tussen de samenstellingen van de analyse-, onderzoek- en OVG-bestanden bestaan de nodige verschillen. De oorzaken zijn verschillen in gezinsgrootte-verdelingen (zie par. 5.2.), ondervertegenwoordiging van kleine gemeenten in de substeekproef (zie par. 4.3.) en het feit dat het subonderzoek betrekking heeft op de periode november-december 1978, terwijl het OVG een continu onderzoek is. Dit heeft tot gevolg dat de resultaten van de analyses niet voor het hele OVG hoeven te gelden.

In het subonderzoek moesten uitsluitend gegevens verzameld worden over afgelegde routes van ritten die op de laatste invuldag van het OVG gemaakt waren. Personen die op een zaterdag met het invullen van afgelegde verplaatsingen begonnen, moesten ook gegevens over de zondag en de maandag invullen. Zodoende kunnen er in het subonderzoek dus over de zondag geen routegegevens verzameld zijn. Achteraf is echter geconstateerd dat er toch gegevens over ritten gemaakt op een zondag in het onderzoek- en analysebestand aanwezig zijn. Een mogelijke verklaring is dat interviewers gegevens verzameld hebben over meer dan één invuldag, bijvoorbeeld bij een middernachtelijke verplaatsing.

Een andere afwijking van het OVG wordt veroorzaakt door het praktische probleem dat vooraf niet te voorzien is of personen routes afleggen in gebieden buiten de eigen woongemeente. Voor een deel is dat ondervangen door naast stadsplattegronden ook provinciale- en landelijke wegenkaarten te gebruiken. Dit probleem heeft geleid tot het feit dat een beperkt deel van de afgelegde routes (circa 5%) niet of slechts ten dele op kaarten zijn ingetekend en derhalve niet zijn nagemeten. Het betekent een verdere vertekening van de steekproef van het subonderzoek.

Door de koppelingsproblematiek van nagemeten afstanden aan het OVG-bestand zijn, als gesteld in de vorige paragraaf, verschillende verplaatsingen buiten het onderzoekbestand gebleven. In tabellen 2 t/m 5 is voor de unieke verplaatsingen (netto-ritten) nagegaan of hiervan een belangrijke vertekening is te verwachten. Volgens de berekende χ^2 -waarde is alleen voor fietsverplaatsingen de verdeling over de klassen van opgegeven afstand significant verschillend voor wel en niet gekoppelde verplaatsingen.

6.3. De inspectie van de gegevens

In deze fase zijn de opgegeven afstanden tegen de nagemeten afstanden afgezet in zogenaamde plots. Per vervoerswijze zijn twee plots gemaakt. De eerste plot heeft betrekking op de feitelijke waarden en de tweede plot bevat de waarden nadat de worteltransformatie heeft plaatsgevonden. Bij de vier in de analyse betrokken vervoerswijzen is sprake van grote verschillen tussen opgegeven en nagemeten afstanden, waaronder het overschatten van de afgelegde afstanden, maar vooral ook toevallige fouten een rol spelen.

De "loopritten" zijn afgebeeld als Afbeelding 1A en 1B in Bijlage 2. Plot 1A laat zien dat de afgelegde afstand doorgaans wordt overschat en dat de spreiding in het schatten aanzienlijk is. Blijkens de lineaire regressie-analyse bedraagt de systematische overschatting gemiddeld ongeveer 300 m. De systematische overschatting neemt licht toe met de grootte van de afgelegde afstand, wat kan worden afgeleid uit het feit dat de richtingscoëfficiënt van de regressielijn groter is dan 1.

De fietsritten (Afbeelding 2A en 2B in Bijlage 2) vertonen een beeld dat niet veel afwijkt van dat van de loopritten. Het lineaire model voor de relatie tussen de opgegeven en de nagemeten afstand is gezien de grotere actieradius van de fiets bij een verklaarde variantie van 0,71 wat relevanter. Hier staat tegenover de tot 1500 m toegenomen standaard schattingsfout. De systematische overschatting van de afgelegde afstanden is toegenomen tot gemiddeld 600 m.

Het aantal bromfietsritten (Afbeelding 3A en 3B in Bijlage 2) is beperkt. Vier enquêtewaarnemingen met nagemeten afstanden van circa tien kilometer blijken uitzonderlijk te zijn. Deze waarnemingen dragen sterk bij tot het resultaat van de lineaire regressie-analyse. Daardoor lijkt het analyse-resultaat in details af te wijken van het tot dusver gegeven beeld.

Verdere interpretatie blijft derhalve achterwege.

Voor wat betreft de personenauto (Afbeelding 4A en 4B in Bijlage 2) blijkt uit plot 4A dat de standaard-schattingsfout is opgelopen tot 5000 m. De gemiddelde grootte van de systematische overschatting van de afgelegde afstanden is gestegen tot 1500 m. Ook plot 4B en de bijbehorende regressie-analyse-uitkomsten passen goed bij de resultaten van de andere vervoerswijzen.

De keuze voor een aparte behandeling van de vervoerswijzen waardoor het

mogelijk was per vervoerswijze aparte klassen van opgegeven afstanden te gebruiken wordt door de uitkomsten van de regressie-analyses gerechtvaardigd. Wanneer alle afstanden te zamen geanalyseerd zouden zijn, dan zouden de afzonderlijke verschillen tussen de vervoerswijzen onzichtbaar zijn.

6.4. De resultaten van de CANALS-analyses

In Tabel 6 zijn de vooraf gekozen variabelen opgenomen die voor de analyses het meest relevant geacht werden en die binnen het OVG, zoals dat in 1978 werd uitgevoerd, beschikbaar waren. Uit Tabel 1 in Bijlage 2 blijkt overduidelijk dat bij de loopverplaatsingen de opgegeven afstand het sterkst correleert met het verschil tussen de wortelgetransformeerde opgegeven en nagemeten afstanden. Door de opgegeven afstand tevens als "afhankelijke" variabele toe te voegen naast het verschil tussen de wortelgetransformeerde opgegeven en nagemeten afstanden, ontstaat een effect waarbij de eerdergenoemde correlatie zo veel mogelijk verdisconteerd wordt.

Bij deze hernieuwde CANALS-analyse is de invloed van de opgegeven afstand op het verschil tussen de getransformeerde afstanden dan ook verwerkt. De volgorde waarin de overige variabelen van belang zijn, kan hierbij, zoals in Tabel 1 in Bijlage 2 te zien is, veranderen.

De overwegende indruk uit deze tabel is dat de verschillende correlaties met de nagemeten afstand nogal gering zijn. Dit wordt bovendien zeer duidelijk geïllustreerd in Afbeelding 5 in Bijlage 2. Hiermee wordt tevens duidelijk dat de systematische fout relatief klein is en met name betrekking heeft op de opgegeven afstand. Daarentegen is de toevallige fout groot, waardoor de bruikbaarheid van de uitkomsten beperkt is. Dit heeft tot gevolg dat de gevonden correcties niet zonder meer gebruikt kunnen worden. Wat de loopverplaatsingen betreft hebben van de onderzochte de variabelen: geslacht, reismotief, vervoermiddelenaanwezigheid, leeftijd en bezigheid nog de meeste betekenis voor een systematisch component aan het misschatten van de afgelegde afstanden.

De uitkomsten van de twee CANALS-analyses voor de fietsritten komen in grote mate overeen met de uitkomsten voor de loopverplaatsingen. Dit blijkt ook onder andere uit Afbeelding 6 in Bijlage 2. Echter nu zijn de variabelen: dagtype, burgerlijke staat, leeftijd, bezitduur rijbewijs en opleidingsniveau de belangrijkste variabelen.

Uit Tabel 1 in Bijlage 2 blijkt dat bij de personenautoritten hetzelfde beeld aanwezig is als bij de loop- en fietsritten, hetgeen wordt verduidelijkt door Afbeelding 7 in Bijlage 2. Bij de personenautoritten zijn de belangrijkste variabelen: regio, bezigheid, opleidingsniveau, leeftijd en burgerlijke staat. Alleen de leeftijd komt bij alle drie vervoerswijzen naar voren als een mogelijk belangrijke invloedsvariabele.

6.5. De resultaten van de (co)variantie-analyses

Van de variabelen die bij de loopverplaatsingen als de belangrijkste variabelen uit de CANALS-analyses tevoorschijn kwamen, blijkt volgens Tabel 2 in Bijlage 2 alleen de opgegeven afstand een significant effect te hebben op de systematische component van de schattingsfout.

Bij de fietsritten zijn de opgegeven afstand en dagtype significant van invloed op de grootte van het misschatten.

De opgegeven afstand en in mindere mate de regio hebben bij de personenautoritten een significant effect op de systematische component van de schattingsfout. Daarmee zijn de resultaten van de (co)variantie-analyses geheel conform de resultaten van de CANALS-analyses.

De grootte van de huishoudens heeft geen aantoonbaar effect op de omvang van het systematisch misschatten. Dit is een belangrijke uitkomst, omdat de steekproef van het subonderzoek zoals gesteld in par. 5.2. niet representatief is op dit punt.

6.6. Verschillen tussen analyse- en onderzoekbestand

Het onderzoekbestand omvat zoals bekend 3567 ritten (Tabel 1). Daarvan zijn er 1055 te voet gemaakt, 801 op de fiets en 1420 met de auto. Zodoende bevatte het analysebestand 57% van de loopritten, 52% van de fietsritten en 58% van de personenautoritten.

Verder zijn er 75 bromfietsritten, inclusief de identieke ritten in het onderzoekbestand aanwezig.

Mogelijk bestaat er een lichte tendens dat kortere ritten eerder identieke betreffen, dus vaker op één dag door één persoon worden gemaakt (grootste effect: bij lopen verwachten we op basis van evenredigheid in het analysebestand 155 ritten in de laagste klasse van opgegeven afstand, het werden er 138; in de hoogste klasse verwachten we 56 ritten, het zijn er 69). Overigens hoeven dit niet ook de veelvuldig gemaakte ritten te zijn.

Een en ander blijkt ook uit de algehele gemiddelde opgegeven afstand. Was deze in het analysebestand voor loopritten 1,015 kilometer, in het onderzoekbestand is hij gedaald tot 0,894 kilometer, voor fietsritten is hij afgenomen van 3,213 kilometer tot 3,063 kilometer en voor personenautoritten van 13,087 kilometer tot 12,157 kilometer.

De algehele gemiddelde nagemeten afstand is daarbij echter in nagenoeg gelijke mate gedaald. De berekende gemiddelde correctie voor de opgegeven afstand verschilt zodoende niet voor analyse- en onderzoekbestand. Voor loopritten is de gemiddelde nagemeten afstand afgenomen van 0,672 kilometer tot 0,599 kilometer, voor fietsritten van 2,373 kilometer tot 2,264 kilometer en voor personenautoritten van 11,177 kilometer tot 10,341 kilometer. Dit betekent dat de zgn. unieke ritten gemiddeld een langere afstand hebben.

Zoals bekend zijn systematische effecten aangetoond in het analysebestand.

De grootten van systematische effecten worden bepaald uit het totale onderzoekbestand.

6.7. Het schatten van de grootte van significant bevonden systematische effecten

Als in de variantie-analyses een variabele van invloed is dan zijn er significante verschillen tussen de klassen van die variabele onderling naar de mate waarin er sprake is van systematisch misschatten. Bij het interpreteren van de analyse-uitkomsten is de grens gelegd bij het 1%-niveau. Dit is gedaan om versluierende invloeden op het schatten van afstanden door de geënquêteerde personen (schattingsproces) en het achteraf berekenen van de afstanden (meetproces) zoveel mogelijk uit te sluiten. Ten aanzien van het zgn. schattingsproces kan sprake zijn van wel of geen ervaring met het schatten van afstanden en bekendheid met de afgelegde route (bijvoorbeeld: hoe vaker een bepaalde route afgelegd des te nauwkeuriger schatten van afstanden). Deze variabele is echter niet in het OVG opgenomen omdat zij voor het verzamelen van verplaatsingsgegevens door het CBS minder relevant werd geacht. Een indicatie van mogelijke invloed zou kunnen worden verkregen uit het analyseren van de variabele reismotief. Woon-school of woon-werk verplaatsingen zijn voorbeelden van verplaatsingen die veelvuldig worden gemaakt. Uit de analyses is gebleken

dat het effect van deze variabele vrijwel nihil is. Ten aanzien van het zgn. meetproces kan sprake zijn van het niet precies aan kunnen geven van begin- en eindpunt van een route, een niet 100% nauwkeurig getekende route en van niet 100% nauwkeurig nameten van de ingetekende route. Daarnaast is de keuze van 1%-niveau te rechtvaardigen vanwege de wijze waarop het analysebestand uiteindelijk is gecreëerd en de steekproef ontstaan.

Om de grootte van de systematische schattingsfout met enige betrouwbaarheid te kunnen vaststellen, zullen de klassen in principe niet te kleine aantallen waarnemingen mogen bevatten. Uitgaande van de bij de inspectie van de gegevens gevonden gemiddelde toevallige schattingsfout, kan voor deze aantallen een globaal criterium worden gegeven (zie Annex III bij Bijlage 2). De opgegeven afstanden voor de loopverplaatsingen moeten, zoals uit Tabel 8 blijkt, voor korte opgegeven afstanden met ruim de helft worden opgehoogd om tot de werkelijk afgelegde afstanden te komen. Voor de langste opgegeven afstanden moeten ze gemiddeld juist worden gehalveerd.

Naarmate de opgegeven afstand toeneemt gaat de aanvankelijke onderschatting van de lengte van de gemaakte rit over in een forse overschatting. Een dergelijk beeld is ook bij fietsritten aanwezig (Tabel 9).

Om na te gaan of de effecten van opgegeven afstand en dagtype bij de fietsritten onafhankelijk van elkaar mogen worden behandeld, is over deze variabelen opnieuw een variantie-analyse uitgevoerd. De variantie-analyse levert blijkens Tabel 2 van Bijlage 2 een significante interactie op, wat betekent dat de mate van systematisch misschatten voor werkdagen anders met de opgegeven afstand verband houdt dan voor weekeindedagen. De resultaten van Tabel 11 verduidelijken de aard van het geconstateerde effect. Bij ritten op werkdagen is bij de korte opgegeven afstanden sprake van een onderschatting van de werkelijk afgelegde afstanden van ongeveer 20%. De langere opgegeven afstanden laten een overschatting zien van gemiddeld 17% en 32%.

Voor ritten in het weekeinde geldt dat onderschattingen van werkelijk afgelegde afstanden niet in de klasse van de kortste afstanden begint, maar in de eerstvolgende klasse. De mate van overschatting bij de langste opgegeven afstanden is gemiddeld meer dan 75%. Hierbij moet worden gerea-

liseerd dat het aantal waarnemingen voor de ritten in het weekeinde zeer beperkt is (N = 26 in het analysebestand).

Tegelijk moet worden geconstateerd dat meer dan de helft van de totale opgegeven afstand in de hoogste afstandsklasse zit.

De variantie-analyse-uitkomsten voor de personenautoritten, die in ~~Tabel 2~~ Tabel 2 van Bijlage 2 zijn weergegeven, laten zien dat de opgegeven afstand en de regio geen significante interactie hebben op het verschil tussen de getransformeerde afstanden. Hun effecten kunnen dan ook onafhankelijk van elkaar worden beschouwd.

Tabel 10 laat zien dat de gemiddelde correcties van de opgegeven afstanden variëren van een onderschatting van 9% voor de kortste opgegeven afstanden tot een overschatting van bijna 20% voor de langere opgegeven afstanden. Tevens valt te constateren dat er nauwelijks enige lijn zit in de groottevolgorde van de correcties.

In regio Noord is de gemiddelde overschatting van de opgegeven afstand ca 16%, in regio Oost ca 7%, in regio Zuid ca 23% en in de regio West ca 15%.

Voor Tabel 12 zijn de hiervoor genoemde percentages zodanig herberekend dat de daar gegeven waarden in principe boven op de gemiddeld correcties voor de opgegeven afstand kunnen komen.

Ook hier moet worden geconstateerd dat meer dan de helft van de gesommeerde opgegeven afstand in de hoogste afstandsklasse valt.

6.8. De toevallige misschatting

Uit een hier niet gegeven vergelijking van de berekende spreidingen in de gecorrigeerde opgegeven ten opzichte van de bijbehorende nagemeten afstanden, lijkt een lichte tendens mogelijk dat identieke verplaatsingen met een geringere toevallige fout zouden worden geschat. Dit verschijnsel kan echter zijn oorzaak vinden in de eerder gesignaleerde tendens van kortere afstanden bij identieke ritten. Globaal neemt de spreiding immers met de afstand toe. Bij personenautoritten is echter in de hoogste klassen van opgegeven afstand de spreiding juist groter.

Ter illustratie van de grote mate van afwijkingen tussen opgegeven en nagemeten afstanden in het subonderzoek zij nog vermeld dat voor vijf

loopritten meer dan 1000 m als afgelegde afstand was opgegeven, terwijl er minder dan 50 m werd nagemeten; voor drie fietsritten met een opgegeven lengte boven 2,5 km geldt hetzelfde, evenals voor 24 personenautoritten waarvoor een lengte van meer dan 15 km werd opgegeven, voor 13 hiervan bedroeg de opgegeven afstand zelfs meer dan 25 km.

Deze grote afwijkingen kunnen ook het gevolg zijn van ponsfouten of codeerfouten bij het overbrengen van getallen op ponsstrips.

Vermeldenswaard is nog dat volgens de eerdere analyses de variabele geslacht geen significante invloed uitoefent op de mate van het mis-schatten.

Uit de cijfers van Tabel 13 lijkt het vervolgens hoogst onaannemelijk dat de schattingsvarianties voor vrouwen verschillen van die voor mannen. Anders dan uit literatuuronderzoek bleek (House & Waller, 1971), schatten mannen dus noch nauwkeuriger noch betrouwbaarder dan vrouwen in het onderhavige onderzoek.

6.9. Samenvatting van de resultaten

Voor de vier onderzochte vervoerswijzen: lopen, fiets, bromfiets en personenauto, geldt dat de toevallige schattingsfout in de opgegeven afstanden groot is in verhouding tot de systematische fout. Hieruit moet worden geconcludeerd dat alleen met cijfers betrouwbaar kan worden gewerkt die gebaseerd zijn op grote aantallen ritten. Dit is van belang voor de steekproefneming van een verplaatsingsonderzoek als het OVG en voor het uitgevoerde subonderzoek. Tevens is het van belang voor de mate waarin differentiatie van de verzamelde gegevens kan plaatsvinden. De bromfietsritten zijn vanwege het geringe aantal niet verder geanalyseerd.

Gemiddeld is er van een substantiële systematische overschatting sprake. De grootte van de systematische schattingsfout varieert daarbij met de opgegeven afstand. Deze variatie hangt bij fietsritten tevens met het dagtype samen. Hierbij hoort de aantekening dat het aantal fietsritten in het weekeinde waarop deze samenhang werd gevonden, zeer gering was.

Bij personenautoritten is er een regio-invloed.

Met mogelijke uitzondering van fietsritten in het weekeinde is in het algemeen sprake van het onderschatten van afstanden bij de korte opgegeven afstanden en van het overschatten van afstanden bij de langere opgegeven afstanden.

7. CONCLUSIES EN AANBEVELINGEN

7.1. Conclusies

Het subonderzoek Nauwkeurigheid opgegeven afstanden is het eerste systematische onderzoek naar het schatten van verplaatsingsafstanden in Nederland. Het is in grote lijnen identiek aan het OVG. De substeekproef is getrokken uit de steekproef van het OVG, waarbij het onderzoek is uitgevoerd door andere personen dan die het OVG uitvoeren. De resultaten hebben ondanks de beperkingen van het onderzoek een beduidend inzicht opgeleverd in de betrouwbaarheid en bruikbaarheid van de OVG-cijfers over verplaatsingsafstanden.

Kennis omtrent de systematische component in het misschatten van afgelegde afstanden is daarvan slechts één aspect. Minstens even belangrijk is het verworven beeld van de toevallige schattingsfout en van steekproef en uitvoering van het OVG.

De grootte van de systematische schattingsfout in de opgegeven afstanden is met name ook zichtbaar gemaakt in de vorm van berekende correcties. Met behulp van het in Annex III bij Bijlage 2 beschreven statistische model kan echter worden aangetoond dat deze correcties, gegeven de grote omvang van de toevallige schattingsfout en de relatief kleine aantallen waarnemingen, allerminst nauwkeurig uit de onderhavige steekproef zijn te bepalen, wat ook blijkt uit de grootten van de berekende spreidingen. Zij dienen dan ook eerder betrekkelijk globaal te worden verstaan.

Met de verdere kanttekeningen die er al bij werden geplaatst, is het de vraag of dergelijke correcties ook feitelijk in een OVG-bestand moeten worden doorgevoerd.

Bij gebruik en interpretatie van de cijfers zou wellicht reeds met richting en orde-van-grootte van de correcties voldoende rekening gehouden kunnen worden.

7.2. Aanbevelingen

De beperkingen en tekortkomingen van het uitgevoerde subonderzoek in acht nemend moet geconstateerd worden dat de verschillen tussen opgegeven en

nagemeten afstanden niettemin zo groot zijn dat er niet aan kan worden ontkomen om zich daarvan rekenschap te geven.

Voor veel gebruik van OVG-cijfers over afgelegde afstanden is een goede indruk van grootte en richting van de vertekening door mischatten wellicht voldoende. In deze zin is het hier gevonden onderzoekresultaat goed bruikbaar. Voor een nauwkeuriger vaststelling van de omvang van het mischatten moeten hogere eisen worden gesteld aan de opzet en inrichting van steekproef en steekproeftrekking alsmede aan de omvang van het aantal onafhankelijke verplaatsingen waarop de analyses gebaseerd worden.

De grote toevallige schattingsfout in het OVG legt beperkingen op aan de mate waarin het materiaal nog zinvol onder te verdelen is. Een grotere omvang van het OVG maakt in principe gedetailleerdere opsplitsingen mogelijk. Het is wenselijk de relatie tussen de omvang van het huidige OVG, de grootte van de toevallige schattingsfout en de gebruiksmogelijkheden van het OVG nader te onderzoeken.

LITERATUUR

CBS (1979). Onderzoek Verplaatsingsgedrag Deel 1: Onderzoekopzet. Nota H 427 79-E4. Centraal Bureau voor de Statistiek, Voorburg, 1979.

Harte, D.B. (1975). Estimates of the length of highway guidelines and spaces. Human factors 17(1975) 5(October): 455-460.

House, E.G. & Waller, P.F. (1971). Accuracy of driver's estimates of vehicle mileage driven. University of North Carolina, Chapel Hill, N.C., 1971.

SWOV (J.M.J. Bos) (1979). Analyse van nagemeten en opgegeven afstanden uit de proefenquête ROVIN; Een interim-verslag. SWOV, 1979.

SWOV (F.J. de Bruin) (1982a). Risico-onderzoek verkeersdeelnemers in Nederland (ROVIN) I; Eindrapport over het deelonderzoek naar de non-response in het CBS-onderzoek Verplaatsingsgedrag, + Tabellen en Bijlagen. R-82-30 I t/m III. SWOV, 1982.

