MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

Leiding: Prof. Dr D. van Dantzig
Chef van de Statistische Consultatie: Prof. Dr J. Hemelrijk

Rapport S 108

# REGRESSIEANALYSE VAN HET VERMOGEN DAT EEN SCHEEPSSCHROEF OPNEEMT

door

R. Doornbos.

1953.

## 1. Inleiding.

Uitgaande van de schroefkarakteristieken van modelproeven kan de volgende formule opgesteld worden:

$$(1.1)\qquad \frac{A.P.K.}{(0{,}1\,N)^3} = C_1 S_s + C + bN^{-2},$$

waarin $C_1$, $C$ en $b$ constanten zijn en

A.P.K. = vermogen in pk gemeten aan de as,
$S_s$ = schijnbare slip in % =
$= 100\left(1 - \frac{V \times 1852}{N.H \times 60}\right)\%$
$N$ = toerental van de schroef per minuut,
$V$ = scheepssnelheid in knopen, (1 knoop = 1852 $^{m}$/uur),
$H$ = spoed van de schroef in m.

Het toerental kan zeer nauwkeurig gemeten worden, het vermogen tot op $\pm$ 5% en de schijnbare slip tot op $\pm$ 10% nauwkeurig.

We willen nu de coefficienten $C_1$, $C$ en $b$ schatten, zowel met inachtneming van de te verwachten waarnemingsfouten in A.P.K. en $S_s$, als onder de aanname dat $S_s$ foutloos is. Eenzelfde berekening voeren we uit in de onderstelling $b=0$, om de invloed van de wrijvingsterm $bN^{-2}$ na te gaan.

Als orde van grootte van de verschillende componenten is opgegeven:

$$C_1 S_s \sim 15\%$$
$$C \sim 82\%$$
$$bN^{-2} \sim 3\% .$$

De volgende tabel geeft gemeten waarden van A.P.K., N en $S_s$.

| A.P.K. | N | $S_s$ % |
|---|---|---|
| 2820 | 70,81 | 25,4 |
| 2120 | 63,94 | 24,8 |
| 1730 | 60,49 | 21,9 |
| 2350 | 66,46 | 26,1 |
| 2480 | 66,11 | 31,0 |
| 2330 | 64,36 | 31,6 |
| 1780 | 60,57 | 26,2 |
| 2535 | 68,25 | 24,5 |
| 2830 | 71,86 | 21,3 |
| 2400 | 68,37 | 20,0 |
| 2420 | 68,56 | 18,6 |
| 2310 | 67,78 | 17,2 |
| 2790 | 71,56 | 17,0 |
| 2080 | 64,67 | 16,6 |
| 2110 | 64,93 | 17,5 |
| 2200 | 64,50 | 30,1 |
| 3000 | 71,88 | 27,0 |
| 1795 | 60,22 | 28,0 |
| 1760 | 60,00 | 25,8 |
| 2390 | 67,12 | 25,8 |
| 2405 | 68,49 | 22,4 |
| 2860 | 71,68 | 21,7 |
| 2850 | 72,37 | 20,8 |
| 2490 | 69,17 | 20,3 |
| 2360 | 67,98 | 20,2 |
| 2400 | 68,43 | 18,8 |
| 1240 | 55,20 | 18,1 |
| 2320 | 67,95 | 17,8 |
| 2240 | 67,37 | 16,3 |
| 2800 | 71,98 | 16,5 |
| 2020 | 64,45 | 16,4 |
| 3000 | 74,18 | 19,4 |

We voeren de volgende notatie in:

$$1000\,\text{A.P.K.} = x$$
$$S_s N^3 = z_1$$
$$N^3 = z_2$$
$$N = z_3$$
$$C_1 = \alpha_1$$
$$C = \alpha_2$$
$$b = \alpha_3$$

De verschillende waarnemingen geven we aan met de index $i$. De opgegeven waarden van $x_i$ en $z_{1i}$ zijn aan waarnemingsfouten onderhevig. We noemen de ware waarden van $x_i$ en $z_{1i}$ $\xi_i^{\circ}$ resp. $\zeta_{1i}^{\circ}$ en de waarnemingsfouten $u_i$ resp. $v_i$ . Verder onderstellen we dat $u_i$ en $v_i$ voortkomen uit onderling onafhankelijke normale verdelingen met gemiddelde nul en spreidingen $\sigma_u$ resp. $\sigma$ .

Er bestaan dus de volgende betrekkingen:

$$\underline{x}_i = \xi_i^{\circ} + \underline{u}_i$$

$$\underline{z}_{1i} = \zeta_{1i}^{\circ} + \underline{v}_i$$

Onderstreping geeft aan, dat de betreffende grootheid stochastisch is, dat wil zeggen, dat een dergelijke grootheid een waarschijnlijkheidsverdeling bezit.

Uit vergelijking (1.1) volgt:

(1.2) $$\xi_i^\circ = \alpha_1 \xi_{1i}^\circ + \alpha_2 z_{2i} + \alpha_3 z_{3i} \qquad (i=1,2,\ldots,n).$$

## 2. Behandeling van het geval dat $S_\xi$ foutloos wordt ondersteld.

In dit geval is dus $v_i = 0$ voor iedere $i$, d.w.z. $\xi_{1i}^\circ = z_{1i}$, de waargenomen waarde. Met behulp van de methode der meest aannemelijke schattingen (Engels: method of maximum likelihood) kunnen we de parameters $\alpha_1$, $\alpha_2$, $\alpha_3$ en $\sigma$ schatten. Dit komt hierop neer, dat de simultane verdelingsdichtheid van alle waarnemingen, opgevat als functie van de te schatten parameters maximaal wordt gemaakt. Voor het hier beschouwde geval van normaal verdeelde waarnemingsfouten komt deze methode overeen met de methode der kleinste kwadraten. Een nadere uiteenzetting kan men b.v. vinden in A.M. MOOD [1] [1].

