STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 73 (M 30a)

# Twee parametervrije toetsen voor k steekproeven

1952

Statistische Afdeling
S 73 (M 30a)

## Twee parametervrije toetsen voor k steekproeven.[1)]

Gegeven zijn k onafhankelijke steekproeven van k stochastische grootheden:

$x_{1,1}, x_{1,2}, \ldots, x_{1,n_1}$ van de grootheid $\underline{x}_1$ [2)],

$x_{2,1}, x_{2,2}, \ldots, x_{2,n_2}$ van de grootheid $\underline{x}_2$,

---------------------------------------

$x_{k,1}, x_{k,2}, \ldots, x_{k,n_k}$ van de grootheid $\underline{x}_k$.

De uitgebreidheden van de steekproeven zijn dus $n_1, n_2, \ldots, n_k$; de eerste index van de waarneming $x_{i,j}$ geeft aan, uit welke steekproef deze waarneming afkomstig is, terwijl de tweede index het nummer der waarneming binnen die steekproef aangeeft.

De hypothese $H_0$, die wij wensen te toetsen, luidt dat de waarnemingen van dezelfde stochastische grootheid afkomstig zijn.
Anders uitgedrukt: $H_0$ houdt in, dat de stochastische grootheden $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_k$ onderling onafhankelijk verdeeld zijn en alle dezelfde waarschijnlijkheidsverdeling bezitten.
Om $H_0$ te toetsen kunnen we gebruik maken van de volgende twee toetsingsmethoden.

### 1. De H-toets.

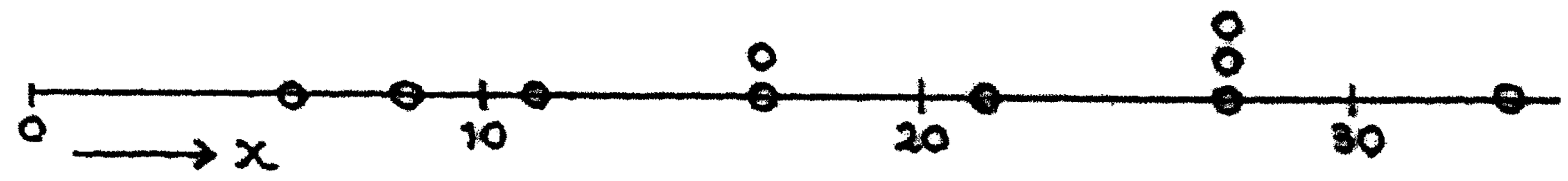
De berekening van de toetsingsgrootheid $\underline{H}$, welke bij de toets gebruikt wordt, geschiedt op de volgende wijze (zie KRUSKAL en WALLIS [1] en [2], RIJKOORT [3] en TERPSTRA [4]).
Alle waarnemingen worden naar opklimmende grootte gerangschikt en vervolgens van een rangnummer voorzien. Indien alle waarnemingen verschillend zijn, zijn dit de rangnummers 1 tot en met $n = n_1+n_2+\ldots+n_k$. Indien t waarnemingen gelijk aan elkaar zijn, wordt aan elk der t waarnemingen eenzelfde rangnummer toegekend. Dit rangnummer is dan gelijk aan het gemiddelde der rangnummers, welke de waarnemingen gekregen zouden

------------------

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.
2) Stochastische grootheden worden door onderstreepte letters aangeduid.

hebben, indien ze alle verschillend waren.

<u>Voorbeeld</u>:




rangnummers: 1, 2, 3, $4\frac{1}{2}$, $4\frac{1}{2}$, 6, 8, 8, 8, 10.

Stelt nu $\underline{R}_i$ de som van de rangnummers in de $i^{\underline{e}}$ steekproef voor en zijn alle waarnemingen onderling verschillend, dan wordt de toetsingsgrootheid $\underline{H}$ gedefiniëerd door

$$\underline{H} = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{\underline{R}_i^2}{n_i} - 3(n+1).$$

Komen onder alle n waarnemingen tezamen r groepen van gelijken voor, waarbij de eerste groep uit $t_1$ waarnemingen bestaat, de tweede groep uit $t_2$ waarnemingen etc., dan wordt bovenstaande uitdrukking voor $\underline{H}$ gedeeld door de factor

$$N = \left[1 - \{(t_1-1)t_1(t_1+1) + \dots + (t_r-1)t_r(t_r+1)\}\right] / (n^3-n).$$