SWOV (F.J. de Bruin) (1982b). Risico-onderzoek verkeersdeelnemers in Nederland (ROVIN) II; Eindrapport over het subonderzoek naar de invloed van zgn. vergeten verplaatsingen in het CBS-onderzoek Verplaatsingsgedrag. R-82-32. SWOV, 1982.

White, S.B. (1976). On the use of annual vehicle miles of travel estimates from vehicle owners. Accid. Anal. & Prev. 8 (1976): 257-261.

TABELLEN 1 T/M 13

Tabel 1. Sankey diagram ROVIN subonderzoek Nauwkeurigheid van opgegeven afstanden

Tabel 2. Aantal (netto) loopverplaatsingen in analysebestand en niet-in-analysebestand naar afstandklasse.

Tabel 3. Aantal (netto) fietsritten in analysebestand en niet-in-analysebestand naar afstandklasse.

Tabel 4. Aantal (netto) bromfietsritten in analysebestand en niet-in-analysebestand naar afstandklasse.

Tabel 5. Aantal (netto) personenautoritten in analysebestand en niet-in-analysebestand naar afstandklasse.

Tabel 6. Variabelen van het analysebestand: Identificatiekenmerken en klassen.

Tabel 7. Klassen van opgegeven afstanden in het analysebestand.

Tabel 8. Omvang systematische en toevallige misschattingen van afgelegde afstanden te voet.

Tabel 9. Omvang systematische en toevallige misschattingen van afgelegde afstanden per fiets.

Tabel 10. Omvang systematische en toevallige misschattingen van afgelegde afstanden per auto.

Tabel 11. Omvang van het dagtype-effect bij fietsverplaatsingen volgens opgegeven afstanden.

Tabel 12. Omvang van het regio-effect van autoverplaatsingen volgens opgegeven afstanden.

Tabel 13. Varianties van de gecorrigeerde opgegeven ten opzichte van de bijbehorende nagemeten afstanden voor mannen en vrouwen.

INGEBRACHT

AFGEVALLEN

GEANALISEERD

3567 verplaatsingen met gekoppelde
nagemeten afstanden
in onderzoekbestand

598 loopverplaatsingen van
403 personen uit 374
huishoudens.

416 fietsverplaatsingen van
301 personen uit 276
huishoudens.

827 autoverplaatsingen van
538 personen uit 448
huishoudens.

44 bromfietsverplaatsingen
van 30 personen uit
30 huishoudens.

124 verplaatsingen met motor, bus,
tram, trein of anders, vanwege
kleine aantallen niet verder
verwerkt.

38 verplaatsingen afgevallen door
onbekende codes (34) of zeer
grote afwijkingen van opgegeven
met nagemeten afstand (4).

1520 verplaatsingen afgevallen omdat
persoon, vervoermiddel, reismotief
en opgegeven en nagemeten
afstanden identiek zijn aan die van
reeds opgenomen verplaatsingen.

Tabel 1. Sankey Diagram ROVIN subonderzoek Nauwkeurigheid van opgegeven afstanden

Afstandklasse	Aantal (netto) loopverplaatsingen	
	In analysebestand	Niet-in-analysebestand
< 250 m	138	33
250 - < 500 m	79	23
500 - < 1.000 m	151	43
1.000 - < 2.500 m	161	46
2.500 - < 5.000 m	52	9
> 5.000 m	17	4
Totaal	598	158

Loopverplaatsingen x
 Afstandklasse $\chi^2 = 10,4$ df = 5 $\chi^2_{-1\%} = 15,1$

Tabel 2. Aantal (netto) loopverplaatsingen in analysebestand en niet-in-analysebestand naar afstandklasse.

Afstandklasse	Aantal (netto) fietsritten	
	In analysebestand	Niet-in-analysebestand
< 750 m	74	28
750 - < 1.500 m	73	32
1.500 - < 2.500 m	103	26
2.500 - < 5.000 m	97	35
5.000 - < 10.000 m	46	15
> 10.000 m	23	7
Totaal	416	143

Fietsritten x
 Afstandklasse $\chi^2 = 14,2$ df = 5 $\chi^2_{-1\%} = 14,1$

Tabel 3. Aantal (netto) fietsritten in analysebestand en niet-in-analysebestand naar afstandklasse.

Afstandklasse	Aantal (netto) bromfietsritten	
	In analysebestand	Niet-in-analysebestand
< 2.500 m	15	4
2.500 - < 5.000 m	11	5
> 5.000 m	18	10
Totaal	44	19

Bromfietsritten x
 Afstandklasse $\chi^2 = 3,7$ $df = 2$ $\chi^2_{-1\%} = 18,5$

Tabel 4. Aantal (netto) bromfietsritten in analysebestand en niet-in-analysebestand naar afstandklasse.

Afstandklasse	Aantal (netto) personenautoritten	
	In analysebestand	Niet-in-analysebestand
< 1.000 m	19	7
1.000 - < 2.500 m	171	60
2.500 - < 5.000 m	162	60
5.000 - < 7.500 m	120	37
7.500 - < 10.000 m	50	17
10.000 - < 25.000 m	93	29
15.000 - < 25.000 m	99	29
> 25.000 m	113	50
Totaal	827	289

Autoritten x
 Afstandklasse $\chi^2 = 10,6$ $df = 7$ $\chi^2_{-1\%} = 18,5$

Tabel 5. Aantal (netto) personenauto-ritten in analysebestand en niet-in-analysebestand naar afstandklasse.

I IDENTIFICATIEKENMERKEN		klassen
volgjaar		geen
volgmaand		geen
volgnummer		geen
regelnummer		geen
II VARIABELEN		Aantal klassen
1	Grootte huishouden	5 1 persoon, 3 en 4 personen, > 5 personen, rest/onbekend
2	Provincie (regio)	4 Noord, Oost, West en Zuid
3	Gemeentegrootte/ urbanisatieklasse	4 3 grootste steden, middel grote steden, kleine steden, platteland
4	Geslacht	2 man, vrouw
5	Burgerlijke staat	5 ongehuwd, gehuwd/samenwonend, gescheiden, weduwe/weduwenaar, onbekend
6	Bezit vervoermiddel	5 auto + andere, auto, fiets + andere, anders, geen
7	Bezitsduur rijbewijs A	5 <2 jaar, 2-5 jaar, 5-10 jaar, >10 jaar, geen
8	Bezigheid	5 werkzaam/militair/scholier, huisvrouw, geen werk, gepensioneerd, anders
9	Opleidingsniveau	4 hoog, middelbaar, laag, onbekend
10	Leeftijd(groep)	9 12-17, 18-24, 25-29, 30-34, 35-44, 45-54, 55-64, 65 en ouder
11	Dagtype	7 zondag, maandag, dinsdag, woensdag, donderdag, vrijdag, zaterdag
12	Dagsoort	3 werk/weekeinde, werkdag, weekeinde
13	Reismotief vertrek/ aankomst adres	4 werken/onderwijs, winkelen, visite/logeren, overig
14	Uurklasse vertrek/ aankomst	4 0-7/19-<0, 7-<10, 10-<16, 16-<19
15	Vervoermiddel	8 auto, fiets, lopen, brom- fiets, bus/tram/metro, trein, motor/scooter, overig
16	Opgegeven afstand	28 per vervoerwijze aantal klassen (zie Tabel 7)
17	Nagemeten afstand	nvt

Tabel 6. Variabelen van het analysebestand: Identificatiekenmerken en klassen.

LOPEN

1. < 250 m
2. 250 - < 500 m
3. 500 - < 1.000 m
4. 1.000 - < 2.500 m
5. 2.500 - < 5.000 m
6. \geq 5.000 m

BROMFIETS

1. < 2.500 m
2. 2.500 - < 5.000 m
3. \geq 5.000 m

FIETS

1. < 750 m
2. 750 - < 1.500 m
3. 1.500 - < 2.500 m
4. 2.500 - < 5.000 m
5. 5.000 - < 10.000 m
6. \geq 10.000 m

PERSONENAUTO

1. < 1.000 m
2. 2.000 - < 2.500 m
3. 2.500 - < 5.000 m
4. 5.000 - < 7.500 m
5. 7.500 - 10.000 m
6. 10.000 - < 15.000 m
7. 15.000 - < 25.000 m
8. \geq 25.000 m

TRAM/BUS/METRO

1. < 2.500 m
 2. 2.500 - < 5.000 m
 3. 5.000 - < 7.500 m
 4. 7.500 - < 15.000 m
 5. \geq 15.000 m
-

Tabel 7. Klassen van opgegeven afstanden in het analysebestand.

Vervoerswijze	Klasse opgegeven afstand					Totaal
	1	2	3	4	5	
<u>LOPEN</u>						
Gemiddelde opgegeven afstand (km)	0,140	0,331	0,558	1,276	3,695	0,894
Gemiddelde nagemeten afstand	0,215	0,340	0,431	0,913	1,651	
Aantal verplaatsingen	273	148	265	271	98	
Gemiddelde correctie voor de opgegeven afstand	+54%	+3%	-23%	-28%	-55%	-33%
Berekende spreiding* van gemiddelde correctie	+12,6	+7,8	+6,8	+7,7	+3,3	
Spreiding* gecorrigeerde opgegeven afstand	0,21	0,23	0,47	1,27	1,06	

* Spreidingen bepaald uit analysebestand (vgl. Aanhangsel bij Bijlage 2)

Tabel 3. Omvang systematische en toevallige misschattingen van afgelegde afstanden te voet.

Vervoerswijze	Klasse opgegeven afstand					Totaal
	1	2	3	4	5	
<u>FIETS</u>						
Gemiddelde opgegeven afstand (km)	0,483	1,004	1,839	3,341	10,112	3,063
Gemiddelde nagemeten afstand	0,570	0,873	1,409	2,250	7,302	
Aantal verplaatsingen	143	137	222	175	124	
Gemiddelde correctie voor de opgegeven afstand	+18%	-13%	-23%	-33%	-28%	-26%
Berekende spreiding* van gemiddelde correctie	+7,7	+6,8	+3,9	+3,2	+4,7	
Spreiding* gecorrigeerde opgegeven afstand	0,31	0,59	0,72	1,03	4,03	

* Spreidingen bepaald uit analysebestand (vgl. Aanhangsel bij Bijlage 2).

Tabel 9. Onvang systematische en toevallige misschattingen van afgelegde afstanden per fiets.

Vervoerswijze	Klasse opgegeven afstand							Totaal
	1	2	3	4	5	6	7	
<u>AUTO</u>								
Gemiddelde opgegeven afstand (km)	1,460	3,431	5,855	8,411	11,606	18,321	51,553	12,157
Gemiddelde nagemeten afstand	1,586	2,952	5,480	6,795	10,202	15,813	42,180	
Aantal verplaatsingen	352	283	208	80	159	162	176	
Gemiddelde correctie voor de opgegeven afstand	+9%	-14%	-6%	-19%	-12%	-14%	-18%	-15%
Berekende spreiding* van gemiddelde correctie	+5,5	+3,6	+2,9	+4,1	+2,7	+3,7	+2,1	
Spreiding* gecorrigeerde opgegeven afstand	1,07	1,53	1,86	2,43	3,01	6,82	12,07	

* Spreidingen bepaald uit analysebestand (vgl. Aangangsel bij Bijlage 2).

Tabel 10. Omvang systematische en toevallige misschattingen van afgelegde afstanden per auto.

<u>FIETS</u>	Klasse opgegeven afstand					Totaal
	1	2	3	4	5	
<u>Werkdagen</u>						
Gemiddelde opgegeven afstand	0,476	1,006	1,836	3,327	9,890	
Gemiddelde nagemeten afstand	0,572	0,838	1,425	2,263	7,461	
Aantal verplaatsingen	134	125	203	167	119	
Gemiddelde correctie voor de opgegeven afstand	+ 20%	- 17%	- 22%	- 32%	- 25%	- 24%
<u>Weekeinde</u>						
Gemiddelde opgegeven afstand	0,589	0,983	1,868	3,625	15,4	
Gemiddelde nagemeten afstand	0,533	1,242	1,237	1,975	3,5	
Aantal verplaatsingen	9	12	19	8	5	
Gemiddelde correctie voor de opgegeven afstand	- 9%	+ 26%	- 34%	- 46%	- 77%	- 52%

Tabel 11. Omvang van het dagtype-effect voor fietsverplaatsingen volgens opgegeven afstanden.

<u>AUTO</u>	Regio			
	Noord	Oost	Zuid	West
Gemiddelde opgegeven afstand	9,321	12,406	10,974	13,856
Gemiddelde nagemeten afstand	8,072	11,515	8,448	11,742
Aantal verplaatsingen	159	370	401	490
Gemiddelde bijkomende regio- correctie voor de opgegeven afstand	+ 1%	+ 8%	- 8%	0

Tabel 12. Omvang van het regio-effect van autoverplaatsingen volgens opgegeven afstanden.

Vervoerswijze		Klasse opgegeven afstand						
		1	2	3	4	5	6	7
<u>LOPEN</u>								
Mannen	VAR (K*opg)(KM ²)	0,03	0,07	0,21	1,00	1,72		
	Aantal verplaatsingen	61	32	66	62	33		
Vrouwen	VAR	0,06	0,04	0,21	2,01	0,58		
	Aantal verplaatsingen	77	47	85	99	36		
<u>FIETS</u>								
Mannen	VAR	0,11	0,60	0,57	1,08	19,07		
	Aantal verplaatsingen	18	25	45	55	37		
Vrouwen	VAR	0,10	0,22	0,48	1,05	7,05		
	Aantal verplaatsingen	56	48	58	42	32		
<u>AUTO</u>								
Mannen	VAR	1,70	1,82	3,62	8,79	8,40	56,92	129,93
	Aantal verplaatsingen	98	90	62	24	55	67	88
Vrouwen	VAR	0,53	3,35	3,37	2,64	9,94	25,40	183,89
	Aantal verplaatsingen	92	72	58	26	38	32	25

Tabel 13. Varianties van de gecorrigeerde opgegeven, ten opzichte van de bijbehorende nagemeten afstanden voor mannen en vrouwen.

BIJLAGEN 1 t/m 5

Bijlage 1.1 t/m 1.3 Verplaatsingsformulier SWOV-proefonderzoek 1976,
Idem, OVG 1978, Idem OVG 1979.

Bijlage 2. Nauwkeurigheid opgegeven verplaatsingsafstanden: Technisch
verslag van het subonderzoek behorende bij het deelonderzoek Validiteit
van het CBS-onderzoek Verplaatsingsgedrag. J.M.J. Bos. SWOV, 1983.

Voorbeeld 1

De heer A. vertrekt op woensdag 6 oktober 's morgens om 2 minuten over zeven van zijn woning, gelegen aan de Euterpestraat 17 te Spijkenisse, om naar de fabriek waar hij werkt te gaan. Hij loopt naar de bushalte in de Noordstraat (200 meter) en komt daar om 7 over zeven aan. Hierbij steekt hij driemaal de rijweg over. De bus vertrekt om 12 minuten over zeven naar het metrostation Hoogvliet (gemeente Rotterdam) en komt daar om 4 minuten over half acht aan. (Afstand 3200 meter, waarvan 2 km binnen bebouwde kom.) De afstand van de bushalte naar het metrostation is ca. 30 meter. De metro vertrekt om 12 over half acht van dit station naar het station Maashaven en komt om 3 minuten voor acht aan. (Afstand 12 km.) Van dit station loopt hij naar zijn fabriek aan de Brielse laan 218 in Rotterdam, aankomsttijd 7 over acht. (Loopafstand 800 meter, waarbij tweemaal de rijweg wordt overgestoken.)

Om 10 over vijf gaat hij naar huis. Hij rijdt mee in de auto van een collega, die ook nog een andere collega naar huis brengt. Zij rijden eerst naar het huis van die andere collega in de Wemeldingestraat 16 te Rotterdam, aankomsttijd 5 voor half zes. (Afstand 3 km, geheel binnen de bebouwde kom.) Nadat deze is uitgestapt rijden zij direct door naar Spijkenisse, Euterpestraat 17, waar zij om 5 over zes aankomen. (Afstand 12 km, waarvan 4 km binnen de bebouwde kom.) Daar de heer A. de volledige medewerking aan dit onderzoek had toegezegd, vroeg hij zijn collega de stand van de kilometerteller van de auto te willen opgeven. De stand bij het wegrijden bij de fabriek was 36.312, in de Wemeldingestraat 36.315 en in de Euterpestraat 36.327.

's Avonds om kwart over tien verlaat hij zijn woning om de hond uit te laten. Hij loopt via Siriusweg en Orionlaan naar het Apolloplein en vandaar via de Planetenweg terug naar de Euterpestraat. Om 5 over half elf is hij weer thuis. Totaal loopafstand 800 meter. In totaal moet hij hierbij 7 maal de rijweg oversteken.

INVULLEN → Het is vandaag: woensdag, 6 oktober 1976 (1) (2) (3) (4) (5) (6) (7) (8) Voorbeeld 1

N.B. LOOPAFSTANDEN VAN MINDER DAN 50 METER WORDEN NIET INGEVULD

Ver- plaat- sing no.	Vertrokken van : (plaats, straat en huisnummer of soort adres zoals halte of station)	Hoe laat ver- trokken	Gegaan naar : (plaats, straat en huisnummer of soort adres zoals halte of station)	Hoe laat daar aan- gekomen	Indien tijd niet langs open- bare weg	Omschrij- ving van dit be- stemmings- adres	Wat gaan doen op bestem- mings- adres ?	
00						ZIE CODE- LIJST A	ZIE CODE- LIJST B	
01	straat, nr.: Euterpestraat 17 plaats: Spijkenisse	07 uur 02 min.	straat, nr.: Bushalte Noordstraat plaats: Spijkenisse	07 uur 02 min.		code van lijst A 1 1	code van lijst B 6 6	
02	straat, nr.: Bushalte Noordstraat plaats: Spijkenisse	07 uur 12 min.	straat, nr.: Metrostation plaats: Hoogvliet	07 uur 34 min.		code van lijst A 1 1	code van lijst B 6 6	
03	straat, nr.: Metrostation plaats: Hoogvliet	07 uur 42 min.	straat, nr.: Metrostation Maashaven plaats: Rotterdam	07 uur 57 min.		code van lijst A 1 1	code van lijst B 6 6	
04	straat, nr.: Metrostation Maashaven plaats: Rotterdam	07 uur 57 min.	straat, nr.: Brielse laan 218 plaats: Rotterdam	08 uur 07 min.		code van lijst A 0 4	code van lijst B 2 1	
05	straat, nr.: Brielse laan 218 plaats: Rotterdam	17 uur 10 min.	straat, nr.: Wemeldingestraat 16 plaats: Rotterdam	17 uur 25 min.		code van lijst A 0 2	code van lijst B 6 6	
06	straat, nr.: Wemeldingestraat 16 plaats: Rotterdam	17 uur 25 min.	straat, nr.: Euterpestraat 17 plaats: Spijkenisse	18 uur 05 min.		code van lijst A 0 1	code van lijst B 2 0	
07	straat, nr.: Euterpestraat 17 plaats: Spijkenisse	22 uur 15 min.	straat, nr.: Apolloplein plaats: Spijkenisse	22 uur 25 min.		code van lijst A 1 3	code van lijst B 3 2	
08	straat, nr.: Apolloplein plaats: Spijkenisse	22 uur 25 min.	straat, nr.: Euterpestraat 17 plaats: Spijkenisse	22 uur 35 min.		code van lijst A 0 1	code van lijst B 2 0	
09	straat, nr.: plaats:	uur min.	straat, nr.: plaats:	uur min.		code van lijst A	code van lijst B	
10	straat, nr.: plaats:	uur min.	straat, nr.: plaats:	uur min.		code van lijst A	code van lijst B	

Opmerkingen :

- Het lopen van de bus naar de metro is niet opgenomen daar deze afstand minder dan 50 meter was.
- Het bestemmingsmotief van de verplaatsingen 1, 2 en 3 is steeds code 66 "rit met ander vervoermiddel vervolgen".
- De verplaatsingen 5 en 6 zijn afzonderlijk genoteerd, omdat het aantal inzittenden in de auto veranderde.
- Bij de metro behoeft geen afstand binnen de bebouwde kom te worden opgegeven.
- Bij het uitlaten van de hond moet de afgelegde afstand in twee delen worden gesplitst, nl. een verplaatsing naar het verste punt, in dit geval het Apolloplein, met als bestemmingsmotief 32 "hond uitlaten" en de weg terug naar huis, met als bestemmingsmotief 20 "wonen". De totale loopafstand wordt over deze beide verplaatsingen verdeeld.
- Denkt u er aan dat bij alle verplaatsingen 02, 03 enzovoorts, het vertrekkadres "Vertrokken van" altijd gelijk moet zijn aan het aankomstadres "Gegaan naar" van de voorafgaande verplaatsing.

Van welke vervoerwijze maakte u bij de verplaat- sing gebruik	INDIEN VERVOERMIDDEL MET KILOMETERTELLER		Hoeveel meter was de afstand voor deze verplaat- sing ?	En hoe- veel meter daarvan in de bebouw- de kom ?	INDIEN LOPEN					ALLEEN INDIEN MET AUTO, MOTOR, SNOR/BROMFIETS OF FIETS	
	Tellerstand bij :				Aantal malen de rijweg overgestoken					Was u de bestuurder of was u meerrijder ? (KRUISJE ZETTEN)	Hoeveel perso- nen (zelf inbegrepen) zaten erin of erop ?
	vertrek	aankomst			to- taal	K	L	M	N		
omschrijving: lopen code 4 0			200 meter	200 meter	3		1	2		bestuurder (8) <input type="checkbox"/>	<input type="checkbox"/>
omschrijving: bus code 5 2			3200 meter	2000 meter						bestuurder (6) <input type="checkbox"/>	<input type="checkbox"/>
omschrijving: metro code 5 5			12000 meter							bestuurder (8) <input type="checkbox"/>	<input type="checkbox"/>
omschrijving: lopen code 4 0			800 meter	600 meter	2	2				bestuurder (8) <input type="checkbox"/>	<input type="checkbox"/>
omschrijving: auto code 4 5	36.312	36.315	3000 meter	3000 meter						bestuurder (8) <input type="checkbox"/>	3 <input type="checkbox"/>
omschrijving: auto code 4 5	36.315	36.327	12000 meter	4000 meter						bestuurder (8) <input type="checkbox"/>	2 <input type="checkbox"/>
omschrijving: lopen code 4 0			400 meter	400 meter	4		2	2		bestuurder (8) <input type="checkbox"/>	<input type="checkbox"/>
omschrijving: lopen code 4 0			400 meter	400 meter	3		2	1		bestuurder (8) <input type="checkbox"/>	<input type="checkbox"/>
omschrijving: code										bestuurder (6) <input type="checkbox"/>	<input type="checkbox"/>
omschrijving: code										bestuurder (8) <input type="checkbox"/>	<input type="checkbox"/>

ALLEREERSTE

VERTREKPUNT

1 Plaats van vertrek:

- 1 van huis
 2 van een ander adres (graag zo volledig mogelijk)

Straat:

Gemeente:

2 Wat voor soort adres of vertrekpunt is dat?

- 1 Woning, pension
 2 Winkel (centrum), kapper, markt
 3 Kantoor, bank
 4 Fabriek, werkplaats
 5 Park, vrije natuur, strand
 6 Ander soort adres of vertrekpunt, nl.:

.....