De verdelingsdichtheid van $\underline{u}_i$ stellen we $f(u_i)$. Dus

$$f(u_i) = \frac{1}{\sqrt{2\pi}\,\sigma_u} e^{-\frac{u_i^2}{2\sigma_u^2}}$$

(normale verdeling met gemiddelde nul en spreiding $\sigma_u$ ).

De $\underline{u}_i$'s zijn onderling onafhankelijk, dus de simultane verdelingsdichtheid $f(u_1, \ldots, u_n)$ van $\underline{u}_1, \ldots, \underline{u}_n$ is:

$$f(u_1,\ldots,u_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma_u} e^{\frac{-u_i^2}{2\sigma_u^2}} .$$

Vullen wij in deze formule $u_i = x_i - \xi_i^\circ$ in, waarin $x_i$ de voor $\underline{x}_i$ waargenomen waarden voorstellen, en substitueren we verder (1.2), dan wordt $f(u_1, \ldots, u_n)$ uitgedrukt in de waarnemingen en de onbekende parameters $\alpha_1$, $\alpha_2$, $\alpha_3$ en $\sigma_u$. De zo verkregen functie, opgevat als functie van de onbekende parameters, wordt nu gemaximaliseerd. Hiertoe nemen we eerst de logarithme, welke gewoonlijk door $L$ wordt aangegeven en de <u>aannemelijkheidsfunctie</u> genoemd wordt. Dus

$$\ln f(u_1,\ldots,u_n) = L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln \sigma_u^2 - \frac{1}{2\sigma_u^2}\sum_{i=1}^{n} u_i^2 =$$

$$= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_u^2 - \frac{1}{2\sigma_u^2}\sum_{i=1}^{n}(x_i - \alpha_1 z_{1i} - \alpha_2 z_{2i} - \alpha_3 z_{3i})^2.$$

------

[1]) Nummers tussen vierkante haken verwijzen naar de litteratuuropgave aan het einde van dit verslag.

Door differentiatie naar $\alpha_1$ , $\alpha_2$ , $\alpha_3$ en $\sigma_u^2$ vinden we:

(2.1) $$\frac{\partial L}{\partial \alpha_1} = \frac{1}{\sigma_u^2} \sum_i z_{1i}(x_i - \alpha_1 z_{1i} - \alpha_2 z_{2i} - \alpha_3 z_{3i})$$

(2.2) $$\frac{\partial L}{\partial \alpha_2} = \frac{1}{\sigma_u^2} \sum_i z_{2i}(x_i - \alpha_1 z_{1i} - \alpha_2 z_{2i} - \alpha_3 z_{3i})$$

(2.3) $$\frac{\partial L}{\partial \alpha_3} = \frac{1}{\sigma_u^2} \sum_i z_{3i}(x_i - \alpha_1 z_{1i} - \alpha_2 z_{2i} - \alpha_3 z_{3i})$$

(2.4) $$\frac{\partial L}{\partial \sigma_u^2} = -\frac{u}{2\sigma_u^2} + \frac{1}{2\sigma_u^4} \sum_i (x_i - \alpha_1 z_{1i} - \alpha_2 z_{2i} - \alpha_3 z_{3i})^2 .$$

Stellen we deze vergelijkingen nul en lossen we vervolgens $\alpha_1$ , $\alpha_2$ , $\alpha_3$ en $\sigma_u^2$ hieruit op, dan vinden we de meest aannemelijke schattingen $\hat{a}_1$ , $\hat{a}_2$ , $\hat{a}_3$ en $\hat{s}_u^2$ van resp. $\alpha_1$ , $\alpha_2$ , $\alpha_3$ en $\sigma_u^2$ . (Het kapje geeft aan, dat de schattingen "meest aannemelijke schattingen" zijn).

We definieren:

(2.5) $$\sum_i x_i z_{ji} = y_j \qquad (j=1,2,3)$$

(2.6) $$\sum z_{ji} z_{ki} = a_{jk} \qquad (j=1,2,3\,;\, k=1,2,3).$$

(2.1), (2.2) en (2.3) gaan nu over in de 3 vergelijkingen:

(2.7) $$\begin{cases} y_1 = \hat{a}_1 a_{11} + \hat{a}_2 a_{12} + \hat{a}_3 a_{13} \\ y_2 = \hat{a}_1 a_{12} + \hat{a}_2 a_{22} + \hat{a}_3 a_{23} \\ y_3 = \hat{a}_1 a_{13} + \hat{a}_2 a_{23} + \hat{a}_3 a_{33} \, . \end{cases}$$

De elementen van de inverse van de matrix $\|a_{jk}\|$ noemen we $a^{jk}$ . Dan is de oplossing voor de $\hat{a}$ 's

(2.8) $$\begin{cases} \hat{a}_1 = a^{11} y_1 + a^{12} y_2 + a^{13} y_3 \\ \hat{a}_2 = a^{12} y_1 + a^{22} y_2 + a^{23} y_3 \\ \hat{a}_3 = a^{13} y_1 + a^{23} y_2 + a^{33} y_3 \end{cases}$$

Uit (2.4) volgt:

(2.9) $$\hat{s}_u^2 = \frac{1}{n} \sum_i (x_i - \hat{a}_1 z_{1i} - \hat{a}_2 z_{2i} - \hat{a}_3 z_{3i})^2 .$$

Beschouwen wij de waarnemingen weer als stochastische grootheden (dus grootheden, die een waarschijnlijkheidsverdeling bezitten), dan worden ook de schattingen stochastisch.

