

MATHEMATISCH CENTRUM,
2de Boerhaavestraat 49,
A m s t e r d a m - O.
Rapport S 145 (M 52)
door Ph. van Elteren.

A

Constructie van betrouwbaarheidsintervallen
met behulp van de toets van WILCOXON.

1. Betrouwbaarheidsinterval bij 2 steekproeven.

Gegeven zijn twee steekproeven:

$x_1, \dots, x_n$ van de stochastische variabele $\underline{x}$ en
 $y_1, \dots, y_m$ " " " " $\underline{y}$.

Er wordt ondersteld dat $\underline{x}$ en $\underline{y}$ verdelingen hebben, die alléén in hun gemiddelden kunnen verschillen. Dus:

$$\mathcal{E}\underline{x} - \mathcal{E}\underline{y} = \delta$$

en $\underline{x}$ en $\underline{y} + \delta$ hebben dezelfde verdeling.

Gevraagd wordt op grond van de steekproeven een betrouwbaarheidsinterval met onbetrouwbaarheidsdrempel α voor δ te bepalen.

Deze vraag kan worden beantwoord met behulp van de toets van WILCOXON, welke wordt behandeld in memorandum S 145 (M 7) van het Mathematisch Centrum. Daarin wordt de toetsingsgrootte $\underline{U}$ van WILCOXON voor steekproeven van de uitgebreidheid n van $\underline{x}$ en van de uitgebreidheid m van $\underline{y}$ gedefinieerd en de verdeling van $\underline{U}$ beschreven onder de hypothese H_0 dat $\underline{x}$ en $\underline{y}$ dezelfde verdeling hebben. Als de verdeling van $\underline{U}$ onder H_0 bekend is, kan de kritieke waarde U_ϵ worden bepaald, gedefinieerd als de grootste waarde, die $\underline{U}$ kan aannemen waarvoor geldt:

$$P \left\{ \underline{U} \leq U_\epsilon \mid H_0 \right\}^1 \leq \epsilon.$$

Voor de bepaling van het betrouwbaarheidsinterval gaan wij nu als volgt te werk:

1. Bepaal alle $m \cdot n$ verschillen van een waarneming van $\underline{x}$ en een waarneming van $\underline{y}$, die men uit het waarnemingsmateriaal kan vormen,
2. Rangschik deze verschillen naar opklimmende grootte,
3. Bepaal voor ieder verschil $x_i - y_j$ de volgende aantallen:
 - a = het aantal verschillen, gelijk aan dat verschil.
 - b = het aantal verschillen, groter dan dat verschil.

Als wij de grens van een naar boven begrensde eenzijdige betrouwbaarheidsinterval willen bepalen,

- 1) Hieronder te verstaan de kans dat $\underline{U} \leq U_\epsilon$ is, als H_0 geldt.

zoeken wij 2 opeenvolgende verschillen: d_1 en d'_1 ($d_1 < d'_1$) zódanig dat $b_1 + \frac{1}{2}a_1 > U_\alpha$ en $b'_1 + \frac{1}{2}a'_1 \leq U_\alpha$ is, waarbij a_1 en b_1 resp. a'_1 en b'_1 de bij d_1 resp. d'_1 behorende aantallen a en b gedefinieerd in punt 3 zijn en α de gekozen onbetrouwbaarheidsdrempel is. Is nu $b_1 \leq U_\alpha$, dan wordt het betrouwbaarheidsinterval $\delta \leq d_1$ en is dus d_1 een bovengrens, die bij het interval behoort. Is echter voor d_1 : $b_1 > U_\alpha$, dan wordt het interval

$\delta < d_1$ en is dus d_1 een bovengrens die niet bij het interval behoort. Als wij echter de grens van een naar beneden begrensd éénzijdig betrouwbaarheidsinterval willen bepalen, zoeken wij 2 opeenvolgende verschillen d_2 en d'_2 ($d_2 > d'_2$) (met aantallen a en b , in 3 gedefinieerd, gelijk aan a_2 en b_2 resp. a'_2 en b'_2) zódanig dat: $b_2 + \frac{1}{2}a_2 < mn - U_\alpha$ en $b'_2 + \frac{1}{2}a'_2 \geq mn - U_\alpha$ is.