3 Wat deed u op het vertrekpunt?

- 1 Wonen, thuis zijn
 2 Werken
 3 Visite, logeren
 4 Onderwijs volgen
 5 Winkelen, boodschappen doen
 6 Iets anders,

nl.:

VERPLAATSING

1 Datum van de verplaatsing 5 januari 19792 Tijdstip van vertrek: 7 uur — minuten

3 Waar ging u heen?

- 1 naar huis
 2 ergens anders heen (adres graag zo volledig mogelijk)

Straat: Muntstraat 10Gemeente: Brakelen 3 zomaar wat rond rijden, een blokje om. → naar vraag 6

4 Wat voor soort adres of punt van aankomst is dat?

- 1 Woning, pension
 2 Winkel (centrum), kapper, markt
 3 Kantoor, bank
 4 Fabriek, werkplaats
 5 Park, vrije natuur, strand
 6 Ander soort punt van aankomst

nl.:

5 Wat deed u op het punt van aankomst?

- 1 Wonen, thuis zijn
 2 Werken
 3 Visite, logeren
 4 Onderwijs volgen
 5 Winkelen, boodschappen doen
 6 Iets anders,

nl.:

6 Tijdstip van aankomst: 8 uur 10 minuten

7 Op welke manier maakte u deze verplaatsing?

(kruis alle vervoerswijzen aan die u gebruikte en vermeld de bijbehorende afstand)

Vervoerswijze	Eerst	Daarna	Dan	Dan	Afstand
1 Te voet			X		<u>0.5</u> km
2 Bestuurder van auto				 km
3 Passagier van auto				 km
4 Fiets	X				<u>3</u> km
5 Bromfiets				 km
6 Motor of scooter				 km
7 Trein		X			<u>8</u> km
8 Bus, tram, metro				 km
9 Taxi				 km
10 Overig nl.:				 km

totaal 11.5 km8 Indien u gebruik maakte van de auto:
Gingen er andere personen met u mee?

- 2 nee
 1 ja, personen

met een toepassing

BIJLAGE 2 bij

RISICO-ONDERZOEK VERKEERSDEELNEMERS IN NEDERLAND (ROVIN) III

NAUWKEURIGHEID OPGEGEVEN VERPLAATSINGSAFSTANDEN

Technisch verslag van het subonderzoek behorende bij het deelonderzoek
Validiteit van het CBS-onderzoek Verplaatsingsgedrag

J.M.J. Bos

Leidschendam, 1983

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV

1. ALGEMEEN

Teneinde bij te dragen tot het verwerven van inzicht in de kwaliteit en de bruikbaarheid van OVG-gegevens zijn systematische en toevallige fouten in opgegeven schattingen van afgelegde afstanden nagegaan, alsmede de factoren die op de grootte van deze fouten van invloed zijn en van de betekenis hiervan voor het gebruik van OVG-gegevens.

Het materiaal betreft de substeekproef huishoudens uit het OVG november-december 1978 met per huishouden één persoon (of twee personen als het een huishouden in een vervangende gemeente betrof), en van deze perso-(o)n(en) de verplaatsingen over de laatste van twee (of drie als het startgesprek voor de enquête op vrijdag viel) enquêtedagen. Er zijn dus geen zondagse verplaatsingen bij, de maandagse komen van de derde enquêtedag. Van deze verplaatsingen zijn de afstanden achteraf aan de hand van op plattegronden ingetekende route-informatie nagemeten, standaardmeetfout daarbij ca. 3%.

De gegevens zijn niet representatief, noch voor algemeen verplaatsingsgedrag (seizoen-, dagsoort-, leeftijd- en korte en lange ritten-problematieken), noch binnen OVG (nameetproblematiek van diverse soorten kaartmateriaal, afvallende gemeenten, meer personen per gezin, de keuze van personen binnen een huishouden, de impliciete huishoudens-grootte-selectie, de uitvoering van het OVG in de onderzoeksperiode door de NSS in plaats van het CBS, de wijzigingen sindsdien in het OVG-vragenformulier, de koppelingsproblematiek van OVG-gegevens en nagemeten afstanden)(vgl. ook nonresponseproblematiek).

Omdat relatief veel verplaatsingen en afstandsschattingen van dezelfde personen afkomstig zijn, zijn verplaatsingen sterker onderling afhankelijk dan in de bevolking. Teneinde error-correlaties te verkleinen zijn identieke verplaatsingen uit het analysebestand verwijderd.

2. ANALYSES

De standaard nameetfout is ondergeschikt aan de schattingsfout. Het verschil tussen de opgegeven en de als werkelijk afgelegd beschouwde, nagemeten afstand geldt als schattingsfout. Deze schattingsfout heeft een systematische en een toevallige component. De grootte van de toevallige component is afhankelijk van de werkelijk afgelegde afstand.

In een proefonderzoek in 1976 is gevonden dat variantiestabilisatie (homoscedasticiteit) hier redelijk wordt bereikt door de worteltransformatie van afstanden. Als afgeleid (zie Annex I) hoort deze transformatie bij een schattingsproces waarvan de error-component evenredig is met die bij een Poissonproces. Variantiestabilisatie wordt vereist bij gebruik van beoogde technieken van (co)variantie-analyse.

Na de voorbereidingsfase waarin gegevens zijn geselecteerd en variabelen geconstrueerd, getransformeerd, gehercodeerd en geklasseerd, bestaat de eigenlijke analyse van verschillen tussen opgegeven en nagemeten verplaatsingsafstanden uit vier fasen:

1. een algemene data-inspectie;
2. een CANALS-analyse om de belangrijkste variabelen op te sporen;
3. (co)variantie-analyses op variabelen die het belangrijkste zijn gebleken, en
4. een fase waarin, nu voor de totale nameetsteekproef, de grootten en hun spreidingen geschat worden van significant bevonden systematische effecten.

Vanwege het onderscheid in hun normale actieradius doorlopen de vervoerswijzen de vier analysefasen apart.

Uitvoeringsfase 1

De opgegeven afstanden worden tegen de nagemeten geplot. Tevens vindt een lineaire-regressie-analyse plaats. Een enkele extreme opgegeven of nagemeten afstand blijft daarbij om beeldvertekening te voorkomen buiten beschouwing. Vervolgens worden de verschillen tussen de wortelgetransformeerde opgegeven en nagemeten afstanden geplot tegen de nagemeten afstand en vindt ook op deze datapunten een lineaire-regressie-analyse plaats.

Uitvoeringsfase 2

Met 44 blijkt het aantal bromfietsverplaatsingen van het onderzoekbestand te gering voor voortgezette analyse. De vervoerswijzen lopen, fiets en auto blijven over.

Op deze datasets worden CANALS-analyses uitgevoerd. Van alle variabelen heeft de opgegeven afstand verreweg de sterkste samenhang met het verschil tussen de wortelgetransformeerde opgegeven en nagemeten afstanden. In een nieuwe serie CANALS-analyses is deze variabele dan ook "uitgepar-tialiseerd".

Uitvoeringsfase 3

De "optimale" herschaling van nominale en ordinale variabelen vormt een essentieel onderdeel van de CANALS-routine, waarna de aanpak in principe die is van een multi-pele-regressie-analyse zonder interactietermen. Dit leidt tot een goede verwantschap met "klassieke" variantie-analyse.

Toetsingscriteria voor de effecten ontbreken echter omdat homoscedastici-teit niet wordt verondersteld. De onderhavige data zouden door de toe-gepaste worteltransformatie wel aan die voorwaarde voldoen. Over de be-langrijkst gebleken variabelen kan dan ook een (co)variantie-analyse wor-den gedaan. SPSS beperkt het aantal in te brengen variabelen tot vijf, waarbij de opgegeven afstand nog als "covariaat" kan meelopen.

Tussen de uitkomsten van de CANALS- en (co)variantie-analyses zijn wat verschillen mogelijk als gevolg van het feit dat de variantie-analyse-techniek gevoeliger is voor het "unbalanced" en "incomplete" design van het onderhavige soort datamatrices.

Uitvoeringsfase 4

Tenslotte worden in het totale steekproefbestand voor de klassen van significant bevonden variabelen de sommen van opgegeven en nagemeten afstanden bepaald en met elkaar vergeleken. Van de aldus geschatte grootten van significante effecten wordt uit het analysebestand de spreiding geschat.

3. RESULTATEN

3.1. Analysefase 1

De resultaten voor deze fase zijn afgebeeld in zogenaamde plots en weergegeven als Afbeelding 1 t/m 4.

Loopverplaatsingen

Afbeelding 1A laat zien dat de afgelegde afstanden doorgaans worden overschat en dat de spreiding in het schatten daarbij aanzienlijk is. De systematische overschatting bedraagt blijkens de lineaire-regressie-analyse gemiddeld ongeveer 300 m. Zij neemt toe met de grootte van de afgelegde afstand, wat kan worden afgeleid uit het feit dat de richtingscoëfficiënt van de regressielijn groter is dan 1. Bij een standaard schattingsfout van 700 m en een R^2 van 0,47 is de relevantie van het lineaire model bij het voorspellen van individuele schattingen echter maar matig.

Dat de toegepaste worteltransformatie inderdaad tot de redelijke stabilisatie van de variantie leidt toont Afbeelding 1B. Uit Annex II blijkt overigens dat de toegepaste worteltransformatie, afkomstig van het proefonderzoek, op inhoudelijke gronden niet onaannemelijk en visueel aanvaardbaar, er toch niet geheel in slaagt om homoscedasticiteit te bereiken, al worden de verschillen tussen de varianties er wel wezenlijk door verminderd. Daarbij moet ook het steekproefkarakter van resultaten in het oog worden gehouden, wat betekent dat het analyseresultaat niet erg onder druk zal staan.

Fietsverplaatsingen

Afbeeldingen 2A en 2B vertonen een beeld dat niet essentieel afwijkt van dat van loopverplaatsingen.

De grotere actieradius van de fiets brengt met zich mee dat het lineaire model voor de relatie tussen opgegeven en nagemeten afstanden, gezien een R^2 van 0,71, wat relevanter wordt. Daar staat echter de tot 1.500 m toegenomen standaard schattingsfout tegenover. De systematische overschatting van de afgelegde afstanden is toegenomen tot 600 m gemiddeld.

Bromfietsverplaatsingen

In Afbeeldingen 3A en 3B lijken vier enquêtewaarnemingen met nagemeten afstanden van rond de 10 km nogal uitzonderlijk. Gegeven het overigens geringe aantal datapunten dragen deze waarnemingen stevig bij tot het resultaat van de regressie-analyse, dat daardoor in details lijkt te kunnen afwijken van het beeld tot dusver. Verdere interpretatie blijft derhalve achterwege.

Personenautoverplaatsingen

De nog weer grotere actieradius van de personenauto zorgt ervoor, zoals Afbeelding 4A onmiddellijk duidelijk maakt dat, hoewel de standaard schattingsfout tot 5.000 m is opgelopen, de relevantie van het lineaire model, met een R^2 van 0,89, verder is toegenomen. De gemiddelde grootte van de systematische overschatting van afgelegde afstanden is gestegen tot 1.500 m.

Ook Afbeelding 4B en de desbetreffende regressie-analyse-uitkomsten passen goed in het ontstane totale plaatje.

De aparte behandeling van de vervoerswijzen is hiermee terecht gebleken.

3.2. Analysefase 2

Loopverplaatsingen

De opgegeven afstand correleert overduidelijk het sterkst met het verschil tussen de wortelgetransformeerde opgegeven en nagemeten afstanden. Door deze variabele tevens als "afhankelijke" variabele toe te voegen aan de criteriumvariabele ontstaat een effect van "uitpartialiseren". Uit de tweede dimensie van de CANALS-analyse-oplossing is de invloed van de opgegeven afstand op de criteriumvariabele verdwenen. De volgorde waarin de overige variabelen van belang zijn, kan hierbij veranderen.

De overwegende indruk is dat de onderscheidene correlaties met de criteriumvariabele nogal gering zijn, iets wat in Afbeelding 5 wordt gevisualiseerd.

De variabelen: geslacht, reismotief, aanwezigheid vervoermiddel, leeftijd(groep) en bezigheid zouden nog de meeste betekenis hebben voor een systematische component aan het misschatten van afgelegde afstanden.

Fietsverplaatsingen

Het beeld van de analyse-uitkomsten is in grote trekken conform dat voor loopverplaatsingen, zoals ook Afbeelding 6 laat zien.

Echter zijn nu: dagtype, burgerlijke staat, opnieuw leeftijd(groep), bezitsduur rijbewijs en opleidingsniveau de belangrijkste variabelen.

Autoverplaatsingen

Opnieuw in beginsel eenzelfde beeld, zoals ook Afbeelding 7 verduidelijkt.

Belangrijkste variabelen zijn nu: regio, opnieuw bezigheid, opnieuw opleidingsniveau, opnieuw leeftijd(groep) en opnieuw burgerlijke staat.

3.3. Analysefase 3

Loopverplaatsingen

Van de variabelen die als belangrijkste uit de CANALS-analyses tevoorschijn kwamen, blijkt alleen de opgegeven afstand in de uitgevoerde (co)variantie-analyse een significant effect te hebben op de systematische component van de schattingsfout.

Fietsverplaatsingen

Hier zijn zowel de opgegeven afstand als het dagtype significant van invloed op de grootte van het misschatten.

Autoverplaatsingen

De opgegeven afstand, en in mindere mate de regio, hebben significant effect op de waarde van de criteriumvariabele.

Daarmee zijn de resultaten van de variantie-analyses geheel conform die van de CANALS-analyses.

3.4. Analysefase 4

Is in de variantie-analyses een variabele van invloed gebleken, dan zijn er significante verschillen tussen de klassen van die variabele onderling naar de mate waarin er van systematisch misschatten sprake is. Om de grootte van de systematische schattingsfout met enige betrouwbaarheid te kunnen vaststellen, zullen de klassen in principe niet te kleine aantallen waarnemingen mogen bevatten.

Uitgaande van de in analysefase I gevonden gemiddelde toevallige schattingsfout kan voor deze aantallen een globaal criterium worden gegeven (zie Annex III).

De verdere analyse heeft plaatsgevonden op het totale steekproefbestand, zij het dat de spreidingen bekend werden over het analysebestand, waarin als bekend de afhankelijkheden geringer zijn (zie Aangangsel).

Loopverplaatsingen

De opgegeven afstanden moeten voor korte opgegeven afstanden met ruim de helft worden opgehoogd om tot de werkelijk afgelegde afstanden te komen. Voor de langste opgegeven afstanden moeten ze gemiddeld juist worden gehalveerd.

Aanvankelijke onderschatting van de lengte van de gemaakte verplaatsing gaat geleidelijk over in forse overschatting naarmate de opgegeven afstand toeneemt.

De berekende steekproefspreadingen zijn niet zo groot dat zij dit beeld redelijkerwijze kunnen verstoren.

De omvang van de toevallige schattingsfout blijkt duidelijk uit het feit dat de spreiding van de gecorrigeerde opgegeven afstanden in dezelfde orde van grootte ligt als de opgegeven afstanden zelf.

Fietsverplaatsingen

Om na te gaan of de effecten van opgegeven afstand en dagtype onafhankelijk van elkaar mogen worden behandeld, is over deze variabelen opnieuw een variantie-analyse uitgevoerd. Deze levert een significante interactie op, wat betekent dat de mate van systematisch misschatten voor werkdagen anders met de opgegeven afstand verband houdt dan voor weekeinddagen.

Korte opgegeven afstanden van werkdagverplaatsingen onderbepalen de werkelijk afgelegde afstanden gemiddeld met 20%. Langere opgegeven afstanden zijn overschattingen van een gemiddelde omvang tussen 17 en 32%.

Daarentegen zouden voor weekeindverplaatsingen de onderschattingen niet de klasse van kortste opgegeven afstanden gelden, maar de eerstvolgende klasse. Bovendien zouden de langste opgegeven afstanden gemiddeld zelfs tot tegen de 80% te hoog zitten.

Ofschoon zowel het dagtype als zijn interactie met de opgegeven afstand significant effect hebben op de mate van het misschatten, moet worden geconstateerd dat de weekeindcorrecties slechts op grond van erg kleine

aantallen waarnemingen zijn geschat. Los van het interactie-effect beschouwd gaat het, tegenover een totaal effect van -24% voor de werkdagverplaatsingen, bij de weekeindverplaatsingen nochtans om 53 (waarvan slechts 26 unieke) waarnemingen en een totale correctie van -52%. Tegelijk moet echter ook worden vastgesteld dat meer dan de helft van de totale opgegeven afstand in de hoogste klasse zit.

Autoverplaatsingen

De variantie-analyse-uitkomsten laten zien dat de opgegeven afstand en de regio geen significante interactie hebben op de criteriumvariabele. Hun effecten kunnen dan ook onafhankelijk van elkaar worden beschouwd. De gemiddelde correcties van de opgegeven afstanden variëren van +9% voor de kortste tot -6 en -19% voor de langere verplaatsingen, overigens zonder dat er lijn zit in hun grootte-volgorde.

In regio Noord is de gemiddelde correctie voor de opgegeven afstand -6%, in regio Oost -7%, in regio Zuid -23% en in regio West -15%. Deze correcties zijn zodanig herberekend dat de gegeven waarden in principe boven op de correcties voor de opgegeven afstand kunnen komen.

Ook hier moet worden geconstateerd dat meer dan de helft van de gesommeerde opgegeven afstand in de hoogste klasse valt.

Vooraf door de grotere aantallen waarnemingen en de grote verplaatsingsafstanden vallen de berekende steekproefspreidingen ten opzichte van die bij loopverplaatsingen nog weer minder ongunstig uit dan die van fietsverplaatsingen. Ook de spreiding van de gecorrigeerde opgegeven afstanden blijft verder achter bij de grootte van de opgegeven afstanden zelf.

4. INTERPRETATIE

Voor alle onderzochte verplaatsingswijzen geldt dat de toevallige schattingsfout in de opgegeven afgelegde afstanden groot is in verhouding tot de systematische fout. Uit dit feit moet worden begrepen dat met cijfers die niet op zeer behoorlijke aantallen waarnemingen zijn gebaseerd niet betrouwbaar kan worden gewerkt.

Enerzijds heeft dit betekenis voor de steekproefomvang van een verplaatsingsonderzoek, anderzijds voor de mate waarin differentiatie van het verzamelde materiaal kan plaatsvinden.

Gemiddeld is er van een substantiële systematische overschatting sprake. De grootte van de systematische schattingsfout varieert daarbij met de opgegeven afstand. Deze variatie hangt bij fietsverplaatsingen tevens met het dagtype samen. Bij autoverplaatsingen is er enige regio-invloed. De werkelijke verplaatsingslengten zullen de geënquêteerden doorgaans onbekend zijn geweest. De eigen schattingen treden dan kennelijk als referentie op, zodat niet de nagemeten maar de opgegeven afstand gaat samenhangen met de mate van het misschatten.

In het algemeen zullen de verplaatsingsduur en de schaal, de richtingborden en een kilometerteller een referentie voor de afgelegde afstand kunnen opleveren.

Het is denkbaar dat in het bijzonder bij de fietsverplaatsingen van het weekeinde, die misschien wat vaker over minder bekende routes voeren, dit mechanisme wat anders zou uitwerken. Juist hier echter zijn de bevindingen van de analyse gebaseerd op uitgesproken kleine aantallen waarnemingen.

Een samenhang met reismotief zou bovendien niet minder aannemelijk geweest hoeven zijn. Reismotief blijkt echter volgens de CANALS-analyses zowel voor fiets- als voor autoverplaatsingen duidelijker met de opgegeven afstand te correleren, waarmee zijn effect dan ook impliciet kan zijn ingebracht.

Ten aanzien van autoverplaatsingen geldt iets dergelijks voor het effect van regio en urbanisatiegraad.

Algemeen, echter met een onregelmatigheid voor fietsverplaatsingen in het weekeinde, zijn korte opgegeven afstanden, gemiddeld genomen, onderschat-

tingen en langere afstanden overschattingen van de werkelijke verplaatsingslengten. Terwijl een ingetekende route korter dan 0,05 km resulteert in een nagemeten afstand van 0,0 km, is de kortste opgegeven afstand in het onderzoekbestand 0,1 km. Zitten verplaatsingen met een afgeronde opgegeven afstand van 0,0 km wél in het verplaatsingsbestand, dan zal er voor korte opgegeven afstanden niettemin een onderschattingseffect blijven.

Het afrondingsverschijnsel zelf zou zijn grootste invloed kunnen hebben voor loopverplaatsingen in de laagste klasse van opgegeven afstand. Bij een gemiddelde nagemeten afstand van 0,215 km en een standaard meetfout van 0,006 km (3%) is een afrondingsfout van 0,05 km niet gering. Toch is er geen reden om te veronderstellen dat deze fout bij redelijke aantallen waarnemingen niet voldoende zou "uitmiddelen". Tenzij de verdeling van verplaatsingslengten minder net zou zijn. Komen er bijvoorbeeld onevenredig veel heel korte verplaatsingen voor, in het analysebestand zijn het er evenwel minder dan 20 op de 138, dan is de grootte van het onderschattingseffect onderbepaald, hier gaat het echter hooguit om de bijtelling van nog eens +3% bij een correctie van +54%.

De onder- en overschattingen zijn in Afbeeldingen 8, 9 en 10 ingetekend. Daarbij valt de regelmatige liggen van de punten voor de loopverplaatsingen op. Eerst de onderschatting van de afgelegde afstand, daarna geleidelijk de overgang naar steeds grotere overschattingen, bijna te gaaf voor een steekproefuitkomst.

Voor de fietsverplaatsingen is eenzelfde verloop herkenbaar, maar er lijkt een oneffenheid mogelijk, die het duidelijkst tot uiting komt in Afbeelding 9B, van de gemiddelde correcties voor opgegeven afstanden. Tussen afstanden van 0,5 en 1 km slaat de onderschatting om in overschatting, die eerst geleidelijk toeneemt, maar voor de hoogste klasse weer vermindert. Voorstelbaar is dat bij afstanden boven de 5 km andere afstandsreferentie zouden optreden. Uit de grootten van de berekende spreidingen blijkt echter dat het hier evengoed kan gaan om toevallige steekproefverschillen.