Indien alle waarnemingen verschillend zijn is $N = 1$.
<u>In het algemeen</u> geldt dus:

$$\underline{H} = \left[\frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{\underline{R}_i^2}{n_i} - 3(n+1)\right] / N.$$

Voor het geval van twee steekproeven ($k = 2$) is $\underline{H}$ gelijk aan de genormeerde toetsingsgrootheid van WILCOXON [6] voor twee steekproeven. De H-toets kan dus beschouwd worden als een generalisatie van de toets van WILCOXON.

Beschouwen wij nu de verzameling van alle bij het experiment mogelijke uitkomsten, dan bezit $\underline{H}$ op deze verzameling een waarschijnlijkheidsverdeling. Indien de hypothese $H_0$, inhoudende dat alle waarnemingen uit dezelfde verdeling afkomstig zijn, juist is, dan bezit $\underline{H}$ voor grote waarden van $n_1, n_2, \dots, n_k$ bij benadering een $\chi^2$-verdeling met k-1 vrijheidsgraden.

Indien de hypothese $H_0$ niet vervuld is, bezit $\underline{H}$ gemiddeld grotere waarden, dan indien $H_0$ wel vervuld is. De hypothese $H_0$ wordt daarom verworpen, indien de uit de steekproef bepaalde waarde van $\underline{H}$ groter is dan een kritieke waarde $H_\alpha$, welke gegeven wordt door

$$P[\underline{H} \geq H_\alpha | H_0] = P[\chi^2_{k-1} \geq H_\alpha | H_0] = \alpha.$$

Hierin is $\alpha$ de voorafgekozen onbetrouwbaarheidsdrempel: indien de gevonden waarde H groter is dan $H_\alpha$ wordt de hypothese $H_0$ verworpen met een onbetrouwbaarheid $\alpha$.
De waarde $H_\alpha$ kan gemakkelijk bepaald worden uit tabellen en nomogrammen van de $\chi^2$-verdeling, evenals de overschrijdingskans van de gevonden waarde H, welke gedefinieerd wordt door:

$$P[\underline{H} \geqq H | H_0] = P[\chi^2_{k-1} \geqq H | H_0].$$

Een benadering voor kleine steekproeven.

Indien de steekproeven niet voldoende groot zijn, wordt de exacte verdeling van $\underline{H}$ niet meer goed benaderd door de $\chi^2$-verdeling.
Voor het geval van 3 steekproeven met $n_i \leqq 5$ $(i=1,2,3)$ zijn de kritieke waarden van $\underline{H}$, corresponderende met de onbetrouwbaarheden $\alpha = 0{,}10$, $0{,}05$ en $0{,}01$, exact berekend (zie [2] en [3]).
Aan de hand van deze uitkomsten is aangetoond, dat de exacte verdeling van $\underline{H}$ voor kleine steekproeven zeer goed benaderd wordt door de F-verdeling. Deze verdeling wordt gekarakteriseerd door 2 grootheden $\nu_1$ en $\nu_2$, de zgn. aantallen graden van vrijheid (zie b.v. [5]).
De uit de steeproefwaarden te berekenen grootheid F wordt dan gegeven door

$$\underline{F} = \frac{\underline{H}(M-k+1)}{(k-1)(M-H)},$$

waarin

$$M = \frac{n^3 - \sum_{i=1}^{k} n_i^3}{n(n+1)}.$$

Verder worden $\nu_1$ en $\nu_2$ gegeven door

$$\nu_1 = 2(k-1)\,\frac{(k-1)(M-k+1) - \mathcal{V}}{M\mathcal{V}}$$

en

$$\nu_2 = 2(M-k+1)\,\frac{(k-1)(M-k+1) - \mathcal{V}}{M\mathcal{V}} = \frac{M-k+1}{k-1}\cdot \nu_1,$$

waarin

$$\mathcal{V} = 2(k-1) - \frac{2[3k^2 - 6k + n(2k^2 - 6k + 1)]}{5n(n+1)} - \frac{6}{5}\sum_{i=1}^{k}\frac{1}{n_i}.$$

De kritieke waarden voor $\underline{F}$ en de overschrijdingskans van de gevonden waarde F kunnen worden bepaald uit tabellen en nomogrammen van de F-verdeling. Grote waarden van F leiden tot verwerping van $H_0$.