Volgens A.M. MOOD (blz.304) heeft nu de grootheid

$$\underline{t} = \frac{\hat{\underline{a}}_p - \alpha_p}{\sqrt{a^{pp} n \underline{\hat{s}}_u^2 / n-3}}$$

voor $p=1,2$ en $3$ een "STUDENT"-verdeling met $n-3$ vrijheidsgraden. Met behulp hiervan kunnen we dus een betrouwbaarheidsinterval voor de $\alpha_p$ opstellen. (Zie memorandum S 47(M 18))[1].

Het geval $\alpha_3 = 0$ kan geheel analoog behandeld worden.

## 3. Behandeling van het geval, dat zowel in A.P.K. als $S_s$ waarnemingsfouten voorkomen.

We nemen nu aan, dat $\underline{u}_i$ en $\underline{v}_i$ normaal en onderling onafhankelijk verdeeld zijn met gemiddelde 0 en spreidingen $\sigma_u$ resp. $\sigma$ . Om de methode der meest aannemelijke schattingen te kunnen toepassen, moeten we onderstellen dat de verhouding $\kappa = \frac{\sigma_u}{\sigma}$ bekend is. Als $\sigma_u$ en $\sigma$ onafhankelijk kunnen variëren, heeft de aannemelijkheidsfunctie namelijk geen absoluut maximum. Men vergelijke hiervoor bv. van DANTZIG [2] .

De verdelingsdichtheid van $\underline{u}_1, \ldots, \underline{u}_n$ , $\underline{v}_1, \ldots, \underline{v}_n$ is nu:

$$\prod_{i=1}^{n} \frac{1}{2\pi\kappa\sigma^2} e^{-\frac{u_i^2}{2\kappa^2\sigma^2} - \frac{v_i^2}{2\sigma^2}},$$

dus

$$L = -n\ln(2\pi\kappa) - n\ln\sigma^2 - \frac{1}{2\kappa^2\sigma^2}\sum_i u_i^2 - \frac{1}{2\sigma^2}\sum_i v_i^2 =$$

$$= -n\ln(2\pi\kappa) - n\ln\sigma^2 - \frac{1}{2\kappa^2\sigma^2}\sum_i (x_i - \alpha_1\zeta_{1i}^o - \alpha_2 z_{2i} - \alpha_3 z_{3i})^2 +$$

$$-\frac{1}{2\sigma^2}\sum_i (z_{1i} - \zeta_{1i}^o)^2.$$

Door te differentieren naar $\zeta_{1i}^o$ , $\alpha_1$, $\alpha_2$ , $\alpha_3$ en $\sigma^2$ vinden we:

$$(3.1) \quad \frac{\partial L}{\partial \zeta_{1i}^o} = \frac{\alpha_1}{\kappa^2\sigma^2}(x_i - \alpha_1\zeta_{1i}^o - \alpha_2 z_{2i} - \alpha_3 z_{3i}) + \frac{1}{\sigma^2}(z_{1i} - \zeta_{1i}^o)$$

$$(3.2) \quad \frac{\partial L}{\partial \alpha_1} = \frac{1}{\kappa^2\sigma^2}\sum_i \zeta_{1i}^o (x_i - \alpha_1\zeta_{1i}^o - \alpha_2 z_{2i} - \alpha_3 z_{3i})$$

------

[1]) Dit memorandum is aan het einde van dit rapport bijgevoegd.

(3.3) $\frac{\partial L}{\partial \alpha_2} = \frac{1}{\kappa^2\sigma^2} \sum_i z_{2i}(x_i - \alpha_1 \zeta^o_{1i} - \alpha_2 z_{2i} - \alpha_3 z_{3i})$

(3.4) $\frac{\partial L}{\partial \alpha_3} = \frac{1}{\kappa^2\sigma^2} \sum_i z_{3i}(x_i - \alpha_1 \zeta^o_{1i} - \alpha_2 z_{2i} - \alpha_3 z_{3i})$

(3.5) $\frac{\partial L}{\partial \sigma^2} = -\frac{n}{\sigma^2} + \frac{1}{2\kappa^2\sigma^4} \sum_i (x_i - \alpha_1 \zeta^o_{1i} - \alpha_2 z_{2i} - \alpha_3 z_{3i})^2 +$

$+ \frac{1}{2\sigma^4} \sum_i (z_{1i} - \zeta^o_{1i})^2.$

Stellen we deze vergelijkingen nul en lossen we vervolgens $\alpha_1$, $\alpha_2$, $\alpha_3$, $\sigma^2$ en de $\zeta^o_{1i}$ hieruit op, dan vinden we de meest aannemelijke schattingen $\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$, $\hat{s}^2$ en $\hat{z}^o_{1i}$ van resp. $\alpha_1$, $\alpha_2$, $\alpha_3$, $\sigma^2$ en $\zeta^o_{1i}$. Na substitutie van deze waarden in de rechter leden van (3.1) t/m (3.5) vinden we door (3.1) met (3.2), (3.3) en (3.4) te combineren:

(3.6) $\sum_i \hat{z}^o_{1i}(z_{1i} - \hat{z}^o_{1i}) = 0$ of $\sum_i \hat{z}^{o2}_{1i} = \sum_i \hat{z}^o_{1i} z_{1i}$

(3.7) $\sum_i z_{2i}(z_{1i} - \hat{z}^o_{1i}) = 0$ of $\sum_i \hat{z}^o_{1i} z_{2i} = \sum_i z_{1i} z_{2i}$

(3.8) $\sum_i z_{3i}(z_{1i} - \hat{z}^o_{1i}) = 0$ of $\sum_i \hat{z}^o_{1i} z_{3i} = \sum_i z_{1i} z_{3i}.$