Het beeld voor de fietsverplaatsingen van het weekeinde vertoont enkele merkwaardigheden, die voor het gevonden dagtype-effect verantwoordelijk zullen zijn. Allereerst zijn de correcties in de laagste twee klassen van opgegeven afstand tegengesteld aan die voor werkdagen, en verder is de

correctie bij langere afstanden wel zéér groot. Gewezen is echter al op de uiterst kleine substeekproef waar het hier om gaat. In de hoogste klasse zitten slechts vijf waarnemingen en dat zijn geweldige overschattingen. Uiteraard komen dit soort overschattingen in een verplaatsingsonderzoek steevast voor. De indruk bestaat evenwel dat zij hier bij toeval teveel accent zullen hebben gekregen en daardoor de beeldvorming ten onrechte overheersen, te meer waar voor de substeekproef in de andere klassen samen nog niet de helft van de totale opgegeven afstand valt. Merkwaardig is bovendien dat er zelfs meer zondagse dan zaterdagse verplaatsingen voorkomen, terwijl ze juist dienden te mankeren. Voor zover er voor fietsverplaatsingen toch van een dagtype-invloed sprake zou zijn, lijkt deze het aannemelijkst te kunnen bestaan in een wat grotere mate van overschatting van in het weekeinde afgelegde afstanden.

Voor personenautoverplaatsingen is het verloop in de ligging van de punten weliswaar herkenbaar, maar grillig genoeg om vragen op te werpen. Toch blijkt uit de grootten van de berekende spreidingen dat deze grilligheid aan steekproeftoeval te wijten kan zijn. Alsdan is zij een goede illustratie van de betrouwbaarheid van dergelijke steekproefuitkomsten. Als bij de fietsverplaatsingen geeft het ook hier te denken dat meer dan de helft van de totale afstand in de hoogste klasse zit. De mate waarin langere afstanden worden overschat heeft derhalve, gemeten naar de aantallen gemaakte verplaatsingen, onevenredig grote invloed op het totale resultaat.

Een vraag naar de veelvuldigheid waarmee steekproefpersonen bepaalde verplaatsingen maakten, waarvan te veronderstellen is dat deze nauwkeuriger kunnen worden geschat, ontbrak in deze enquête. Deze variabele is dan ook niet toegankelijk voor analyse, al komen in het onderzoekbestand wel identieke verplaatsingen voor van personen op hun toevallige enquêtedag. Wellicht zou een eventueel belang van deze variabele met betrekking tot de problematiek van het misschatten van afgelegde afstanden het eerst zijn gebleken in de betekenis van een variabele als: reismotief.

Tussen de samenstellingen van analyse-, onderzoek- en OVG-bestanden bestaan de nodige verschillen. Met name zal de huishoudensgrootte-verdeling van het analysebestand nogal kunnen afwijken, in die zin dat vanwege de gerealiseerde huishoudenssteekproef verplaatsingen van personen uit grotere huishoudens ondervertegenwoordigd zijn.

Voor variabelen die bij de analyse niet van invloed zijn gebleken op de mate van het misschatten zijn dergelijke verschillen, mits niet al te extreem, echter in principe niet relevant.

Bij een variabele die wèl invloed heeft is de verdeling van het aantal waarnemingen over zijn klassen bepalend voor de grootte van het totale effect. Vandaar dat per klasse een correctie is gegeven.

Van variabelen als seizoen, die niet in het analysebestand voorkomen, kan uiteraard de betekenis voor het misschatten niet worden vastgesteld. Dat er nog erg bruikbare en wezenlijke tussen zouden zitten lijkt wat twijfelachtig.

5. SLOTBESCHOUWING

Resumerend moet worden vastgesteld dat het subonderzoek Nauwkeurigheid opgegeven afstanden, ondanks een aantal tekortkomingen, een beduidend inzicht heeft opgeleverd in de betrouwbaarheid en bruikbaarheid van de OVG-cijfers over verplaatsingsafstanden.

Kennis omtrent de systematische component in het misschatten van afgelegde afstanden is daarvan slechts één aspect.

Minstens even belangrijk is het verworven beeld van de toevallige schattingsfout en van steekproef en uitvoering van het OVG.

De grootte van de systematische schattingsfout in de opgegeven afstanden is met name ook zichtbaar gemaakt in de vorm van berekende correcties.

Met behulp van het in Annex III beschreven statistische model kan echter worden aangetoond dat deze correcties, gegeven de grote omvang van de toevallige schattingsfout en de relatief kleine aantallen waarnemingen, allerminst nauwkeurig uit de onderhavige steekproef zijn te bepalen.

Zij dienen dan ook eerder betrekkelijk globaal te worden verstaan.

Met de verdere kanttekeningen die er al bij werden geplaatst, lijkt het derhalve gewenst voorzichtigheid te betrachten bij het feitelijke doorvoeren van dergelijke correcties in een OVG-bestand.

Wel is het uiteraard van belang bij gebruik en interpretatie van de cijfers met richting en orde-van-grootte van de correcties ernstig rekening te houden.

Uit de grote betekenis van de toevallige fout moet bovendien worden geconcludeerd dat een grote OVG-steekproef voor enigszins betrouwbaar resultaat noodzakelijk is.

AFBEELDINGEN 1 T/M 10

Afbeelding 1A. Scatterdiagram opgegeven tegen nagemeten loopafstanden.

Afbeelding 1B. Scatterdiagram verschil tussen opgegeven en nagemeten afstanden na worteltransformatie tegen nagemeten loopafstanden.

Afbeelding 2A. Scatterdiagram opgegeven tegen nagemeten fietsafstanden.

Afbeelding 2B. Scatterdiagram verschil tussen opgegeven en nagemeten afstanden na worteltransformatie tegen nagemeten fietsafstanden.

Afbeelding 3A. Scatterdiagram opgegeven tegen nagemeten bromfietsafstanden.

Afbeelding 3B. Scatterdiagram verschil tussen opgegeven en nagemeten afstanden na worteltransformatie tegen nagemeten bromfietsafstanden.

Afbeelding 4A. Scatterdiagram opgegeven tegen nagemeten personenautoafstanden.

Afbeelding 4B. Scatterdiagram verschil tussen opgegeven en nagemeten afstanden na worteltransformatie tegen nagemeten personenautoafstanden.

Afbeelding 5. Resultaten CANALS-analyse loopverplaatsingen.

Afbeelding 6. Resultaten CANALS-analyse fietsverplaatsingen.

Afbeelding 7. Resultaten CANALS-analyse personenautoverplaatsingen.

Afbeelding 8A. Relatie gemiddelde opgegeven en gemiddelde nagemeten loopafstanden.

Afbeelding 8B. Gemiddelde percentage correctie bij gemiddelde opgegeven loopafstanden.

Afbeelding 9A. Relatie gemiddelde opgegeven en gemiddelde nagemeten fietsafstanden.

Afbeelding 9B. Gemiddelde percentage correctie bij gemiddelde opgegeven fietsafstanden.

Afbeelding 10A. Relatie gemiddelde opgegeven en gemiddelde nagemeten personenauto-afstanden.

Afbeelding 10B. Gemiddelde percentage correctie bij gemiddelde opgegeven personenauto-afstanden.

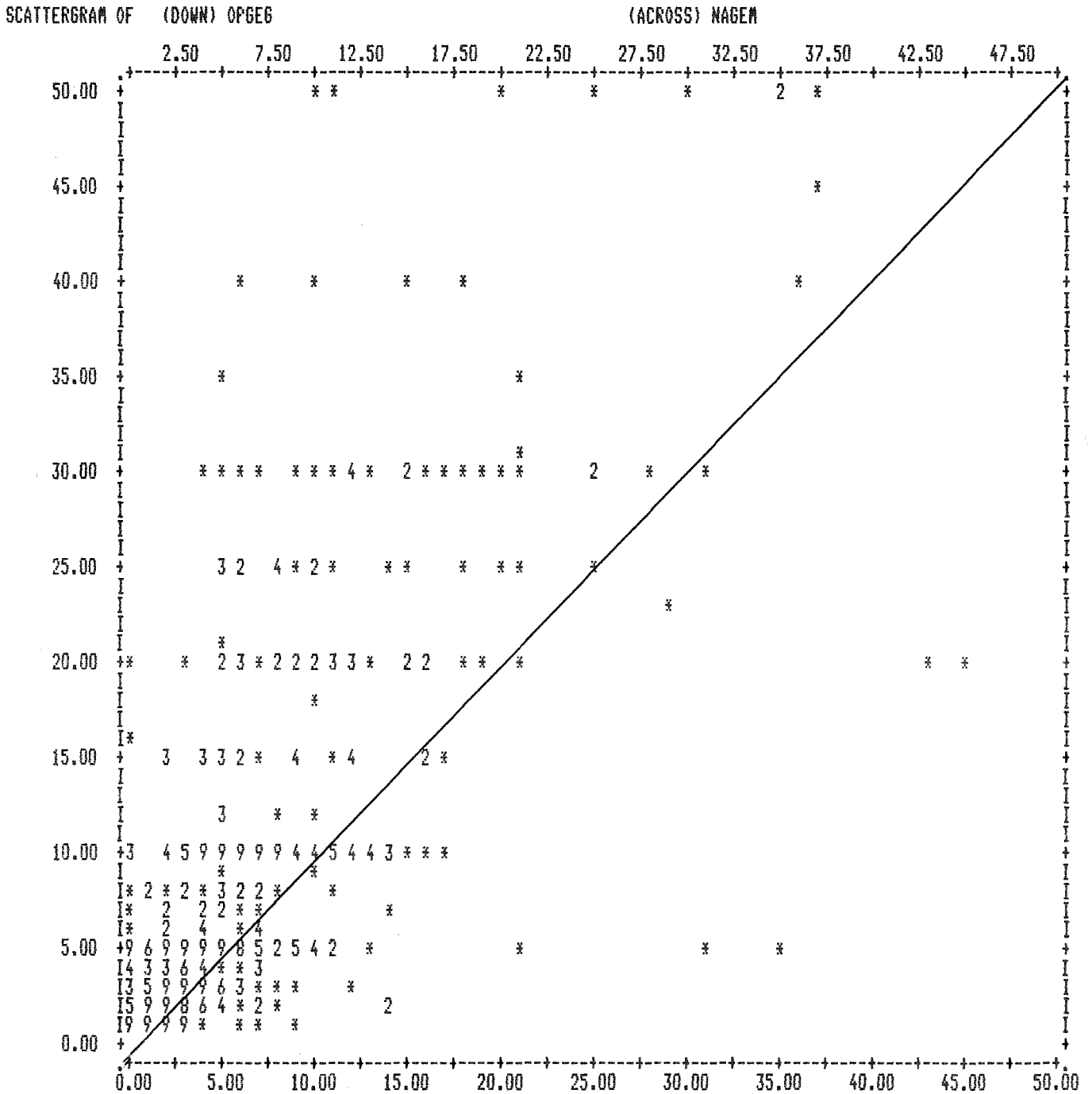

```

SPSS-11
RUN NAME      ROVIN-NAMET 78  SCATRGRAM LOPEN  OPGE-NAGEM 2-83
VARIABLE LIST OPGEG,NAGEM,VVSQRT
INPUT MEDIUM  DR1:RECR1.DAT
N OF CASES    598
INPUT FORMAT  REAL(19X,F4.0,1X,F4.0,16X,F9.4)
              3 variable name(s) actually read
READ INPUT DATA
              598 observations have been read with 7 variables (including the 4 system variables) in each observation
SCATTERGRAM   OPGEG(0,50) WITH NAGEM(0,50)
OPTIONS       4
STATISTICS    ALL

```

ROVIN-NAMET 78 SCATRGRAM LOPEN OPGE-NAGEM 2-83

02-81



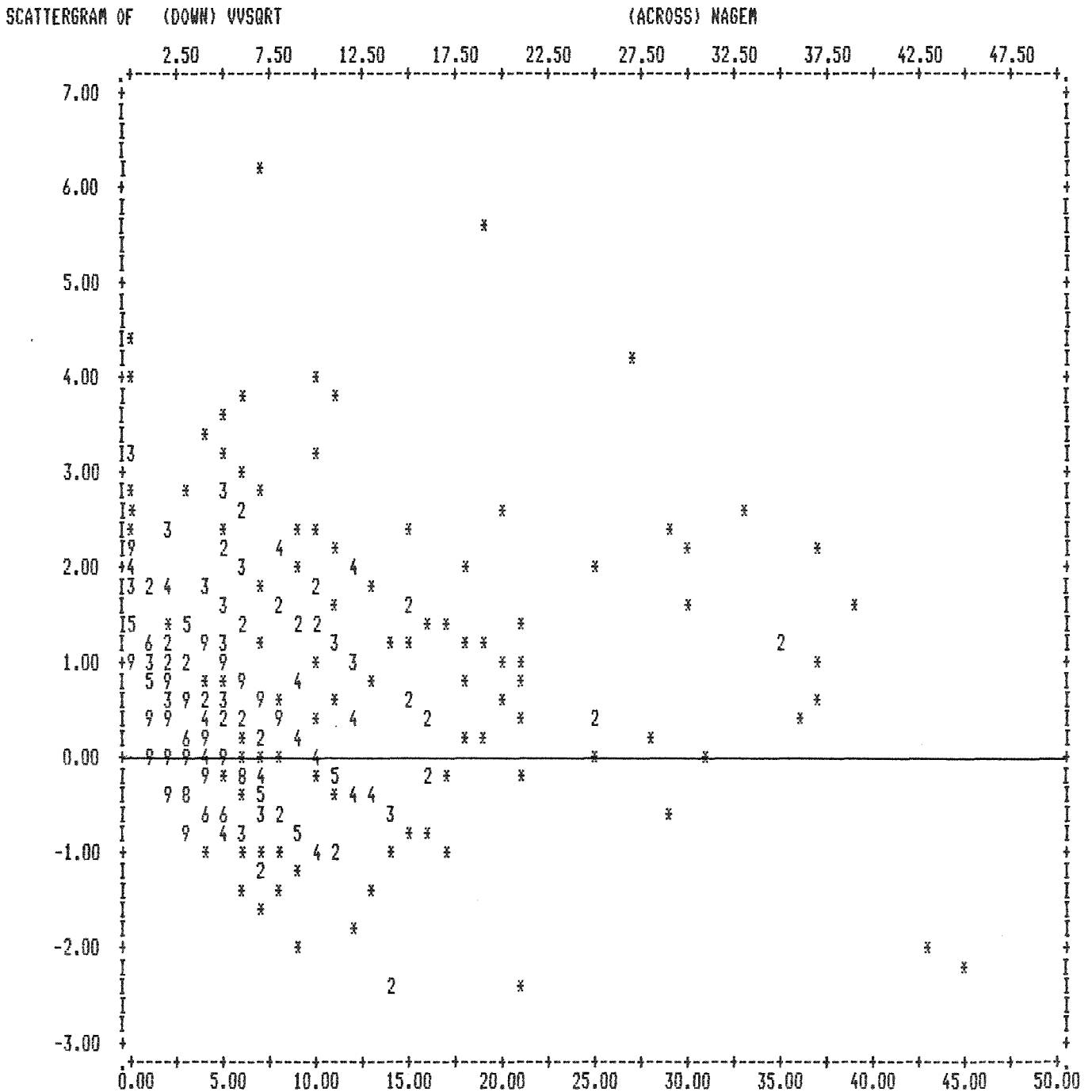
ROVIN-NAMET 78 SCATRGRAM LOPEN OPGEG-NAGEM 2-83

STATISTICS..

CORRELATION (R)-	0.68246	R SQUARED	-	0.46575	SIGNIFICANCE	-	0.00000
STD ERR OF EST -	7.04903	INTERCEPT (A) -		2.85812	SLOPE (B)	-	1.01721
PLOTTED VALUES -	587	EXCLUDED VALUES-		11	MISSING VALUES -		0

SPSS-11
 RUN NAME ROVIN-NAMET 78 SCTRGRM LOPEN (ROVIN-NAMET 78-83)
 VARIABLE LIST OPGEV,NAGEM,UVSQRT
 INPUT MEDIUM DR1:RECR1.DAT
 N OF CASES 598
 INPUT FORMAT REAL(19X,F4.0,1X,F4.0,16X,F9.4)
 3 variable name(s) actually read
 READ INPUT DATA
 598 observations have been read with 7 variables (including the 4 system variables) in each observation
 SCATTERGRAM UVSQRT(-3,7) WITH NAGEM(0,50)
 OPTIONS 4
 STATISTICS ALL

ROVIN-NAMET 78 SCTRGRM LOPEN (ROVIN-NAMET 78-83)



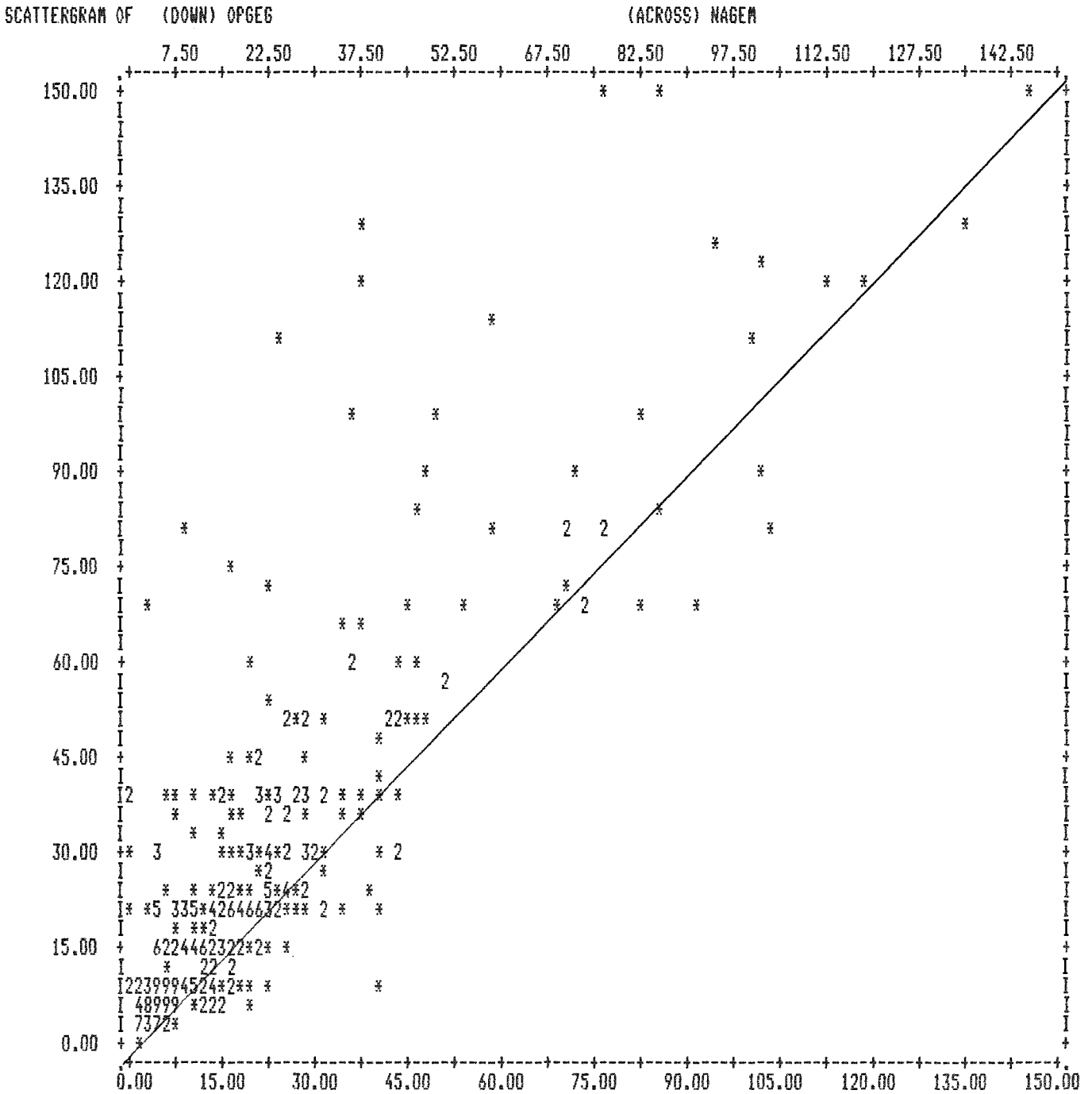
ROVIN-NAMET 78 SCTRGRM LOPEN (ROVIN-NAMET 78-83)

STATISTICS..

CORRELATION (R)-	-0.01791	R SQUARED	-	0.00032	SIGNIFICANCE	-	0.33167
STD ERR OF EST -	1.08722	INTERCEPT (A) -		0.57750	SLOPE (B)	-	-0.00286
PLOTTED VALUES -	593	EXCLUDED VALUES-		5	MISSING VALUES -		0

SPSS-11
 RUN NAME ROVIN-NAMET 78 SCATRGRAM FIETS
 VARIABLE LIST OPGEG,NAGEM,VVSQRT
 INPUT MEDIUM DR1:RECR3.DAT
 N OF CASES 416
 INPUT FORMAT REAL(19X,F4.0,1X,F4.0,16X,F9.4)
 3 variable name(s) actually read
 READ INPUT DATA
 416 observations have been read with 7 variables (including the 4 system variables) in each observation
 SCATTERGRAM OPGEG(0,150) WITH NAGEM(0,150)
 OPTIONS 4
 STATISTICS ALL

ROVIN-NAMET 78 SCATRGRAM FIETS



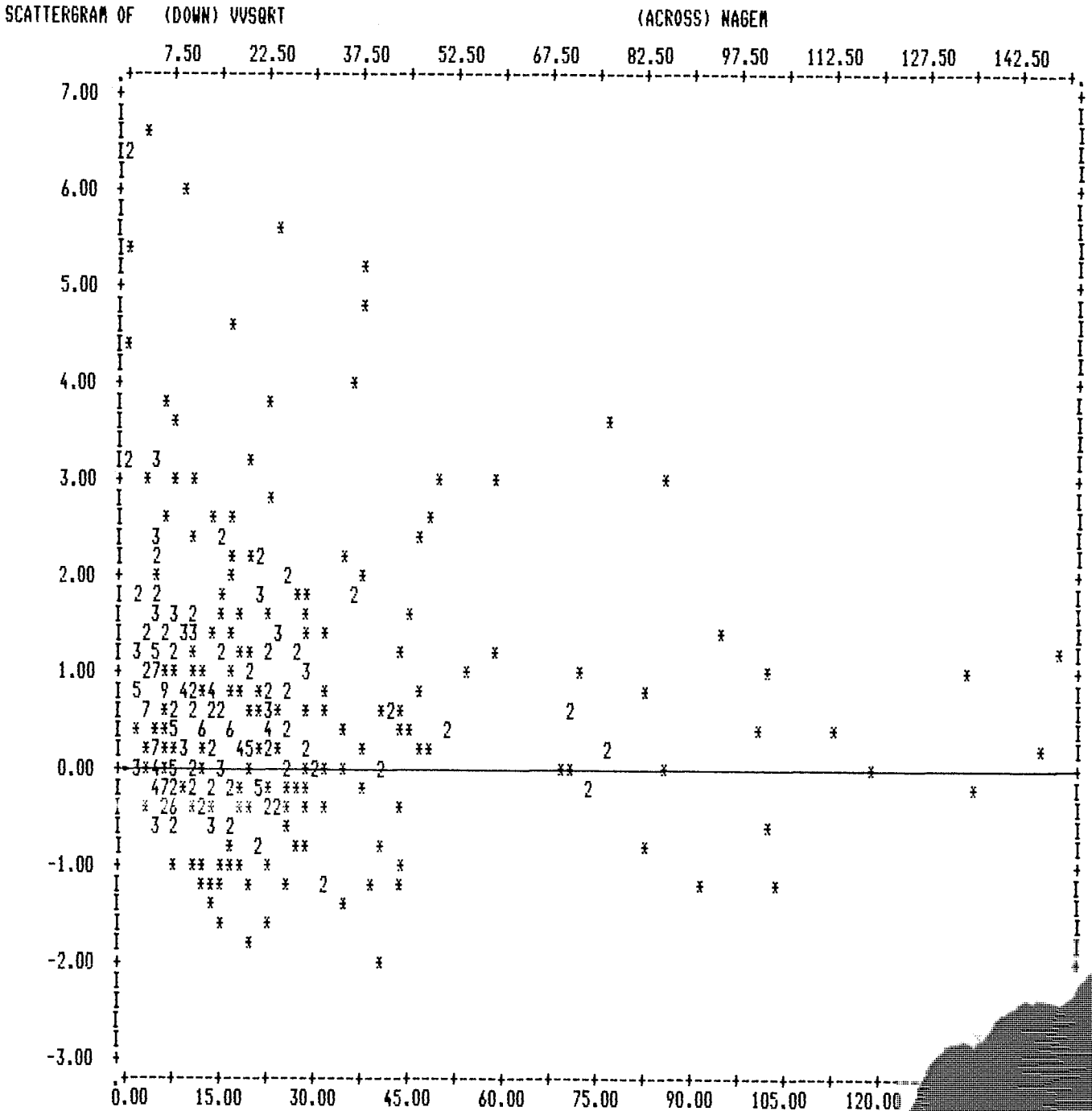
ROVIN-NAMET 78 SCATRGRAM FIETS OPGEG-NAGEM 2-83

STATISTICS..