2. De $T^2$-toets.

De bij deze toets gebruikte toetsingsgrootheid $\underline{T}^2$ wordt als volgt berekend (TERPSTRA [4]).
Evenals bij de H-toets, worden allereerst alle waarnemingen naar opklimmende grootte gerangschikt (voor gelijke waarnemingen geldt hetzelfde als onder 1), en wordt voor iedere steekproef de som der rangnummers $\underline{R}_i$ $(i = 1,2,\ldots,k)$ bepaald.
Bovendien wordt nog ieder tweetal van steekproeven onderling vergeleken. Dit gaat het gemakkelijkst door eerst de steekproeven willekeurig te nummeren.
De $h^e$ en $j^e$ steekproef $(h \leq j)$ worden dan vergeleken door de $n_h+n_j$ waarnemingen uit de twee steekproeven naar opklimmende grootte te rangschikken en van een rangnummer te voorzien (voor gelijke waarnemingen geldt weer hetzelfde als onder 1). Voor de $h^e$ steekproef wordt vervolgens de som der rangnummers bepaald, welk aantal we $\underline{R}_h^{(j)}$ noemen.
De toetsingsgrootheid $\underline{T}^2$ wordt dan gedefinieerd door

$$\underline{T}^2 = 12 \sum_{h<j} \frac{\underline{\tilde{U}}_{h,j}^2}{n_h n_j} - \frac{12}{n+1} \sum_{i=1}^{k} \frac{\underline{\tilde{U}}_i^2}{n_i},$$

waarin $$\underline{\tilde{U}}_{h,j} = \underline{R}_h^{(j)} - \tfrac{1}{2} n_h (n_h + n_j + 1),$$

en $$\underline{\tilde{U}}_i = \underline{R}_i - \tfrac{1}{2} n_i (n+1).$$ 3)

De toetsingsgrootheid $\underline{T}^2$ kan ook geschreven worden als

$$\underline{T}^2 = 12 \sum_{h<j} \frac{\underline{\tilde{U}}_{h,j}^2}{n_h n_j} - n\underline{H},$$

--------------------

3) De grootheden $\underline{\tilde{U}}_i$ en $\underline{\tilde{U}}_{h,j}$ staan in verband met de toetsingsgrootheid $\underline{U}$ van WILCOXON [6], zoals ze gedefinieerd is door MANN en WHITNEY [7]. Beschouwen we 2 steekproeven van de uitgebreidheden n en m, dan is $\underline{U}$ het aantal keren, dat een waarneming uit de tweede steekproef kleiner is dan een waarneming uit de eerste steekproef (het zgn. aantal inversies) en is $\underline{\tilde{U}} = \underline{U} - \frac{1}{2}nm$. De grootheden $\underline{\tilde{U}}_i$ en $\underline{\tilde{U}}_{h,j}$ hebben betrekking op de $i^e$ steekproef en de overige steekproeven tezamen, resp. op de $h^e$ en de $j^e$ steekproef.

waarin $\underline{H}$ gegeven wordt door de teller van de algemene uitdrukking onder 1e.
Voor het geval van 2 steekproeven ($k = 2$) is $\underline{T}^2$ gelijk aan de genormeerde toetsingsgrootheid $\underline{U}$ van WILCOXON [6] voor 2 steekproeven.

Indien de hypothese $H_0$, inhoudende dat alle waarnemingen uit dezelfde verdeling afkomstig zijn, juist is, bezit de toetsingsgrootheid $\underline{T}^2$ voor grote waarden van $n_1, n_2, \ldots, n_k$ bij benadering een $\chi^2$-verdeling met $\frac{k(k-1)}{2}$ vrijheidsgraden.
Indien $H_0$ niet vervuld is, bezit $\underline{T}^2$ gemiddeld grotere waarden dan onder $H_0$.
De hypothese $H_0$ wordt daarom verworpen indien de gevonden waarde $T^2$ groter is dan een kritieke waarde $T^2_\alpha$, welke gegeven wordt door

$$P\left[\underline{T}^2 \geqq T^2_\alpha \mid H_0\right] = P\left[\underline{\chi}^2_{\frac{k(k-1)}{2}} \geqq T^2_\alpha \mid H_0\right] = \alpha .$$