Dus kunnen we (3.2), (3.3) en (3.4) schrijven:

(3.9) $\sum_i \hat{z}^o_{1i}(x_i - \hat{a}_1 z_{1i} - \hat{a}_2 z_{2i} - \hat{a}_3 z_{3i}) = 0$

(3.10) $\sum_i z_{2i}(x_i - \hat{a}_1 z_{1i} - \hat{a}_2 z_{2i} - \hat{a}_3 z_{3i}) = 0$

(3.11) $\sum_i z_{3i}(x_i - \hat{a}_1 z_{1i} - \hat{a}_2 z_{2i} - \hat{a}_3 z_{3i}) = 0.$

Uit (3.1) volgt:

(3.12) $\hat{z}^o_{1i} = \frac{\hat{a}_1(x_i - \hat{a}_2 z_{2i} - \hat{a}_3 z_{3i}) + \kappa^2 z_{1i}}{\hat{a}_1^2 + \kappa^2}.$

Noemen we de minoren van $\|a_{pq}\|$ $A_{pq}$, dan volgt uit (3.7) en (3.8) met (2.5) en (2.6):

(3.13) $$\hat{a}_2 = \frac{A_{12}\hat{a}_1 + y_2 a_{33} - y_3 a_{23}}{A_{11}}$$

(3.14) $$\hat{a}_3 = \frac{A_{13}\hat{a}_1 + y_3 a_{22} - y_2 a_{23}}{A_{11}} .$$

Substitueren we nu (3.9), (3.10) en (3.11) in (3.6), en stellen wij $X^2 = \sum_{i=1}^{n} x_i^2$, dan krijgen we de volgende vierkantsvergelijking voor $\hat{a}_1$ :

(3.15) $$\hat{a}_1^2(y_1A_{11} + y_2A_{12} + y_3A_{13}) +$$

$$+ \hat{a}_1\left[\kappa^2(a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}) + y_2^2 a_{33} - 2y_2y_3a_{23} + y_3^2 a_{22} - X^2A_{11}\right] +$$

$$- \kappa^2(y_1A_{11} + y_2A_{12} + y_3A_{13}) = 0 .$$

De eerste en de derde term hebben tegengesteld teken. Er is dus een positieve en een negatieve wortel. De positieve wortel correspondeert met het maximum van L, zoals door berekening is te verifiëren. Uit physische overwegingen komt een negatieve eerste term vanzelfsprekend ook niet in aanmerking.

Stellen we het rechter lid van (3.5) nul en substitueren we hierin de geschatte waarden voor $\sigma^2$, $\alpha_1$, $\alpha_2$, $\alpha_3$ en $\zeta_{1i}^{\circ}$ en (3.9), dan is:

$$\hat{s}_u^2 = \kappa^2 \hat{s}^2 .$$

Voor het geval dat $\alpha_3 = 0$ vinden we in plaats van (3.12)·

(3.17) $$\hat{a}_1^2(y_1a_{22} - y_2a_{12}) + \hat{a}_1\left[\kappa^2(a_{11}a_{22} - a_{12}^2) + y_2^2 - X^2a_{22}\right] +$$

$$- \kappa^2(y_1a_{22} - y_2a_{12}) = 0,$$

terwijl $$\hat{a}_2 = \frac{y_2 - \hat{a}_1a_{12}}{a_{22}} .$$

We beschikken in dit geval niet over een methode om precieze betrouwbaarheidsintervallen voor de parameters op te stellen. Wel kunnen we een schatting van de spreidingen van de geschatte parameterwaarden berekenen. Voor grote n zijn de meest aannemelijke schattingen van de parameters namelijk bij benadering simultaan normaal verdeeld. Zijn $\theta_1, \ldots, \theta_k$ de geschatte parameters, dan zijn de varianties (kwadraat van de spreiding = variantie) hiervan, de corresponderende diagonaalelementen uit de matrix:

$$\|M\| = -\|\mathcal{E}\{\frac{\partial^2 L}{\partial\theta_r \partial\theta_s}\}\|^{-1} \qquad (r,s = 1,\ldots,k)$$

(vergelijk voor de theorie b.v. M.G. KENDALL [3]).

In bovenstaande formule stelt het symbool $\mathcal{E}$ de mathematisch verwachting voor. In ons geval zijn de parameters $\zeta_{1i}^o$, $\alpha_1$, $\alpha_2$, $\alpha_3$ en $\sigma^2$.

Uit de vergelijkingen (3.1) tot (3.5) volgt:

$$\frac{\partial^2 L}{\partial \zeta_{1i}^{o2}} = \frac{-\alpha_1^2}{\kappa^2\sigma^2} - \frac{1}{\sigma^2} = -\frac{1}{\sigma^2}\left[\frac{\alpha_1^2}{\kappa^2} + 1\right],$$

$$\frac{\partial^2 L}{\partial\zeta_{1i}^o \partial\zeta_{1j}^o} = 0 \quad \text{voor } i \neq j,$$

$$\frac{\partial^2 L}{\partial(\sigma^2)^2} = \frac{n}{\sigma^4} - \frac{1}{\kappa^2\sigma^6} \sum_i (x_i - \alpha_1\zeta_{1i}^o - \alpha_2 z_{2i} - \alpha_3 z_{3i})^2 +$$

$$- \frac{1}{\sigma^6} \sum_i (z_{1i} - \zeta_{1i}^o)^2,$$

dus

$$\mathcal{E}\left(\frac{\partial^2 L}{\partial(\sigma^2)^2}\right) = \frac{n}{\sigma^4} - \frac{n}{\sigma^4} - \frac{n}{\sigma^4} = -\frac{n}{\sigma^4},$$

$$\mathcal{E}\left(\frac{\partial^2 L}{\partial\alpha_1 \partial\zeta_{1i}^o}\right) = \frac{-\alpha_1}{\kappa^2\sigma^2} \zeta_{1i}^o,$$

$$\mathcal{E}\left(\frac{\partial^2 L}{\partial\alpha_2 \partial\zeta_{1i}^o}\right) = \frac{-\alpha_1}{\kappa^2\sigma^2} z_{2i},$$

$$\mathcal{E}\left(\frac{\partial^2 L}{\partial\alpha_3 \partial\zeta_{1i}^o}\right) = \frac{-\alpha_2}{\kappa^2\sigma^2} z_{3i},$$

