S 190 (M 75)

De symmetrietoets van Wilcoxon

Stichting Mathematisch Centrum
Statistische Afdeling
1956

MATHEMATISCH CENTRUM,
2e Boerhaavestraat 49,
Amsterdam-O.
Rapport S 190 (M 75).

De symmetrietoets van WILCOXON [1)]

Met behulp van deze toets kan men de hypothese H_o toetsen dat een aantal onderling onafhankelijke grootheden $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_m$ alle symmetrisch ten opzichte van 0 verdeeld zijn. Over de vorm van de verdelingen van $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_m$ behoeft niets ondersteld te worden en deze verdelingen behoeven niet identiek te zijn.

De toets berust op één waarneming van ieder der grootheden, dus op m waarnemingen $z_1, z_2, \ldots, z_m$ en de toetsingsgrootheid wordt als volgt berekend:

De waarnemingen die =0 zijn worden weggelaten. De absolute waarden van de overblijvende n waarnemingen worden naar opklimmende grootte gerangschikt en van rangnummers voorzien; hierbij krijgen gelijken als rangnummer het gemiddelde van de rangnummers die zij zouden krijgen als zij ongelijk waren. De toetsingsgrootheid T is nu gelijk aan de som van de rangnummers afkomstig van de positieve waarnemingen verminderd met de som van de rangnummers afkomstig van de negatieve waarnemingen. B.v.

$$z_i : \quad 3 \quad 7 \quad 0 \quad -2 \quad -1 \quad 5 \quad 10 \quad 1 .$$

Hier is dus m=8 en n=7. De overblijvende waarnemingen zijn

$$3 \quad 7 \quad -2 \quad -1 \quad 5 \quad 10 \quad 1 .$$

Rangschikking van deze waarnemingen naar opklimmende grootte van hun absolute waarde geeft:

$$1 \quad -1 \quad -2 \quad 3 \quad 5 \quad 7 \quad 10 .$$

Als men nu overgaat op rangnummers krijgt men

$$1\tfrac{1}{2} \quad 1\tfrac{1}{2} \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 ,$$

dus

---

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

$$T = 1\tfrac{1}{2} + 4 + 5 + 6 + 7 - (1\tfrac{1}{2}+3) = 19 .$$

Als de hypothese H_o juist is dan is de verwachting van de grootheid $\underline{T}$:

(1) $$\mathcal{E}\{\underline{T}|H_o\} = 0 .$$

Indien de absolute waarden van de waarnemingen gesplitst kunnen worden in k groepen gelijken ter grootte $t_1, t_2, \ldots, t_k$ (zodat de absolute waarde van twee waarnemingen, behorende tot twee verschillende groepen niet gelijk zijn), dan is dus

$$n = \sum_{i=1}^{k} t_k{}^3 .$$

Noem

$$D = \sum_{i=1}^{k} t_k{}^3 .$$

In het bovengegeven voorbeeld is k=6; er komt één groep van 2 gelijken voor en 5 groepen van ieder één. De variantie van $\underline{T}$ onder H_o is nu

(2) $$\sigma^2\{\underline{T}|H_o\} = \frac{n^3 - D + 3n(n+1)^2}{12} = \frac{1}{6}\, n(n+1)(2n+1) - \frac{1}{12}\,(D-n) .$$

Als er geen gelijken voorkomen dan wordt D=n en gaat (2) over in

(3) $$\sigma^2\{\underline{T}|H_o\} = \frac{1}{6}\, n(n+1)(2n+1) .$$

Deze waarde benevens die van $\sigma\{\underline{T}|H_o\}$ is voor $20 \leq n \leq 100$ getabelleerd in [1].