CORRELATION (R)-	0.84096	R SQUARED	-	0.70721	SIGNIFICANCE	-	0.00000
STD ERR OF EST -	14.68681	INTERCEPT (A) -		5.94121	SLOPE (B)	-	1.04743
PLOTTED VALUES -	409	EXCLUDED VALUES-		7	MISSING VALUES -		0

SPSS-11
 RUN NAME ROVIN-NAMET 78 SCTRGRM FIETS
 VARIABLE LIST OPGEV,NAGEM,VVSQRT
 INPUT MEDIUM DR1:RECR3.DAT
 N OF CASES 416
 INPUT FORMAT REAL(19X,F4.0,1X,F4.0,16X,F9.4)
 3 variable name(s) actually read
 READ INPUT DATA
 416 observations have been read with 7 variables (including the 4 system variables) in each observation
 SCATTERGRAM VVSQRT(-3,7) WITH NAGEM(0,150)
 OPTIONS 4
 STATISTICS ALL

ROVIN-NAMET 78 SCTRGRM FIETS



ROVIN-NAMET 78 SCTRGRM FIETS

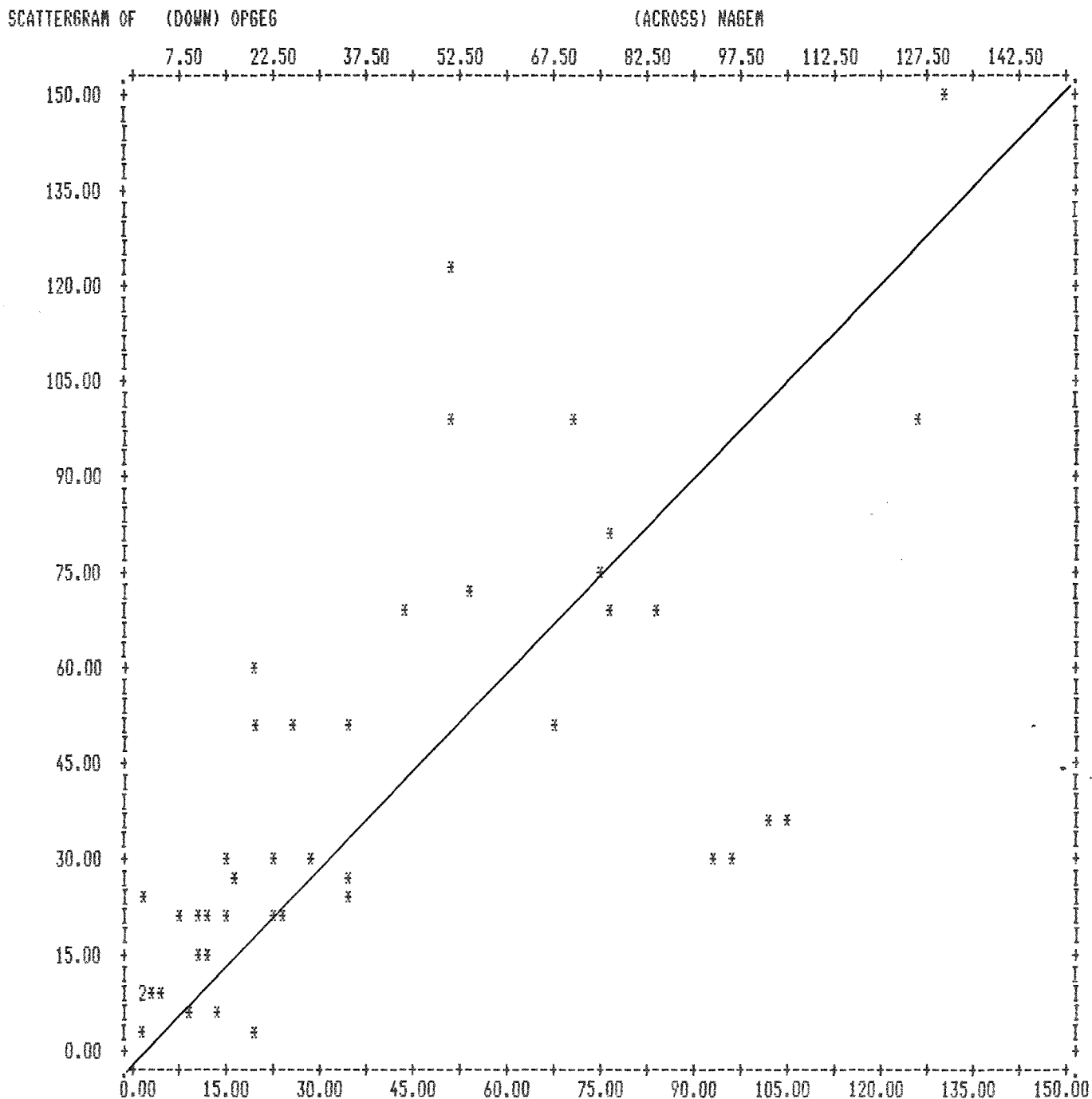
STATISTICS..

CORRELATION (R)-	-0.08439	R SQUARED	-	0.00712	SIGNIFICANCE	-
STD ERR OF EST -	1.28025	INTERCEPT (A) -		0.80374	SLOPE (B)	-
PLOTTED VALUES -	410	EXCLUDED VALUES-		6	MISSING VALUES -	

```

SPSS-11
RUN NAME      ROVIN-NAMET 78  SCATRGRAM BROMFTS  OPGEG-NAGEM 2-83
GET FILE      DR1:RECO.SPD
2009 observations have been read with 29 variables (including the 4 system variables) in each observation
SELECT IF     (V21 EQ 4)
COMPUTE       OPGEG=V22
COMPUTE       NAGEM=V24
SCATTERGRAM   OPGEG(0,150) WITH NAGEM(0,150)
OPTIONS       4
STATISTICS    ALL
    
```

ROVIN-NAMET 78 SCATRGRAM BROMFTS OPGEG-NAGEM 2-83



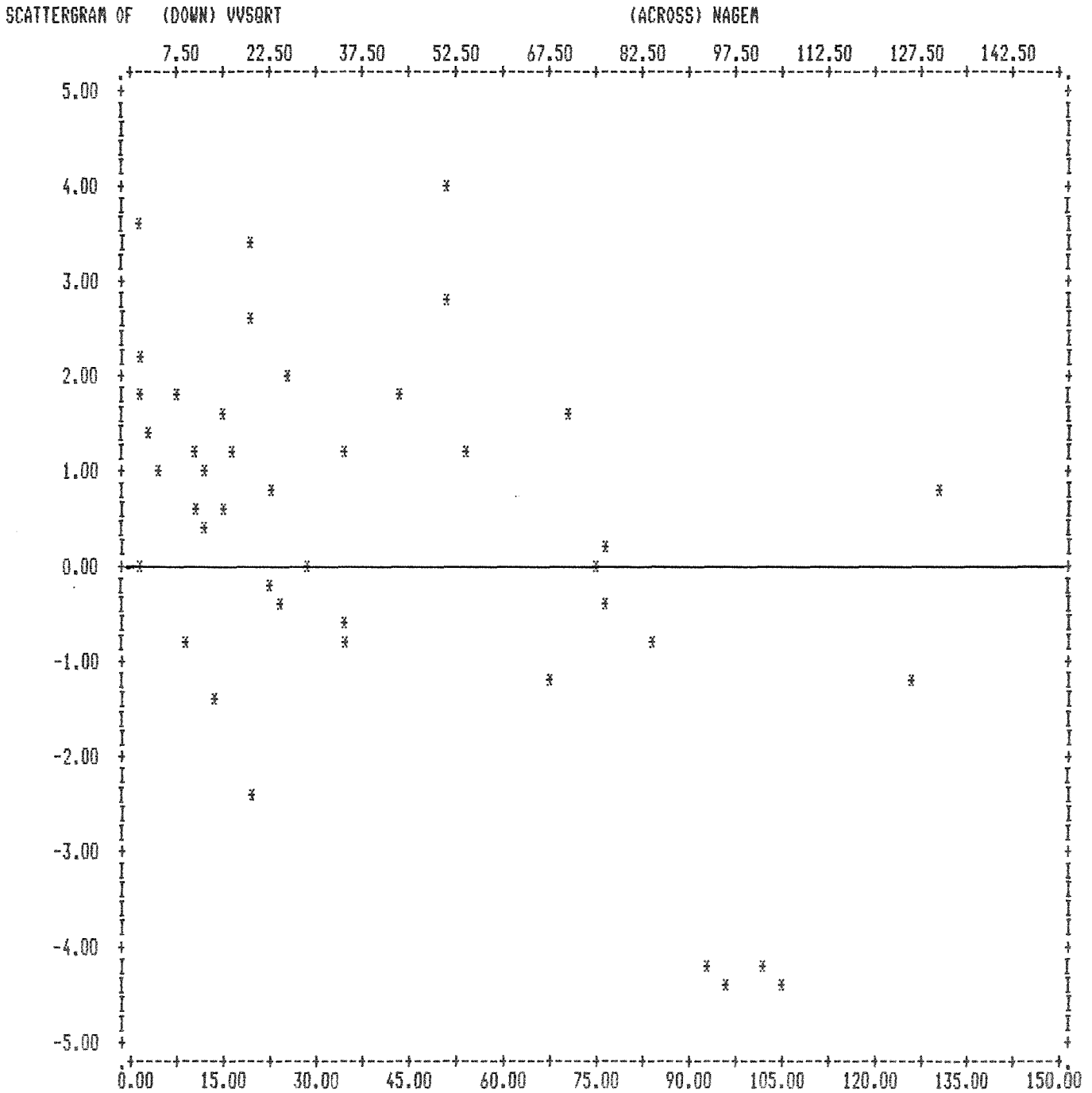
ROVIN-NAMET 78 SCATRGRAM BROMFTS OPGEG-NAGEM 2-83

STATISTICS..

CORRELATION (R)-	0.68212	R SQUARED	-	0.46529	SIGNIFICANCE	-	0.00000
STD ERR OF EST -	25.72375	INTERCEPT (A) -		15.97398	SLOPE (B)	-	0.64737
PLOTTED VALUES -	43	EXCLUDED VALUES-		1	MISSING VALUES -		0

SPSS-11
 RUN NAME ROVIN-NAMET 78 SCTRGRM BROWFTS WORTL(0-N)-NAGEM 3-83
 VARIABLE LIST OPPEG,NAGEM
 INPUT MEDIUM DR1:REC04.DAT
 N OF CASES 44
 INPUT FORMAT REAL(19X,F4.0,1X,F4.0)
 2 variable name(s) actually read
 READ INPUT DATA
 44 observations have been read with 6 variables (including the 4 system variables) in each observation
 COMPUTE VVSQRT=SQRT(OPPEG)-SQRT(NAGEM)
 SCATTERGRAM VVSQRT(-5,5) WITH NAGEM(0,150)
 OPTIONS 4
 STATISTICS ALL

ROVIN-NAMET 78 SCTRGRM BROWFTS WORTL(0-N)-NAGEM 2-83



ROVIN-NAMET 78 SCTRGRM BROWFTS V0-VN NAGEM 2-83

STATISTICS..

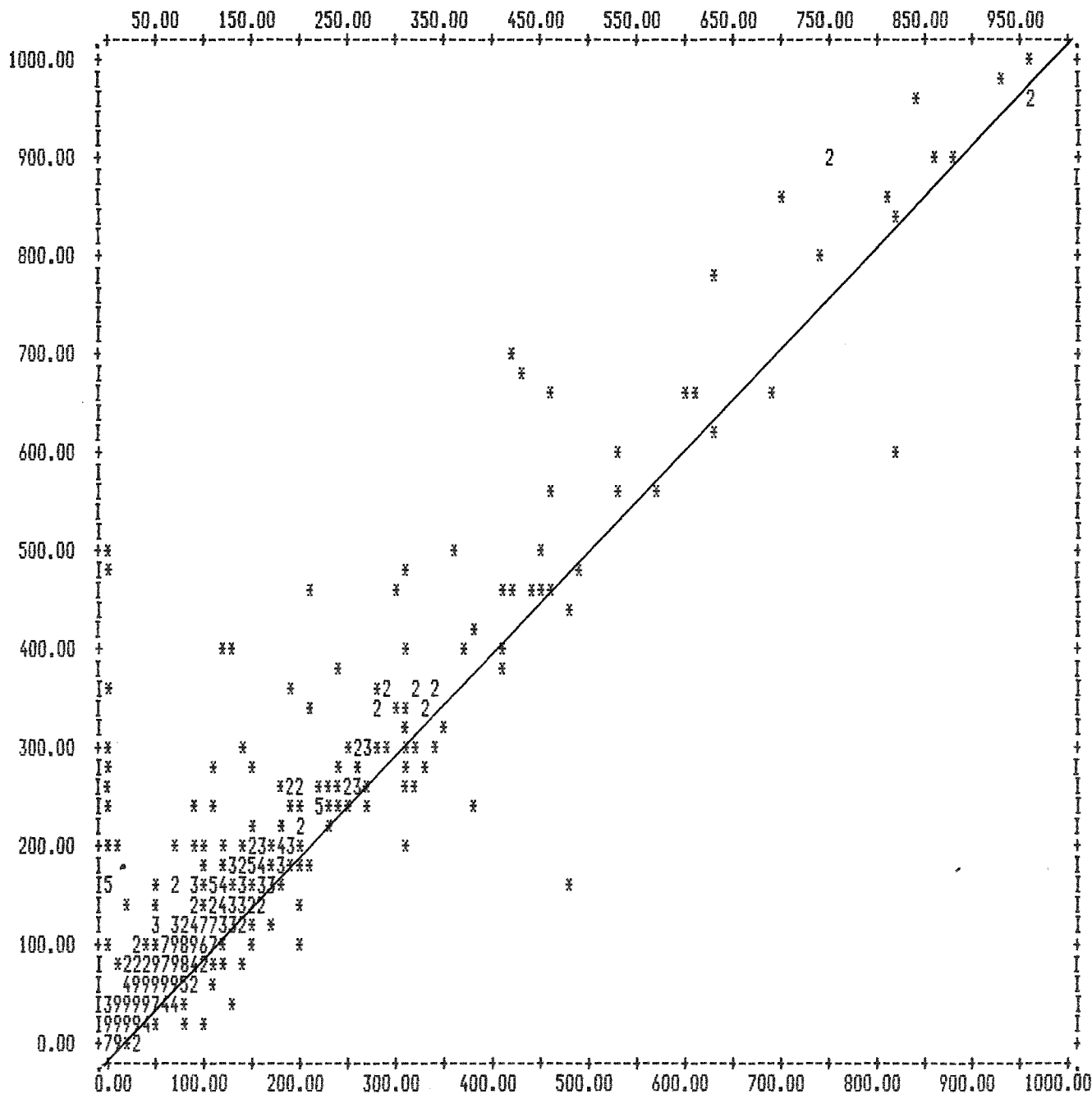
CORRELATION (R)-	-0.51868	R SQUARED	-	0.26903	SIGNIFICANCE	-	0.00018
STD ERR OF EST -	1.74612	INTERCEPT (A) -	-	1.45783	SLOPE (B) -	-	-0.02858
PLOTTED VALUES -	43	EXCLUDED VALUES-	-	1	MISSING VALUES -	-	0

SPSS-11
 RUN NAME ROVIN-NAMET 78 SCATRGRAM AUTO
 VARIABLE LIST OPGEG,NAGEM,VVSBRT
 INPUT MEDIUM DR1:RECR2.DAT
 N OF CASES 827
 INPUT FORMAT REAL(19X,F4.0,1X,F4.0,16X,F9.4)
 3 variable name(s) actually read
 READ INPUT DATA
 827 observations have been read with 7 variables (including the 4 system variables) in each observation
 SCATTERGRAM OPGEG(0,1000) WITH NAGEM(0,1000)
 OPTIONS 4
 STATISTICS ALL

ROVIN-NAMET 78 SCATRGRAM AUTO

SCATTERGRAM OF (DOWN) OPGEG

(ACROSS) NAGEM



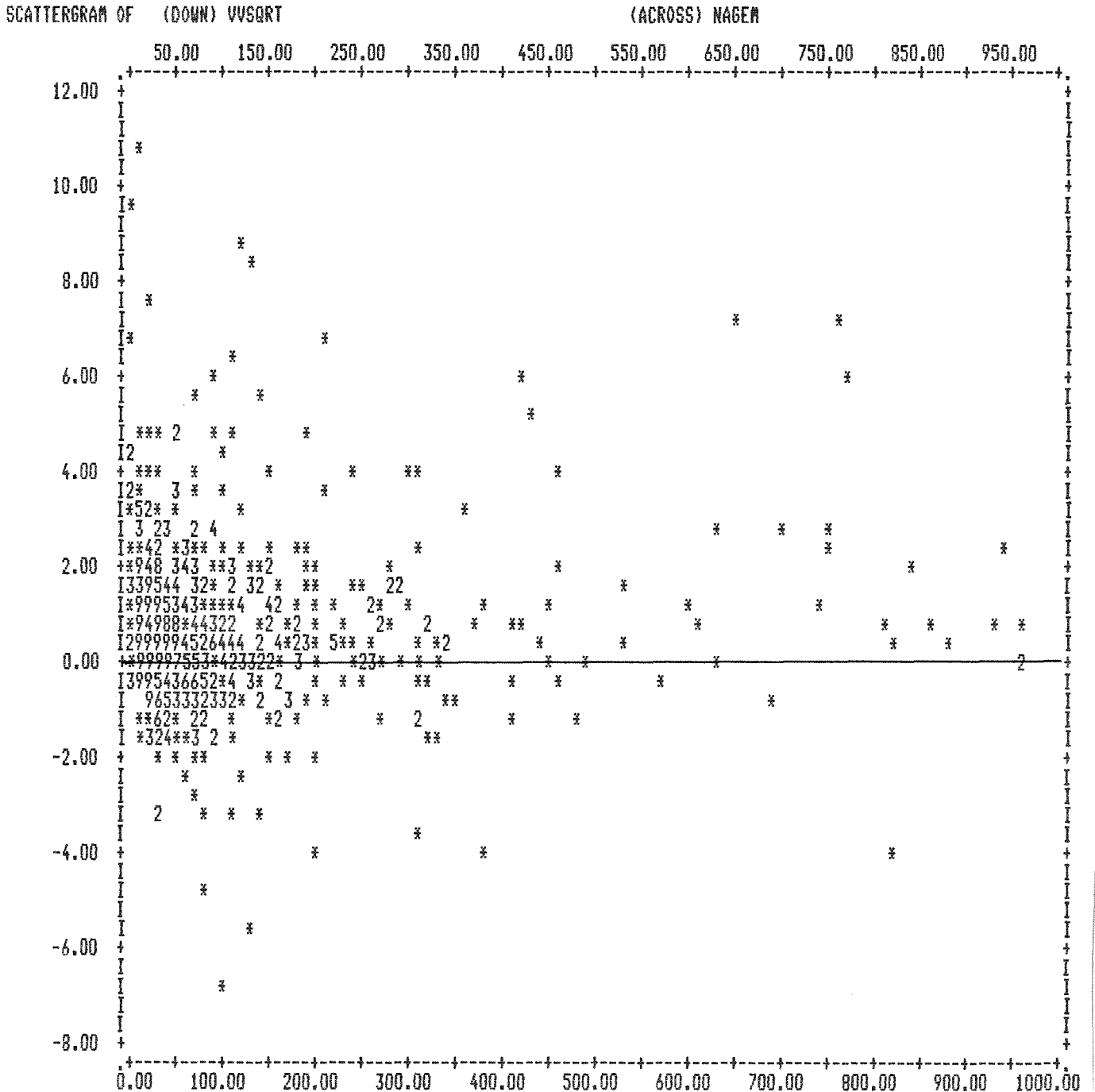
ROVIN-NAMET 78 SCATRGRAM AUTO OPGEG-NAGEM 2-83

STATISTICS..

