

ANALYSE VAN NAGEMETEN EN OPGEGEVEN AFSTANDEN UIT DE PROEFENQUÊTE  
ROVIN; EEN INTERIM-VERSLAG

R-79-57

J.M.J. Bos

Voorburg, 1979

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV

### 1. Doelstelling

Met de analyse wordt beoogd na te gaan of er sprake is van een systematisch onder- of overschatten van de lengten van afgelegde afstanden; welke in de enquête opgenomen variabelen op een dergelijke systematiek van invloed zouden zijn en hoe groot deze systematiek danwel is.

### 2. Voorbereiding

Van de in totaal 8.820 verplaatsingen uit de enquête waren er slechts 3.434 analyseerbaar. Vanwege de geringe aantallen konden daarna nog de vrachtauto- en de brommerverplaatsingen niet aan een verdere analyse worden onderworpen. Bij nog eens 19 verplaatsingen ontbrak een code voor enige analysevariabele, zodat ten slotte moest worden voortgegaan met 1.423 verplaatsingen per personenauto, 841 verplaatsingen per fiets en 1.024 verplaatsingen te voet. Deze verplaatsingen werden gemaakt door 531, respectievelijk 386 en 527 personen, zodat de schattingen van de afgelegde afstanden niet alle onafhankelijk van elkaar zullen zijn (zie bijlagen 1 en 2).

### 3. Werkwijze

De verplaatsingen zijn geanalyseerd met behulp van variantieanalyse. Deze techniek was evenwel niet rechtstreeks toepasbaar, omdat hij veronderstelt, dat de errors niet alleen ongecorrleerd, maar tevens normaal-verdeeld zijn en gelijke varianties hebben. De ongecorrleerdheid werd nagestreefd, ongetwijfeld zonder ze geheel te bereiken zoals hierboven reeds werd aangegeven, door bij meervoudige opgave van eenzelfde verplaatsing deze slechts eenmaal in het gegevensbestand op te nemen. Voor normale verdelingen en gelijke varianties moet worden gezorgd door de nagemeten en opgegeven afstanden aan een transformatie te onderwerpen. Deze transformatie dient eerst te worden gezocht.

#### 4. Transformatie

De keuze van een transformatie voor de afstanden uit de enquête wordt bepaald door de verdeling van de schattingsfout. Over deze verdeling zijn verschillende veronderstellingen mogelijk.

In het geval de spreiding van de schattingsfout onafhankelijk is van de grootte van de te schatten afstand, wordt de te analyseren score eenvoudig het verschil van nagemeten en opgegeven afstanden. Neemt de spreiding lineair toe met de geschatte afstand, dan wordt de te analyseren score berekend als het verschil van de logaritmische-getransformeerde nagemeten en opgegeven afstanden.

In veel gevallen, zoals ook bij Poisson-verdeelde grootheden, neemt de spreiding van de schattingsfout proportioneel toe met de wortel uit de schatting. In ons geval zou de te analyseren score dan het verschil zijn van de wortelgetransformeerde nagemeten en opgegeven afstanden (zie het hierover handelende en als bijlage 3 toegevoegde memo 72549).

Welke van de genoemde veronderstellingen over de verdeling van de schattingsfout het aannemelijkst is, zal beoordeeld worden aan de hand van daartoe gemaakte plots.

#### 5. Transformatiekeuze

Om de transformatie te kunnen zoeken die de variantie het beste stabiliseert, zijn de verschillen van zowel de niet-, de logaritmische-, als de wortel-getransformeerde nagemeten en opgegeven afstanden geplot tegen de nagemeten afstand.

Bij de plots van de verschillen tussen de nagemeten en opgegeven afstanden neemt de grootte van het spreidingsgebied toe naarmate de afgelegde afstand groter is.

De logaritmische-transformatie heeft echter tot gevolg dat het spreidingsgebied juist kleiner wordt, als de afgelegde afstand toeneemt.

Bij de worteltransformatie komt de spreiding van de te analyseren score het meest overeen met wat wordt nagestreefd. Deze transformatie zal danook verder toegepast worden.

Het beeld, dat uit de plots ontstaat, is voor de verschillende vervoerswijzen in essentie gelijk.

Opgemerkt zij nog, dat er volgens de plots van een systematische overschatting van de lengte van de afgelegde weg sprake zou kunnen zijn (zie de bijgevoegde plots 1 t/m 9).

## 6. Analysevariabelen

Vastgesteld is dus, dat de score die zal worden geanalyseerd het verschil is van de wortelgetransformeerde nagemeten en opgegeven afstanden.

De analyses zullen plaatsvinden voor elk van de vervoerswijzen: personenauto, fiets en te voet afzonderlijk.

Van de volgende variabelen zal de invloed op de zuiverheid van het schatten van afgelegde afstanden worden nagegaan:

- de nagemeten afstand, die daartoe op grond van de plots, met het oog op een redelijke verdeling van de aantallen te analyseren verplaatsingen, in drie klassen is opgedeeld. De klassegrenzen verschillen per vervoerswijze.
- leeftijd van de geënquêteerde (6 klassen).
- geslacht van de geënquêteerde (2 klassen).
- grootte van de gemeente waarin de geënquêteerde woont (5 klassen).
- bestuurder (bij verplaatsingen te voet uiteraard niet van toepassing) (2 klassen).
- reismotief (4 klassen).
- dagsoort (4 klassen). (Zie voor bijzonderheden en aantallen verplaatsingen tabellen 1 t/m 7).

Bij de analyses zal tevens naar interactie-effecten worden gezocht van de genoemde variabelen.

## 7. Analysemodel

In veel gevallen zal er in het analysemodel sprake zijn van onderling onafhankelijke variabelen. De totale variantie van de te analyseren score wordt dan zodanig opgesplitst, dat de varian-

tiecomponent die bij een te beschouwen variabele hoort kan worden aangeduid als het effect van het verschil in gemiddelde score tussen de onderscheidene klassen van die variabele. De variabelen van de proefenquête zullen echter stellig samenhangen, terwijl bovendien de aantallen verplaatsingen als gevolg van het toevalskarakter van de steekproef niet gelijkelijk over de klassen van de variabelen verdeeld zullen zijn. Nochtans blijft het ten dele mogelijk variantieporties te berekenen en die aan de variabelen toe te wijzen. Doordat er interacties tussen de variabelen optreden, wordt de totale variantie echter niet slechts in variantie-, maar ook in 'covariantie'-porties gedecomposeerd. Daarmee zou dan iets over de mate van samenhang van de variabelen gezegd kunnen worden. Mede evenwel omdat niet met alle mogelijke interacties voldoende rekening behoeft gehouden te kunnen worden, moet de interpretatie voorzichtiger zijn. Voor een beter begrip van deze methodische complicaties zij verwezen naar de literatuur dienaangaande (bv. Searl: Linear Models).

## 8. Uitgevoerde analyses

In eerste instantie hebben los van elkaar twee analyses op het enquêtemateriaal plaatsgevonden:

De eerste analyse betrof de variabele: nagemeten afstand en de variabelen die de geënquêteerde kenmerken: leeftijd, geslacht en gemeentegrootte.

De tweede analyse betrof de variabele: nagemeten afstand en de variabelen die de verplaatsing kenmerken: bestuurder, reismotief en dagsoort.

Van deze analyses zijn de variantie-covariantie-matrices, die betrekking hebben op de hoofdeffecten van de genoemde variabelen, gegeven in tabel 8.

Van de verdere resultaten zijn slechts de effecten van de belangrijkste 1e-orde interacties in tabel 14 opgenomen. Tabellen 15 en 16 vermelden nog gemiddelde scores.

De volgende analyse die op het materiaal plaatsvond bevatte alle variabelen en al hun 1e-orde interacties. Deze analyse is nog

niet geheel afgerond. Het belangrijkste resultaat staat in tabel 10.

### 9. Resultaten: invloedsvariabelen

De variantie-covariantie-matrices van tabel 8 geven op de diagonaal de grootte van de variantie, die door de variabele van de desbetreffende kolom 'verklaard' wordt, als alle variabelen van die matrix worden gefit. De andere elementen van de matrices bevatten de covarianties tussen de variabelen, weer als alle desbetreffende hoofdeffecten gefit worden. Variantie en covarianties bepalen samen de maximale betekenis van een variabele. Uit tabel 8 is zo berekend hoe groot voor de onderscheidene variabelen het hoofdeffect is als zij elk afzonderlijk worden gefit. Tabel 9 geeft een overzicht van de uitkomsten. Voor alle drie de beschouwde vervoerswijzen is de variabele: nagemeten afstand significant van invloed op de schattingsystematiek. Voor verplaatsingen te voet geldt dit eveneens voor de variabele: gemeentegrootte. Tabel 11 bevat de waarden van de F-toets. Hoe significant de gevonden effecten ook mogen zijn, zoals tabel 12 laat zien gaat het nochtans slechts om een marginale vermindering van de totale variantie. Tabel 10, met daarin de gezamenlijke hoofdeffecten van alle variabelen, overtuigt op dit punt zonodig nog meer.

Toch is nog onderzocht, of bij de verplaatsingen te voet de variabelen: nagemeten afstand en gemeentegrootte niet zodanig correleren, dat de variabelen weliswaar elk afzonderlijk significant zijn, maar niet beide tegelijk. Tabel 13 maakt duidelijk, dat dit niet het geval is.

Voortgezette analyse bracht ook aan het licht, dat er belangrijke effecten van 1e-orde interacties van de variabelen optraden. Het gaat hier, zoals tabel 14 te zien geeft, vooral om de interactie van de variabelen: nagemeten afstand en gemeentegrootte. Bij alle drie de beschouwde vervoerswijze is deze interactie significant, terwijl de door dit effect 'verklaarde' variantie ook in alle gevallen groter is dan die waarvoor de eerdere variabele: nagemeten afstand stond, zoals een vergelijking met de getalwaarden van tabel 9 duidelijk maakt.

## 10. Schattingsonzuiverheid

Zowel bij de vervoerswijze: personenauto als bij de vervoerswijze: fiets is er sprake van een significante overschatting van de afgelegde afstand. De overschatting bij verplaatsingen te voet is niet significant, zoals blijkt uit tabel 15.

Met welke kilometerwaarden deze overschattingen overeenkomen is nog niet bekend.

Ten opzichte van de spreiding in de waarnemingen zijn de effecten echter niet opvallend groot, terwijl ze bovendien sterk verband houden met de samenstelling van de steekproef, gegeven dat bepaalde variabelen op de mate van overschatting een significante invloed hebben.

