MATHEMATICH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
STATISTISCHE AFDELING

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

Rapport S 131

Toets voor de gelijkheid van twee kleine kansen met behulp van even grote steekproeven en het onderscheidingsvermogen van deze toets

door

Gerda Klerk-Grobben

1954

1. Inleiding.

Op bepaalde gebieden van onderzoek doet zich het probleem voor van het vergelijken van kleine waarschijnlijkheden. Wij hebben daarbij b.v. het oog op geologische onderzoekingen, waarbij analyses van bodemlagen worden verricht aan de hand van de soorten stuifmeelkorrels, die daarin voorkomen. Stellen we de fractie van een bepaalde soort stuifmeelkorrels, die in een bepaalde laag voorkomen, gelijk aan θ_1 en die in een andere laag aan θ_2, dan gaat het er om de hypothese

$$(1) \qquad H_0 : \quad \theta_1 = \theta_2$$

te toetsen, op grond van de in twee steekproeven gevonden aantallen. De statistische methoden hiervoor zijn bekend en elders uitvoerig beschreven (zie [1], eerste deel [1]).

In dit rapport wordt een variant van één dezer methoden (en wel een zeer eenvoudige) besproken, welke toepasbaar is wanneer θ_1 en θ_2 klein zijn.

Een der belangrijkste vragen bij dit soort problemen is echter, hoe groot men de omvang van de beide steekproeven moet nemen, om bij gegeven waarden van θ_1 en θ_2, die ongelijk zijn, te bereiken dat men deze ongelijkheid, behoudens een bepaalde kleine waarschijnlijkheid, ook ontdekt. Bij het voorbeeld van de stuifmeelkorrels komt dit dus neer op de vraag hoeveel korrels men uit ieder der twee te vergelijken lagen moet tellen. Deze vraag zal in paragraaf 3 nader worden gepreciseerd.

In de volgende paragrafen worden de toets zelf en grafieken voor de bepaling van de steekproefgrootte besproken en aan de hand van enkele voorbeelden nader toegelicht. De appendix bevat wiskundige afleidingen.

2. Toets voor $\theta_1 = \theta_2$, als θ_1 en θ_2 klein zijn [2].

Het waarnemingsmateriaal bestaat uit twee steekproeven, waarvan wij eerst onderstellen, dat ze even groot zijn en wel beide uit een groot aantal, N, exemplaren bestaan. (Dus b.v. twee steekproeven van ieder N stuifmeelkorrels.) In elk der steekproeven wordt geteld, hoeveel exemplaren er in voorkomen van een in beide steekproeven zeldzame soort A. Laten dit er n_1 en n_2

---

1) Cijfers tussen vierkante haken zijn een verwijzing naar de literatuurlijst aan het einde van dit rapport.

2) Zie voor de algemene gang van zaken bij het toetsen van hypothesen: memorandum S 47 (M 6), dat aan het einde van dit rapport is bijgevoegd.

zijn; we stellen

(2) $$n_1 + n_2 = m$$

Bij de toetsing van de hypothese $H_0\,(\theta_1 = \theta_2)$ handelen we nu alsof we m maal met een munt geworpen hebben, waarvan we de zuiverheid willen toetsen, en alsof van deze m worpen n_1 worpen het resultaat "kruis" hebben gegeven. Voor de toepassing van de toets is een tabel van kritieke waarden voor n_1 bij verschillende onbetrouwbaarheidsdrempels [2] toegevoegd (tabel I). De opgegeven onbetrouwbaarheidsdrempels behoren bij éénzijdige toetsing (zie hieronder).

Leidt de toepassing van deze toets tot verwerping, dan wordt dus de hypothese $H_0\,(\theta_1 = \theta_2)$ verworpen. De grootte van N heeft op de toets geen invloed; de toets is echter slechts van toepassing als N veel groter dan m is.

<u>Een- en tweezijdige toetsing</u>. We spreken van <u>tweezijdige</u> toetsing indien het kritieke gebied bestaat uit grote zowel als kleine waarden van de toetsingsgrootheid. In ons geval moet tweezijdige toetsing worden gebruikt, indien we tot verwerping van $H_0\,(\theta_1 = \theta_2)$ willen komen wanneer $\theta_1 < \theta_2$ en ook wanneer $\theta_1 > \theta_2$. Weten we echter reeds van te voren, dat alleen $\theta_1 \leqq \theta_2$ kan optreden, zodat $\theta_1 > \theta_2$ zeker niet juist is, dan kunnen we een <u>éénzijdige</u> toets gebruiken; het kritieke gebied zal dan alleen uit kleine waarden van n_1 bestaan. (Is alleen $\theta_1 \geqq \theta_2$ mogelijk, dan wordt het kritieke gebied uit grote waarden van n_1 gevormd.) Het gebruik van een éénzijdige toets is ook dan gerechtvaardigd, indien het voor het onderzoek uitsluitend van belang is na te gaan of $\theta_1 < \theta_2$, terwijl $\theta_1 = \theta_2$ en $\theta_1 > \theta_2$ voor de onderzoeker (en voor de te nemen beslissing) op hetzelfde neerkomen. Men toetst dan eigenlijk de hypothese $\theta_1 \geqq \theta_2$ tegen $\theta_1 < \theta_2$ en doet dit met een linker-éénzijdig kritiek gebied.

In tabel I zijn linkeréénzijdige kritieke gebieden opgegeven. De rechtséénzijdige kritieke waarden (voor toetsing van $H_0 : \theta_1 \leqq \theta_2$ tegen $\theta_1 > \theta_2$) zijn in deze tabel te vinden door verwisseling van n_1 en n_2 of door de in de tabel opgegeven waarden van m af te trekken.

Ook bij tweezijdige toetsing kan de tabel gebruikt worden. De kritieke zone bestaat dan uit het opgegeven gebied èn het overeenkomstige gebied van hoge waarden. In dit laatste geval moeten

---

2) Zie voor de algemene gang van zaken bij het toetsen van hypothesen: memorandum S 47 (M 6), dat aan het einde van dit rapport is bijgevoegd.

de opgegeven onbetrouwbaarheidsdrempels verdubbeld worden.