CORRELATION (R)-	0.94143	R SQUARED	-	0.88630	SIGNIFICANCE	-	0.00000
STD ERR OF EST -	52.77370	INTERCEPT (A) -		14.58928	SLOPE (B)	-	1.02249
PLOTTED VALUES -	817	EXCLUDED VALUES-		10	MISSING VALUES -		0

SPSS-11
 RUN NAME ROVIN-NAMET 78 SCTRGRM AUTO VORT(10-N)-NAGEM 2-83
 VARIABLE LIST OPGEV,NAGEM,VVSQRT
 INPUT MEDIUM DR1:RECR2.DAT
 N OF CASES 827
 INPUT FORMAT REAL(19X,F4.0,1X,F4.0,16X,F9.4)
 3 variable name(s) actually read
 READ INPUT DATA
 827 observations have been read with 7 variables (including the 4 system variables) in each observation
 SCATTERGRAM VVSQRT(-8,12) WITH NAGEM(0,1000)
 OPTIONS 4
 STATISTICS ALL

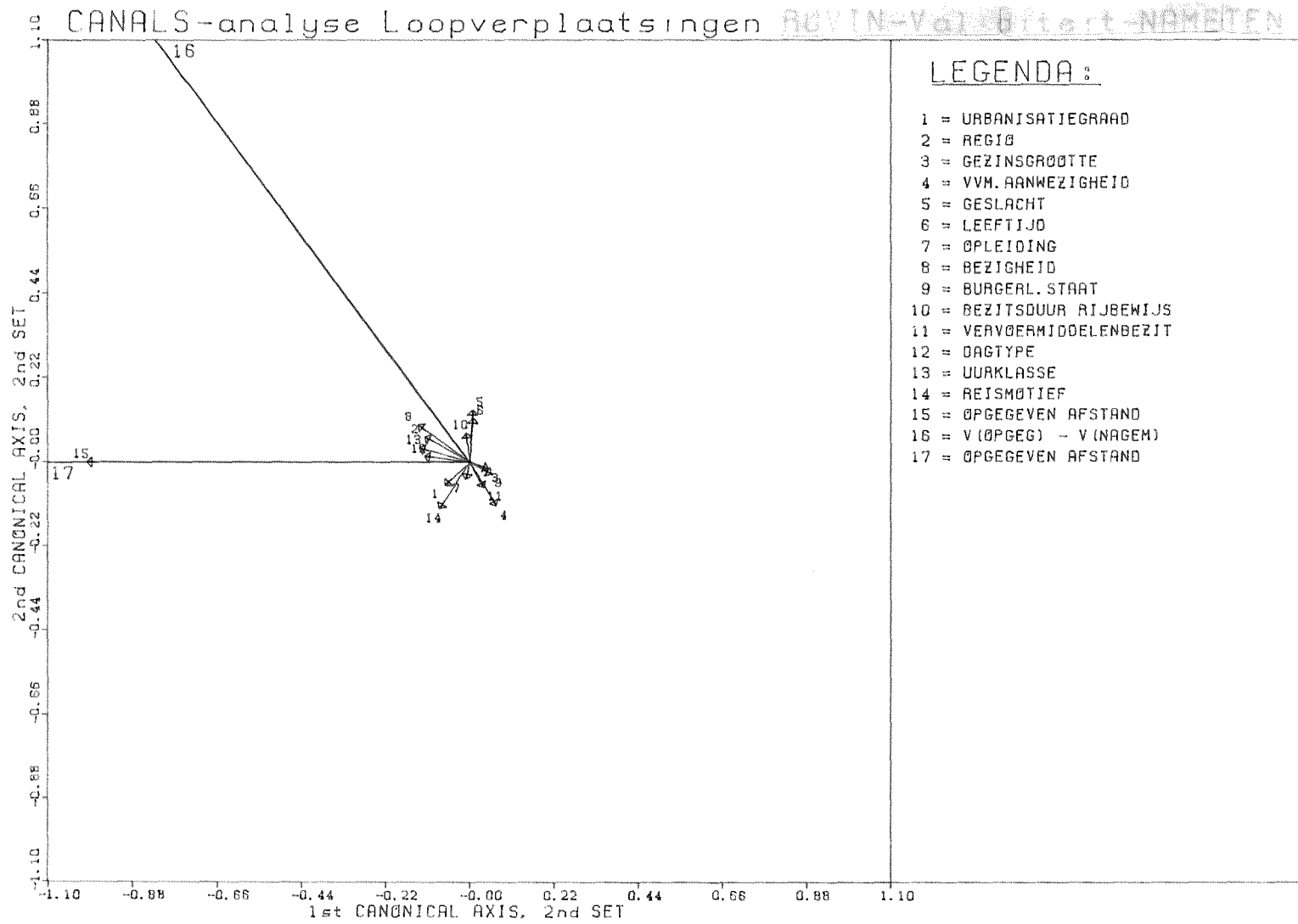
BOVIN-NAMET 78 SCTRGRM AUTO VORT(10-N)-NAGEM 2-83



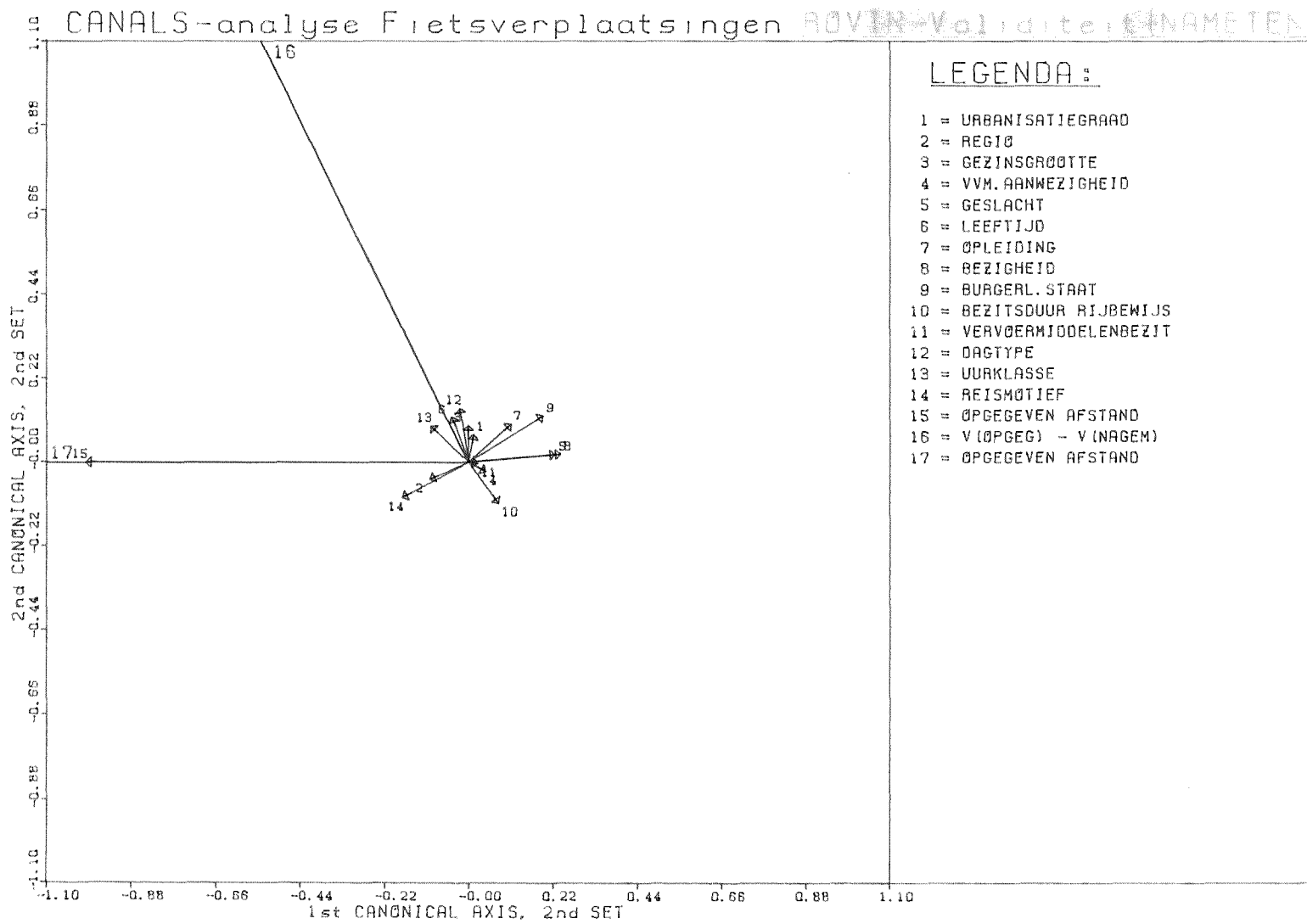
ROVIN-NAMET 78 SCTRGRM AUTO VORT(10-N)-NAGEM 2-83

STATISTICS..

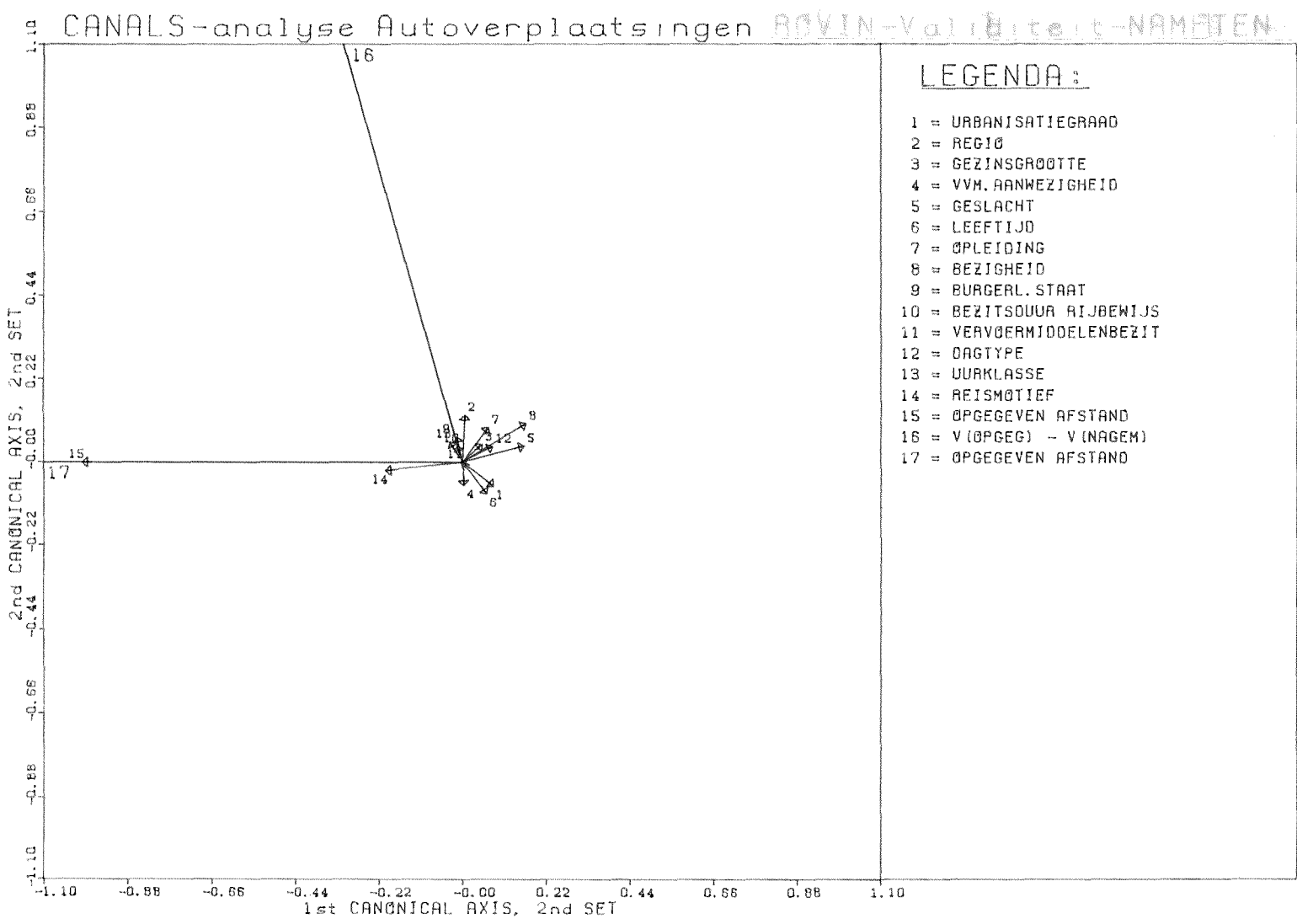
CORRELATION (R)-	0.05353	R SQUARED -	0.00287	SIGNIFICANCE -	0.06432
STD ERR OF EST -	1.66459	INTERCEPT (A) -	0.61871	SLOPE (B) -	0.00059
PLOTTED VALUES -	807	EXCLUDED VALUES-	20	MISSING VALUES -	0



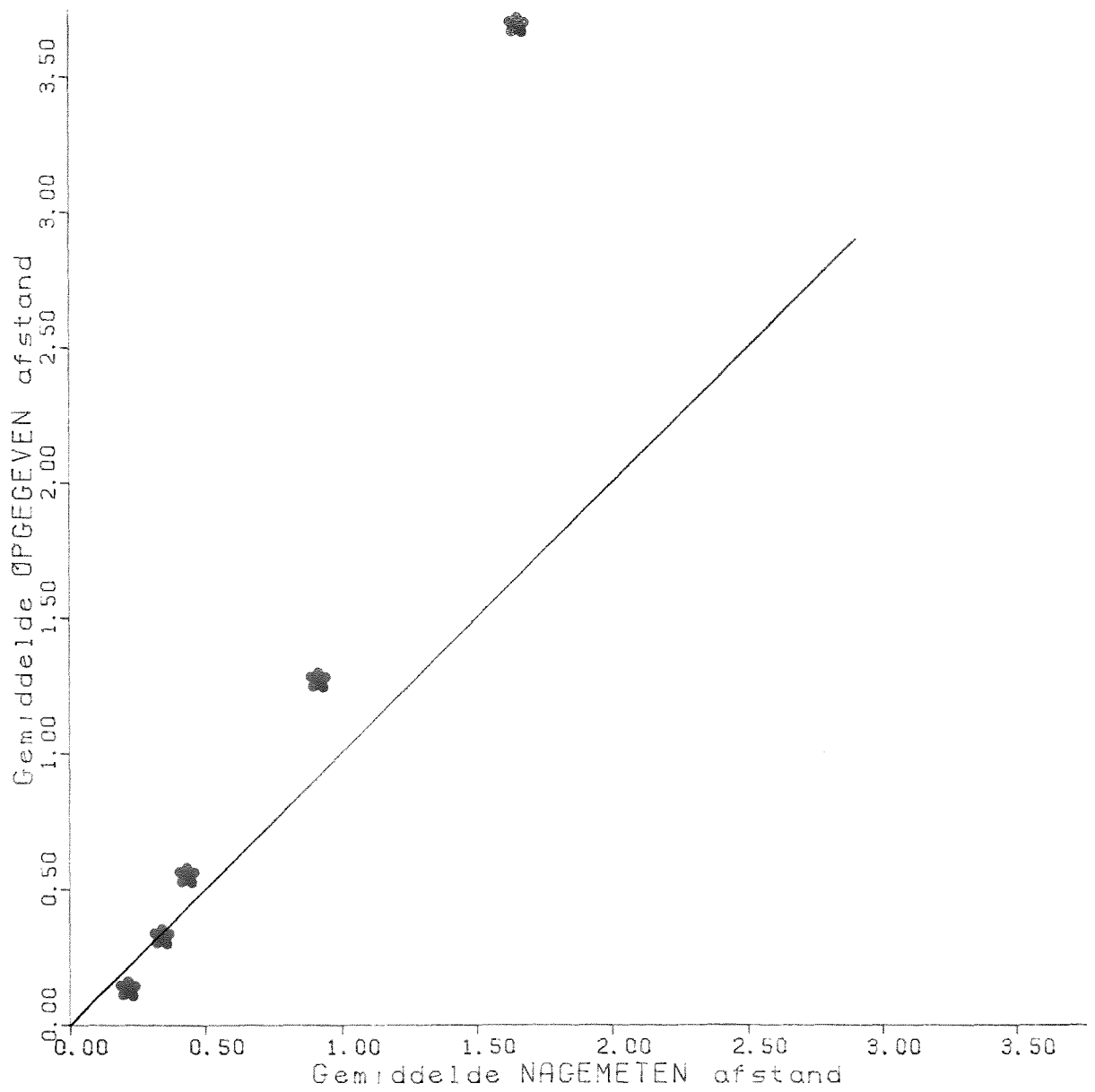
Afbeelding 5. Resultaten CANALS-analyse loopverplaatsingen.



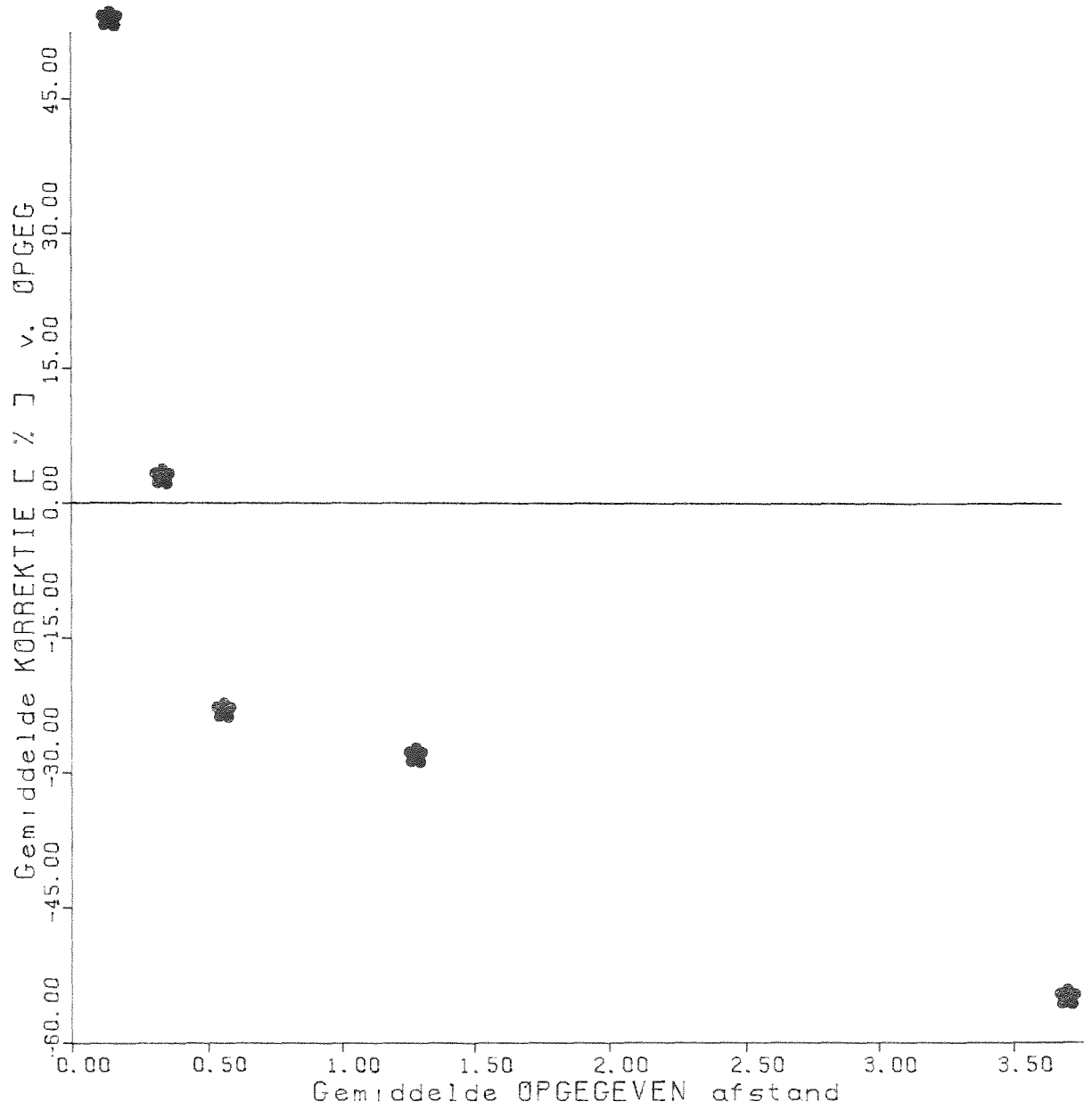
Afbeelding 6. Resultaten CANALS-analyse fietsverplaatsingen.



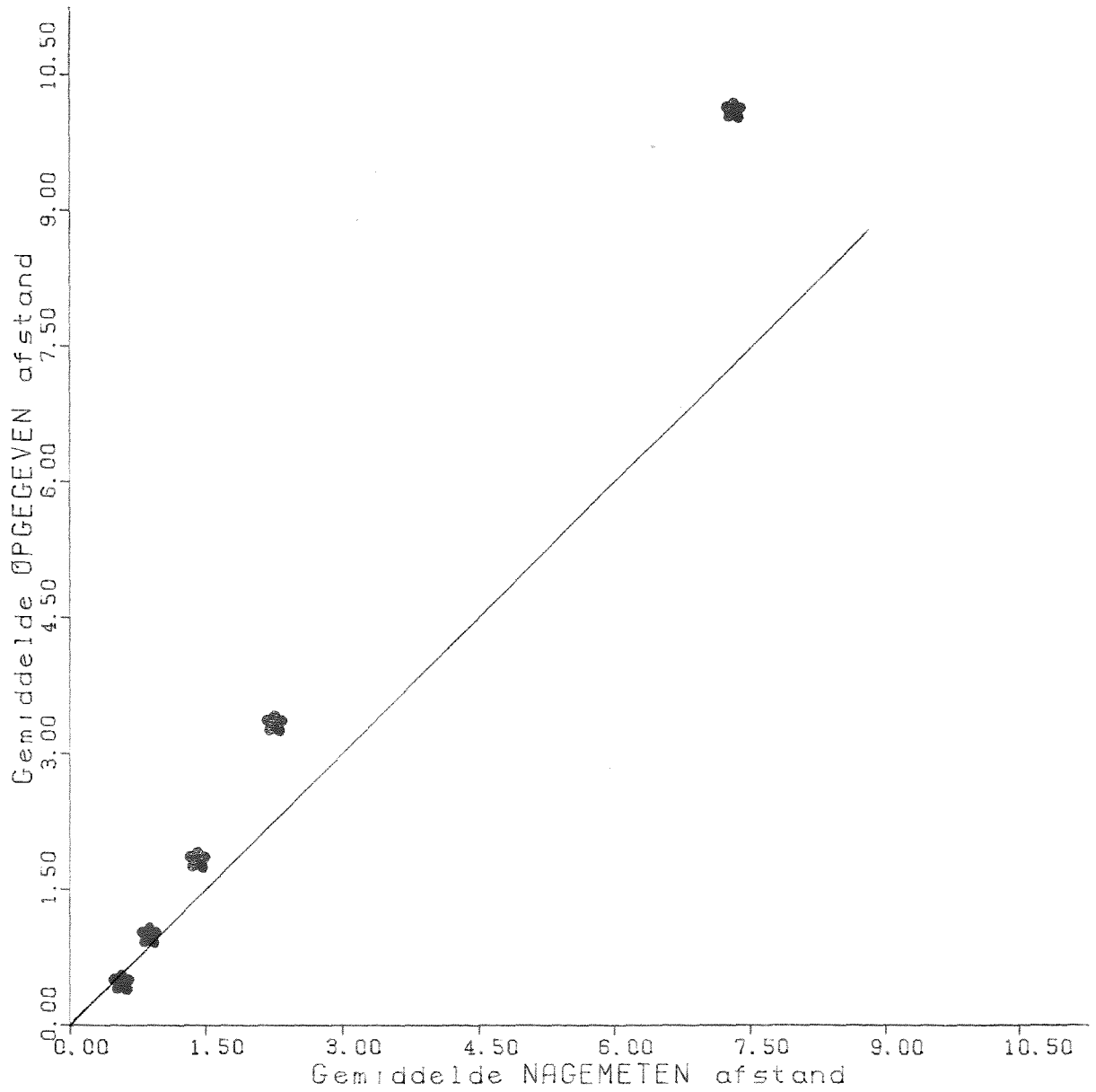
Afbeelding 7. Resultaten CANALS-analyse personenautoverplaatsingen.