In tabel 16 staan nog de gevonden klassegemiddelde scores voor de significant bevonden variabele: nagemeten afstand. De waarden zijn in grafiek 1 ingetekend. Gegeven de geringe variantie die deze variabele weet te 'verklaren', kan aan dergelijke gemiddelden, die evenals het algemeen gemiddelde tevens samenhangen met de steekproefsamenstelling, niet veel betekenis worden toegekend.

## 11. Conclusies

Oschoon er wel variabelen zijn gevonden die op de schattingssystematiek significante invloed uitoefenen, met name de variabele: nagemeten afstand en de interactie van de variabelen: nagemeten afstand en gemeentegrootte, is hun feitelijke betekenis slechts marginaal. Het steekproefmodel zorgt daarbij nog voor interpretatieproblemen: doordat het 'design' niet ortogonaal is hangen de variabelen met alle andere samen.

Gevonden is ook, dat er behalve bij verplaatsingen te voet sprake is van een significante, maar eveneens tamelijk geringe overschatting van de afgelegde afstand. Dit resultaat hangt evenwel sterk af van de samenstelling van de steekproef. Voor de juiste waarden en een goede interpretatie is hier bovendien nodig, dat ook meervoudig opgegeven verplaatsingen in de berekeningen betrokken zijn. De desbetreffende uitkomsten zijn nog niet beschikbaar.

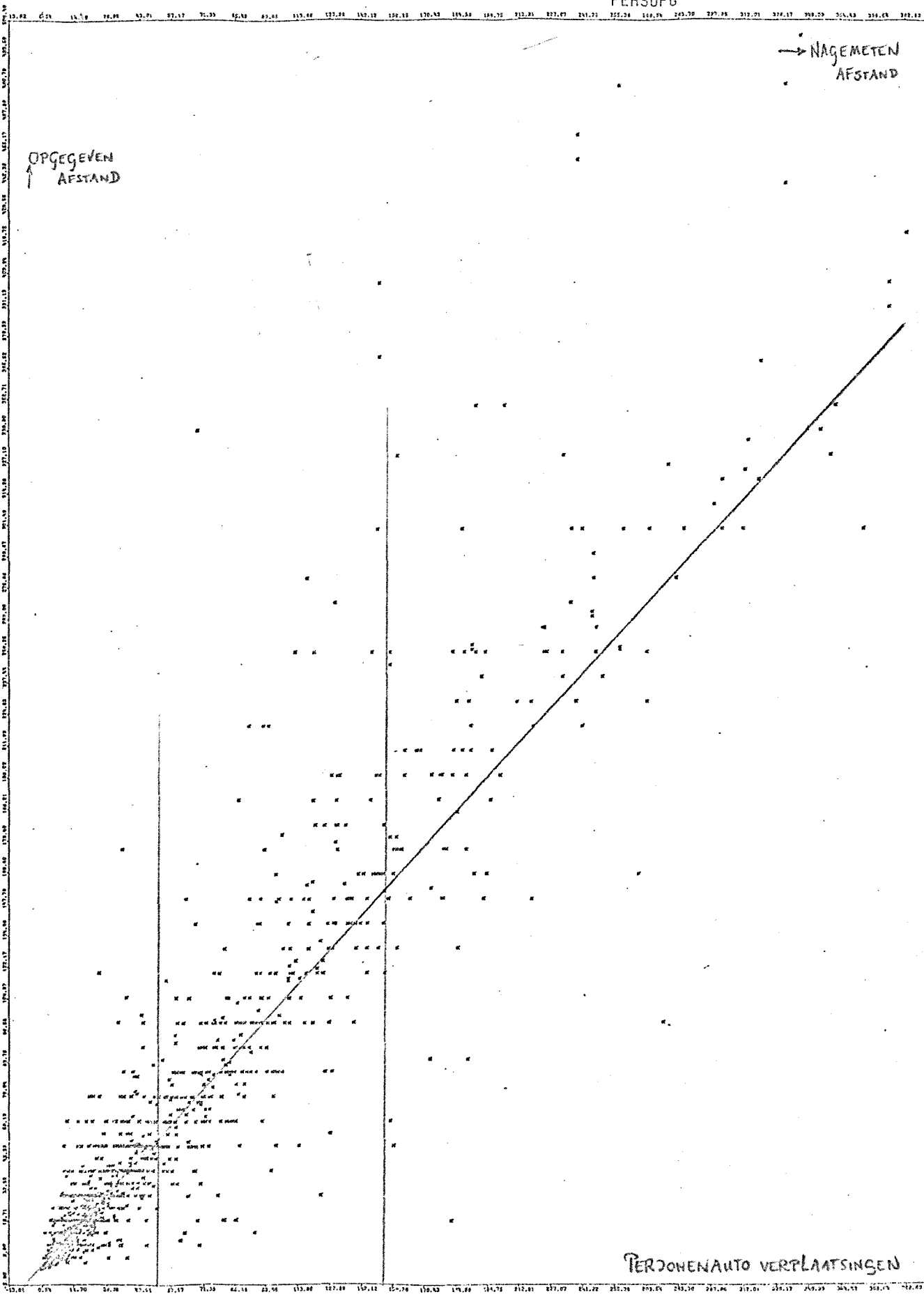
Tenslotte moet eraan worden herinnerd, dat de gegeven conclusies zijn bereikt met minder dan de helft van het totale enquêtemateriaal. De betekenis van de bevindingen moet danook danig worden gerelativeerd.

## 12. Correctiefactoren

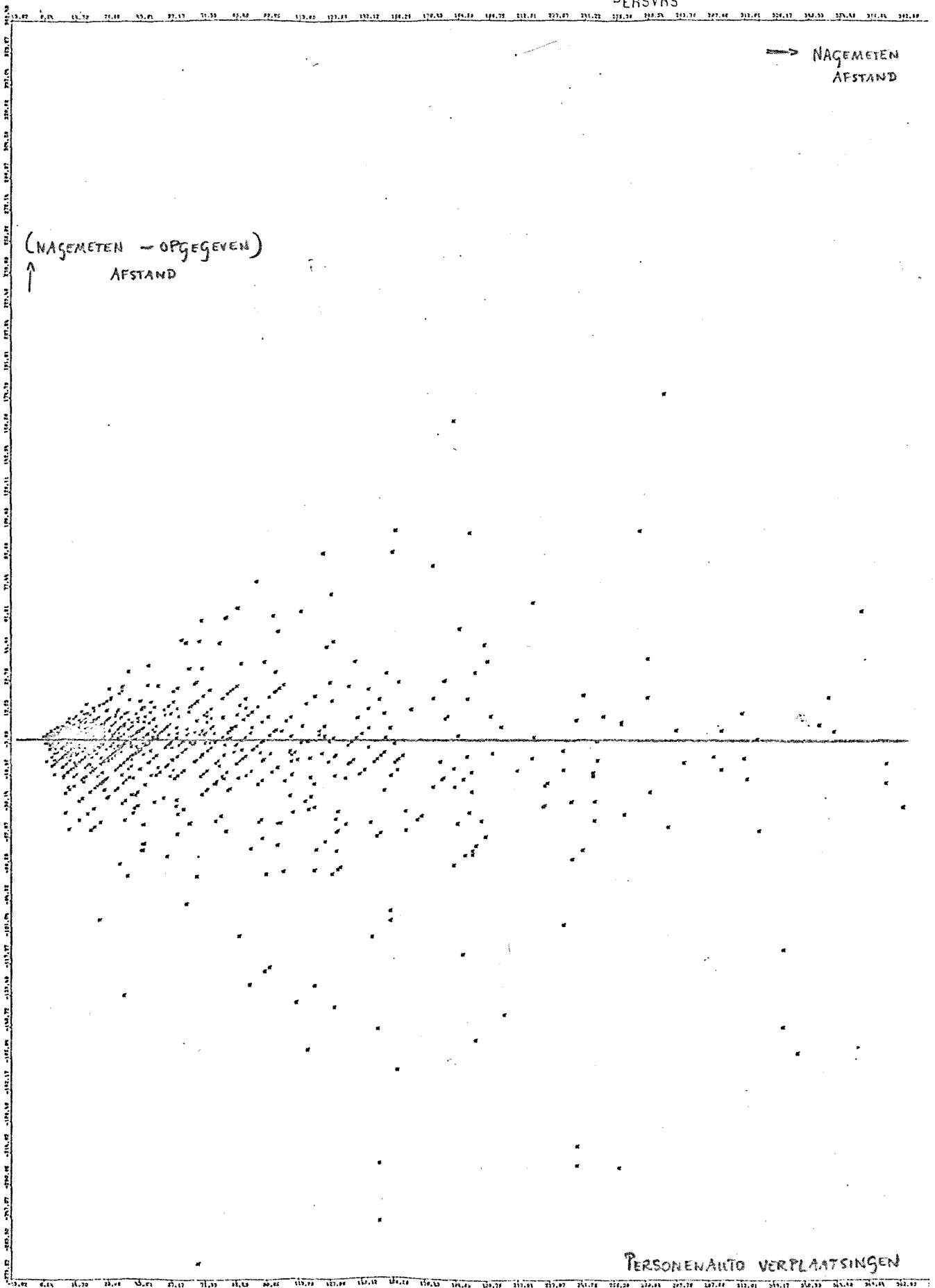
Uit het voorgaande moge duidelijk zijn, dat correcties op de afstandsschattingen uit de CBS-enquête, met behulp van proefenquêteresultaten, niet aan te bevelen zijn. Volgens de analyse-uitkomsten in engere zin, zou het effect van dergelijke correcties bovendien niet meer dan marginale relevantie kunnen hebben. Op zich is het trouwens geenszins zonder belang de omvang van systematische fouten te kennen, al zal altijd de nodige voorzichtigheid betracht dienen te worden met het aanpassen van waarnemingen. Naast de reeds genoemde argumenten zijn er nog andere die voornog tegen het doorvoeren van correctie pleiten. Met name zij genoemd, dat de proefenquête in november en december 1976 plaatsvond, terwijl de CBS-enquête continu is, dat vragenlijsten verschillen, enquêteerprocedures verschillen en dat non-responsen en derhalve responsen verschillen (zie ook memo 68767).  
Op een aantal punten zal het validiteitsonderzoek van november en december 1978, dat nauwer bij de CBS-steekproef aansluit, stellig een verbetering kunnen zijn.)