Als nu $b'_2 \geq mn - U_\alpha$ wordt het betrouwbaarheidsinterval $\delta \geq d_2$ en is dus d_2 een benedengrens, die bij het interval behoort, is daarentegen $b'_2 < mn - U_\alpha$, dan wordt het interval $\delta > d'_2$ en is dus d'_2 een benedengrens, die niet bij het interval behoort.

Indien men een tweezijdig begrensd betrouwbaarheidsinterval wil verkrijgen, bepaalt men als boven aangegeven een bovengrens en een benedengrens, maar dan ieder bij een onbetrouwbaarheidsdrempel $\frac{1}{2}\alpha$, als α de voorgeschreven onbetrouwbaarheid is.

Voorbeeld:

Waarnemingen van $\underline{x}$: 18,2; 16,2; 20,3; 17,4; 18,9; 21,4

" " $\underline{y}$: 16,7; 12,5; 15,2; 14,3; 15,9; 17,8

Gevraagd wordt een tweezijdig begrensd betrouwbaarheidsinterval voor $\delta = \bar{c}_x - \bar{c}_y$ bij een onbetrouwbaarheidsdrempel $\alpha = 0,05$.

Oplossing: Wij vinden in een tabel van de verdeling van de toetsingsgrootte U van WILCOXON (zie b.v. [1] van de literatuurlijst):

$$U_{0,025} = 5$$

$$mn - U_{0,025} = 36 - 5 = 31$$

Wij rangschikken nu de x - en de y -waarnemingen naar opklimmende grootte en geven de verschillen $x_i - y_j$ in onderstaande tabel:

Tabel der verschillen $x_i - y_j$ in het voorbeeld

		waarnemingen van x					
		16,2	17,4	18,2	18,9	20,3	21,4
waar- nemingen van y	12,5	+3,7	+4,9	+5,7	+6,4	+7,8	+8,9
	14,3	+1,9	+3,1	+3,9	+4,6	+6,0	+7,1
	15,2	+1,0	+2,2	+3,0	+3,7	+5,1	+6,2
	15,9	+0,3	+1,5	+2,3	+3,0	+4,4	+5,5
	16,7	-0,5	+0,7	+1,5	+2,2	+3,6	+4,7
	17,8	-1,6	-0,4	+0,4	+1,1	+2,5	+3,6

Wij rangschikken nu de verschillen naar opklimmende grootte en bepalen de aantallen a en b boven vermeld:

	a	b		a	b		a	b		a	b
-1,6	1	35	+1,9	1	25	+4,4	1	12	+6,4	1	3
-0,5	1	34	+2,2	2	23	+4,6	1	11	+7,1	1	2
-0,4	1	33	+2,3	1	22	+4,7	1	10	+7,8	1	1
+0,3	1	32	+2,5	1	21	+4,9	1	9	+8,9	1	0
+0,4	1	31	+3,0	2	19	+5,1	1	8			
+0,7	1	30	+3,1	1	18	+5,5	1	7			
+1,0	1	29	+3,6	2	16	+5,7	1	6			
+1,1	1	28	+3,7	2	14	+6,0	1	5			
+1,5	2	26	+3,9	1	13	+6,2	1	4			

Wij vinden nu:

$$\text{voor } d_1 = 6,0: b_1 + \frac{1}{2}a_1 = 5\frac{1}{2}$$

$$\text{voor } d_1' = 6,2: b_1' + \frac{1}{2}a_1' = 4\frac{1}{2}$$

$$b_1 = 5 = U_{0,025} \quad \text{dus } \delta \leq d_1 = 6,0$$

$$\text{voor } d_2 = 0,7: b_2 + \frac{1}{2}a_2 = 30\frac{1}{2}$$

$$\text{voor } d_2' = 0,4: b_2' + \frac{1}{2}a_2' = 31\frac{1}{2}$$

$$b_2' = 31 = mn - U_{0,025} \quad \text{dus } \delta \geq d_2 = 0,7$$

Het tweezijdig begrensde betrouwbaarheidsinterval wordt dus:

$$+0,7 \leq \delta \leq +6,0.$$

N.B. De feitelijke eenzijdige overschrijdingskans van $U = 5$ bij $m = n = 6$ is 0,02. Bij een onbetrouwbaarheidsdrempel 0,01 zouden wij dus hetzelfde interval gevonden hebben. Het verdient aanbeveling om deze kleinst mogelijke onbetrouwbaarheid

op te geven, aangezien deze aanzienlijk lager kan zijn dan de voorgeschreven waarde.