Voor kleine waarden van n kan men, als er geen gelijken voorkomen, de exacte verdeling van $\underline{T}$ onder de hypothese H_o eenvoudig berekenen met behulp van een recursieformule. Tabellen van kritieke waarden voor dit geval vindt men, voor n = 5,6,...,20 in [1]. Als n groot is kan men de verdeling van $\underline{T}$ onder H_o benaderen met een normale verdeling met gemiddelde en variantie volgens (1) en (2). Tabellen van benaderende kritieke waarden vindt men voor n = 21,22,...,100 in [1].

De links (resp. rechts) eenzijdige kritieke zone bestaat uit kleine (resp. grote) waarden van T en de tweezijdige kritieke zone bestaat uit kleine en grote waarden van T, dus uit grote waarden van $|T|$.

Indien men de kritieke zones en overschrijdingskansen bepaalt met behulp van de normale benadering dan passe men een continuïteitscorrectie ter grootte 1 toe.

Opmerkingen

1. Als men de hypothese H_0 wil toetsen dat $\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_m$ alle symmetrisch ten opzichte van a verdeeld zijn, waarbij a een gegeven getal is, dan past men de toets toe op $z_1-a, z_2-a, \ldots, z_m-a$.
2. De toetsingsgrootheid T hangt nauw samen met de grootheid W van WILCOXON's tweesteekproeventoets (zie b.v. [2]). Stel n_1 is het aantal positieve en n_2 het aantal negatieve waarnemingen. Als nu W de toetsingsgrootheid van WILCOXON's tweesteekproeventoets is, toegepast op de n_1 positieve waarnemingen als eerste en de absolute waarden van de n_2 negatieve waarnemingen als tweede steekproef, dan geldt

$$T = W - n_1 n_2 + \tfrac{1}{2}(n+1)(n_1-n_2) \ .$$

3. Als de absolute waarde van alle waarnemingen gelijk zijn, dus als k=1 en $t_1=n$, dan krijgen deze waarnemingen alle hetzelfde rangnummer $\frac{1}{2}(n+1)$. Als n_1 weer het aantal positieve waarnemingen is en n_2 het aantal negatieve waarnemingen, dan wordt in dit geval de toetsingsgrootheid

$$T = \tfrac{1}{2}(n+1)(n_1-n_2) \ ,$$

dus T is een lineaire functie van de toetsingsgrootheid n_1 van de tekentoets, zodat de toets identiek is met de tekentoets.
4. De toets wordt vaak gebruikt als men een aantal grootheden tweemaal heeft waargenomen, voor en na een bepaalde gebeurtenis, als men wil nagaan of deze gebeurtenis invloed op de grootheden heeft uitgeoefend. Noemen wij de grootheden vóór het optreden der gebeurtenis z_i' en erna z_i'', dan zijn de grootheden $z_i=z_i'-z_i''$ $(i=1,2,\ldots,m)$, als de gebeurtenis geen invloed heeft, alle symmetrisch ten opzichte van 0 verdeeld.
5. WILCOXON gebruikt een iets andere grootheid, nl. de som van de rangnummers afkomstig van positieve waarnemingen. Als wij deze grootheid T_W noemen, dan geldt:

$$T = 2T_W - \tfrac{1}{2}n(n+1) \ .$$

In het voorbeeld is $T_W = 23\frac{1}{2}$ en dus

$$T = 2.23\tfrac{1}{2} - \tfrac{1}{2}.7(7+1) = 19 \ .$$

Tabellen van kritieke waarden van T_W vindt men voor n=6,7,...,25 in [3].

Literatuur

[1] BENARD, A. en Constance van EEDEN,
Handleiding voor de symmetrietoets van Wilcoxon,
Rapport S 208 (M 76) van de Statistische Afdeling
van het Mathematisch Centrum, Amsterdam 1956.

[2] WABEKE, Ir Doraline en Constance van EEDEN,
Handleiding voor de toets van Wilcoxon,
Rapport S 176 (M 65) van de Statistische Afdeling
van het Mathematisch Centrum, Amsterdam 1955.

[3] WILCOXON, F.,
Some rapid approximated statistical procedures,
New York, 1949, p. 5-6.