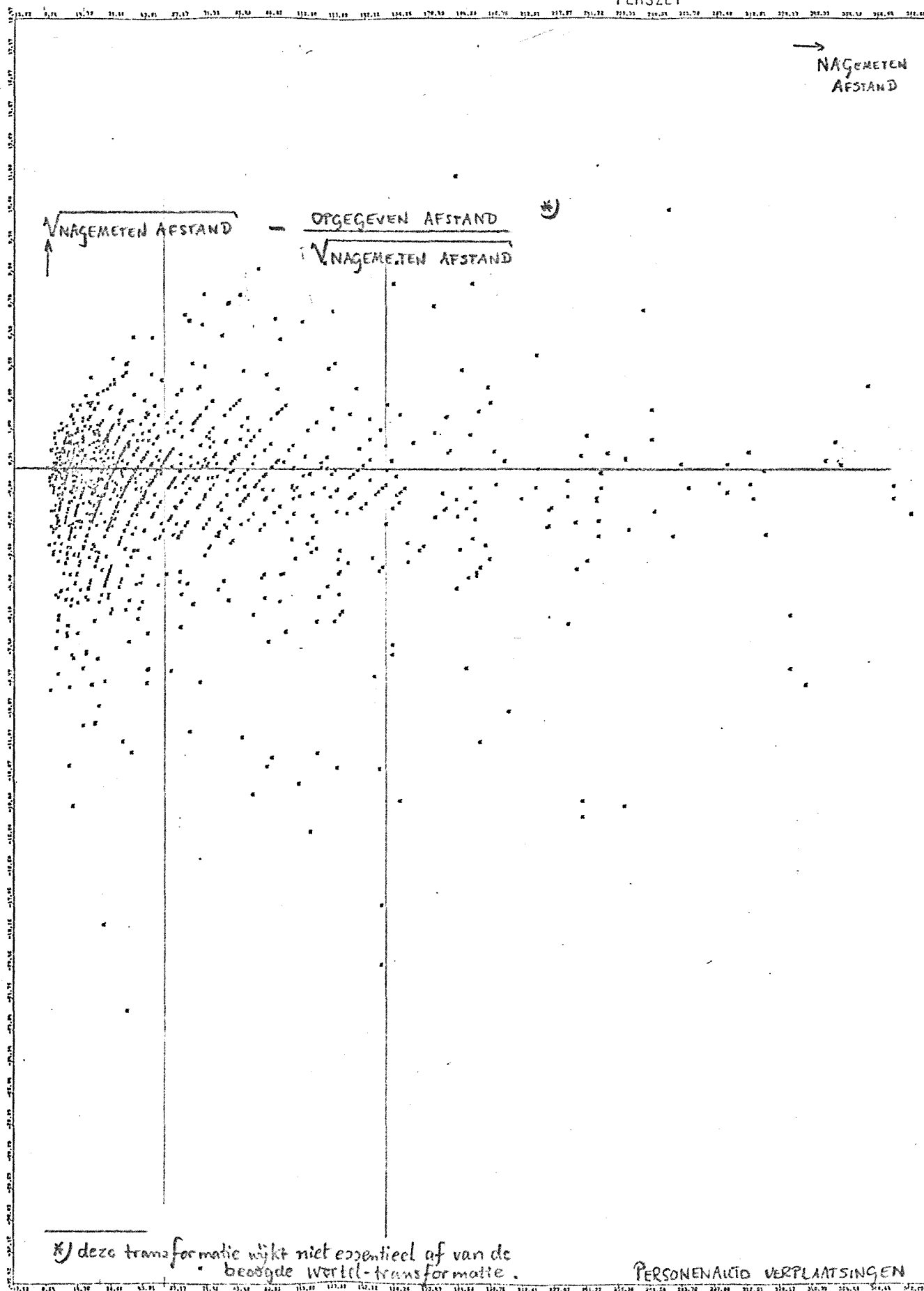
PERSOOP



PERSVRS



PERSZET



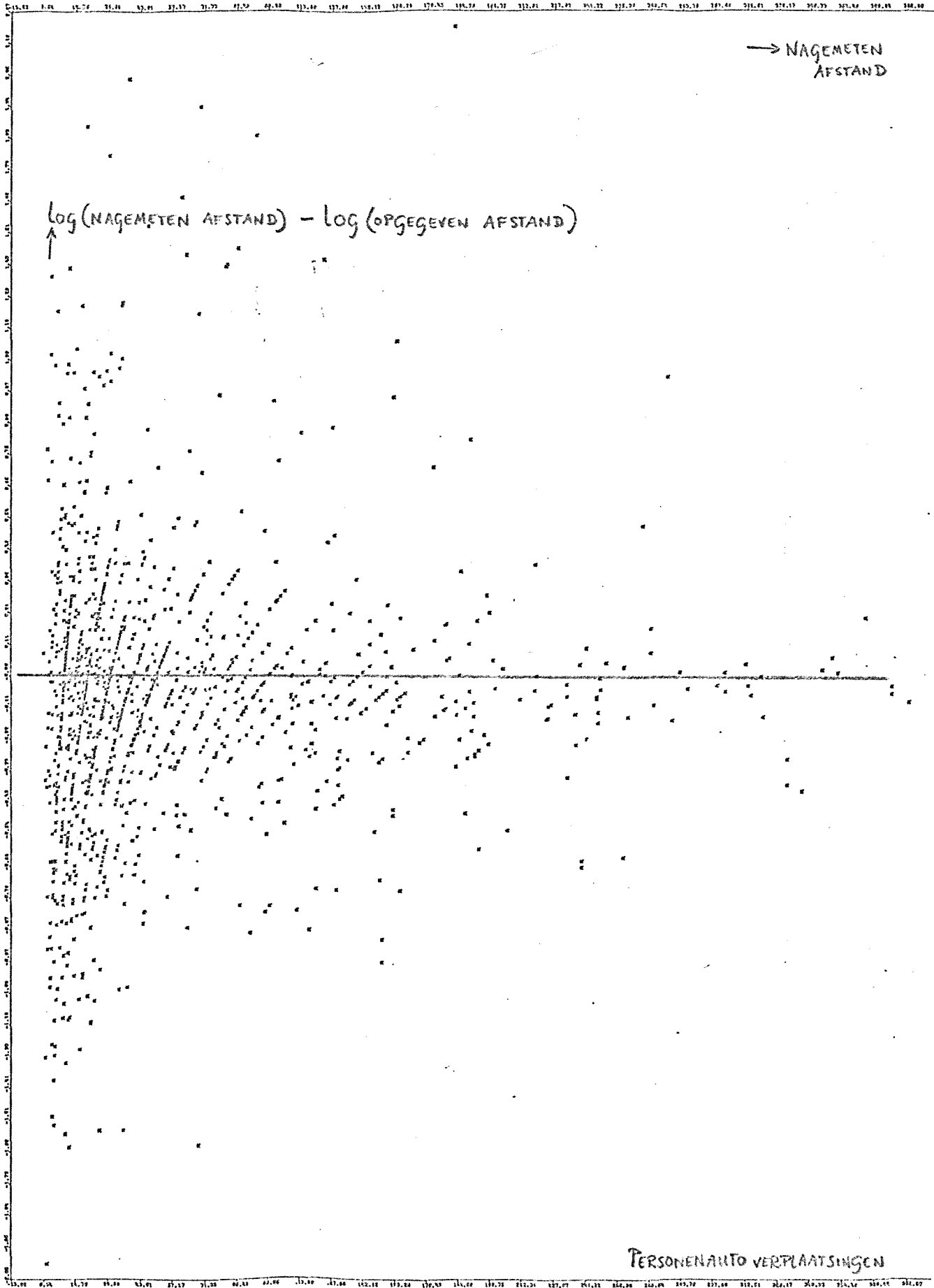
→ NAGEMETEN AFSTAND

NAGEMETEN AFSTAND  
↑

OPGEGEVEN AFSTAND  
NAGEMETEN AFSTAND

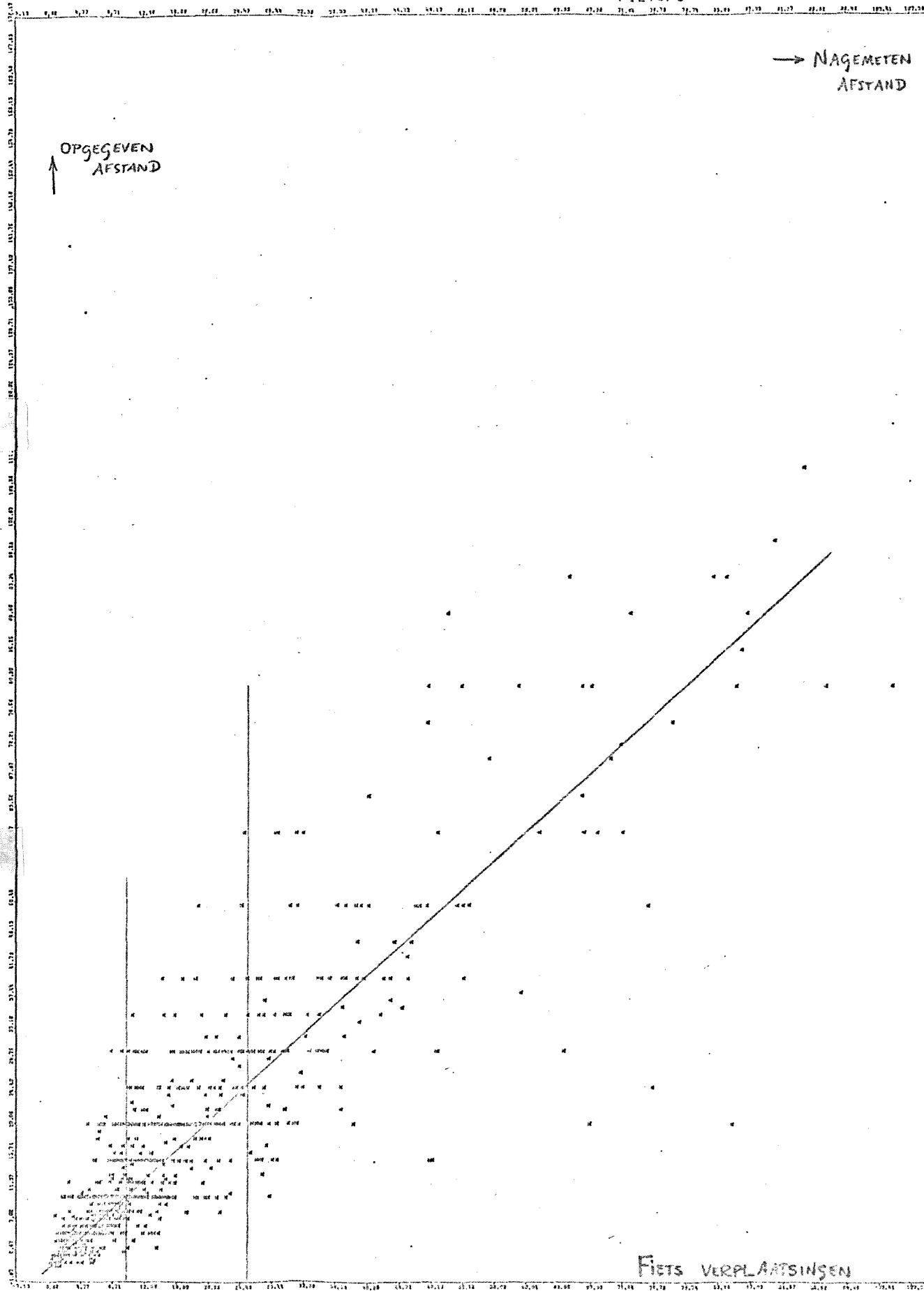
\*) deze transformatie wijkt niet essentieel af van de beoogde wortel-transformatie.