De onbetrouwbaarheid van de uitspraak is dan gelijk aan $\alpha$.
De overschrijdingskans van de gevonden waarde $T^2$ wordt gegeven door

$$P\left[\underline{T}^2 \geqq T^2 \mid H_0\right] = P\left[\underline{\chi}^2_{\frac{k(k-1)}{2}} \geqq T^2 \mid H_0\right].$$

Opmerkingen

1e. Uit het voorgaande blijkt, dat de $T^2$-toets bewerkelijker is dan de H-toets. Dit zal in vele gevallen echter geen bezwaar zijn, daar de $T^2$-toets eventueel bestaande verschillen tussen de k steekproeven vermoedelijk scherper aantoont dan de H-toets, d.w.z. dat hij een groter onderscheidend vermogen bezit.
2e. Dat de grootheid $\underline{T}^2$ voor grote waarden van $n_1, n_2, \ldots, n_k$ bij benadering een $\chi^2$-verdeling bezit, is strikt genomen alleen bewezen voor het geval, dat alle waarnemingen verschillend zijn. Indien echter niet te veel gelijken voorkomen, kan bovenstaande methode zonder bezwaar toegepast worden.
3e. Een onderzoek naar de aard van de verdeling van $\underline{T}^2$ voor kleine steekproeven is nog gaande. Vermoedelijk zal, evenals bij de H-toets, de exacte verdeling van $\underline{T}^2$ goed benaderd kunnen worden met behulp van de F-verdeling.

## 3. Voorbeeld van de H-toets en de $T^2$-toets.

Gegeven zijn de volgende 5 steekproeven, elk van de uitgebreidheid 10.

1.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $R_1$ | 1 | 2 | 4 | 6 | 9 | 13 | 16 | 20 | 23 | 28 |
| $R_1^{(2)}$ | 1 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 15 | 17 |
| $R_1^{(3)}$ | 1 | 2 | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 |
| $R_1^{(4)}$ | 1 | 2 | 3 | 4 | 5 | 7 | 8 | 10 | 12 | 14 |
| $R_1^{(5)}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 11 |

2.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $R_2$ | 3 | 5 | 8 | 11 | 15 | 19 | 24 | 29 | 33 | 36 |
| $R_2^{(3)}$ | 1 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 |
| $R_2^{(4)}$ | 1 | 2 | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 |
| $R_2^{(5)}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 9 | 11 | 12 |

3.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $R_3$ | 7 | 10 | 14 | 17 | 21 | 27 | 31 | 34 | 39 | 43 |
| $R_3^{(4)}$ | 1 | 2 | 4 | 5 | 7 | 10 | 12 | 13 | 16 | 18 |
| $R_3^{(5)}$ | 1 | 2 | 3 | 4 | 5 | 7 | 8 | 10 | 12 | 15 |

4.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $R_4$ | 12 | 18 | 22 | 25 | 30 | 35 | 38 | 41 | 44 | 46 |
| $R_4^{(5)}$ | 1 | 2 | 3 | 4 | 6 | 8 | 10 | 12 | 14 | 16 |

5.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $R_5$ | 26 | 32 | 37 | 40 | 42 | 45 | 47 | 48 | 49 | 50 |

De steekproeven zijn op de in paragraaf 1 en 2 gegeven wijzen van verschillende rangnummers voorzien. Uit het schema vinden we:

| $R_i$ | $R_h^{(j)}$ | $\tilde{U}_{h,j}$ | $R_i^2$ | $\tilde{U}^2_{h,j}$ | $R_i^2/n_i$ | $\tilde{U}^2_{h,j}/n_h n_j$ |
|---|---|---|---|---|---|---|
| $R_1=122$ | | | $R_1^2=14884$ | | 1488,4 | |
| | $R_1^{(2)}=89$ | $\tilde{U}_{1,2}=-16$ | | 256 | | 2,56 |
| | $R_1^{(3)}=76$ | $\tilde{U}_{1,3}=-29$ | | 841 | | 8.41 |
| | $R_1^{(4)}=66$ | $\tilde{U}_{1,4}=-39$ | | 1521 | | 15,21 |
| | $R_1^{(5)}=56$ | $\tilde{U}_{1,5}=-49$ | | 2401 | | 24,01 |
| $R_2=183$ | | | $R_2^2=33489$ | | 3348,9 | |
| | $R_2^{(3)}=91$ | $\tilde{U}_{2,3}=-14$ | | 196 | | 1,96 |
| | $R_2^{(4)}=76$ | $\tilde{U}_{2,4}=-29$ | | 841 | | 8,41 |
| | $R_2^{(5)}=60$ | $\tilde{U}_{2,5}=-45$ | | 2025 | | 20,25 |