Voorbeeld. Stel we willen $H_0\ (\theta_1 \geqq \theta_2)$ toetsen tegen $\theta_1 < \theta_2$ en vinden in de steekproeven respectievelijk $n_1 = 6$ en $n_2 = 14$ exemplaren van de desbetreffende soort. Hierbij is dus $n_1 + n_2 = m = 20$. In tabel I vinden we dan dat bij $m = 20$ en onbetrouwbaarheidsdrempel $\alpha = 0{,}05$ de kritieke waarde voor $n_1 = 6$ is. D.w.z. dat, indien de nulhypothese $\theta_1 = \theta_2$ waar is, de kans om voor n_1 een waarde $\leqq 6$ te vinden hoogstens 0,05 is, vinden we een dergelijke "onwaarschijnlijke" waarde van n_1, dan zullen we de nulhypothese verwerpen ten gunste van $\theta_1 < \theta_2$; zoals eens in ons voorbeeld. Kiezen we echter als onbetrouwbaarheidsdrempel $\alpha = 0{,}025$, dan bestaat het kritieke gebied uit waarden $n_1 \leqq 5$; in dit geval zullen we dus op grond van bovenstaande steekproeven niet tot verwerping van de hypothese $\theta_1 = \theta_2$ overgaan. Voor tweezijdige toetsing (H_0: $\theta_1 = \theta_2$ tegen $\theta_1 < \theta_2$ èn $\theta_1 > \theta_2$) bestaat het kritieke gebied voor $\alpha = 0{,}05$ en $m = 20$ uit waarden voor n_1 : $\leqq 5$ en $\geqq 20-5 = 15$ (in tabel I de kolom onder $\alpha = 0{,}025$ gebruiken). Bij tweezijdige toetsing zou de nulhypothese volgens bovenstaande steekproeven en onbetrouwbaarheidsdrempel $\alpha = 0{,}05$ dus niet verworpen worden.

Steekproeven van ongelijke omvang. Tot nu toe hebben we ondersteld dat de steekproeven beide uit evenveel exemplaren (N) bestaan. Is dit echter niet het geval, bestaan dus de steekproeven uit N_1 resp. N_2 exemplaren met $N_1 \neq N_2$, dan kan als N_1 en N_2 sterk verschillen, de beschreven toets niet meer op n_1 en n_2 worden toegepast. (De toets berust nl. op een binomiale verdeling met gelijke kansen, terwijl nu een binomiale verdeling met kansen $\frac{N_1}{N_2+N_1}$ en $\frac{N_2}{N_1+N_2}$ gebruikt behoort te worden, of een benadering hiervan; zie appendix, § 4.) Zijn N_1 en N_2 echter groot en verschillen ze slechts weinig (bv. $N_1 = 1000$ en $N_2 = 985$), dan zal de fout, die gemaakt wordt wanneer we bovenstaande toets toch toepassen op n_1 en n_2, slechts gering zijn. Op het geval van ongelijke steekproeven gaan we, behoudens nog enkele opmerkingen in de appendix, verder niet in.

Enkele opmerkingen.

1) De in deze paragraaf beschreven toets kan alleen gebruikt worden indien de kansen θ_1 en θ_2 klein zijn (b.v. $< 0{,}1$) en N groot is. In andere gevallen is de in de appendix beschreven benadering van de binomiale verdelingen (met kansen θ_1 en $1-\theta_1$ resp. θ_2 en $1-\theta_2$) door een Poisson-verdeling niet goed meer en kunnen beter de gebruikelijke methoden der dubbele dichotomie (2 x 2-tabel)

gebruikt worden (zie [1]).
2) Ook bij meer dan twee steekproeven kunnen methoden, zij het minder eenvoudige, voor de toetsing van de gelijkheid van kleine kansen worden ontwikkeld (zie b.v. [2]).

3. De keuze van N.

In deze paragraaf behandelen we de vraag naar de grootte van N. We behandelen hierbij alleen het geval $N_1 = N_2 = N$, daar de in paragraaf 2 beschreven toets alleen hiervoor geldt en daar dit bovendien tot de minste waarnemingen leidt. Is $N_1 \neq N_2$ dan moet het totale aantal waarnemingen, $N_1 + N_2$, groter zijn dan wanneer $N_1 = N_2$ om dezelfde doeltreffendheid te bereiken en dit wordt sterker naarmate N_1 en N_2 meer verschillen. Men zal dus zo mogelijk even grote steekproeven nemen.

In principe berust de keuze van N in problemen als het onderhavige, op de wens de ongelijkheid van θ_1 en θ_2 met minstens een bepaalde kans te ontdekken, indien θ_1 en θ_2 bepaalde van elkaar verschillende waarden bezitten. Deze kans, β, om tot verwerping van de getoetste hypothese, $\theta_1 = \theta_2$, te komen, wordt het onderscheidingsvermogen van de toets genoemd en het gaat dus eigenlijk om kennis van dit onderscheidingsvermogen.

Voor twee gevallen, nl. voor de eenzijdige toets met onbetrouwbaarheidsdrempel $\alpha = 0{,}05$ en voor de eenzijdige toets met $\alpha = 0{,}025$ is het onderscheidingsvermogen berekend als functie van de alternatieve hypothesen. Voor tweezijdige toetsing is dit het onderscheidingsvermogen bij respectievelijk $\alpha = 0{,}10$ en $\alpha = 0{,}05$. Bij andere onbetrouwbaarheidsdrempels verlopen de berekeningen volkomen analoog; grotere waarden van α zullen tot een groter en kleinere waarden van α tot een kleiner onderscheidingsvermogen bij dezelfde alternatieve hypothese leiden.

Het blijkt dat het onderscheidingsvermogen β bepaald wordt door de grootheden:

(3) $$\mu_1 \overset{def}{=} N\theta_1 \qquad \text{en} \qquad \mu_2 \overset{def}{=} N\theta_2$$

waarin N de (nog te kiezen) omvang van de steekproeven is. Verder blijkt het nuttig de grootheid:

(4) $$k \overset{def}{=} \frac{\theta_2}{\theta_1} = \frac{\mu_2}{\mu_1}$$

in te voeren. De in de appendix (paragraaf 5) beschreven berekeningen voeren tot de in de figuren 1a en 1b samengevatte numerieke resultaten. In deze figuren is het onderscheidingsvermogen β, bij gegeven k uitgezet als functie van μ_1.

Het gebruik van de figuren voor de bepaling van N is het eenvoudigste te beschrijven aan de hand van een voorbeeld.

Voorbeeld. We onderstellen dat vroegere ervaring of een vooronderzoek geleerd hebben, dat de fracties θ_1 en θ_2 van de orde 0,01 zijn en dat men nu met kans 0,90 wil kunnen onderscheiden tussen fracties $\theta_1 = 0{,}02$ en $\theta_2 = 0{,}005$. Is al van te voren bekend welke van de twee steekproeven bij de grootste fractie zal horen, indien deze ongelijk zijn (dus welke bodemlaag in dat geval de grootste fractie stuifmeelkorrels van de soort A bezit), dan toetsen we eenzijdig (zie § 2). Toetsen we met $\alpha = 0{,}05$, dan moet dus figuur 1b met $k = \frac{\theta_2}{\theta_1} = \frac{0{,}005}{0{,}02} = \frac{1}{4}$ gebruikt worden. Uit deze figuur lezen we dan af, dat om een onderscheidingsvermogen $\beta = 0{,}90$ te bereiken bij $k = \frac{1}{4}$, μ_1 minstens ongeveer 18 moet bedragen.

