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 1955-23(9)

Verdeling van de mediaan van steekproeven van eèn oneven aantal waarnemingen uit een gegeven verdeling.

door

Ph. van Elteren

1955

Rapport S 1955-23(9)

Verdeling van de mediaan van steekproeven van een oneven aantal waarnemingen uit een gegeven verdeling.

door Ph. van Elteren.

1. Inleiding; afleiding van de verdeling.

Wij beschouwen steekproeven van $n=2k+1$ waarnemingen uit een gegeven verdeling met verdelingsfunctie $F(x)$. De waarnemingen worden gerangschikt naar opklimmende grootte, x_r stelt de r^e waarneming in deze rij voor; de mediaan m van de steekproef is dan de middelste waarneming, dus x_{k+1}.

Voorbeelden:

$n = 5$: waarnemingen: 1,3; 3,0; 5,4; 2,7; 2,1,
gerangschikt: 1,3; 2,1; 2,7; 3,0; 5,4,
$k = 2$ dus $m = x_3 = 2{,}7$.

$n = 7$: waarnemingen: 2; 5; 5; 4; 3; 3; 3,
gerangschikt: 2; 3; 3; 3; 4; 5; 5,
$k = 3$ dus $m = x_4 = 3$.

Als de mediaan van de steekproef niet groter is dan m zullen hoogstens k waarnemingen groter zijn dan m. De kans dat een willekeurige waarneming groter is dan m is

$$P[x > m] = 1 - P[x \leq m] = 1 - F(m).$$

De kans, dat precies i waarnemingen groter zijn dan wordt dus:

$$\binom{n}{i}\{1 - F(m)\}^{i}\{F(m)\}^{n-i},$$

en de kans dat dus hoogstens k waarnemingen groter zijn dan m, (dus dat de mediaan van de steekproef niet groter is dan m) wordt:

$$(1) \quad P[m \leq m] = \sum_{i=0}^{k}\binom{n}{i}\{1 - F(m)\}^{i}\{F(m)\}^{n-i}.$$

(Zie b.v. literatuurlijst: D. VAN DANTZIG (1947-1949)

Het rechterlid van (1) kan worden bepaald met een tabel van de Binomiale verdeling of een tabel van de onvolledige Bêta-functie. Men kan ook gemakkelijk kansen $P[m \geq m]$ of $P[m=m]$ (de laatste alleen bij discrete verdelingen) afleiden. Wij zullen het een en ander toelichten aan een aantal voorbeelden. (zie par.3)

Opmerking Uit formule 1 kan nog worden afgeleid, dat de verdeling van de mediaan van steekproeven uit een symmetrische verdeling, zelf symmetrisch is met hetzelfde symmetriepunt Immers: als c het symmetriepunt van de verdeling is, geldt:

$$F(m) = P[\underline{x} \leqq m] = P[\underline{x} \geqq 2c-m]$$

$$P[\underline{m} \leqq m] = \sum_{i=0}^{k} \binom{n}{i} \{1-F(m)\}^{i} \{F(m)\}^{n-i}$$

$$P[\underline{m} \geqq 2c-m] = \sum_{i=0}^{k} \binom{n}{i} \{1-P[\underline{x} \geqq 2c-m]\}^{i} \{P[\underline{x} \geqq 2c-m]\}^{n-i}$$

$$= \sum_{i=0}^{k} \binom{n}{i} \{1-F(m)\}^{i} \{F(m)\}^{n-i}.$$

2 Gegroepeerde steekproeven.

Men verkrijgt gegroepeerde steekproeven uit een continue verdeling, door het interval (a_0, a_N) der waarden, die door de variabele worden aangenomen te verdelen in intervallen (a_0, a_1), (a_1, a_2), . . . , (a_{N-1}, a_N) en dan aan iedere waarneming uit het interval $(a_\ell, a_{\ell+1})$ $(\ell = 0, 1, \ldots, N-1)$ eenzelfde waarde toe te kennen. De toegekende waarde, voor het interval $(a_\ell, a_{\ell+1})$ geven wij aan met w_ℓ; gewoonlijk wordt hiervoor het midden $\frac{1}{2}(a_\ell + a_{\ell+1})$ van het interval gekozen; als de verdeling zich uitstrekt van $-\infty$ tot $+\infty$, wordt $a_0 = -\infty$, $a_N = +\infty$ en moeten w_0 en w_N afzonderlijk gedefinieerd worden.