We definieren

$$(3.18) \qquad \sum_i (\zeta_{1i}^o)^2 = \alpha_{11}^o$$

$$(3.19) \qquad \sum_i \zeta_{1i}^o z_{ki} = \alpha_{1k}^o \quad (k=2,3),$$

dan is:

$$\frac{\partial^2 L}{\partial\alpha_1^2} = \frac{-1}{\kappa^2\sigma^2} \sum_i (\zeta_{1i}^o)^2 = \frac{-1}{\kappa^2\sigma^2} \alpha_{11}^o ;$$

verder is volgens (3.7)

$$\frac{\partial^2 L}{\partial\alpha_1 \partial\alpha_2} = \frac{-1}{\kappa^2\sigma^2} \sum_i \zeta_{1i}^o z_{2i} = \frac{-1}{\kappa^2\sigma^2} a_{12},$$

en volgens (3.8)

$$\frac{\partial^2 L}{\partial\alpha_1\partial\alpha_3} = \frac{-1}{K^2\sigma^2}\sum_i \zeta^{\circ}_{1i} z_{3i} = \frac{-1}{K^2\sigma^2} a_{13},$$

terwijl

$$\frac{\partial^2 L}{\partial\alpha_i\partial\alpha_j} = \frac{-1}{K^2\sigma^2} a_{ij} \qquad (i,j=1,2,3).$$

Tenslotte is:

$$\mathcal{E}\left(\frac{\partial^2 L}{\partial\sigma^2\partial\alpha_j}\right) = 0 \qquad (j=1,2,3)$$

en

$$\mathcal{E}\left(\frac{\partial^2 L}{\partial\zeta^{\circ}_{1i}\partial\sigma^2}\right) = 0 \qquad (i=1,\dots,32).$$

Geven we grootheden, waarin $\zeta^{\circ}_{1i}$ staat in plaats van $z_{1i}$,
aan met de bovenindex $^{\circ}$ , dan wordt de matrix $\left\|\mathcal{E}\left(\frac{\partial^2 L}{\partial\theta_r\partial\theta_s}\right)\right\|$
in ons geval:

$$\|T\| = \left\|\begin{array}{cccccccccc}
\frac{-n}{\sigma^4} & 0 & 0 & 0 & 0 & - & - & - & 0 \\
0 & \frac{-1}{K^2\sigma^2}\alpha^{\circ}_{11} & \frac{-1}{K^2\sigma^2}a_{12} & \frac{-1}{K^2\sigma^2}a_{13} & \frac{-\alpha_1}{K^2\sigma^2}z^{\circ}_{1,1} & - & - & - & \frac{-\alpha_1}{K^2\sigma^2}z^{\circ}_{1,32} \\
0 & \frac{-1}{K^2\sigma^2}a_{12} & \frac{-1}{K^2\sigma^2}a_{22} & \frac{-1}{K^2\sigma^2}a_{23} & \frac{-\alpha_1}{K^2\sigma^2}z_{2,1} & - & - & - & \frac{-\alpha_1}{K^2\sigma^2}z_{2,32} \\
0 & \frac{-1}{K^2\sigma^2}a_{13} & \frac{-1}{K^2\sigma^2}a_{23} & \frac{-1}{K^2\sigma^2}a_{33} & \frac{-\alpha_1}{K^2\sigma^2}z_{3,1} & - & - & - & \frac{-\alpha_1}{K^2\sigma^2}z_{3,32} \\
0 & \frac{-\alpha_1}{K^2\sigma^2}z^{\circ}_{1,1} & \frac{-\alpha_1}{K^2\sigma^2}z_{2,1} & \frac{-\alpha_1}{K^2\sigma^2}z_{3,1} & \frac{-1}{\sigma^2}\left[\frac{\alpha_1^2}{K^2}+1\right] & - & - & - & 0 \\
- & - & - & - & - & - & - & - & - \\
0 & \frac{-\alpha_1}{K^2\sigma^2}z^{\circ}_{1,32} & \frac{-\alpha_1}{K^2\sigma^2}z_{2,32} & \frac{-\alpha_1}{K^2\sigma^2}z_{3,32} & 0 & - & - & - & \frac{-1}{\sigma^2}\left[\frac{\alpha_1^2}{K^2}+1\right]
\end{array}\right\|$$

Nu is dus $\sigma^2_{\hat{a}_1} = -\frac{T_{22}}{T}$ , als $T_{22}$ de minor van $t_{22}$ en $T = |t_{pq}|$ is, $\sigma^2_{\hat{a}_2} = -\frac{T_{33}}{T}$ en $\sigma^2_{\hat{a}_3} = -\frac{T_{44}}{T}$ .