Volgens (3) moet dan

$$N \geq \frac{\mu_1}{\theta_1} \approx \frac{18}{0{,}02} = 900$$

zijn.

Is niet van te voren bekend bij welke laag de grootste fractie korrels van de bepaalde soort behoort, dan moet er tweezijdig getoetst worden. Voor dezelfde onbetrouwbaarheidsdrempel $= 0{,}05$, moeten we nu figuur 1a gebruiken. Bij $k = \frac{1}{4}$ vinden we dan bij $\beta = 0{,}90$: $\mu_1 \approx 23$ dus

$$N \geq \frac{\mu_1}{\theta_1} \approx \frac{23}{0{,}02} = 1150.$$

Opmerkingen.

1) De figuren bezitten geen grote nauwkeurigheid, daar de krommen eenvoudig als vloeiende lijnen door enkele, verdikt weergegeven, berekende punten getrokken zijn. Voor de toepassingen behoeft echter het onderscheidingsvermogen in de regel niet met grote precisie bekend te zijn.

2) De berekening van het onderscheidingsvermogen geldt, evenals de toepasbaarheid van de toets, alleen indien θ_1 en θ_2 klein zijn. Voor grotere θ's is het onderscheidingsvermogen van de toetsing met de methode der dubbele dichotomie benaderd door Prof. Dr D.van DANTZIG [3] en P.B.PATNAIK [4].

Appendix.

4. Afleiding van de toets.

Is de kans om uit een populatie van exemplaren (b.v. een bodemlaag met stuifmeelkorrels) een exemplaar van een bepaalde soort A te trekken θ, dan wordt de kans, dat men in een steekproef van N exemplaren er n van soort A aantreft, gegeven door de binomiale verdeling:

$$P[\underline{n} = n] = \binom{N}{n} \theta^{n} (1-\theta)^{N-n} \qquad 3)$$

Bij vergelijking van twee steekproeven uit verschillende collecties zullen we de kansen op A aangeven met θ_1 resp. θ_2, de aantallen A-exemplaren in de steekproeven met n_1 resp. n_2. De hypothese H_0, die we willen toetsen, luidt dan: $\theta_1 = \theta_2$.

Het feit dat de kansen θ in de binomiale verdeling klein zijn bij de door ons bekeken gevallen, maakt een benadering door middel van Poisson-verdelingen mogelijk, temeer daar men grote steekproeven zal moeten nemen om nog een redelijke kans te hebben verschillen tussen twee kleine kansen te ontdekken (zie paragraaf 3 en 5).

De Poisson-benadering van de binomiale verdeling ontstaat door de limiet van de binomiale verdeling te nemen voor $N \to \infty$, waarbij tegelijkertijd θ zodanig naar nul gaat, dat $N\theta$ tot een constant getal, μ, nadert. Voor alle gehele waarden van $n \geq 0$ wordt dan de kans op n A-exemplaren bij benadering gegeven door:

$$P[\underline{n} = n] = \frac{e^{-\mu} \mu^{n}}{n!} \qquad \text{(Poisson-verdeling)}$$

In ons geval is dus voor de eerste steekproef:

$$(5) \qquad P[\underline{n}_1 = n_1] = \frac{e^{-N_1\theta_1} (N_1\theta_1)^{n_1}}{n_1!}$$

---

3) Een stochastische grootheid, d.i. een grootheid, die geen vaste waarde bezit maar aan een waarschijnlijkheidsverdeling onderworpen is, geven we aan door een onderstreepte letter; $\underline{n}$ b.v. is hier het aantal exemplaren A in een te nemen steekproef. Een door $\underline{n}$ aangenomen waarde (in één bepaalde steekproef) geven we aan door n zonder onderstreping. $P[\underline{n} = n]$ betekent: de kans, dat de stochastische grootheid $\underline{n}$ de waarde n aanneemt.

en voor de tweede steekproef:

$$(6) \qquad P[\underline{n}_2 = n_2] = \frac{e^{-N_2\theta_2}(N_2\theta_2)^{n_2}}{n_2!} .$$

Daar $\underline{n}_1$ en $\underline{n}_2$ onderling onafhankelijk verdeeld zijn, wordt de kans op een waarnemingsresultaat (n_1, n_2) gegeven door:

$$P[\underline{n}_1 = n_1 \, , \, \underline{n}_2 = n_2] = P[\underline{n}_1 = n_1] \cdot P[\underline{n}_2 = n_2] .$$

De voorwaardelijke kans op (n_1, n_2) onder de voorwaarde $\underline{n}_1+\underline{n}_2 = m$ is dan:

$$P[\underline{n}_1 = n_1 , \underline{n}_2 = n_2 \,|\, \underline{n}_1 + \underline{n}_2 = m] = P[\underline{n}_1 = n_1 \, , \underline{n}_2 = m - n_1] =$$

$$= \frac{P[\underline{n}_1 = n_1] \cdot P[\underline{n}_2 = n_2]}{P[\underline{n}_1 + \underline{n}_2 = m]} .$$

Nu heeft de som van twee Poisson-verdeelde grootheden weer een Poisson-verdeling met als parameter de som van hun parameters, dus:

$$(7) \qquad P[\underline{n}_1 + \underline{n}_2 = m] = \frac{e^{-N_1\theta_1 - N_2\theta_2}(N_1\theta_1 + N_2\theta_2)^m}{m!} ,$$

zodat:

$$P[\underline{n}_1 = n_1 , \underline{n}_2 = m - n_1] = \frac{e^{-N_1\theta_1}(N_1\theta_1)^{n_1} \, e^{-N_2\theta_2}(N_2\theta_2)^{n_2} \; m!}{n_1! \quad n_2! \; e^{-N_1\theta_1 - N_2\theta_2}(N_1\theta_1 + N_2\theta_2)^m} =$$

(8)

$$= \frac{m!}{n_1! \, n_2!} \left(\frac{N_1\theta_1}{N_1\theta_1 + N_2\theta_2}\right)^{n_1} \left(\frac{N_2\theta_2}{N_1\theta_1 + N_2\theta_2}\right)^{n_2} .$$

Onder de hypothese H_0 $(\theta_1 = \theta_2)$ gaat dit over in

$$(9) \qquad P[\underline{n}_1 = n_1 , \underline{n}_2 = m - n_1 \,|\, H_0] = \binom{m}{n_1} \left(\frac{N_1}{N_1 + N_2}\right)^{n_1} \left(\frac{N_2}{N_1 + N_2}\right)^{n_2}$$

en als we uitgaan van steekproeven met gelijke omvang $(N_1 = N_2 = N)$ dan wordt dit nog vereenvoudigd tot:

$$(10) \qquad P[\underline{n}_1 = n_1 , \underline{n}_2 = m - n_1 / H_0] = \binom{m}{n_1} \left(\tfrac{1}{2}\right)^m ,$$

dus een binomiale verdeling met gelijke kansen.