In vele gevallen kan de waarde van U_ϵ alleen bij benadering bepaald worden (Zie [1]). In die gevallen gelden de betrouwbaarheidsgrenzen uiteraard ook alleen bij benadering.

2. Betrouwbaarheidsinterval bij een aantal paren steekproeven.

Gegeven zijn k paren stochastische grootheden $\underline{x}_i, \underline{y}_i$ ($i = 1, \dots, k$). Wij beschikken over onderling onafhankelijke steekproeven van ieder van de grootheden:

$$\left. \begin{array}{l} x_{i,1}, \dots, x_{i,n_i} \text{ van } \underline{x}_i \\ y_{i,1}, \dots, y_{i,m_i} \text{ van } \underline{y}_i \end{array} \right\} (i = 1, \dots, k)$$

Er wordt ondersteld, dat $\underline{x}_i$ en $\underline{y}_i$ verdelingen hebben, die alleen in hun gemiddelden kunnen verschillen, terwijl deze verschillen dan voor alle i gelijk zijn. Dus:

$$E \underline{x}_i - E \underline{y}_i = \delta \quad (i = 1, \dots, k)$$

en $\underline{x}_i$ en $\underline{y}_i + \delta$ hebben dezelfde verdeling (Deze verdeling behoeft voor verschillende waarden van i niet hetzelfde te zijn).

Gevraagd wordt op grond van de steekproeven een betrouwbaarheidsinterval voor δ , met onbetrouwbaarheidsdrempel α te bepalen.

Deze vraag kan worden beantwoord met behulp van een combinatie van k onderling onafhankelijke toetsen van WILCOXON. De methode van het combineren van onderling onafhankelijke toetsen wordt behandeld in memorandum S 102 (M 17). Men kan daarin het volgende vinden: Zij c_i ($i = 1, \dots, k$) een gegeven constante en U_i ($i = 1, \dots, k$) de toetsingsgrootheid U van WILCOXON voor het paar steekproeven $x_{i,1}, \dots, x_{i,n_i}; y_{i,1}, \dots, y_{i,m_i}$, dan heeft

$$T = \sum_{i=1}^k c_i U_i$$

onder de hypothese H_0 dat voor iedere i $\underline{x}_i$ en $\underline{y}_i$ dezelfde verdeling hebben, bij benadering een normale verdeling met gemiddelde: $\sum_{i=1}^k c_i \mu_i$

$$\text{en variantie } \sum_{i=1}^k c_i^2 \sigma_i^2,$$

waarin $\mu_i = \frac{1}{2} m_i n_i$
en $\sigma_i^2 = \frac{1}{12} m_i n_i (m_i + n_i + 1)$

(eventueel met een correctie van gelijke waarnemingen, zie memorandum S 47 (M 7)), de verwachting resp. de variantie onder H_0 voorstellen van de toetsingsgrootte $\underline{U}_1$. Indien wij kiezen $c_1 = c_2 = \dots = c_k = 1$, kan men hieruit onmiddellijk de verdeling van $\underline{T} = \sum_i \underline{U}_i$ onder de hypothese H_0 afleiden, dus ook de grootste waarde T_ε , die $\underline{T}$ kan aannemen en waarvoor geldt:

$$P\{\underline{T} \leq T_\varepsilon \mid H_0\} \leq \varepsilon.$$

Voor de bepaling van het betrouwbaarheidsinterval gaan wij als volgt te werk:

1. Bepaal alle $\sum_{i=1}^k m_i n_i$ verschillen van een waarneming van $\underline{x}_1$ en een waarneming van de overeenkomstige stochastische grootte $\underline{y}_1$,
2. Rangschik de verschillen naar opklimmende grootte,
3. Bepaal voor ieder verschil $x_{1j} - y_{1l}$:

a = het aantal verschillen, gelijk aan dat verschil,

b = het aantal verschillen, groter dan het verschil.