We vermenigvuldigen de 36$^e$ kolom achtereenvolgens met

$$\frac{\frac{-\alpha_1}{K^2\sigma^2}\zeta^{\circ}_{1,32}}{\frac{-1}{\sigma^2}\left[\frac{\alpha_1^2}{K^2}+1\right]}, \qquad \frac{\frac{-\alpha_1}{K^2\sigma^2}z_{2,32}}{\frac{-1}{\sigma^2}\left[\frac{\alpha_1^2}{K^2}+1\right]} \quad \text{en} \quad \frac{\frac{-\alpha_1}{K^2\sigma^2}z_{3,32}}{\frac{-1}{\sigma^2}\left[\frac{\alpha_1^2}{K^2}+1\right]} \quad \text{en}$$

trekken deze kolom dan resp. af van de 2$^e$, 3$^e$ en 4$^e$ kolom. Van de laatstgenoemde kolommen zijn de laatste elementen nu gelijk aan nul.

Door hetzelfde procédé toe te passen op de 5$^e$ tot en met de 35$^e$ kolom, bereiken we dat van de 2$^e$, 3$^e$ en 4$^e$ kolom vanaf de 5$^e$ rij alle elementen nul worden.

De term $t_{22}$ is dan overgegaan in

$$t'_{22} = \frac{-1}{\kappa^2\sigma^2}\alpha^\circ_{11} + \frac{\frac{\alpha_1^2}{\kappa^4\sigma^4}}{\frac{1}{\sigma^2}\left[\frac{\alpha_1^2}{\kappa^2}+1\right]}\alpha^\circ_{11} = \frac{-1}{\sigma^2(\alpha_1^2+\kappa^2)}\alpha^\circ_{11} = C.\alpha^\circ_{11}\text{, waarin } C = \frac{-1}{\sigma^2(\alpha_1^2+\kappa^2)}$$

Verder is $t'_{23} = \frac{-1}{\sigma^2(\alpha_1^2+\kappa^2)} a_{12} = C.a_{12}$ enz.

Dus

$$(3.18) \qquad \sigma^2_{\hat{a}_1} = \frac{-T_{22}}{T} = -\frac{\begin{vmatrix} Ca_{22} & Ca_{23} \\ Ca_{23} & Ca_{33} \end{vmatrix}}{\begin{vmatrix} Ca^\circ_{11} & Ca_{12} & Ca_{13} \\ Ca_{12} & Ca_{22} & Ca_{23} \\ Ca_{13} & Ca_{23} & Ca_{33} \end{vmatrix}} =$$

$$= -\frac{1}{C}\frac{A_{11}}{|a^\circ|} = \sigma^2(\alpha_1^2+\kappa^2)\frac{A_{11}}{|a^\circ|}$$

$$(3.19) \qquad \sigma^2_{\hat{a}_2} = \sigma^2(\alpha_1^2+\kappa^2)\frac{A^\circ_{22}}{|a^\circ|}$$

$$(3.20) \qquad \sigma^2_{\hat{a}_3} = \sigma^2(\alpha_1^2+\kappa^2)\frac{A^\circ_{33}}{|a^\circ|}.$$

In het geval dat $\alpha_3 = 0$ vinden we:

$$(3.21) \qquad \sigma^2_{\hat{a}_1} = \sigma^2(\alpha_1^2+\kappa^2)\frac{a_{22}}{|a^\circ|}$$

en

$$(3.22) \qquad \sigma^2_{\hat{a}_2} = \sigma^2(\alpha_1^2+\kappa^2)\frac{a_{11}}{|a^\circ|},$$

waarin

$$|a^\circ| = \begin{vmatrix} a^\circ_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}.$$

Om de varianties numeriek te schatten substitueren we de meest aannemelijke schattingen voor $\xi^\circ_{1i}$, $\alpha_1$, $\alpha_2$, $\alpha_3$ en $\sigma^2$ in (3.18), (3.19),...,(3.22).

## 4. Resultaten.

Voor het geval dat $S_s$ niet foutloos is, moet $\kappa = \frac{\sigma_u}{\sigma}$ worden geschat. Nemen we hiervoor, in aansluiting op de verstrekte gegevens:

$$\kappa = \frac{5\% \text{ van de gemiddelde waarde } 1000 \text{ APK}}{10\% \text{ van de gemiddelde waarde } S_s N^3}$$

dan vinden we $\kappa = 0{,}18$. De berekeningen zijn uitgevoerd voor $\kappa = 0{,}15$ en $\kappa = 0{,}2$. De spreidingen van de parameters zijn berekend voor $\kappa = 0{,}15$. We krijgen dan de volgende uitkomsten:

I. $S_s$ foutloos .

| a) met wrijvingsterm | |
|---|---|
| | Betrouwbaarheidsinterval s.pr.0,05 |
| $\hat{a}_1 = 0{,}064$<br>$\hat{a}_2 = 5{,}87$<br>$\hat{a}_3 = 2200$<br>$\hat{s} = 43.10^3$ | $0{,}051 < \hat{a}_1 < 0{,}077$<br>$5{,}15 < \hat{a}_2 < 6{,}59$<br>$-200 < \hat{a}_3 < 4200$ |
| b) zonder wrijvingsterm | |
| | Betrouwbaarheidsinterval s.pr.0,05 |
| $\hat{a}_1 = 0{,}070$<br>$\hat{a}_2 = 6{,}21$<br>$\hat{s} = 46.10^3$ | $0{,}057 < \hat{a}_1 < 0{,}083$<br>$5{,}92 < \hat{a}_2 < 6{,}50$ |