Bij iedere waarde van m is nu met behulp van deze binomiale verdeling een kritiek gebied voor $\underline{n}_1$ met gegeven onbetrouwbaarheidsdrempel α te bepalen (zie voor deze en dergelijke in de toetsingstheorie gebruikelijke termen: memorandum S 47 (M 6)). Is de verdeling van de toetsingsgrootheid, n_1, discreet (d.w.z. dat n_1 slechts discrete waarden kan aannemen), dan is het in het algemeen niet mogelijk het kritieke gebied Z_m zo te kiezen, dat de onbetrouwbaarheid precies α wordt. Gewoonlijk wordt dan voor Z_m het grootste gebied gekozen, waarvoor nog geldt:

$$P[\underline{n}_1 \in Z_m \,|\, m, H_0] \leq \alpha . \tag{11}$$

In dit rapport zullen we α de onbetrouwbaarheidsdrempel blijven noemen en voor de waarde van $P[\underline{n}_1 \in Z_m \,|\, H_0]$ de naam voorwaardelijke (onder voorwaarde $\underline{n}_1 + \underline{n}_2 = m$) onbetrouwbaarheid en het symbool α_{m_0} gebruiken. De onvoorwaardelijke onbetrouwbaarheid geven we dan met α_0 aan:

$$\alpha_0 = \sum_{m=0}^{\infty} \alpha_{m0} \, P[\underline{n}_1 + \underline{n}_2 = m \,/\, H_0] .$$

(De onbetrouwbaarheid α_0 hangt dus ook af van N_1 en N_2, zie (7).) Worden de kritieke gebieden Z_m zo gekozen, dat ze aan (11) voldoen, dan blijkt de onvoorwaardelijke onbetrouwbaarheid α_0 belangrijk kleiner dan α te zijn voor de meest voorkomende waarden van m ($m < 100$). Door een iets gewijzigde keuze van de kritieke zones Z_m is het nu mogelijk de onbetrouwbaarheidsdrempel α beter te benaderen (en hierdoor het onderscheidingsvermogen van de toets te vergroten, zie § 5). We kiezen dan nl. bij iedere m, Z_m zo, dat α_{m0} de onbetrouwbaarheidsdrempel α het beste benadert, waarbij nu dus in een aantal gevallen wel $\alpha_{m0} > \alpha$ kan worden. Dit komt er op neer dat aan een aantal kritieke gebieden Z_m één waarde van n_1 wordt toegevoegd (nl. als α_{m0} zeer veel kleiner dan α was). Enkele van deze gevallen zijn in tabel II vermeld.

Tabel II

Vergelijking van het kritieke gebied, Z_m, voor $\underline{n}_1$ volgens oude en volgens nieuwe methode bepaald (zie tekst)

m	linkseenzijdige toetsing				tweezijdige toetsing			
	oude methode $\alpha_{m0} \leq \alpha = 0{,}05$		nieuwe methode $\alpha_{m0} \simeq \alpha = 0{,}05$		oude methode $\alpha_{m0} \leq \alpha = 0{,}05$		nieuwe methode $\alpha_{m0} \simeq \alpha = 0{,}05$	
	Z_m	α_{m0}	Z_m	α_{m0}	Z_m	α_{m0}	Z_m	α_{m0}
7	0	0,008	≤ 1	0,063	0; 7	0,016	0; 7	0,016
8	≤ 1	0,035	≤ 1	0,035	0; 8	0,008	≤ 1; ≥ 7	0,070
10	≤ 1	0,011	≤ 2	0,055	≤ 1; ≥ 9	0,022	≤ 1; ≥ 9	0,022
11	≤ 2	0,033	≤ 2	0,033	≤ 1; ≥ 10	0,012	≤ 2; ≥ 9	0,065
32	≤ 10	0,025	≤ 11	0,055	≤ 9; ≥ 23	0,020	≤ 10; ≥ 22	0,050
41	≤ 14	0,030	≤ 15	0,059	≤ 13; ≥ 28	0,028	≤ 14; ≥ 27	0,060

De kritieke gebieden voor $\underline{n}_1$, welke voor $m = 1$ tot 100 in tabel I zijn opgegeven, zijn volgens dit laatste voorschrift gekozen. Voor een aantal waarden van μ_1 ($= N\theta_1 = N\theta_2$ onder H_0) werd de onvoorwaardelijke onbetrouwbaarheid α_0 berekend (zie tabel IV, § 5); deze bleek steeds kleiner dan de onbetrouwbaarheidsdrempel α te zijn. Ter vergelijking is in tabel III voor enkele waarden van $(1+k)\mu_1$ de onbetrouwbaarheid α_0 bij gebruik van kritieke zones volgens (11) opgegeven.

Tabel III

Onbetrouwbaarheid α_0 van de toets met kritieke zones volgens (11)

$(1+k)\mu_1$	eenzijdige toetsing onbetrouwbaarheidsdrempel $\alpha = 0{,}025$	$\alpha = 0{,}05$
5	0,005	0,012
10	0,011	0,023
15	0,014	0,030
30	0,015	0,036

Vergelijken we dit met de overeenkomstige waarden uit tabel IV, dan zien we dat de gewijzigde zones een veel betere benadering van α geven, terwijl de verwachting gerechtvaardigd is, dat de onbetrouwbaarheid α_0 ook voor andere waarden van μ_1 slechts weinig van de onbetrouwbaarheidsdrempel α zal verschillen. Op deze manier is dus bij dezelfde onbetrouwbaarheidsdrempel het kritieke gebied en daarmee het onderscheidingsvermogen vergroot. (Onder de nulhypothese, waaronder we α_0 berekenen, is $\mu_1 = \mu_2$ dus $k = 1$; in de 1e kolom van tabel III staat dus de waarde van $2\mu_1$. Om in overeenstemming met tabel IV te blijven, is dit toch met $(1+k)\mu_1$ aangegeven.)