PERSONENAUTO VERPLAATSINGEN

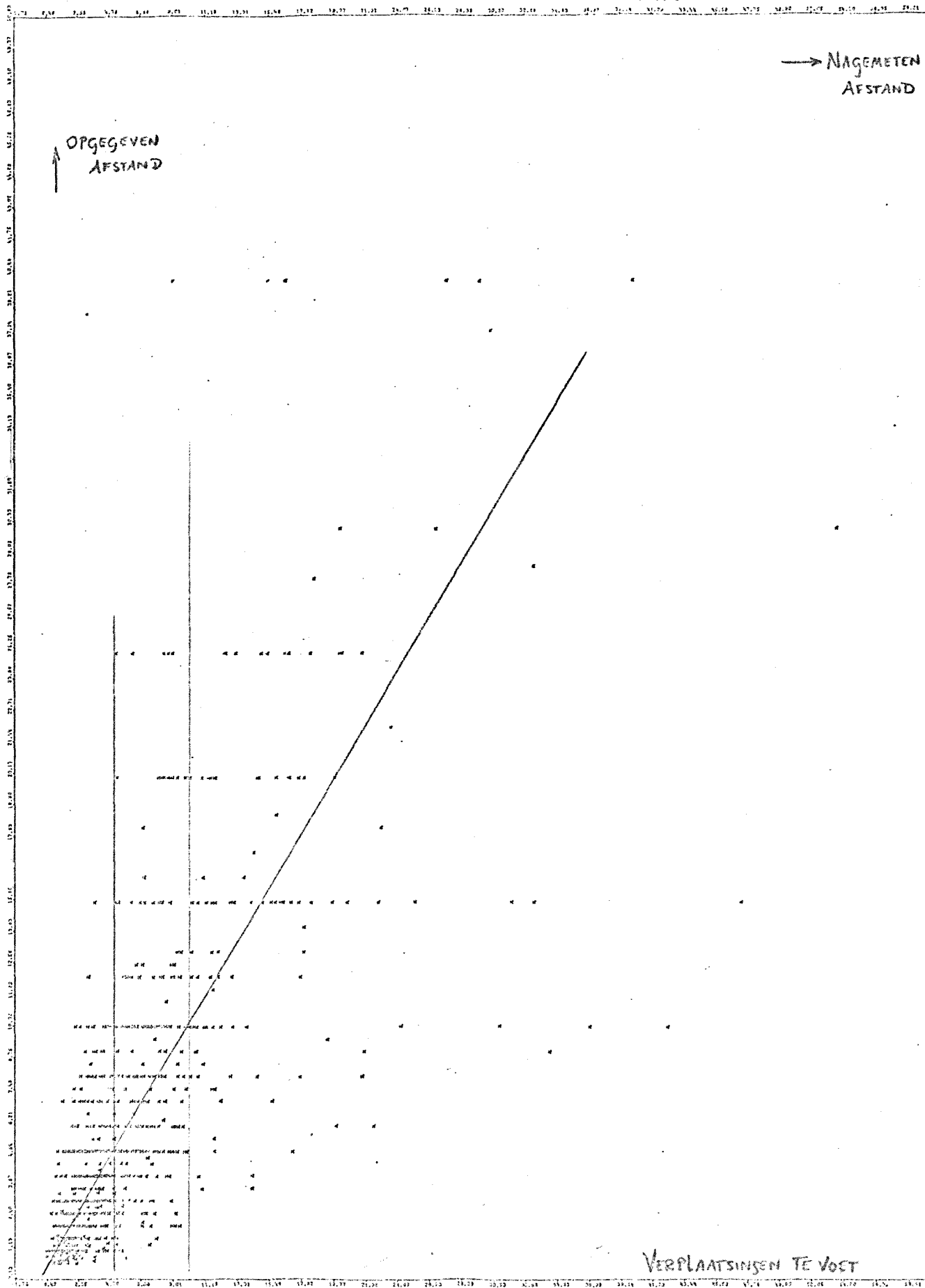


PERSONENAUTO VERPLAATSINGEN

F I E T O P G



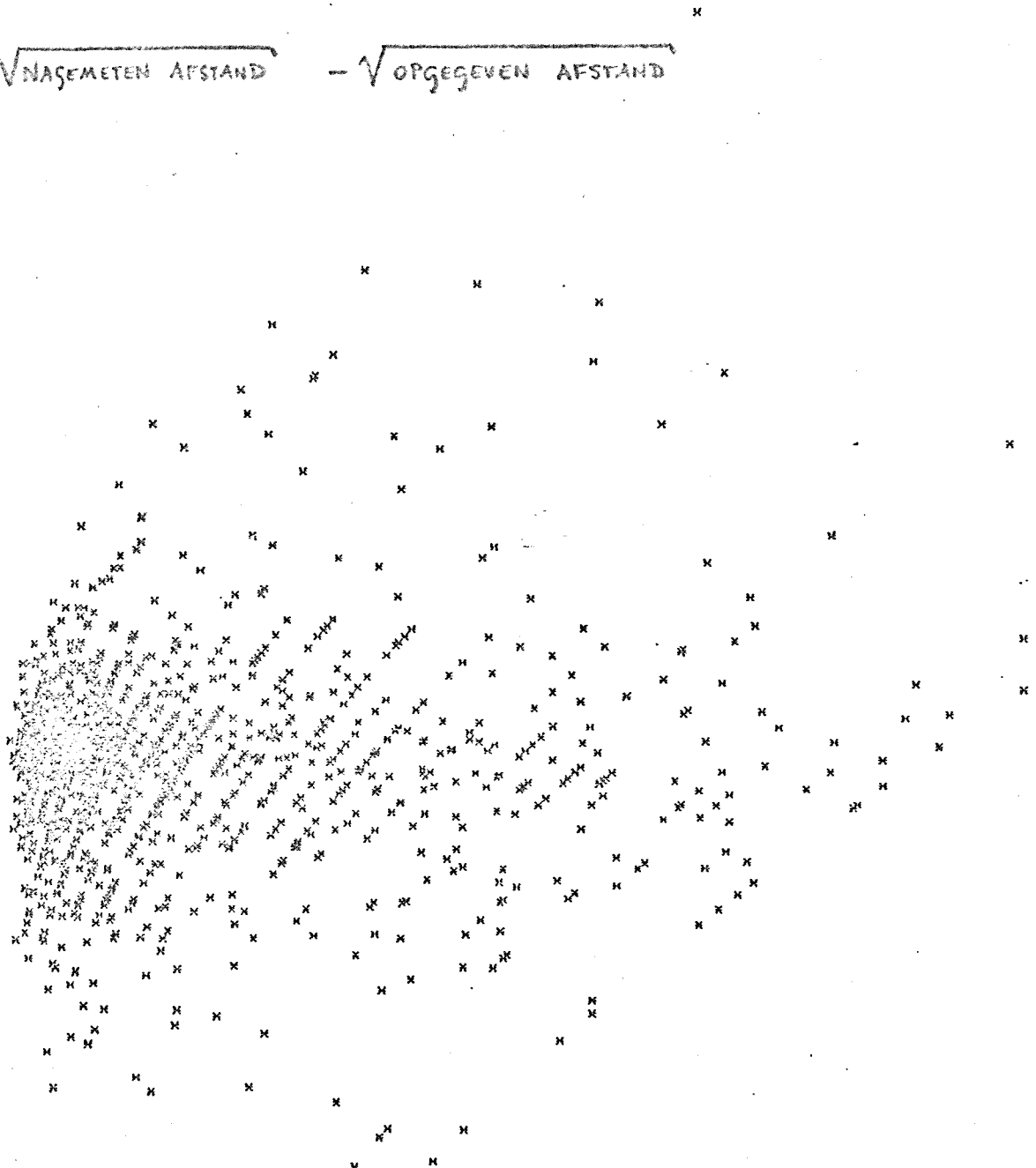
FIETS VERPLAATSINGEN



VERPLAATSINGEN TE VOET

→ NAGEMETEN AFSTAND

$\sqrt{\text{NAGEMETEN AFSTAND}}$  -  $\sqrt{\text{VOPGEGEVEN AFSTAND}}$



→ NAGEMETEN  
AFSTAND

↑  $\sqrt{\text{NAGEMETEN AFSTAND} - \text{VOPGEGEVEN AFSTAND}}$



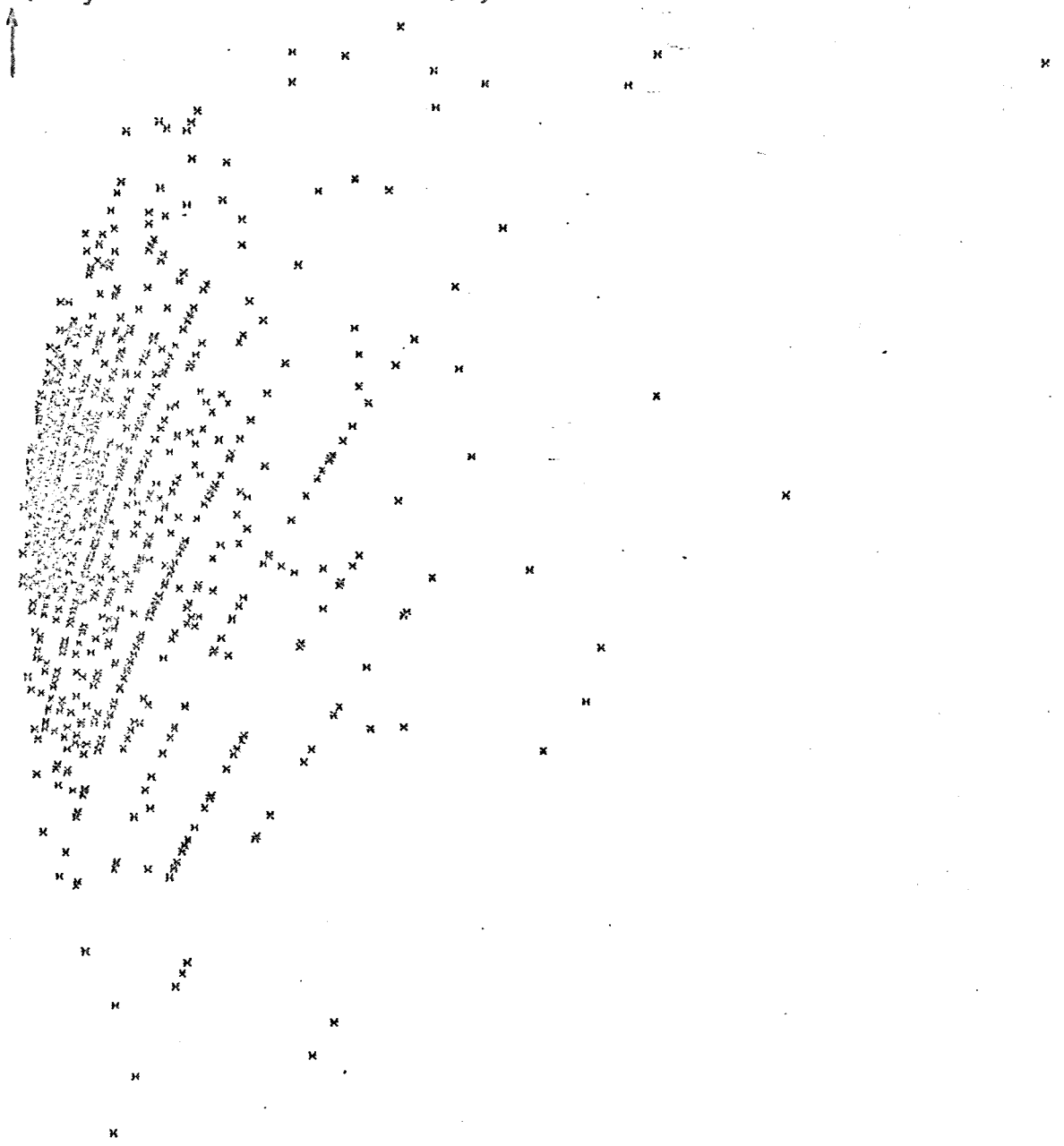
FIETS VERPLAATSINGEN



VOETSORT

→ NAGEMETEN  
AFSTAND

$\sqrt{\text{NAGEMETEN AFSTAND}} - \sqrt{\text{OPGEGEVEN AFSTAND}}$



VERLAATSIJGEN TE VOET

Aantallen verplaatsingen	Nagemeten afstand:		
	≤ 5 km	> 5 en ≤ 15 km	> 15 km
Personenauto	897	339	187
	≤ 1 km	> 1 en ≤ 2.5 km	> 2.5 km
Fiets	358	319	164
	≤ .5 km	> .5 en ≤ 1 km	> 1 km
Te voet	602	291	131

TABEL 1. Klasseindelingen en bijbehorende aantallen verplaatsingen voor de variabele: nagemeten afstand.

Aantallen verplaatsingen	Leeftijd:					
	≤ 9 jr.	10 t/m 14 jr.	15 t/m 19 jr.	20 t/m 24 jr.	25 t/m 49 jr.	> 50 jr.
Personenauto	158	57	51	192	653	312
Fiets	94	135	146	86	232	148
Te voet	208	113	96	51	306	250

TABEL 2. Klasseindeling (als in proefenquête) en bijbehorende aantallen verplaatsingen voor de variabele: leeftijd.

Aantallen verplaatsingen	Geslacht:	
	man	vrouw
Personenauto	872	551
Fiets	374	467
Te voet	449	575

TABEL 3. Aantallen verplaatsingen voor de variabele: geslacht.

Aantallen verplaatsingen	Grootte van de gemeente van inwoning volgens in- wonersaantal:				
	<12.500	12.500-	25.000-	75.000-	≥250.000
		25.000	75.000	250.000	
Personenauto	154	329	273	498	169
Fiets	102	229	146	273	91
Te voet	92	224	153	342	213

TABEL 4. Klasseindeling (als in proefenquête) en bijbehorende aantallen verplaatsingen voor de variabele: gemeentegrootte.

Aantallen verplaatsingen	Bestuurder:	
	bestuurder	passagier
Personenauto	870	553
Fiets	774	67
Te voet	n.v.t.	

TABEL 5. Aantallen verplaatsingen voor de variabele: bestuurder.

Aantallen verplaatsingen	Reismotief:			
	werken	onderwijs volgen	sociaal, recreatief	overig
Personenauto	208	21	639	555
Fiets	94	108	268	371
Te voet	40	110	453	421

TABEL 6. Klasseindeling (als in proefenquête) en bijbehorende aantallen verplaatsingen voor de variabele: reismotief.

Aantallen verplaatsingen	Dagsoort:			
	ma. t/m do.	vr.	za.	zo.
Personenauto	533	151	405	334
Fiets	483	86	183	89
Te voet	525	117	194	188

TABEL 7. Klasseindeling (als in proefenquête) en bijbehorende aantallen verplaatsingen voor de variabele: dagsoort.

Variantie- covariantie- matrices	Variabele:				Residuele variantie
	nagemeten afstand	leeftijd	geslacht	gemeente- grootte	
Personenauto	2.1659				
	.0159	2.4538			
	-.0041	.0102	.0624		
	.0446	-.0232	.0063	1.8283	
	-.0000	.0000	.0000	.0000	328.2845
Fiets	1.1653				
	.0631	.4481			
	.0534	-.0170	.1821		
	-.0393	.0439	-.0051	.4961	
	.0001	-.0000	.0000	.0000	60.8064
Te voet	.9418				
	-.0165	.6054			
	.0158	.0080	.0687		
	.0142	.0437	.0060	.5872	