Als nu de mediaan van een steekproef een waarde is uit het interval $(a_\ell, a_{\ell+1})$, wordt de mediaan van de gegroepeerde steekproef w_ℓ. De kans, dat de mediaan van de gegroepeerde steekproef gelijk is aan w_ℓ, is dus gelijk aan de kans dat de mediaan van de niet gegroepeerde steekproef in het interval $(a_\ell, a_{\ell+1})$ ligt.

Men verkrijgt dus de verdeling van de mediaan van gegroepeerde steekproeven met gegeven omvang en gegeven groepering, uit een gegeven continue verdeling, door de verdeling van de mediaan van ongegroepeerde steekproeven te berekenen en daarin de gegeven groepering aan te brengen.

Voor toepassing van dit principe zie men par. 3, voorbeeld 4.

3. Voorbeelden

Voorbeeld 1:

$\underline{x}$ is normaal verdeeld met gemiddelde 0 en spreiding 1, $n = 5$ dus $k = 2$. Willen wij nu b.v de kans uitrekenen, dat $\underline{m} \leqq 0{,}5$ is, dan gaan wij als volgt te werk.

$$P[\underline{m} \leqq 0{,}5] = \sum_{i=0}^{2} \binom{5}{i} \{1 - F(0{,}5)\}^{i} \{F(0{,}5)\}^{5-i}$$

Volgens de tabel van de normale verdeling is:

$$F(0{,}5) = 0{,}691$$

dus is:

$$P[\underline{m} \leqq 0{,}5] = \sum_{i=0}^{2} \binom{5}{i} 0{,}309^{i}\, 0{,}691^{5-i}$$

Een tabel van de Binomiale verdeling (NATIONAL BUREAU OF STANDARDS (1950)) geeft:

$$\sum_{i=3}^{5} \binom{5}{i} 0{,}309^{i}\, 0{,}691^{5-i} = 0{,}175 \qquad \text{(lineaire interpolatie).}$$

Wij vinden dus: $P[\underline{m} \leqq 0{,}5] = 1 - 0{,}175 = 0{,}825.$

Voor de overgang op de tabel van de onvolledig Bêta-functie (K PEARSON (1934)) maken wij gebruik van de relatie:

$$\sum_{s=r}^{n} \binom{n}{s} p^{s} (1-p)^{n-s} = I_p\,(r, n-r+1)$$

Dus:

$$\sum_{i=3}^{5} \binom{5}{i} 0{,}309^{i} \cdot 0{,}691^{5-i} = I_{0{,}309}\,(3,3) = 0{,}175$$

en $P[\underline{m} \leqq 0{,}5] = 1 - 0{,}175 = 0{,}825$

Voorbeeld 2:

$\underline{x}$ heeft de binomiale verdeling voor het aantal successen onder 10 experimenten met kans $\frac{1}{2}$ op succes.

$n = 7$ dus $k = 3$.

Willen wij nu b.v. de kans berekenen, dat $\underline{m} \geqq 6$ is, dan geschiedt dit als volgt:

$$\begin{aligned} P[\underline{m} \geqq 6] &= 1 - P[\underline{m} \leqq 5] \\ &= 1 - \sum_{i=0}^{3} \binom{7}{i} \{1 - F(5)\}^{i} \{F(5)\}^{5-} \\ &= 1 - \sum_{i=0}^{3} \binom{7}{i} \cdot \left(\tfrac{1}{2}\right)^{10} = \sum_{i=4}^{7} \binom{7}{i} \left(\tfrac{1}{2}\right)^{1} = \frac{1}{2} \end{aligned}$$

Om $P[\underline{m} = 6]$ te vinden bepalen wij b.v.

$$P[\underline{m} \geqq 5] - P[\underline{m} \geqq 6]$$

$$P[\underline{m} \geqq 5] = 1 - P[\underline{m} \leqq 4] = \sum_{i=4}^{7} \binom{7}{i} \{1 - F(4)\}^{i} \{F(i)\}^{7-i}$$

$$F(4) = \sum_{j=0}^{4} \binom{10}{j} \left(\frac{1}{2}\right)^{10} = 1 - \sum_{j=5}^{10} \binom{10}{j} \left(\frac{1}{2}\right)^{10} =$$

$$= 1 - 0{,}623 = 0{,}377$$

Dus:

$$P[\underline{m} \geqq 5] = \sum_{i=4}^{7} \binom{7}{i} 0{,}623^{i}\, 0{,}377^{7-i}$$