Is $N_1 \neq N_2$ dan geldt voor de toetsingsgrootheid uitdrukking (9); dit is dus een binomiale verdeling met als kansen $\frac{N_1}{N_1+N_2}$ en $\frac{N_2}{N_1+N_2}$, die beide uit het waarnemingsmateriaal bekend zijn. Bij de bepaling van de kritieke gebieden moeten dus tabellen van deze verdelingen gebruikt worden. Zijn deze niet beschikbaar, dan kan, wanneer N_1 en N_2 niet veel uiteenlopen, toch op n_1 en n_2 de beschreven toets worden toegepast. Zijn N_1 en N_2 wel sterk verschillend, is dus één van de kansen $\frac{N_1}{N_1+N_2}$ en $\frac{N_2}{N_1+N_2}$ klein, dan kan de verdeling (9) beter met een Poisson-verdeling, zoals we in het begin van deze paragraaf bespraken, worden benaderd. Wij gaan op deze kwesties niet nader in; zij worden uitvoerig besproken in [1].

5. Berekening van het onderscheidingsvermogen.

Zoals reeds in paragraaf 3 werd vermeld, is het onderscheidingsvermogen alleen berekend voor het geval $N_1 = N_2 = N$. We zullen beginnen met een wat preciezere definitie:

Het onderscheidingsvermogen, β, van een toets ten opzichte van een alternatieve hypothese, welke bepaalde ongelijke waarden aan θ_1 en θ_2 toekent, is de kans op een juiste beslissing. D.w.z. bij eenzijdige toetsing is β de kans dat de toetsingsgrootheid een waarde in het kritieke gebied aanneemt, indien deze bepaalde alternatieve hypothese juist is; bij tweezijdige toetsing is β de kans dat de toetsingsgrootheid een waarde aanneemt van dat deel van de twee waaruit het kritieke gebied bestaat, dat tot een juiste conclusie omtrent θ_1 en θ_2 leidt. B.v.: is de alternatieve hypothese $\theta_1 = 0{,}02$ en $\theta_2 = 0{,}005$ dan is β de kans, dat $\underline{n}_1$ in het rechter deel van het kritieke gebied zal liggen, berekend onder deze hypothese, zodat de toets tot verwerping van $H_0(\theta_1 = \theta_2)$ ten gunste van $\theta_1 > \theta_2$ zal leiden.

Uit deze definities van het onderscheidingsvermogen volgt dat β bij tweezijdige toetsing met onbetrouwbaarheidsdrempel α gelijk is aan β bij eenzijdige toetsing met onbetrouwbaarheidsdrempel $\frac{1}{2}\alpha$.

Het onderscheidingsvermogen zal afhangen van θ_1, θ_2 en N en wel zoals zal blijken van de waarden van $N\theta_1$ en $N\theta_2$. Bovendien zal het onderscheidingsvermogen afhangen van de waarde van α, die immers de grootte van het kritieke gebied bepaalt: bij een kleinere waarde van α zal een kleinere waarde van β behoren.

De nu volgende afleiding van het onderscheidingsvermogen is ook te vinden bij J.PRZYBOROWSKI en H.WILENSKI [5], die β berekenden voor dezelfde toets, maar gebaseerd op kritieke gebieden welke aan (11) voldoen (zie paragraaf 4), waardoor β dus kleiner wordt dan bij de hier gevolgde methode.

Evenals in paragraaf 3, definiëren we:

$$(12) \qquad \mu_1 = N\theta_1 \; ; \mu_2 = N\theta_2 \; ; \; k = \frac{\theta_2}{\theta_1} \text{ en } \underline{m} = \underline{n}_1 + \underline{n}_2 \, .$$

De voorwaardelijke kans (voorwaarde $\underline{m} = m$) op een waarnemingsresultaat (n_1, n_2) (zie (8)), gaat hierdoor over in

$$P[\underline{n}_1 = n_1 , \underline{n}_2 = m - n_1] = \binom{m}{n_1}\left(\frac{1}{1+k}\right)^{n_1}\left(\frac{k}{1+k}\right)^{m-n_1},$$

een binomiale verdeling met kansen $\frac{1}{1+k}$ en $\frac{k}{1+k}$. Door sommatie van deze kansen $P[\underline{n}_1 = n_1 , \underline{n}_2 = m - n_1]$ over het voorwaardelijke kritieke gebied Z_m, vinden we het voorwaardelijk onderscheidingsvermogen β_m . Het onvoorwaardelijke onderscheidingsvermogen β is dan:

$$\beta = \sum_{m=0}^{\infty} \beta_m \, P[\underline{m} = m] \tag{13}$$

waarin volgens (7) en (12):

$$P[\underline{m} = m] = \frac{e^{-(1+k)\mu_1} \{(1+k)\mu_1\}^m}{m!}$$

geldt (een Poisson-verdeling met parameter $(1+k)\mu_1$). De grootheden μ_1 en k zijn hierin door de alternatieve hypothese en N bepaald.

Met behulp van tabellen van de Poisson-verdeling en van binomiale verdelingen met ongelijke kansen (zie [6] en [7]) is nu bij gegeven μ_1, k en α het onderscheidingsvermogen numeriek te bepalen. De resultaten zijn verzameld in tabel IV.

Tabel IV

Onderscheidingsvermogen β van de toets voor gelijkheid van kleine kansen

éénzijdige toetsing, α = 0,025 (tweezijdige toetsing α = 0,05)											
$(1+k)\mu_1$	onbetrouwbaarheid α_0	k = 2/3		k = 1/2		k = 1/3		k = 1/4		k = 1/5	
		μ_1	β	μ_1	β	μ_1	β	μ_1	β	μ_1	β
5	0,012	3	0,03	3,33	0,07	3,75	0,13	4	0,18	4,167	0,22
10	0,020	6	0,08	6,67	0,16	7,50	0,32	8	0,46	8,33	0,55
15	0,023	9	0,11	10,00	0,24	11,25	0,48	12	0,65	12,50	0,75
20	0,023	12	0,13	13,33	0,31	15,00	0,61	16	0,78	16,67	0,86
25	0,022	15	0,15	16,67	0,38	18,75	0,70	20	0,86	20,83	0,92
30	0,022	18	0,18	20,00	0,44	22,50	0,78	24	0,91	25,00	0,95
35	0,023	21	0,21	23,33	0,51	26,25	0,84	28	0,94	29,17	0,97
40	0,024	24	0,23	26,67	0,57	30,00	0,89	32	0,96	33,33	0,98
45	0,024	27	0,26	30,00	0,63	33,75	0,92	36	0,97	37,50	0,98
50	0,024	30	0,29	33,33	0,68	37,50	0,94	40	0,98	41,67	0,98

éénzijdige toetsing α = 0,05 (tweezijdige toetsing α = 0,10)											
$(1+k)\mu_1$	onbetrouwbaarheid α_0	k = 2/3		k = 1/2		k = 1/3		k = 1/4		k = 1/5	
		μ_1	β	μ_1	β	μ_1	β	μ_1	β	μ_1	β
5	0,030	3	0,07	3,33	0,13	3,75	0,22	4	0,29	4,167	0,35
10	0,042	6	0,13	6,67	0,25	7,50	0,45	8	0,59	8,33	0,67
15	0,047	9	0,18	10,00	0,36	11,25	0,62	12	0,77	12,50	0,84

vervolg tabel IV (eenz. toetsing, $\alpha = 0,05$).