	-.0000	-.0000	.0000	.0000	48.5118
Personenauto		nagemeten	bestuur-	reismo-	dagsoort
		afstand	der	tief	
	2.1090				
	.1448	1.6927			
	-.0012	-.3261	.8064		
-.0536	-.1740	.1092	.8455		
.0001	-.0006	.0000	.0000	330.0433	
Fiets	1.3224				
	.0392	.1890			
	.0364	.0246	.1562		
	-.0673	.0006	.0492	.7843	
	-.0000	.0000	-.0000	.0000	60.6787
Te voet	1.0911				
	-.0427		.2160		
	-.0199	n.v.t.	-.0084	.2782	
	-.0000		-.0000	.0000	49.4136

TABEL 8. De belangrijkste variantie-covariantie-matrices.

'Verklaarde' varianties	Hoofdeffect van de afzonderlijke variabele:						
	nagemeten afstand	leef- tijd	ge- slacht	gemeente- grootte	bestuur- der	reis- motief	dag- soort
Personenauto	2.30 <sup>*)</sup>	2.48	.09	1.88	1.04	.22	.61
Fiets	1.34 <sup>*)</sup>	.65	.24	.50	.33	.39	.75
Te voet	.97 <sup>*)</sup>	.68	.13	.72 <sup>*)</sup>	n.v.t.	.12	.22
Aantal vrij- heidsgraden	2	5	1	4	1	3	3

**TABEL 9.** Door de hoofdeffecten 'verklaarde' varianties, als zij afzonderlijk worden gefit.

\*) volgens F-toets op 99%-sniveau significant (zie tabel 11).

'Verklaarde' varianties	Alle hoofdeffecten tegelijk	Alle hoofdeffecten èn hun 1e-orde interacties met nagemeten afstand tegelijk	Totale variantie	Aantal vrijheids- graden
Personenauto	9.52	20.94	334.89	1422
Fiets	3.43	7.78	63.30	840
Te voet	nog niet beschikbaar		50.86	1023

**TABEL 10.** Door de hoofdeffecten 'verklaarde' varianties, als zij alle tegelijk en als zij alle inclusief hun 1e-orde interacties met nagemeten afstand tegelijk worden gefit. Ter vergelijking is tevens de totale variantie vermeld.

F-waarden	Hoofdeffect van de afzonderlijke variabele:						
	nagemeten afstand	leef- tijd	ge- slacht	gemeente- grootte	bestuur- der	reis- motief	dag- soort
Personenauto	4.91 <sup>*)</sup>	2.11	.38	2.00	4.43	.31	.86
Fiets	9.06 <sup>*)</sup>	1.73	3.18	1.66	4.39	1.73	3.35
Te voet	9.93 <sup>*)</sup>	2.76	2.61	3.66 <sup>*)</sup>	n.v.t.	.80	1.47
Aantal vrij- heidsgraden	2	5	1	4	1	3	3

TABEL 11. Waarden van de F-toets voor de hoofdeffecten, als zij afzonderlijk worden gefit.

\*) op 99%-sniveau significant (zie tabel 9).

% en 'verklaarde' variantie	Hoofdeffect van de afzonderlijke variabele:						
	nagemeten afstand	leef- tijd	ge- slacht	gemeente- grootte	bestuur- der	reis- motief	dag- soort
Personenauto	.69	.74	.03	.56	.31	.07	.18
Fiets	2.12	1.03	.38	.79	.52	.62	1.18
Te voet	1.91	1.34	.26	1.42	n.v.t.	.24	.43

TABEL 12. Percentages door de hoofdeffecten 'verklaarde' variantie, als zij afzonderlijk worden gefit.

Additioneel hoofdeffect van de variabele: gemeentegrootte, gegeven dat de variabele: nagemeten afstand is gefit:		
Additioneel 'verklaarde' variantie	Aantal vrijheids- graden	F- waarde
Te voet	.69 <sup>*)</sup>	4
		3.57 <sup>*)</sup>

TABEL 13. Bij vervoerswijze: te voet door het hoofdeffect van de variabele: gemeentegrootte additioneel 'verklaarde' variantie, nadat het hoofdeffect van de variabele: nagemeten afstand is gefit.