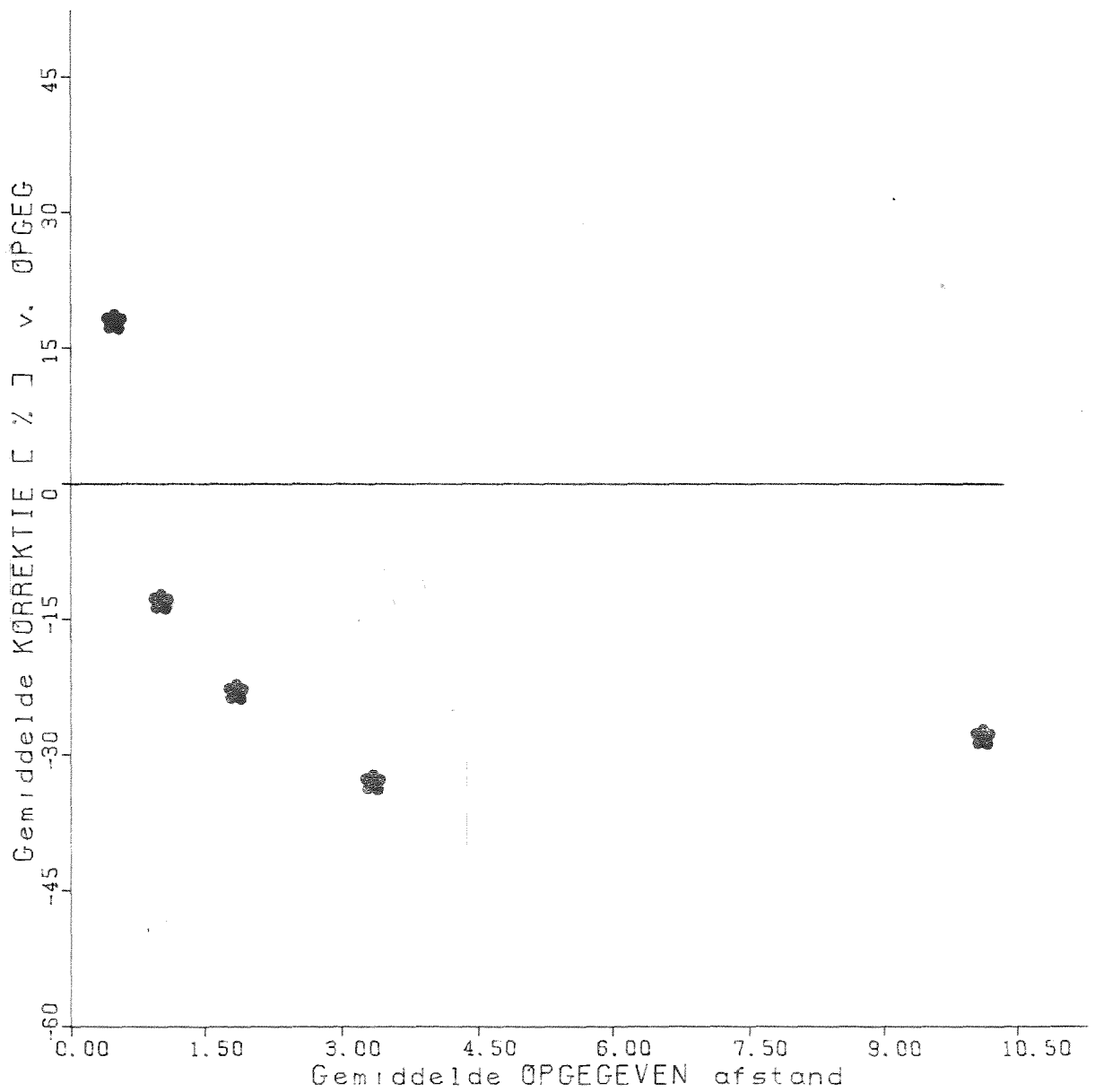
Abbeelding 8A. Relatie gemiddelde opgegeven en gemiddelde nagemeten loopafstanden.



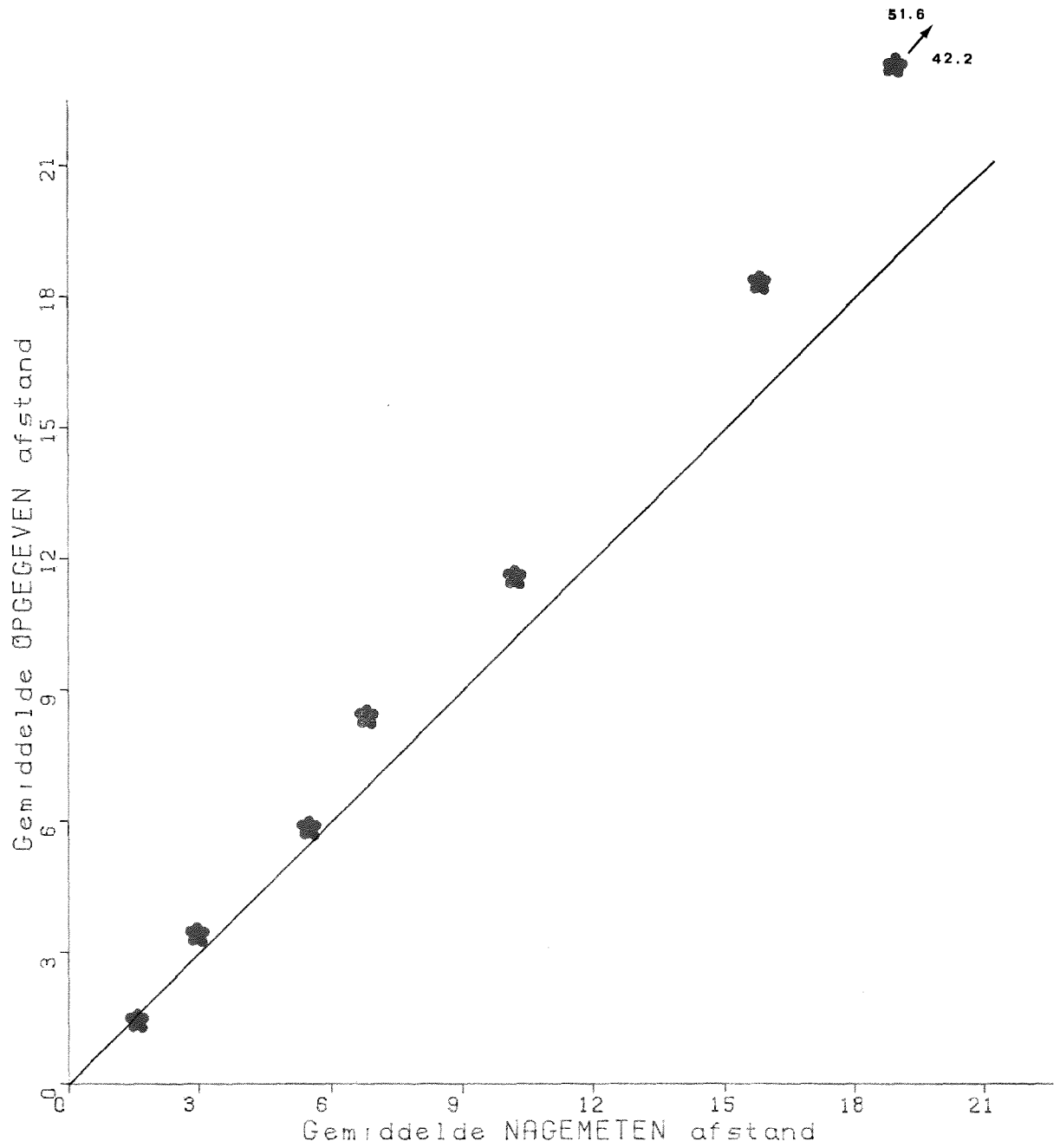
Afbeelding 3B. Gemiddelde percentage correctie bij gemiddelde opgegeven loopafstanden.



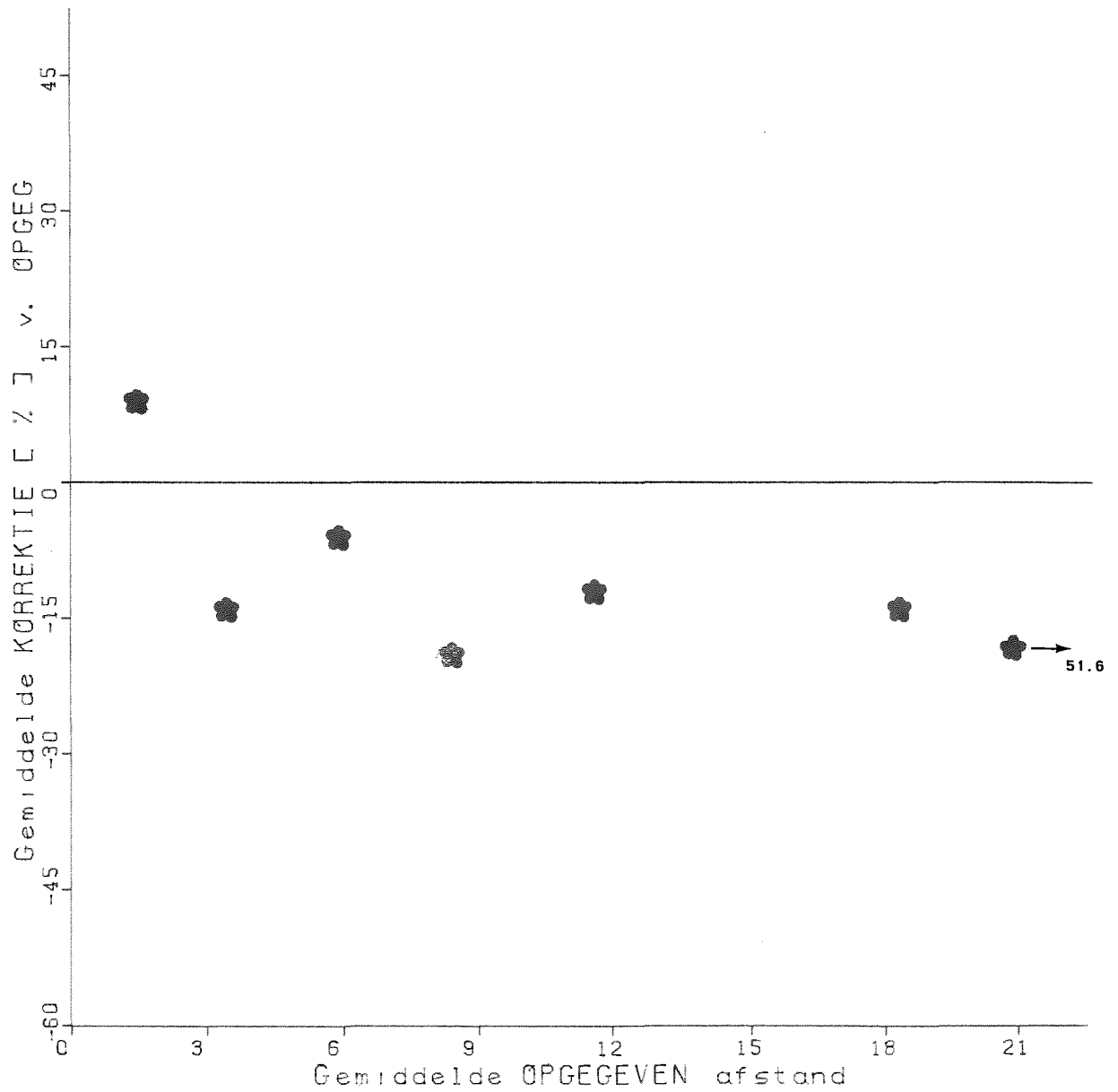
Afbeelding 9A. Relatie gemiddelde opgegeven en gemiddelde nagemeten fietsafstanden.



Afbeelding 9B. Gemiddelde percentage correctie bij gemiddelde opgegeven fietsafstanden.



Afbeelding 10A. Relatie gemiddelde opgegeven en gemiddelde nagemeten personenauto-afstanden.



Afbeelding 10B. Gemiddelde percentage correctie bij gemiddelde opgegeven personenauto-afstanden.

A

Correlations between the optimally scaled variables of the first set and the canonical variate of the second set (= $\sqrt{V_{opgegeven}} - \sqrt{V_{nagem.}}$ *).

	LOPEN	FIETS	AUTO
1 = urbanisatieklasse	0,027	-0,054	0,047
2 = regio	-0,113	0,000	-0,020
3 = grootte huish.	-0,020	0,051	-0,028
4 = aanw. vervoerm.	0,122	-0,067	0,039
5 = geslacht	-0,100	0,086	0,029
6 = leeftijd	0,033	0,003	-0,081
7 = opleiding	0,052	0,012	-0,046
8 = bezigheid	0,050	-0,047	-0,027
9 = burgerl. staat	-0,074	-0,044	-0,071
10 = bezitsduur rijbewijs	0,008	0,098	-0,026
11 = bezit vervoerm.	-0,111	0,036	-0,006
12 = dagtype	-0,098	-0,125	-0,014
13 = uurklasse	-0,081	-0,125	-0,051
14 = reismotief	-0,028	-0,072	-0,018
15 = opgegeven afstand*	-0,593	-0,444	-0,298
16 = nagemeten afstand	0,039	0,047	0,070

*) ordinale (i.p.v. nominale) schalings-eis

B

Correlations between the optimally scaled variables of the first set and the canonical variates of the second set for each dimension.

	LOPEN		FIETS		AUTO	
1 = urbanisatieklasse	-0,065	-0,062	-0,012	-0,073	0,080	-0,064
2 = regio	-0,113	0,066	0,102	0,045	0,005	0,126
3 = grootte huish. *	0,049	-0,019	0,002	-0,097	0,051	0,048
4 = aanw. vervoerm. *	0,071	-0,116	-0,046	0,025	0,002	-0,062
5 = geslacht	0,008	0,136	-0,227	-0,021	0,162	0,044
6 = leeftijd	0,009	0,114	0,044	-0,119	0,063	-0,084
7 = opleiding	-0,010	-0,046	-0,111	-0,101	0,068	0,092
8 = bezigheid	-0,135	0,098	-0,241	-0,021	-0,167	0,104
9 = burgerl. staat	0,057	-0,032	-0,196	-0,123	-0,017	0,068
10 = bezitsduur rijbewijs	-0,010	0,077	-0,079	0,107	-0,015	0,044
11 = bezit vervoerm. *	0,037	-0,068	-0,024	0,004	-0,005	0,002
12 = dagtype	-0,119	0,015	0,024	-0,143	0,079	0,043
13 = uurklasse	-0,132	0,038	0,100	-0,096	-0,035	0,055
14 = reismotief	-0,080	-0,124	0,175	0,093	-0,204	-0,022
15 = opgegeven afstand*	-1,000	0,001	1,000	-0,001	-1,000	0,000

*) ordinale (i.p.v. nominale) schalings-eis

Correlations between the optimally scaled variables of the second set and the canonical variates of the second set for each dimension.

	LOPEN		FIETS		AUTO	
16 = $\sqrt{V_{opgeg.}} - \sqrt{V_{nagem.}}$ *	-0,598	0,801	0,443	-0,897	-0,275	0,961
17 = opgegeven afstand	-1,000	-0,000	1,000	0,001	-1,000	0,000

*) ordinale (i.p.v. nominale) schalings-eis

Tabel 1. CANALS-analyse.

Source of variation	Sum of squares	Df	Mean square	F	Signif. of F
<u>LOPEN</u>					
Covariate	300,608	1	300,608	312,639	0,000
Opgeg.	300,608	1	300,608	312,639	0,000
Main effects	18,721	14	1,337	1,391	0,152
Geslacht	2,733	1	2,733	2,843	0,092
Reismotief	1,374	3	0,458	0,476	0,699
Aanw. vervoermid.	3,581	2	1,790	1,862	0,156
Leeftijd	4,369	6	0,728	0,757	0,604
Bezigheid	0,129	2	0,064	0,067	0,935
Explained	319,329	15	21,289	22,141	0,000
Residual	559,603	582	0,962		
Total	878,932	597	1,472		
<u>FIETS</u>					
Covariate	98,575	1	98,575	54,022	0,000
Opgeg.	98,575	1	98,575	54,022	0,000
Main effects	62,862	15	4,191	2,297	0,004
Dagtype	22,663	2	11,332	6,210	0,002
Burgerl. staat	6,811	1	6,811	3,733	0,054
Leeftijd	24,741	6	4,124	2,260	0,037
Bezitsduur rijbew.	11,819	3	3,940	2,159	0,092
Opleiding	2,762	3	0,921	0,505	0,679
Explained	161,438	16	10,090	5,530	0,000
Residual	728,060	399	1,825		
Total	889,498	415	2,143		
<u>AUTO</u>					
Covariate	270,958	1	270,958	44,218	0,000
Opgeg.	270,958	1	270,958	44,218	0,000
Main effects	183,121	15	12,208	1,992	0,014
Regio	66,882	3	22,294	3,638	0,013
Bezigheid	15,228	2	7,614	1,243	0,289
Opleiding	51,513	3	17,171	2,802	0,039
Leeftijd	37,767	5	7,553	1,233	0,292
Burgerl. staat	13,255	2	6,628	1,082	0,340
Explained	454,082	16	28,380	4,631	0,000
Residual	4963,082	810	6,128		
Total	5417,531	826	6,559		

Due to empty cells or a singular matrix, higher order interactions have been suppressed.

Tabel 2. Variantie-analyse $\sqrt{\text{Opgeg.}}$ - $\sqrt{\text{Nagem.}}$

Source of variation	Sum of squares	Df	Mean square	F	Signif. of F
<u>FIETS</u>					
Main effects	178,847	5	35,769	23,340	0,000
Dagsoort*)	25,603	1	25,603	16,706	0,000
Opgeg. afst.	156,792	4	39,198	25,577	0,000
2-way interaction	88,463	4	22,116	14,431	0,000
Dagsoort*opgeg. afst.	88,463	4	22,116	14,431	0,000
Explained	267,310	9	29,701	19,380	0,000
Residual	622,219	406	1,533		
Total	889,529	415	2,143		
*)werkdag - weekeinddag					
<u>AUTO</u>					
Main effects	613,239	9	68,138	11,569	0,000
Regio	54,727	3	18,242	3,097	0,026
Opgeg. afst.	550,735	6	91,789	15,585	0,000
2-way interaction	98,876	18	5,493	0,933	0,539
Regio*opgeg. afst.	98,876	18	5,493	0,933	0,539
Explained	712,115	27	26,375	4,478	0,000
Residual	4705,659	799	5,889		
Total	5417,773	826	6,559		

Tabel 2. (vervolg). Variantie-analyse $\sqrt{\text{Opgeg.}}$ - $\sqrt{\text{Nagem.}}$

ANNEX I. Over datatransformaties t.b.v. variantiestabilisatie

1. Stel dat het schatten van een gegeven afgelegde afstand δ_g zodanig geschiedt dat geldt:

$$\mathcal{E}(\delta_g) \sim N(0, \sigma_g),$$

zodat $\int u_{2r+1} = 0$ en $\int u_{2r} = \frac{2r!}{2^r \cdot r!} \sigma^{2r}$, $\mathcal{E} \int u_1 / u_2 = 0$ enz.

Het meest aannemelijk is dat de onnauwkeurigheid van het schatten proportioneel met de grootte van de geschatte afstand samenhangt: $\sigma_g = c \delta_g$, $c \ll 1$.

Naast deze schattingsonnauwkeurigheid $\mathcal{E}(\delta_g)$ bevat de schatting nog een systematische component, waarvan in eerste aanleg kan worden verondersteld dat hij eveneens evenredig is met de grootte van de afgelegde afstand: $s(\delta_g) = k_0 + k_1 \delta_g$.

Het model is aldus geworden: $d_g = \delta_g + s(\delta_g) + \mathcal{E}(\delta_g)$.

2. Voor de transformatie T waarmee variantiestabilisatie zou worden bereikt moet gelden:

$$T(\delta) = \int \frac{1}{\delta} d\delta := \log \delta.$$

Nagegaan wordt hierna omtrent de verdelingseigenschappen van deze loggetransformeerde afstand.

1. $\mathcal{E} \log d_g = \mathcal{E} \log(\delta_g + s(\delta_g))(1 + \Delta_g)$ met $|\Delta_g| = \left| \frac{\mathcal{E}(\delta_g)}{\delta_g + s(\delta_g)} \right| \ll 1$,

zodat met Taylor geldt: $\log(1 + \Delta) = \Delta - \frac{1}{2} \Delta^2 + \frac{1}{3} \Delta^3 - \dots$, waardoor:

$$\mathcal{E} \log d_g = \mathcal{E} \log(\delta_g + s(\delta_g)) - \mathcal{E} \sum \frac{1}{2n} \Delta_g^{2n} + \mathcal{E} \sum \frac{1}{2n-1} \Delta_g^{2n-1} \approx$$

$$\approx \log \delta_g + \frac{s(\delta_g)}{\delta_g} - \mathcal{E} \frac{1}{2} \Delta_g^2 \text{ met } \left| \frac{s(\delta_g)}{\delta_g} \right| \ll 1, \text{ zodat tenslotte}$$

wordt gevonden:

$$\mathcal{E} \log d_g \approx \log \delta_g + \frac{k_0}{\delta_g} + k_1 - \frac{1}{2} c^2.$$

$$\begin{aligned}
 2. \quad E (\log d_g - E \log d_g)^2 &= E \log^2 d_g - E^2 \log d_g = \\
 &= -E^2 (\log(\delta_{g+s}(\delta_g)) - \sum \frac{1}{2n} \Delta_g^{2n}) + E (\log(\delta_{g+s}(\delta_g)) - \\
 &- \sum \frac{1}{2n} \Delta_g^{2n} + \sum \frac{1}{2n-1} \Delta_g^{2n-1})^2 = \\
 &= E \sum \frac{1}{2n} \Delta_g^{2n} - E^2 \sum \frac{1}{2n} \Delta_g^{2n} + E \sum \frac{1}{2n-1} \Delta_g^{2n-1} - \\
 &- 2 E (\sum \frac{1}{2n} \Delta_g^{2n} \times \sum \frac{1}{2n-1} \Delta_g^{2n-1}) \approx \\
 &\approx E \frac{1}{4} \Delta_g^4 - E^2 \frac{1}{2} \Delta_g^2 + E (\Delta_g^2 + \frac{2}{3} \Delta_g^4), \text{ waarmee met } \left| \frac{s(\delta_g)}{\delta_g} \right| \ll 1
 \end{aligned}$$

tenslotte wordt gevonden:

$$\text{VAR } \log d_g \approx c^2 \left(1 - \frac{2s(\delta_g)}{\delta_g}\right) + \frac{5}{2} c^4 \approx c^2 \left(1 - \frac{2k_0}{\delta_g} - 2k_1\right) + \frac{5}{2} c^4.$$

Mag, wat misschien nog waarschijnlijker is, worden aangenomen dat geldt:

$$\sigma_g = c d_g \approx c (\delta_{g+s}(\delta_g)), \text{ dan wordt deze betrekking vereenvoudigd}$$

tot:

$$\text{VAR } \log d_g \approx c^2 + \frac{5}{2} c^4.$$

3. Onder handhaving van alle andere aannamen kan alternatief worden verondersteld, dat de onnauwkeurigheid van het schatten niet samenhangt met de grootte, maar met de wortel uit de grootte van de te schatten afstand: $\sigma_g = c \sqrt{\delta_g}$.