Aangezien de tabel van het National Bureau of Standards de waarden $\sum_{s=r}^{n} \binom{n}{r} p^{s}(1-p)^{n-s}$ voor $p \leqq 0{,}5$ bevat, herleiden wij het rechterlid tot:

$$\sum_{j=0}^{3} \binom{7}{j} 0{,}377^{j}\, 0{,}623^{7-j} = 1 - \sum_{j=4}^{7} \binom{7}{j} 0{,}377^{j}\, 0{,}623^{7-j} =$$

$$= 1 - 0{,}247 = 0{,}753.$$

(Hierbij is gebruik gemaakt van lineaire interpolatie). Dus:

$$P[\underline{m} = 6] = P[\underline{m} \geqq 5] - P[\underline{m} \geqq 6] = 0{,}253.$$

Met behulp van formule (1) kan men eveneens kritieke waarden van de verdeling van $\underline{m}$ bepalen. Dit wordt toegelicht in het volgende voorbeeld.

Voorbeeld 3:

$\underline{x}$ is normaal verdeeld (gemiddelde 0, spreiding 1)
Wij bepalen b.v. de waarde m van $\underline{m}$ met $P[\underline{m} \geqq m] = 0{,}025$:

$$0{,}025 = P[\underline{m} \geqq m] = \sum_{i=5}^{9} \binom{9}{i} \{1 - F(m)\}^{i} \{F(m)\}^{9-i}.$$

Door terug te zoeken in de tabel van de Binomiale verdeling vindt men $1 - F(m) = 0{,}212$ (lineaire interpolatie) dus $F(m) = 0{,}788$.

Deze waarde zoekt men terug in een tabel van de normale verdeling; men vindt: $m = 0{,}80$

In het volgende voorbeeld bepalen wij de verdeling van de mediaan van gegroepeerde steekproeven uit een continue verdeling:

Voorbeeld 4:

Steekproeven van 7 waarnemingen uit een normale verdeling met gemiddelde 1 en spreiding 2 worden gegroepeerd, doordat de steekproefwaarden worden afgerond op 1 decimaal. Wij berekenen de kans, dat de mediaan van een gegroepeerde steekproef gelijk is aan -0,7. Een waarneming -0,7 uit een gegroepeerde steekproef, komt overeen met een waarneming tussen -0,65 en 0,75 uit een ongegroepeerde steekproef. De kans, dat de mediaan van een gegroepeerde steekproef -0,7 is, is dus gelijk aan de kans, dat de mediaan van een ongegroepeerde steekproef ligt tussen - 0,65 en -0,75. Wij bepalen dus:

$$P[-0{,}75 \leqq \underline{m} \leqq -0{,}65] = P[\underline{m} \leqq -0{,}65] - P[\underline{m} \leqq -0{,}75]$$

$$F(-0{,}65) = \phi\left(\frac{-0{,}65-1}{2}\right) = \phi(-0{,}825) = 0{,}205.$$

Hierin is $\phi(x)$ de verdelingsfunctie van de normale verdeling met gemiddelde 0 en spreiding 1.

$$P[\underline{m} \leqq -0{,}65] = \sum_{i=4}^{7} \binom{7}{i} . (0{,}205)^i (0{,}795)^{7-i} = 0{,}036$$

$$F(-0{,}75) = \phi\left(\frac{-0{,}75-1}{2}\right) = \phi(-0{,}875) = 0{,}191$$

$$P[\underline{m} \leqq -0{,}75] = \sum_{i=4}^{7} \binom{7}{i} \; 0{,}191^i \; 0{,}809^{7-i} = 0{,}028$$

Dus

$$P[-0{,}75 \leqq \underline{m} \leqq -0{,}65] = 0{,}008$$

Uit het bovenstaande volgt, dat de kans, dat de mediaan van de gegroepeerde steekproef 0,7 is, gelijk is aan 0,008.

Literatuur

D. VAN DANTZIG, (1947-1949) Kadercursus Mathematische Statistiek, Hoofdstuk 6 par. 7.(Math.Stat. p. 276, Whr p. 365).

NATIONAL BUREAU OF STANDARDS, (1950) Tables of the binomial probability distribution, Appl.Math.Series 6, Washington D.C. U.S.A.

K PEARSON, (1934) Tables of the incomplete Bêta-function, Biometrika Office, London.