| $R_i$ | $R_h^{(j)}$ | $\tilde{U}_{h,j}$ | $R_i^2$ | $\tilde{U}_{h,j}^2$ | $R_i^2/n_i$ | $\tilde{U}_{h,j}^2/n_h n_j$ |
|---|---|---|---|---|---|---|
| $R_3=243$ | | | $R_3^2=59049$ | | 5904,9 | |
| | $R_3^{(4)}=88$ | $\tilde{U}_{3,4}=-17$ | | 289 | | 2,89 |
| | $R_3^{(5)}=67$ | $\tilde{U}_{3,5}=-38$ | | 1444 | | 14,44 |
| $R_4=311$ | | | $R_4^2=96721$ | | 9672,1 | |
| | $R_4^{(5)}=76$ | $\tilde{U}_{4,5}=-29$ | | 841 | | 8,41 |
| $R_5=416$ | | | $R_5^2=173056$ | | 17305,6 | |

Hieruit volgt:

$$\sum_{i=1}^{5} \frac{R_i^2}{n_i} = 37719{,}90.$$

en

$$\sum_{h<j\leq 5} \frac{\tilde{U}_{h,j}^2}{n_h n_j} = 106{,}55.$$

We vinden dus

$$H = \frac{12}{50 \times 51}\, 37719{,}90 - 3\times 51 = 177{,}51 - 153 = 24{,}51$$

en

$$T^2 = 12 \times 106{,}55 - 50 \times 24{,}51 = 53{,}10.$$

De grootheden $\underline{H}$ en $\underline{T}^2$ bezitten onder de hypothese $H_0$ bij benadering $\chi^2$-verdelingen met resp. 5-1 = 4 en $\frac{5 \times 4}{2}$ = 10 graden van vrijheid.
Voor de berekende waarden van H en $T^2$ vinden we in een $\chi^2$-tabel of nomogram overschrijdingskansen, welke beide kleiner zijn dan $1.10^{-4}$.
De steekproeven zijn dus volgens beide toetsingsmethoden duidelijk systematisch verschillend en $H_0$ moet verworpen worden.

4. Literatuur.


[1] W.H.KRUSKAL, A non-parametric test for the several sample problem, Ann. Math. Stat. 23, no. 4 (1952).

[2] W.H.KRUSKAL and W.A.WALLIS, Use of ranks in one criterion analysis of variance, Journ. Am. Stat. Ass., 47 (1952), pp. 538-621.

[3] P.J.RIJKOORT, A generalization of Wilcoxon's test, Kon. Ned. Ak. v. Wet. Proc. A 55, Indagationes Mathematicae 14 (1952), pp. 394-404.

[4] T.J.TERPSTRA, A non-parametric k-sample-test and its connection with the H-test, Report S 92 (VP 2), Mathematisch Centrum, Amsterdam.

[5] H.CRAMÈR, Mathematical Methods of Statistics, 1946.

[6] F.WILCOXON, Individual comparisons by ranking methods, BiometricsBull. 1, 80-83 (1945).

[7] H.B.MANN and D.R.WHITNEY, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. 18, 50-60 (1947).

5. Tabellen en nomogrammen van de $\chi^2$-verdeling.

1. M.G.KENDALL, The Advanced Theory of Statistics, I, 1947, p. 444-446.
2. H.CRAMÈR, Mathematical Methods of Statistics, 1946, p. 559.
3. Statistica 1 (1946), p. 109.

6. Tabellen en nomogrammen van de F-verdeling.

1. Maxime Merrington and Catherine M.Thompson, Tables of percentage points of the inverted Beta (F) distribution, Biometrika 33 (1943), 73-88.