De gezochte variantiestabilisatie wordt in dit geval bereikt met een $\sqrt{\quad}$ -transformatie, waarbij de verdelingseigenschappen als volgt zijn:

$$1. \quad E \sqrt{d_g} = E (\sqrt{\delta_{g+s}(\delta_g)} \times \sqrt{1 + \Delta_g}), \text{ zodat met Taylor:}$$

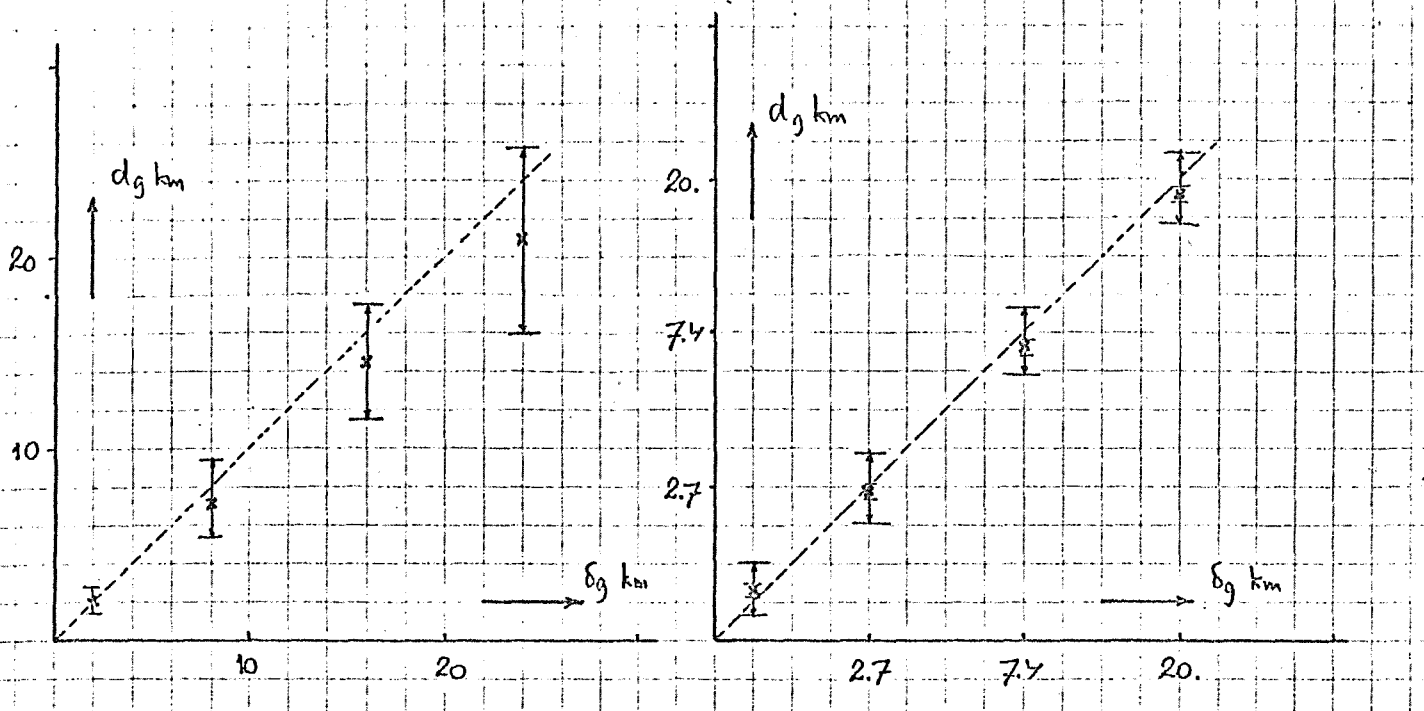
$$\sqrt{1 + \Delta} = 1 + \frac{1}{2} \Delta - \frac{1}{8} \Delta^2 + \frac{1}{16} \Delta^3 - \frac{5}{128} \Delta^4 + \dots$$

$$\begin{aligned}
 E \sqrt{d_g} &\approx \sqrt{\delta_{g+s}(\delta_g)} - \frac{c^2}{8 \sqrt{\delta_{g+s}(\delta_g)}} - \frac{15c^4}{128 \delta_g \sqrt{\delta_g}} \approx \\
 &\approx \sqrt{\delta_{g+s}(\delta_g)} - \frac{c^2}{8 \sqrt{\delta_g}} \approx \sqrt{\delta_g} + \frac{4s(\delta_g) - c^2}{8 \sqrt{\delta_g}}.
 \end{aligned}$$

$$2. \quad \text{VAR } \sqrt{d_g} = E d_g - E^2 \sqrt{d_g} \approx \frac{1}{4} c^2 + \frac{c^4}{10 \delta_g}.$$

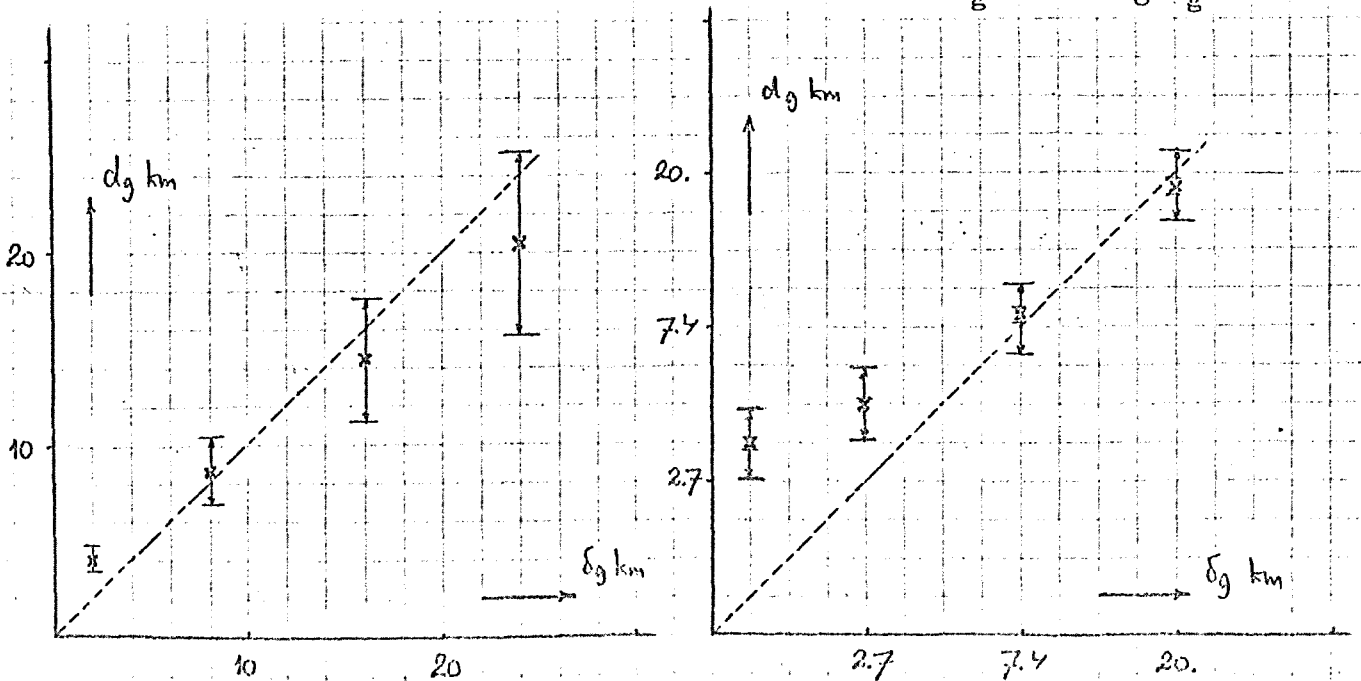
Met $c = c_0 \sqrt{c_1}$ kan een schaalfactor worden ingebracht.

Het log-model.



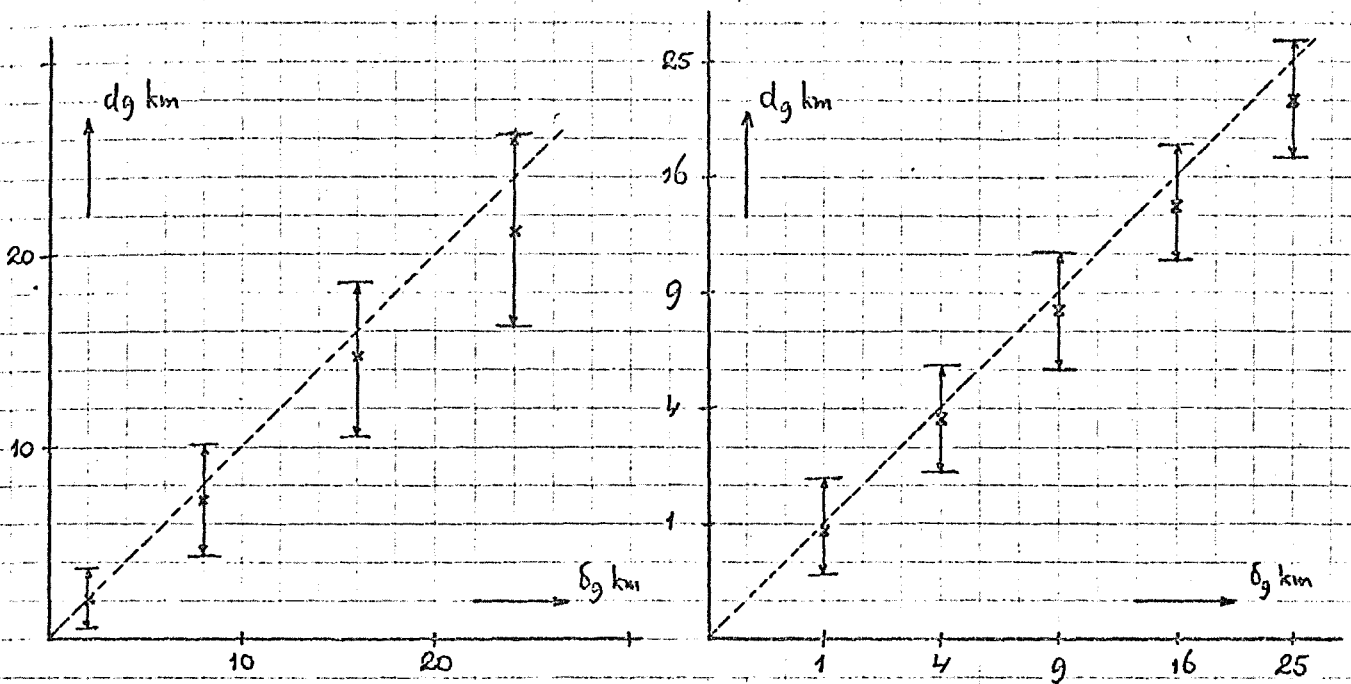
Verwachte ligging van de datapunten als het geschetste model zou voldoen. Links zijn de datapunten zelf weergegeven, rechts de log-data. Tevens staan de grenzen van het 2σ -gebied ingetekend.

Als voorbeeld is het schatten bij door personenauto's afgelegde kilometers gekozen, waarbij $c = .1$ is gesteld. Voor de systematische onder- en overschatting is na globale inspectie van wat data gesteld: $s(\delta_g) = \frac{1}{4} - \frac{1}{8} \delta_g$.



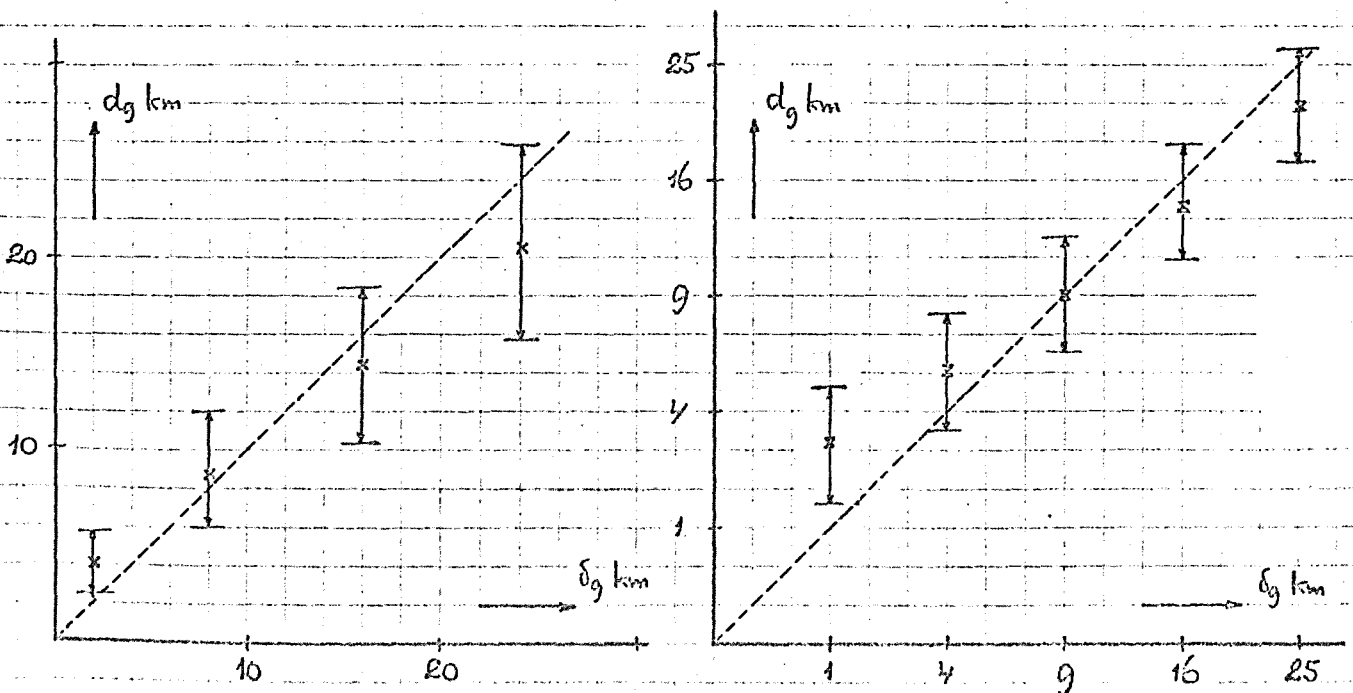
Dezelfde grafiek als boven, echter nu met: $s(\delta_g) = 2\frac{1}{2} - \frac{1}{4} \delta_g$.

Het $\sqrt{\quad}$ -model.



Dezelfde grafieken als boven, echter nu niet voor het log-, maar voor het $\sqrt{\quad}$ -model.

Daarbij is $c = \frac{1}{2}$ gesteld; $s(\delta_g) = \frac{1}{4} - \frac{1}{8} \delta_g$.



Dezelfde grafiek als boven, echter nu met: $s(\delta_g) = 2\frac{1}{2} - \frac{1}{4} \delta_g$.

N.B. Bij kleinere waarden van δ_g en relatief grotere van $s(\delta_g)$ gelden niet de benaderingsformules.

ANNEX II

Vergelijking van de varianties van de verschillen tussen opgegeven en nagemeten afstanden voor en na de variantie-stabiliserende transformatie.

	Klasse van opgegeven afstand						
	1	2	3	4	5	6	7
<u>LOOPVERPLAATSINGEN:</u>							
VAR (OPG - NAG)	.044	.053	.221	1.624	3.747		
VAR ($\sqrt{\text{OPG}}$ - $\sqrt{\text{NAG}}$)	.045	.055	.090	.156	.180		
<u>FIETSVERPLAATSINGEN:</u>							
(voor)	.099	.346	.516	1.254	20.205		
(na)	.040	.091	.100	.189	.526		
<u>AUTOVERPLAATSINGEN:</u>							
(voor)	1.131	2.545	3.468	5.802	9.036	46.274	154.626
(na)	.140	.205	.158	.345	.252	1.415	2.044

ANNEX III

Over de betrouwbaarheid van de schatting voor de grootte van de systematische component in de mischatting van afgelegde verplaatsingsafstanden.

Een maat voor de omvang van de systematische schattingsfout bij het schatten van de lengten van gemaakte verplaatsingen is de verhouding tussen werkelijk afgelegde en opgegeven afstanden (hun verschil zou een absolute maat zijn). Werkelijk afgelegde afstanden worden daarbij geschat uit nameting van op kaartmateriaal ingetekende opgegeven routes.

Zijn opg een opgegeven en nag de bijbehorende nagemeten afstand, \overline{opg} de gemiddelde opgegeven afstand in een beschouwde klasse van opg en N het aantal waarnemingen in die klasse, en is de gemiddelde standaard-toevallige schattingsfout s in opg veel groter dan de, verder derhalve te verwaarlozen, standaard-nameteefout in nag .

Voor de verhouding k tussen nagemeten en opgegeven afstanden binnen de beschouwde klasse wordt zodoende geschreven:

$$k = \frac{\sum nag}{\sum opg}$$

Geldt voor de verwachtingswaarde van de getorrigeerde opgegeven afstanden $E(k * opg) = nag$ en is bij onderling onafhankelijke waarnemingen binnen de beschouwde klasse de error als NORMAAL-verdeeld op te vatten, dan volgt voor de variantie van k :

$$\hat{VAR}(k) \approx \frac{s}{N * \overline{opg}^2}, \text{ waarbij } s = \hat{VAR}(k * opg).$$

Bij een groter aantal waarnemingen komt dus onder overigens gelijke omstandigheden een grotere betrouwbaarheid van de op grond van de steekproef bepaalde waarde van k .

Uitgaande van continuïteit in de schaal, kunnen de grenzen van de beschouwde klasse daardoor zo worden verlegd dat die betrouwbaarheid in principe aan zekere criteria voldoet.

De geschatte waarde van k zal de werkelijke waarde, gegeven de steekproef, slechts met kans α meer dan ter grootte Δk ontlopen als geldt:

$$\Delta k = U_{\frac{1}{2}\alpha} \sqrt{VAR(k)}.$$

Stel zo bijvoorbeeld voor autoverplaatsingen $\Delta k = 0.1$ bij waarden $k = 1$, $s = 5 \text{ km}$, $\overline{opg} = 10 \text{ km}$ en stel $\alpha = 0.2$ zodat ingevolge de normaliteitsassumptie $U_{\frac{1}{2}\alpha} = 1.3$, dan is voorwaarde: $N > 40$.

Omgekeerd kan door de waarde van s uit de steekproef te berekenen voor elke klasse van opg de grootte van $VAR(k)$ worden geschat.

Is als bij de onderzoeksresultaten de correctie $k^* [\%] = 100(k-1)$ dan wordt: $VAR(k^*) = 100^2 VAR(k)$, waarmee de gezochte betrouwbaarheid bekend is.

Aanhangsel

OVER STEEKPROEVEN

In de periode van het ROVIN-onderzoek is in het kader van het OVG een steekproef van N gezinnen gedurende in principe twee dagen geënkquêteerd. Hiervan horen er n tot het over één dag gerealiseerde subonderzoek Nauwkeurigheid opgegeven afstanden.

Bij een gemiddeld aantal personen g boven 11 jaar per huishouden en een gemiddeld aantal verplaatsingen a per persoon per dag, gaat het dus in principe in het OVG om $2 \cdot a \cdot g \cdot N$ en in het subonderzoek Nauwkeurigheid opgegeven afstanden om $a \cdot N$ verplaatsingen.

Eventuele regio-, dagtype- en vervoerswijze-verschillen blijven gemakshalve, evenals aannemelijke toevalsverschillen, buiten beschouwing.

De resultaten, verkregen uit de analyse van de nagemeten verplaatsingen, moeten nu in principe worden toegepast op de OVG-verplaatsingen, die los gezien van seizoenseffecten een zekere fractie zouden vertegenwoordigen van de, zijnde N_T het totaal aantal Nederlandse huishoudingen, zodoende $365 \cdot a \cdot g \cdot N_T$ in totaal in Nederland per jaar gemaakte verplaatsingen. Uiteindelijk is het in het OVG in principe begonnen om dit laatste aantal en om de bijbehorende afgelegde afstanden.

Theoretisch zijn er verschillende manieren denkbaar om tot een goede schatting van die gegevens te komen.

Allereerst kan uit het totaal van alle verplaatsingen een random steekproef worden getrokken. Ook is echter een random steekproef mogelijk uit het totaal van alle afgelegde afstanden. Beide steekproeven leveren zonder veel complicaties het gezochte verplaatsingsbeeld op. De praktische haalbaarheid van dergelijke steekproeven is hier overigens geen punt van overweging. Tevens blijven daarmee allerlei steekproeftechnische finesses uit de sfeer van randomisatie en stratificatie, vooral bedoeld ter verhoging van de efficiency, achterwege.

In plaats van deze rechtstreekse benadering van het gezochte gegeven kan ook voor een indirecte methode worden gekozen. Deze ontstaat vanuit de

gedachte dat een verplaatsing steeds de verplaatsing van een verkeersmiddel betreft, lopen hierbij gezien als een vervoerswijze, zodat als alternatief een random steekproef kan worden getrokken uit het totaal van alle verkeersmiddelen.

De verplaatsing van de beschouwde verkeersmiddelen is evenwel bedoeld om de verplaatsing van personen. Dus kan ook uit de bevolking een random steekproef worden getrokken, die rekening houdend met de bezettingsgraad van de verkeersmiddelen, eveneens het gezochte beeld van de gemaakte verplaatsingen oplevert.

Tenslotte past het OVG een random steekproef van gezinnen toe, waarbij alle gezinsleden boven 11 jaar tot de te realiseren, getrapte bevolkingssteekproef behoren.

Met het indirecter worden van de invalshoek is er echter in toenemende mate sprake van de introductie van afhankelijkheden. Als gevolg hiervan is niet op grond van de gerealiseerde steekproefaantallen exact de grootte van de te verwachten steekproefspreadings aan te geven.

Zo is de spreiding in de correcties voor het misschatten gebaseerd op de aantallen gemelde verplaatsingen, ofschoon per persoon gemiddeld een aantal verplaatsingen zijn gemaakt en afstanden geschat in het subonderzoek Nauwkeurigheid opgegeven afstanden.

Om deze reden is het twijfelachtig of de ophoging van de aantallen verplaatsingen met de identieke, tot een vermindering zou leiden van de spreiding in het gevonden analyseresultaat.

Het heeft dan ook de voorkeur om de vermoedelijk toch al niet te groot berekende spreidingen te handhaven ten aanzien van de uitkomsten uit het totale steekproefbestand.

Overigens zij opgemerkt dat afhankelijkheden uiteraard ook in de werkelijkheid aanwezig zijn.