II. Met waarnemingsfout in $S_s$ .

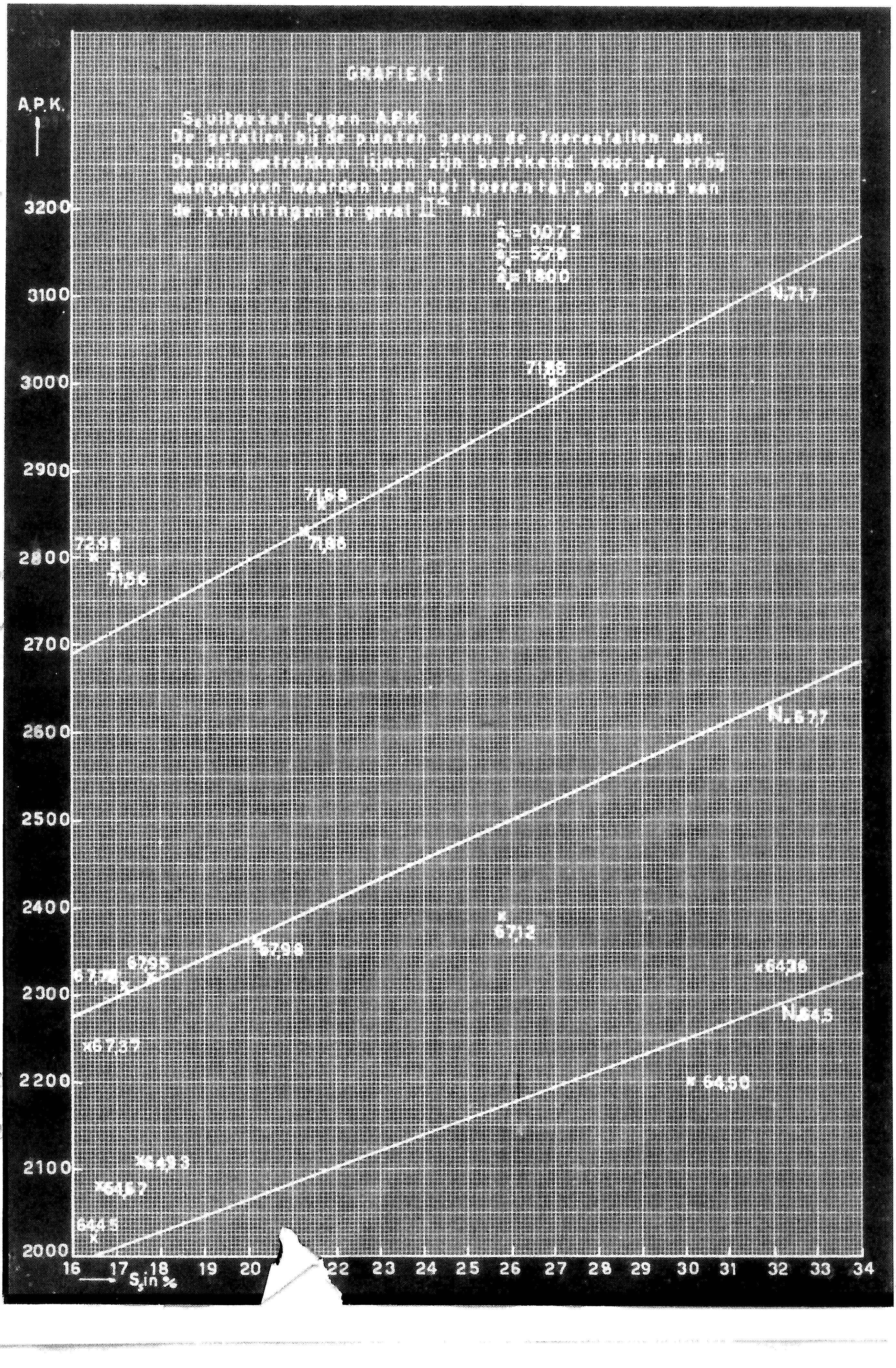
| a) met wrijvingsterm | | |
|---|---|---|
| $\kappa = 0{,}15$ | | $\kappa = 0{,}2$ |
| $\hat{a}_1 = 0{,}072$<br>$\hat{a}_2 = 5{,}79$<br>$\hat{a}_3 = 1800$<br>$\hat{s}_u = 28.10^3$<br>$\hat{s} = 185.10^3$ | $\sigma_{\hat{a}_1} = 0{,}005$<br>$\sigma_{\hat{a}_2} = 0{,}16$<br>$\sigma_{\hat{a}_3} = 700$ | $\hat{a}_1 = 0{,}069$<br>$\hat{a}_2 = 5{,}82$<br>$\hat{a}_3 = 1900$<br>$\hat{s}_u = 29.10^3$<br>$\hat{s} = 144.10^3$ |
| b) zonder wrijvingsterm | | |
| $\hat{a}_1 = 0{,}073$<br>$\hat{a}_2 = 6{,}15$<br>$\hat{s}_u = 29.10^3$<br>$\hat{s} = 194.10^3$ | $\sigma_{\hat{a}_1} = 0{,}004$<br>$\sigma_{\hat{a}_2} = 0{,}10$ | $\hat{a}_1 = 0{,}072$<br>$\hat{a}_2 = 6{,}17$<br>$\hat{s}_u = 30.10^3$<br>$\hat{s} = 152.10^3$ |

Voor de grootte van de drie componenten vinden we in geval $I^a$: $C_1 S_s = 18\%$, $C = 77\%$ en $bN^{-2} = 5\%$

$I^b$: $C_1 S_s = 20\%$, $C = 80\%$

$II^a$: ($\kappa = 0{,}15$) $C_1 S_s = 20\%$, $C = 75\%$, $bN^{-2} = 5\%$

$II^b$: ($\kappa = 0{,}15$) $C_1 S_s = 21\%$, $C = 79\%$.