Men gaat verder geheel te werk als beschreven is in par. 1 voor één paar steekproeven, met dien verstande dat men voor U_α moet lezen T_α .

Voorbeeld:

$k = 2$

waarnemingen van $\underline{x}_1$:	5,65	6,20	5,05	5,50	
" "	$\underline{y}_1$:	3,00	3,15	2,85	3,20
" "	$\underline{x}_2$:	7,05	8,40	7,35	7,60
" "	$\underline{y}_2$:	3,95	5,65	4,35	3,95

Gevraagd: wordt weer een tweezijdig begrensde betrouwbaarheidsinterval voor $\delta = \mathcal{E} \underline{x}_i - \mathcal{E} \underline{y}_i$ bij een onbetrouwbaarheid drempel $\alpha = 0,05$.

Oplossing: Wij vinden met behulp van een tabel van de verdeling van de toetsingsgrootte $\underline{U}$ voor de kritieke waarden:

$$T_{0,025} = 4 \quad \text{en} \quad \sum_{i=1}^2 m_i n_i - T_{0,025} = 32 - 4 = 28.$$

(deze kritieke waarden zijn niet bepaald met behulp van boven-
beschreven benadering, doch afgeleid uit de exacte verdeling
van $\underline{U}_1 + \underline{U}_2$, die in dit geval gemakkelijk te bepalen is).

Wij rangschikken de x_i en y_i waarnemingen weer naar op-
klimmende grootte en bepalen de verschillen:

Tabel der verschillen $x_{1j} - y_{1j}$					
		waarnemingen van y_1			
		2,85	3,--	3,15	3,20
waar- nemingen van x_1	5,05	2,20	2,05	1,90	1,85
	5,50	2,65	2,50	2,35	2,30
	5,65	2,80	2,65	2,50	2,45
	6,20	3,35	3,20	3,05	3,--

		waarnemingen van y_2		
		3,95 (2)	4,35	5,65
waar- nemingen van x_2	7,05	3,10 (2)	2,70	1,40
	7,55	3,60 (2)	3,20	1,90
	7,60	3,65 (2)	3,25	1,95
	8,40	4,45 (2)	4,05	2,75

(Waarnemingen en verschillen waarachter het cijfer (2) ver-
meld is komen 2 maal voor).

Wij rangschikken nu de verschillen naar opklimmende
grootte en bepalen de aantallen a en b gedefinieerd in par.1:

	a	b		a	b		a	b		a	b
1,40	1	31	2,30	1	24	2,75	1	16	3,25	1	8
1,85	1	30	2,35	1	23	2,80	1	15	3,35	1	7
1,90	2	28	2,45	1	22	3,--	1	14	3,60	2	5
1,95	1	27	2,50	2	20	3,05	1	13	3,65	2	3
2,05	1	26	2,65	2	18	3,10	2	11	4,05	1	2
2,20	1	25	2,70	1	17	3,20	2	9	4,45	2	0

Wij vinden nu:
voor $d_1 = 3,60 \quad b_1 + \frac{1}{2}a_1 = 6$
" $d'_1 = 3,65 \quad b'_1 + \frac{1}{2}a'_1 = 4$
 $b_1 = 5 > T_{0,025} = 4 \quad \text{dus} \quad \delta < d'_1 = 3,65$

$$\begin{aligned} \text{voor } d_2 &= 1,95 & b_2 + \frac{1}{2}a_2 &= 27\frac{1}{2} \\ " \quad d_2' &= 1,90 & b_2' + \frac{1}{2}a_2' &= 29 \\ b_2' &= 28 = \sum_{i=1}^2 m_i \eta_i - T_{9,025} = 28 & \text{ dus } \delta &\geq d_2 = 1,95. \end{aligned}$$

Het tweezijdig begrensde betrouwbaarheidsinterval wordt dus:

$$1,95 \leq \delta < 3,65$$

Literatuur:

[1] H.R. VAN DER VAART Gebruiksaanwijzing voor de toets van WILCOXON, Rapport S32 (M 4) van het Mathematisch Centrum 2e druk 1952.