\*) op 99%-sniveau significant.

Effect, zonder hoofdeffecten, van de 1e-orde interactie van de variabelen:						
	nagemeten afstand en gemeentegrootte:			leeftijd en gemeentegrootte		
	'Verklaarde' variantie	Aantal vrij- heids- graden	F- waarde	'Verklaarde' variantie	Aantal vrij- heids- graden	F- waarde
Personenauto	5.43 <sup>*)</sup>	8	2.91 <sup>*)</sup>	9.17 <sup>*)</sup>	20	1.97 <sup>*)</sup>
Fiets	1.52 <sup>*)</sup>	8	2.56 <sup>*)</sup>	1.13	20	.75
Te voet	1.82 <sup>*)</sup>	8	4.71 <sup>*)</sup>	1.58	20	1.61

TABEL 14. Door het eerste-orde interactie-effect van de variabelen: nagemeten afstand en gemeentegrootte en door dat van de variabelen: leeftijd en gemeentegrootte 'verklaarde' varianties, als geen hoofdeffecten worden gefit.

\*) op 99%-sniveau significant.



Nulde-orde effect:				
	Algemeen gemiddelde score	Aantal vrijheids- graden	t- waarde ( $\mu=0$ )	Spreiding <sup>1)</sup> van de meetpunten
Personenauto	-.1128 <sup>*</sup> )	1422	8.76 <sup>*</sup> )	.49
Fiets	-.0448 <sup>*</sup> )	840	4.73 <sup>*</sup> )	.27
Te voet	-.0102	1023	1.46	.22

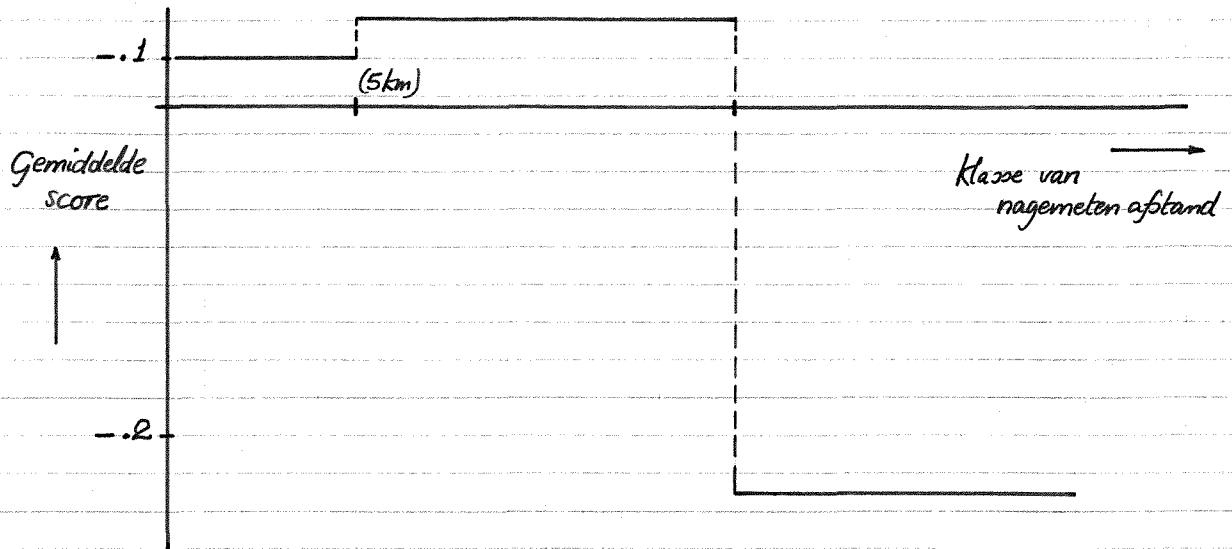
TABEL 15. Algemeen gemiddelde score van het verschil tussen de wortelge-transformeerde enquêtewaarden voor nagemeten en opgegeven afstanden.

<sup>\*</sup>) op 99%-sniveau significant. 1) d.i. de wortel uit de variantie.

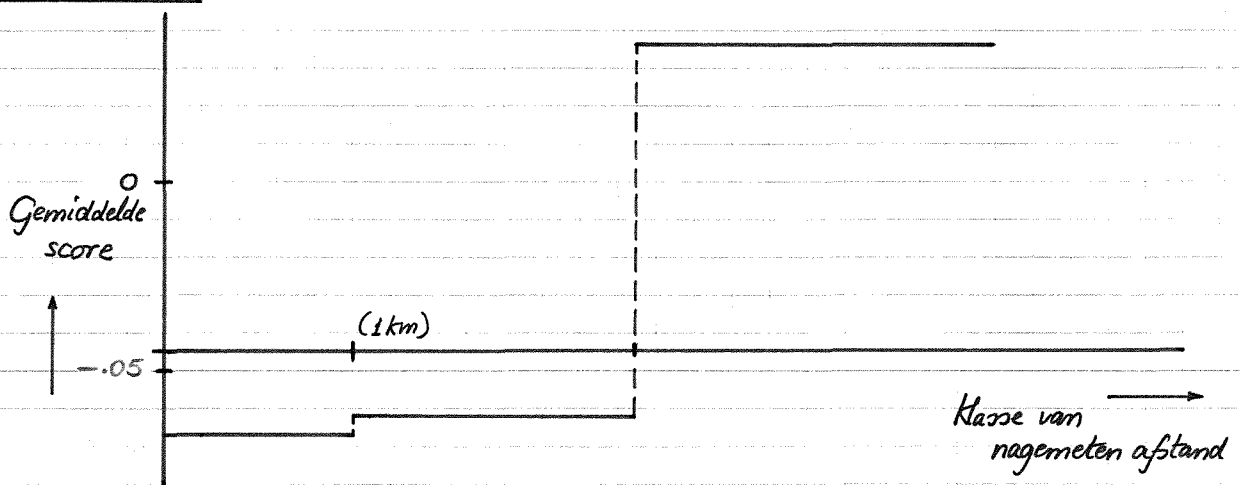
Klassegemiddelde scores t.o.v. het algemeen gemiddelde	Nagemeten afstand:			
	≤ 5 km	>5 en	≤ 15 km	> 15 km
Personenauto	.0126	.0233	-.1028	
	≤ 1 km	>1 en	≤ 2.5 km	>2.5 km
Fiets	-.0214	-.0176	.0810	
	≤ .5 km	>.5 en	≤ 1 km	> 1 km
Te voet	-.0157	-.0032	.0791	

TABEL 16. Klassegemiddelde scores van het verschil tussen de wortelge-transformeerde enquêtewaarden voor nagemeten en opgegeven afstanden, als het hoofdeffect van de variabele: nagemeten afstand wordt gefit. De gemiddelden staan in afwijking van het algemeen gemiddelde (zie tabel 15) en zijn op 99%-sniveau significant (zie tabel 9).

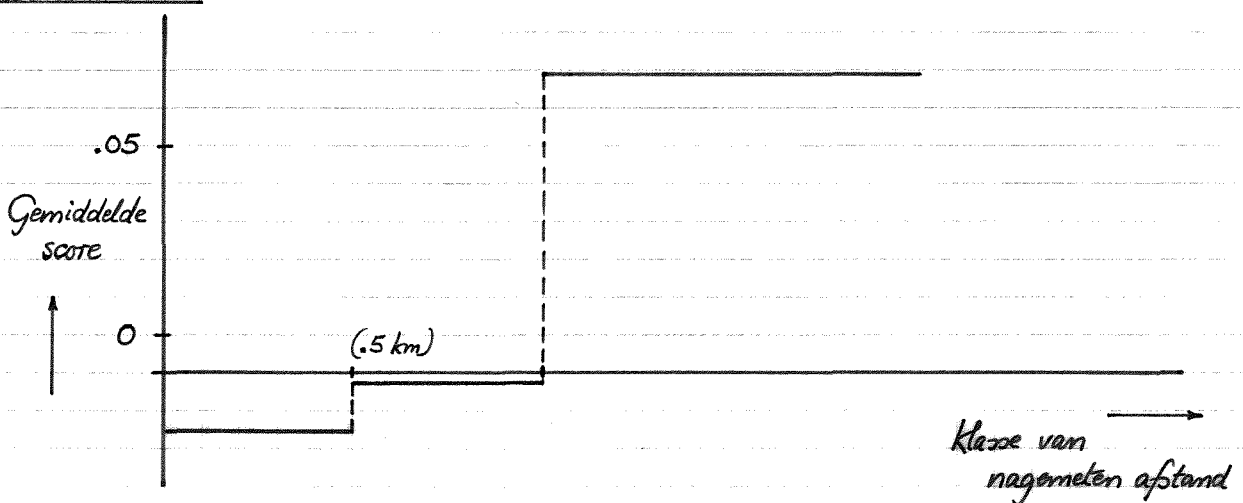
### Personenauto



### Fiets



### Te voet



**GRAFIEK 1.** Klassegemiddelden en algemeen gemiddelde van het verschil tussen de wortelgetransformeerde enquêtewaarden voor nagemeten en opgegeven afstanden, als het hoofdeffect van de variabele: nagemeten afstand wordt gefit (zie tabellen 15 en 16).