GRAFIEK I
S uitgezet tegen A.P.K.
De getallen bij de punten geven de toerentallen aan.
De drie getrokken lijnen zijn berekend voor de erbij aangegeven waarden van het toerental, op grond van de schattingen in geval II a.1
= 0,072
= 579
= 1800
A.P.K.
3200
3100
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100
2000
16
17
18
19
20
22
23
24
25
26
27
28
29
30
31
32
33
34
S in %
N, 71,7
N, 67,7
N, 64,5
71,58
71,69
71,80
72,98
71,56
67,12
67,99
67,95
67,78
67,37
64,76
64,50
64,93
64,57
64,45

Deze resultaten zijn verkregen op grond van het in de inleiding ondersteldе verband, dat door (1.1) gegeven wordt. Het is interessant even na te gaan, in hoeverre de gevonden resultaten met deze onderstelling in overeenstemming zijn.

In de bijgevoegde grafiek 1 is daartoe voor 3 groepen waarnemingen, waarvan de waarden van N dicht bij elkaar liggen, het verband tussen A.P.K. en $S_s$ aangegeven. De getrokken lijnen zijn berekend voor de drie aangegeven waarden van N, waarbij is aangenomen $\hat{a}_1 = 0{,}072$, $\hat{a}_2 = 5{,}79$ en $\hat{a}_3 = 1800$. We zien dat de grafiek, de spreidingen in $S_s$ en A.P.K. in aanmerkingen genomen, geen aanleiding geeft aan het ondersteldе verband tussen A.P.K., $S_s$ en N te twijfelen.

We merken op, dat in geval $I^a$ het betrouwbaarheidsinterval voor de wrijvingsterm de waarde nul bevat. In geval $II^a$ moeten we bedenken, dat de opgegeven spreidingen, dus ook $\sigma_{a_3}$ slechts schattingen zijn. De gebruikte formules zijn slechts geldig voor grote $n$, zoals in § 3 ook reeds werd opgemerkt.

Er bestaat voor het algemene geval ook een parametervrije methode, d.w.z. dat geen onderstelling behoeft te worden gemaakt omtrent de verdeling van de $\underline{u}_i$ en de $\underline{v}_i$. De toepassing van een dergelijke methode heeft het voordeel, dat eventuele afwijkingen van de min of meer uit de lucht gegrepen onderstelling van normaliteit der waarnemingsfouten ons dan geen parten kunnen spelen. Hierbij moeten echter de waarden van $S_s$ naar opklimmende grootte gerangschikt worden. Bij het beschikbare waarnemingsmateriaal zouden we dan slechts 4 stel waarnemingen van de 32 kunnen gebruiken, daar de intervallen te klein zijn vergeleken met de spreiding. Als we bv. de waarden 30,1 en 27,0 hebben, dan kunnen de ware waarden resp. 27,1 en 29,1 zijn, zodat de volgorde niet vaststaat. Deze methode zouden we dus wel kunnen toepassen op voldoend uit elkaar liggende waarnemingen.

LITTERATUUR.

[1] A.M.MOOD, Introduction to the theory of statistics. London 1950, p.152 e.v.

[2] D.v.DANTZIG, Kadercursus Statistiek 1948-49, Hoofdstuk IV, § 1, p.323 e.v.

[3] M.G.KENDALL, The advanced theory of statistics, vol.II London 1948, p.36 e.v.

MATHEMATISCH CENTRUM
Amsterdam.
Statistische Afdeling.
S47(M18)

## Betrouwbaarheidsintervallen (algemeen).[1]

Zij $\underline{x}$ een stochastische grootheid, die een verdelingsfunctie bezit die, op een onbekende parameter $\theta$ na, geheel bekend is ( $\theta$ kan bv. het gemiddelde van $\underline{x}$ zijn, of de spreiding of iets dergelijks), dan kan men de vraag stellen uit een aantal waarnemingen van $\underline{x}$ een schatting voor $\theta$ af te leiden.

Een betrouwbaarheidsinterval $\underline{\mathcal{J}}$ voor $\theta$ is een interval, waarvan de grenzen afhankelijk zijn van de waarnemingen $x_1, \ldots,$ $x_n$ van $\underline{x}$ , en dat de eigenschap bezit, behoudens een zekere gegeven onbetrouwbaarheid $\alpha$, de juiste waarde van $\theta$ te bevatten. Dit betekent, dat bij een serie bepalingen van betrouwbaarheidsintervallen slechts in ongeveer een fractie $\alpha$ van deze gevallen het interval $\underline{\mathcal{J}}$ zo zal uitvallen, dat het $\theta$ niet bevat. Hierbij is dus $\theta$ constant en het interval $\underline{\mathcal{J}}$ veranderlijk (en wel stochastisch). Hierin ligt het grote verschil met een zgn. voorspellingsinterval, d.i. een gegeven vast interval, waar een stochastisch punt met een zekere waarschijnlijkheid in valt.

Het algemene principe ter bepaling van een betrouwbaarheidsinterval is het volgende: zij $\mathcal{T}$ een toets voor de hypothese $\theta = \theta_0$ (vgl.S47(M6)), dan is $\mathcal{J}$ de verzameling van die waarden $\theta_0$ die bij toepassing van $\mathcal{T}$ op grond van de gevonden waarnemingen $x_1, \ldots, x_n$ niet voor verwerping in aanmerking komen. Is $\mathcal{T}$ toegepast met een onbetrouwbaarheidsdrempel $\alpha$ , dan is dit ook de onbetrouwbaarheidsdrempel van het betrouwbaarheidsinterval.

Litteratuur:

M.G. Kendall, The Advanced Theory of Statistics, London 1946, deel II, p.62-84.
A.M. Mood, Introduction to the theory of Statistics, London 1950, p.220.
J. Neyman, First course in probability and statistics, N.Y.1950.

---

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.