$(1+k)\mu_1$	α_0	$k = 2/3$		$k = 1/2$		$k = 1/3$		$k = 1/4$		$k = 1/5$	
		μ_1	β	μ_1	β	μ_1	β	μ_1	β	μ_1	β
20	0,047	12	0,22	13,33	0,44	15,00	0,73	16	0,87	16,67	0,92
25	0,046	15	0,24	16,67	0,51	18,75	0,81	20	0,92	20,83	0,95
30	0,046	18	0,28	20,00	0,57	22,50	0,86	24	0,95	25,00	0,97
35	0,047	21	0,31	23,33	0,63	26,25	0,90	28	0,96	29,17	0,98
40	0,048	24	0,34	26,67	0,69	30,00	0,93	-	-	-	-
45	0,048	27	0,37	30,00	0,73	-	-	-	-	-	-
50	0,049	30	0,40	33,33	0,77	-	-	-	-	-	-

De berekeningen werden uitgevoerd voor $\alpha = 0,05$ en $\alpha = 0,025$ bij éénzijdige toetsing (voor tweezijdige toetsing komt dit dus overeen met $\alpha = 0,10$ en $\alpha = 0,05$).

De sommatie over m (zie (13)) is niet uitgestrekt tot hele grote en kleine waarden van m, maar zodanig afgebroken dat de som van $P[\underline{m} = m]$ voor alle niet gebruikte waarden van m kleiner dan 0,015 is. Het gevonden onderscheidingsvermogen kan dus hoogstens 0,015 kleiner dan de werkelijke waarde zijn ($\beta_m < 1$ voor alle m), afgezien natuurlijk van fouten, die het gevolg zijn van de benadering van een binomiale verdeling door een Poisson-verdeling, die we in het begin van paragraaf 4 uitvoerden.

Opmerking.

Een gebruikelijke definitie van het onderscheidingsvermogen luidt: β is de kans dat H_0 verworpen wordt, indien een gegeven alternatieve hypothese juist is. Dat wil dus zeggen: de kans dat de toetsingsgrootheid in het kritieke gebied valt, berekend onder de alternatieve hypothese. Bij tweezijdige toetsing zou dan ook dat deel van de kritieke zone, dat tot een verkeerde conclusie omtrent de alternatieve hypothese zou leiden, een bijdrage tot het onderscheidingsvermogen leveren. Liever definiëren we dan ook β als de kans om H_0 te verwerpen ten gunste van een juiste beslissing.

Tabel I

Kritieke waarden van de linkseenzijdige toets voor gelijkheid van kleine kansen

m	onbetrouwbaarheidsdrempel				m	onbetrouwbaarheidsdrempel			
	0,005	0,01	0,025	0,05		0,005	0,01	0,025	0,05
1	-	-	-	-	51	16	17	18	19
2	-	-	-	-	52	16	17	18	20
3	-	-	-	-	53	17	18	19	20
4	-	-	-	0	54	17	18	19	20
5	-	-	0	0	55	17	18	20	21
6	-	0	0	0	56	18	19	20	21
7	0	0	0	1	57	18	19	21	22
8	0	0	1	1	58	19	20	21	22
9	0	0	1	1	59	19	20	21	23
10	0	1	1	2	60	20	20	22	23
11	1	1	2	2	61	20	21	22	24
12	1	1	2	3	62	20	21	23	24
13	1	2	2	3	63	21	22	23	24
14	2	2	3	3	64	21	22	24	25
15	2	2	3	4	65	22	23	24	25
16	2	3	4	4	66	22	23	25	26
17	3	3	4	5	67	22	23	25	26
18	3	4	4	5	68	23	24	25	27
19	3	4	5	5	69	23	24	26	27
20	4	4	5	6	70	24	25	26	28
21	4	5	5	6	71	24	25	27	28
22	4	5	6	7	72	25	26	27	28
23	5	5	6	7	73	25	26	28	29
24	5	6	7	7	74	25	26	28	29
25	6	6	7	8	75	26	27	28	30
26	6	7	7	8	76	26	27	29	30
27	6	7	8	9	77	27	28	29	31
28	7	7	8	9	78	27	28	30	31
29	7	8	9	9	79	28	29	30	32
30	7	8	9	10	80	28	29	31	32
31	8	8	9	10	81	28	30	31	33
32	8	9	10	11	82	29	30	32	33
33	9	9	10	11	83	29	30	32	33
34	9	10	11	12	84	30	31	32	34
35	9	10	11	12	85	30	31	33	34
36	10	10	12	13	86	31	32	33	35
37	10	11	12	13	87	31	32	34	35
38	11	11	12	13	88	31	33	34	36
39	11	12	13	14	89	32	33	35	36
40	11	12	13	14	90	32	33	35	37
41	12	13	14	15	91	33	34	36	37
42	12	13	14	15	92	33	34	36	38
43	13	13	15	16	93	34	35	37	38
44	13	14	15	16	94	34	35	37	38
45	13	14	15	16	95	34	36	37	39
46	14	15	16	17	96	35	36	38	39
47	14	15	16	17	97	35	37	38	40
48	15	15	17	18	98	36	37	39	40
49	15	16	17	18	99	36	37	39	41
50	15	16	18	19	100	37	38	40	41

Literatuur.

[1] Mej. C.van EEDEN, Methoden voor het vergelijken, toetsen en schatten van onbekende kansen, Rapport S 115 (M 45) van de Statistische Afdeling van het Mathematisch Centrum (1953), verschijnt in Statistica 3 en 4 (1953).

[2] R.A.FISHER, The significance of deviations from expectations in a Poisson series, Biometrics 6 (1950), pp. 17-24.