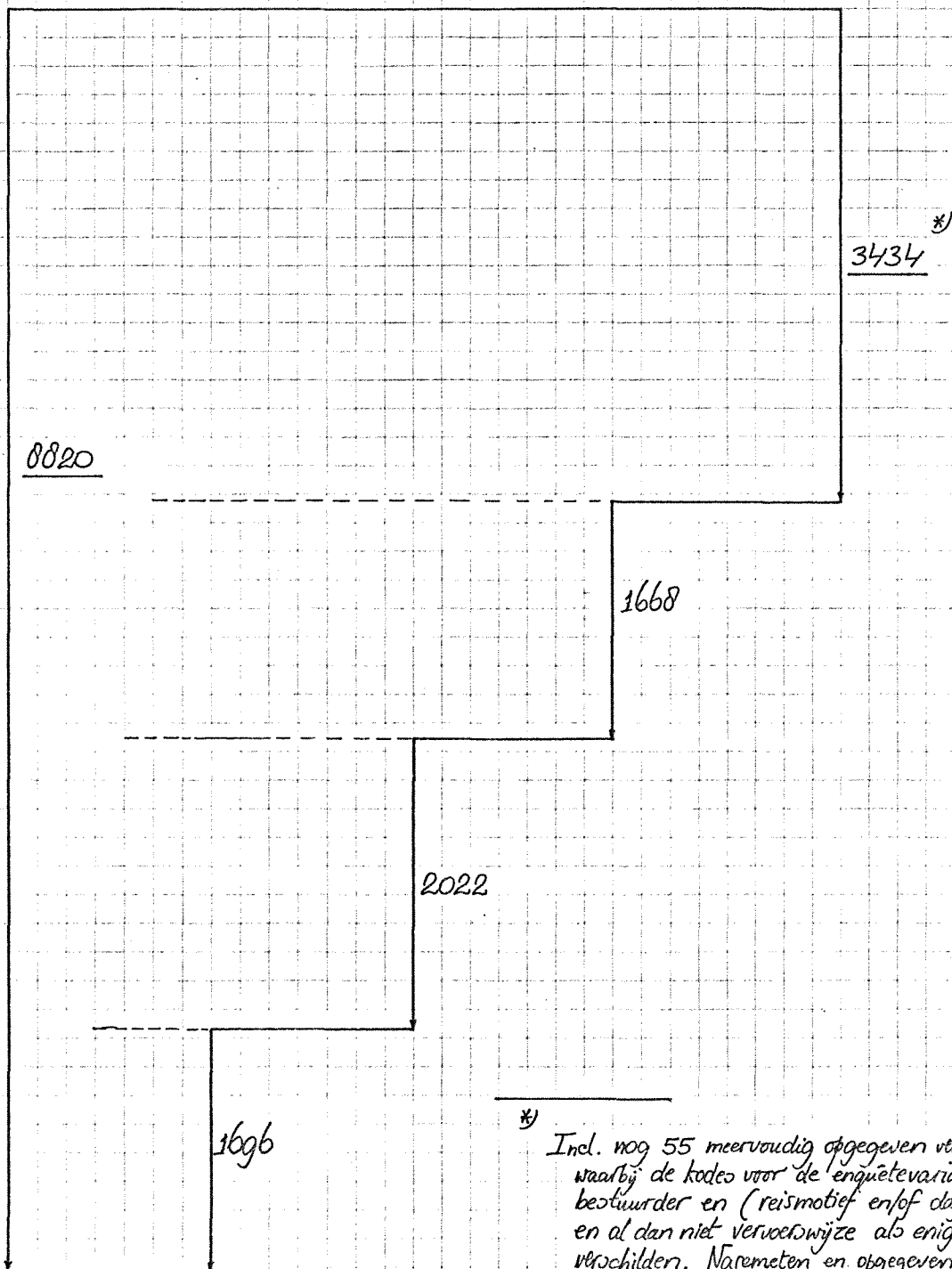
Totaal aantal verplaatsingen.

Afgevallen door andere vervoerswijze.

Afgevallen door onbekende afstanden.

Afgevallen door meervoudige opgave.

Totaal aantal analyseerbare verplaatsingen.



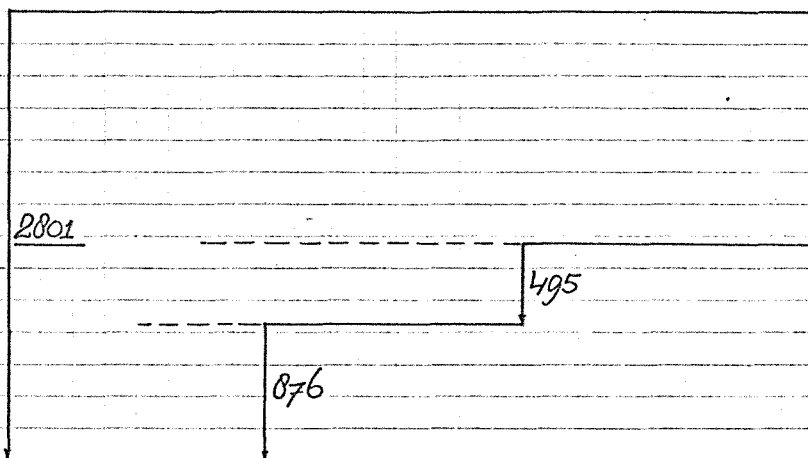
\* Incl. nog 55 meervoudig opgegeven verplaatsingen, waarbij de codes voor de enquêtevariabelen bestuurder en (reismotief en/of dagvoort) en al dan niet vervoerswijze als enige verschillen. Naarmeten en opgegeven afstanden, leeftijd, geslacht en grootte van de gemeente van inwoning zijn dan nog gelijk.

Aantallen  
verplaatsingen.

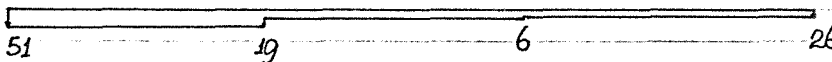
Afgevallen door  
onbekende afstanden.

Afgevallen door  
meervoudige opgave.

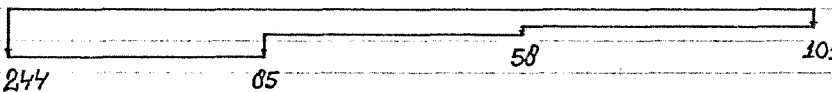
Aantallen  
geanalyseerde verplaatsingen.



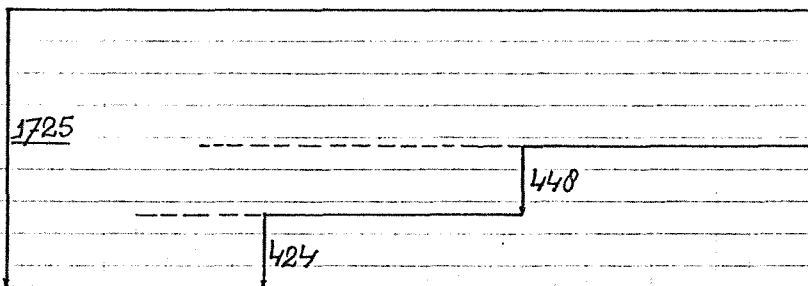
Personenauto-  
verplaatsingen,  
no. nog 7 met  
enige onbestemde  
kode.



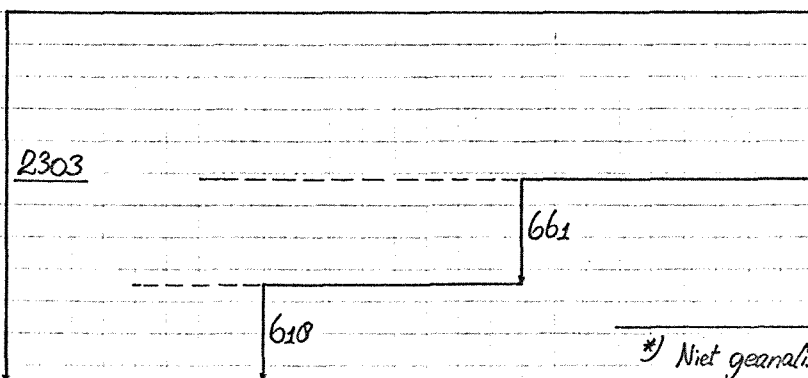
Vrachtauto-  
verplaatsingen.



Brommer-  
verplaatsingen.



Fiets-  
verplaatsingen,  
no. nog 12 met  
enige onbestemde  
kode.



Te voet-  
verplaatsingen.

\*) Niet geanalyseerd vanwege te geringe aantallen.

ROVIN. Over datatransformaties t.b.v. variantiestabilisatie.

1. Stel dat het schatten van een gegeven afgelegde afstand  $\delta_g$  zodanig geschiedt dat geldt:

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

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:

$$\begin{aligned} \mathcal{E} \log d_g &= \mathcal{E} \log(\delta_g + s(\delta_g)) - \mathcal{E} \sum \frac{1}{2n} \Delta_g^{2n} + \mathcal{E} \sum \frac{1}{2n-1} \Delta_g^{2n-1} \approx \\ &\approx \log \delta_g + \frac{s(\delta_g)}{\delta_g} - \mathcal{E} \frac{1}{2} \Delta_g^2 \text{ met } \left| \frac{s(\delta_g)}{\delta_g} \right| \ll 1, \text{ zodat tenslotte} \end{aligned}$$

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:

$\sigma_g = c d_g \approx c(\delta_{g+s}(\delta_g))$ , 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  $\sqrt{\cdot}$ -transformatie, waarbij de verdelingseigenschappen als volgt zijn:

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

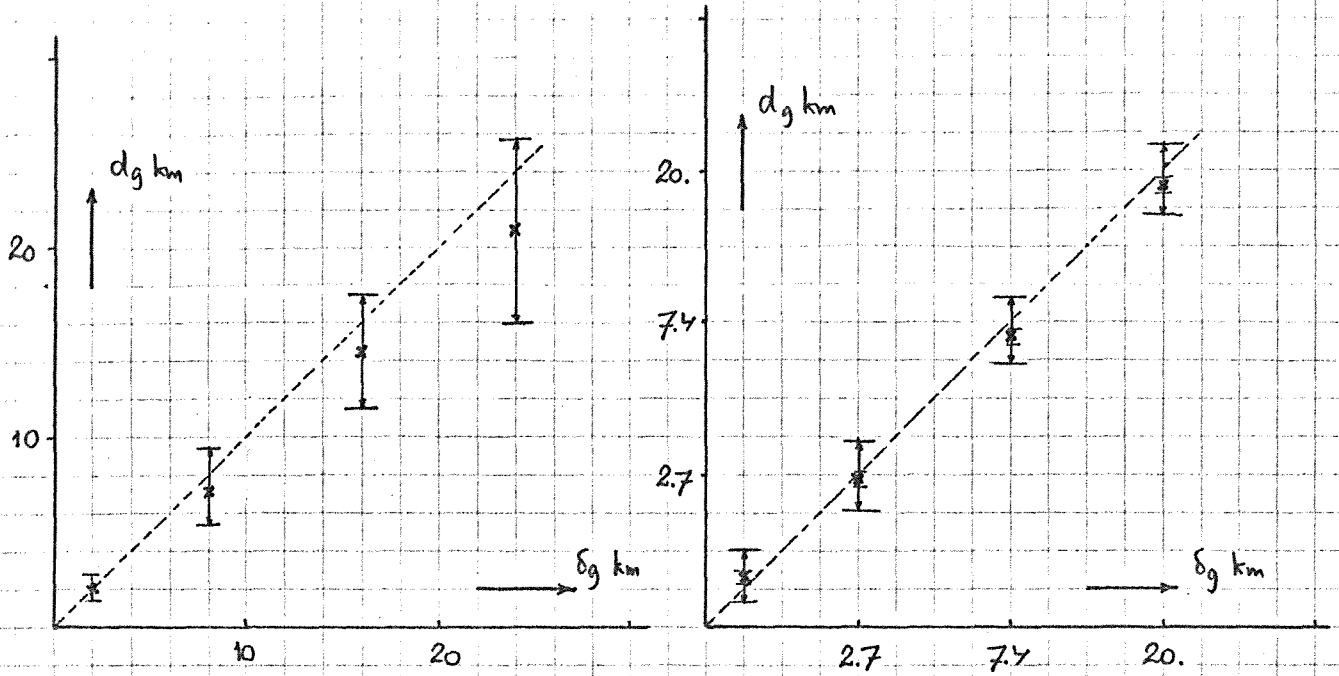
$$\sqrt{1 + \Delta} = 1 + \frac{1}{2} \Delta - \frac{1}{8} \Delta^2 + \frac{1}{16} \Delta^3 - \frac{5}{128} \Delta^4 + \dots$$