[3] Prof. Dr D.van DANTZIG, Kadercursus Mathematische Statistiek, pp. 309-326.

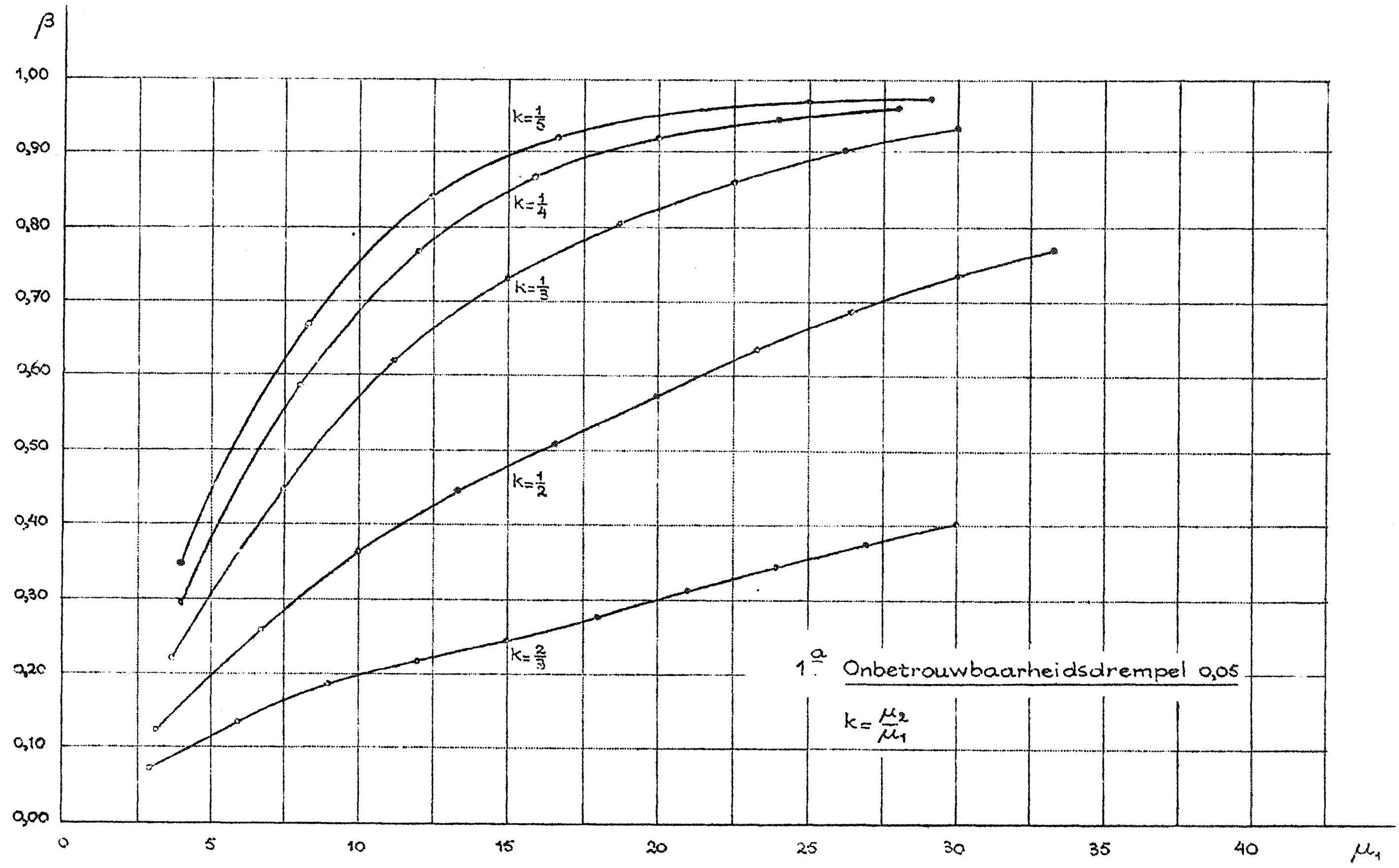
[4] P.B.PATNAIK, The power function of the test for the difference between two proportious in a 2x2-table, Biometrika 35 (1948), pp. 157-175.

[5] J.PRZYBOROWSKI and H.WILENSKI, Homogeneity of results in testing samples from Poisson series with an application to testing clover seed for dodder, Biometrika 31 (1940), pp. 313-323.

[6] E.C.MOLINA, Poisson's exponential binomial limit, D.van Nostrand Company, New York (1945).

[7] -----------, Tables of the binomial probability distribution, National Bureau of Standards, Applied Mathematics, Series 6 (1950).

[8] A.van WIJNGAARDEN, Table of the cumulative symmetric binomial distribution, Proceedings van de Kon. Ned. Akad. v. Wetensch. 53 (1950), pp. 858-868, Indagationes Mathematicae 12 (1950), p. 301-312.



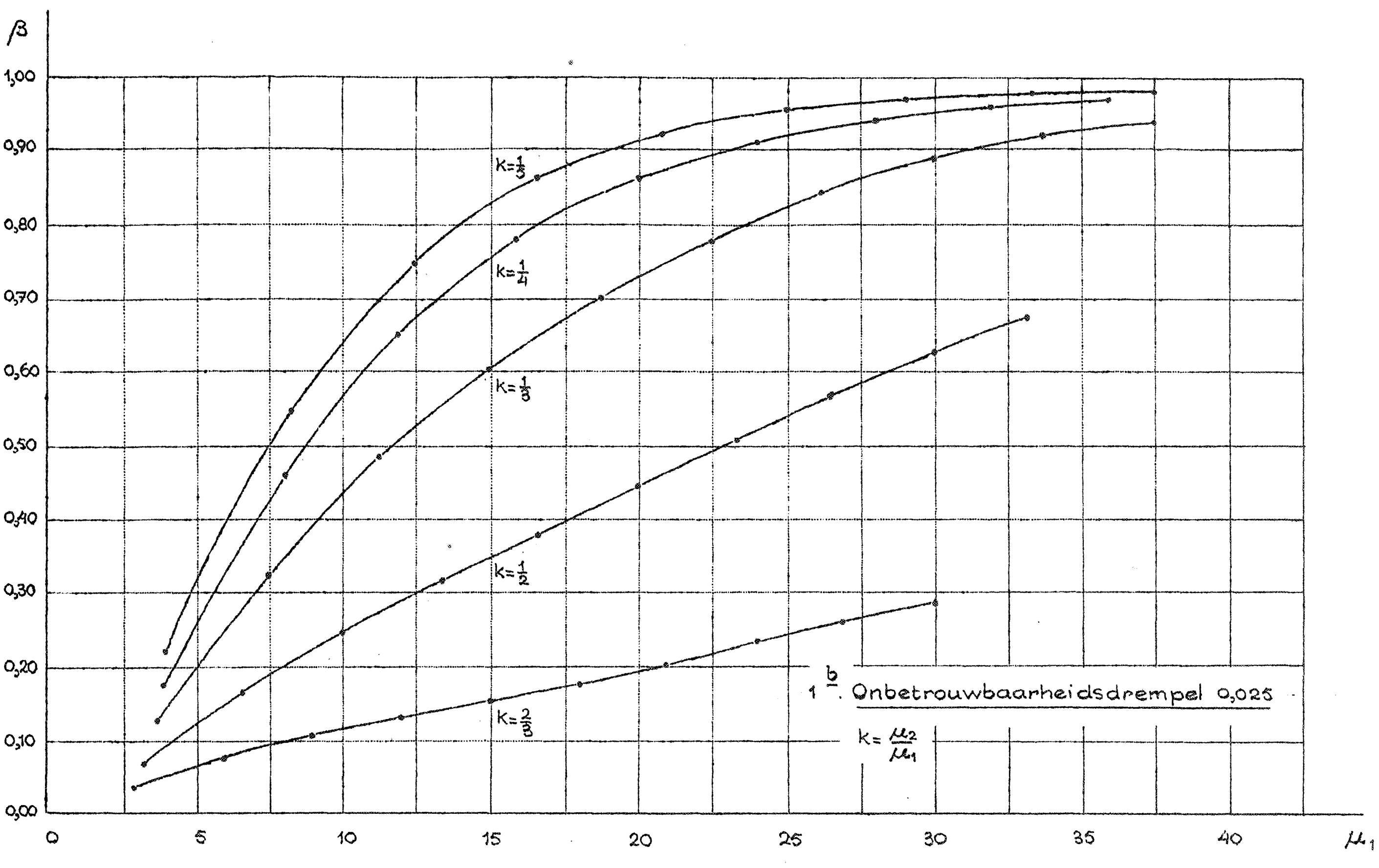


Fig. 1.1 Onderscheidingsvermogen van de eenzijdige toets voor gelijkheid van de gemiddelden van twee Poisson-verdeelde grootheden.

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling.

S47(M6)

Algemene gang van zaken bij het toetsen van een hypothese.[1]

De toetsing van een hypothese H_0 berust steeds op een aantal waarnemingen $x_1, x_2, \dots, x_n$ van één of meer stochastische grootheden[2], of op enige groepen van waarnemingen (bv. twee steekproeven).

Bij een toets behoort een toetsingsgrootheid $\underline{u}$ (soms meer dan één), die een functie is van bovengenoemde stochastische grootheden en die, voor de waargenomen waarden $x_1, x_2, \dots, x_n$ een waarde aanneemt, die berekend kan worden (bv.: het gemiddelde der waarnemingen, of de spreiding, of het verschil van de gemiddelden van twee waarnemingen).

De toetsingsgrootheid wordt steeds zo gekozen, dat men, op grond van de onderstelling, dat H_0 juist is, de waarschijnlijkheidsverdeling van deze grootheid kan berekenen.

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van $\underline{u}$, en wel op zodanige wijze, dat de kans, dat $\underline{u}$ een in Z gelegen waarde aanneemt, onder de hypothese H_0, gelijk is aan een gegeven getal α, zodat Z dus van α afhankelijk is[3]. Z heet de kritieke zône van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwerpt nu H_0 op grond van de waarnemingen $x_1, x_2, \dots, x_n$, indien de bij deze waarnemingen behorende waarde van $\underline{u}$ in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "significant" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien H_0 juist is, gelijk aan α. Derhalve is α de kans op ten onrechte verwerping van de juiste hypothese, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met α = 0,05 resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, waarin de hypothese die men toetst juist is, deze toch verwerpen.

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

2) Een stochastische grootheid is een grootheid, die een waarschijnlijkheidsverdeling bezit, of, anders gezegd, een grootheid, die voor de elementen van een collectie (universum, populatie) gedefinieerd is en daarop allerlei waarden aanneemt. Stochastische grootheden worden aangegeven door onderstreepte letters.

3) Soms kan men slechts bereiken, dat deze kans $\leq \alpha$ is.

De toetsingstheorie biedt in het algemeen geen mogelijkheid om tot aanvaarding van een hypothese te komen. Indien een bepaalde hypothese H_0 niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

Wel zal men vaak in de loop van een statistische analyse bepaalde onderstellingen, die plausibel schijnen en voor de verdere analyse van nut zijn, toetsen, alvorens ze bij de verdere bewerking van het materiaal te gebruiken. Worden zij dan op grond van de toets niet verworpen, dan houdt dit in zo verre een rechtvaardiging van die onderstellingen in, dat een grote afwijking door de toets veelal wel zou zijn ontdekt. Indien men dan verder de onderstellingen gebruikt, verwaarloost men eventueel aanwezige afwijkingen van onbekende grootte, die echter niet zo groot zijn, dat zij door de toets zijn ontdekt.

Vele toetsen gelden zelf alleen onder bepaalde onderstellingen omtrent de waarschijnlijkheidsverdelingen der stochastische grootheden, waarvan waarnemingen zijn verricht. Deze nevenvoorwaarden dienen steeds uitdrukkelijk te worden vermeld en, zo mogelijk, zelf te worden getoetst.

In plaats van de onbetrouwbaarheidsdrempel α wordt vaak bij de uitslag van een toetsing de overschrijdingskans k opgegeven; dit is de kleinste waarde van α, waarbij in het betrokken geval, nog tot verwerping van H_0 zou zijn overgegaan; anders gezegd: de kleinste α, waarvoor de gevonden waarde der toetsingsgrootheid nog juist in de (bij α behorende) kritieke zône Z ligt. Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α, dan wordt verworpen, indien $k \leqq \alpha$ is.

Voor het onderscheid tussen één- en tweezijdige toetsing en de keuze tussen deze twee mogelijkheden vergelijke men bv. de tweede hieronder gegeven litteratuurplaats. Wij moeten hier volstaan met de opmerking, dat éénzijdige toetsing veelal eerder tot verwerping van H_0 leidt, maar dat deze slechts onder bijzondere omstandigheden kan worden toegepast.

Litteratuur:

J.Neyman, First course in probability and statistics, New York, 1950, Chapter 5.

J.Hemelrijk en H.R. van der Vaart, Het gebruik van één- en tweezijdige overschrijdingskansen voor het toetsen van hypothesen, Statistica 4 (1950) p.54-66.