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

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

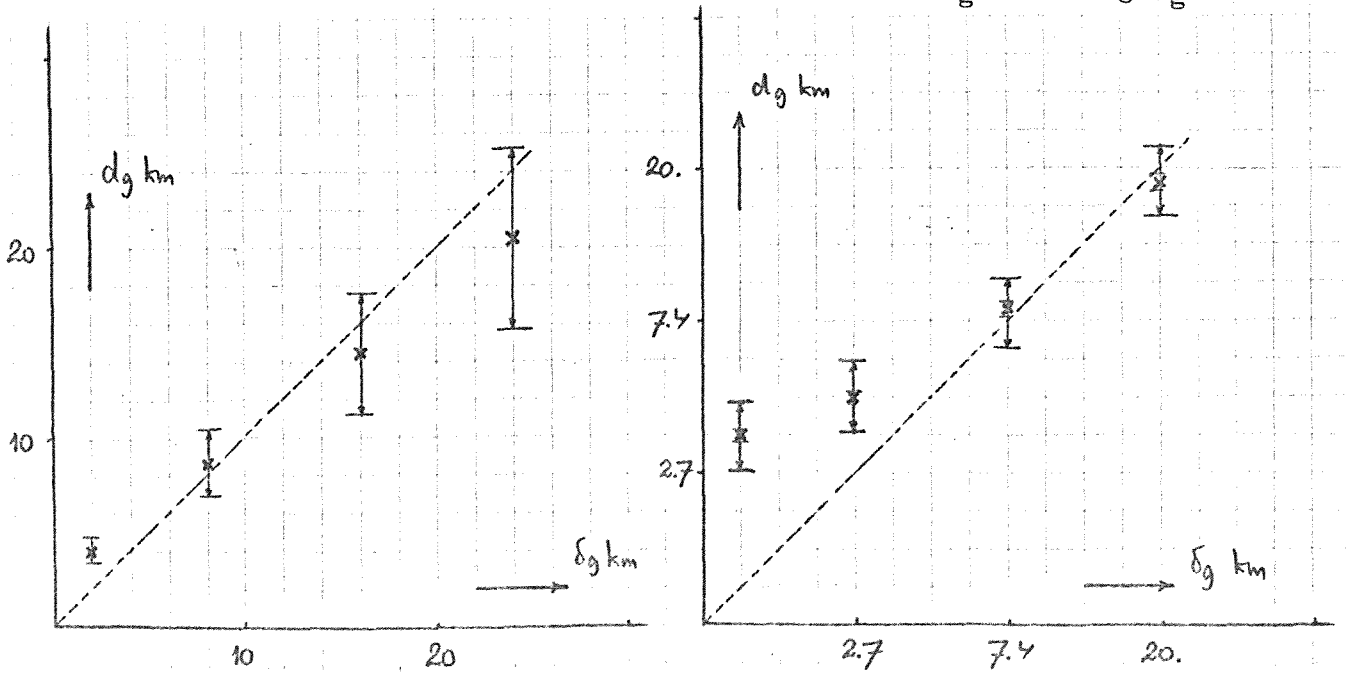
Met  $c = c_0 \sqrt{c_1}$  kan een schaalfactor worden ingebracht.

Het log-model.



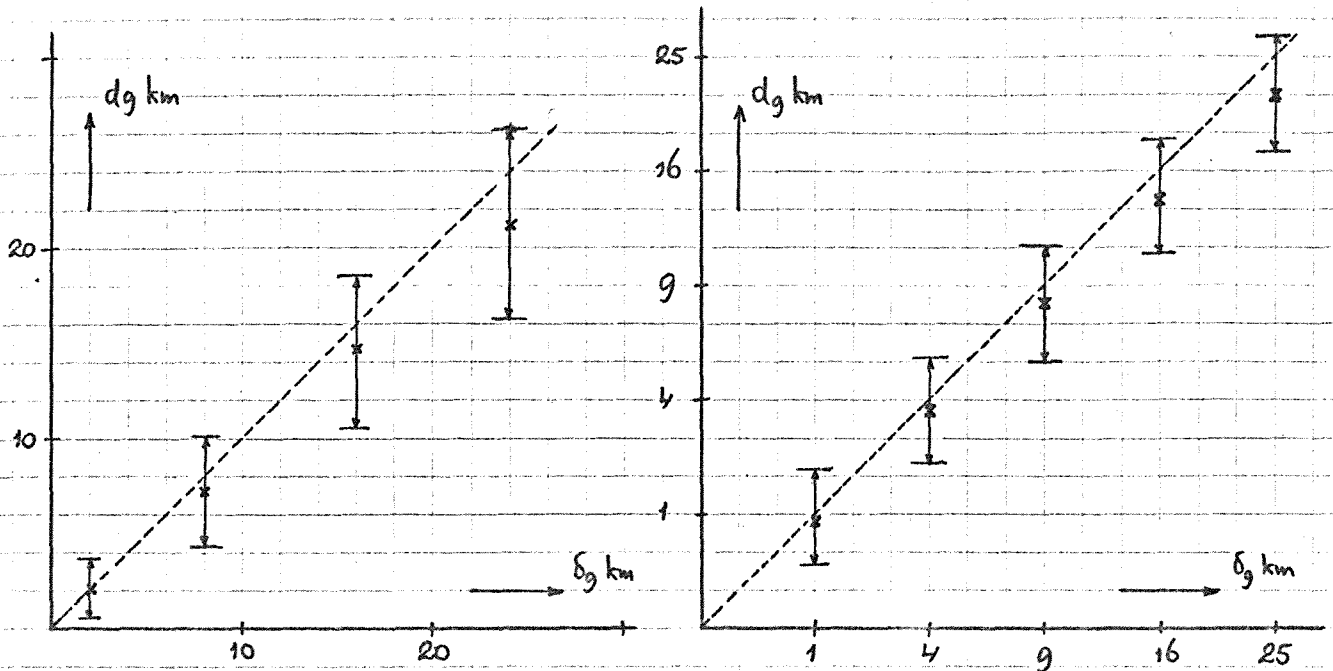
Verwachte ligging van de datapunten als het geschetste model zou voldoen. Links zijn de datapunten zelf weergegeven, rechts de log-data. Tevens staan de grenzen van het  $2\sigma$ -gebied ingetekend.

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



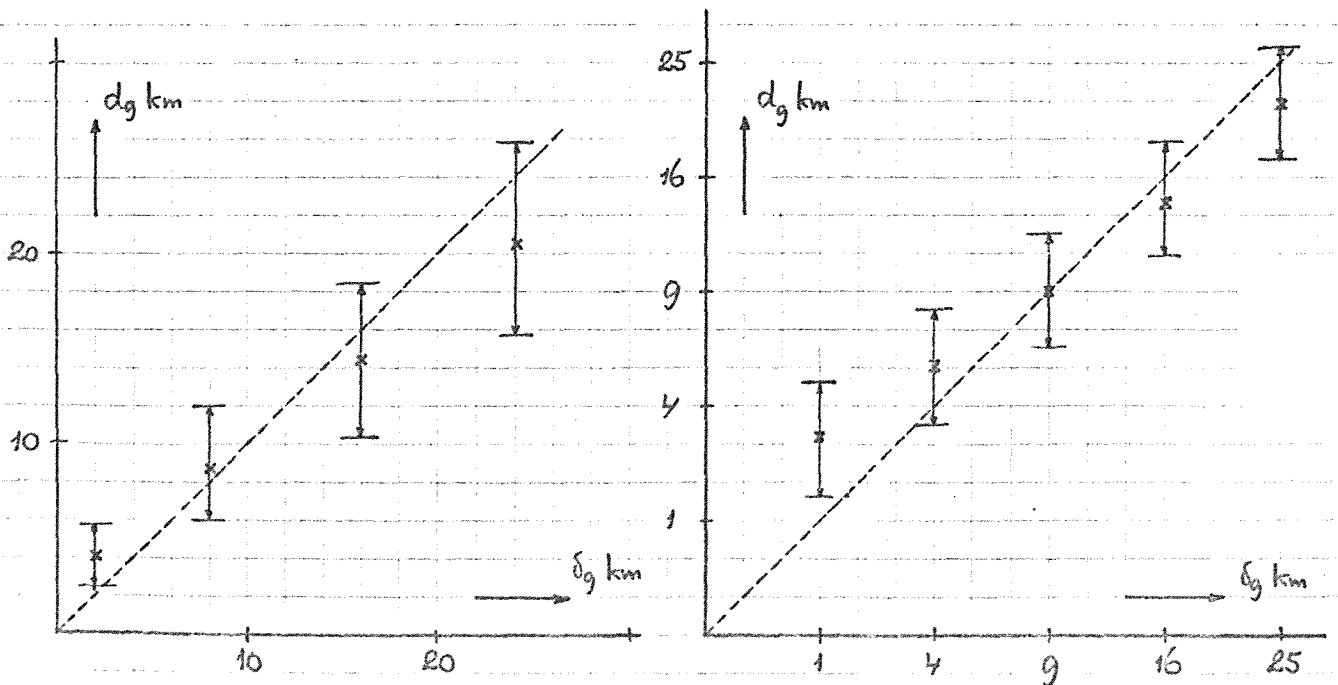
Dezelfde grafiek als boven, echter nu met:  $s(\delta_g) = 2\frac{1}{2} - \frac{1}{4} \delta_g$ .

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  $\delta_g$  en relatief grotere van  $s(\delta_g)$  gelden niet de benaderingsformules.