

## 6.4 Der Kruskal-Wallis Test

Der Test von Kruskal und Wallis, auch H-Test genannt, ist ein Test, mit dem man die Verteilungen von Teilstichproben auf Unterschiede untersuchen kann. Bei diesem Test geht man davon aus, dass  $g$  Teilstichproben mit nicht notwendigerweise gleichen Teilstichprobenumfängen vorliegen. Die  $j$ -te Teilstichprobe soll aus Realisierung  $x_{1j}, x_{2j}, \dots, x_{n_jj}$  von unabhängig und identisch verteilten Zufallsvariablen  $X_{1j}, X_{2j}, \dots, X_{n_jj}$  mit der stetigen Verteilungsfunktion  $F_j$  bestehen. Die Teilstichproben sollen nicht verbunden sein, d.h. auch die Zufallsvariablen  $X_{ij}$  mit  $i = 1, 2, \dots, n_j$  und  $j = 1, 2, \dots, g$  sind (paarweise) unabhängig, die zu verschiedenen Teilstichproben gehören. Es genügt für diesen Test, wenn das Datenniveau mindestens ordinal ist.

In unserem Beispiel handelt es sich um 3 Teilstichproben mit gleichen Teilstichprobenumfängen. In dem Beispiel gehen wir davon aus, dass bei drei Gruppen von Versuchspersonen die Testleistungen gemessen wurden. Allgemein werden folgende Hypothesen getestet:

$H_0: F_1(z) = F_2(z) = \dots = F_g(z)$  für alle  $z$   
gegen

$H_1: \text{nicht alle } g \text{ Verteilungen sind gleich für mindestens ein } z$

Die Daten:

v1	v2	V3
8,7	17,1	17,5
1,1	15,3	9
4,9	14,5	11,3
3,8	12	20,2
7,5	5,8	16,3
16,7	9,3	16,7

Um den Output zu erhalten, müssen zunächst die Variablen v1, v2 und v3 ganz rechts im Menü unter „Welche Spalten sollen verglichen werden:“ ausgewählt werden. Danach können Sie →**Vergleich mehrerer unverbundener Teilstichproben** →**Kruskal Wallis** wählen:

### Kruskal-Wallis

H0: Die Verteilungsfunktionen aller Teilstichproben<sup>(2)</sup> sind identisch gegen

H1: Es existieren (mindestens) zwei Teilstichproben<sup>(2)</sup> mit unterschiedlichen Verteilungsfunktionen

<sup>(2)</sup> Bemerkung: Gemeint sind hier die (theoretischen) Teilstichproben der Zufallsvariablen.

Die Daten:

	Teilstichprobe 1 (Rangzahl in Klammern)	Teilstichprobe 2 (Rangzahl in Klammern)	Teilstichprobe 3 (Rangzahl in Klammern)
Beobachtung 1	8.7 (6)	17.1 (16)	17.5 (17)
Beobachtung 2	1.1 (1)	15.3 (12)	9 (7)
Beobachtung 3	4.9 (3)	14.5 (11)	11.3 (9)
Beobachtung 4	3.8 (2)	12 (10)	20.2 (18)
Beobachtung 5	7.5 (5)	5.8 (4)	16.3 (13)
Beobachtung 6	16.7 (14.5)	9.3 (8)	16.7 (14.5)
Teilstichprobenumfänge $n_j$	6	6	6
Rangsummen	31.5	61	78.5

Umfang Gesamtstichprobe n	18
H-Statistik	6.599415204678
H-Statistik bei Bindungen	6.606232782369
approximativer p-Wert (Freiheitsgrade Chi-Quadrat-Verteilung: 2)	0.0368

Im Beispiel sind die Teilstichprobenumfänge  $n_j$  alle gleich, was aber nicht notwendig ist. Zunächst werden hier wie beim Rangsummentest von Wilcoxon die Ränge für alle Teilstichproben zusammen vergeben. Der kleinste Wert aller Teilstichproben ist 1,1 und kommt einmal vor, womit diese Beobachtung den Rang 1 erhält. Es kommt nur ein Wert doppelt vor, das ist die 16,7.

Wir stellen die Stichprobe noch mal allgemein dar:

Die erste Teilstichprobe ist  $x_{11}, x_{21}, \dots, x_{n_1,1}$ , die zweite Teilstichprobe ist  $x_{12}, x_{22}, \dots, x_{n_2,2}$  und die letzte ist die  $g$ -te Teilstichprobe  $x_{1g}, x_{2g}, \dots, x_{n_g,g}$ . Diese Teilstichproben werden also wie eine einzige Stichprobe betrachtet und dann die Ränge vergeben. Es sei  $r_{ij} = \text{Rang}(x_{ij})$  mit  $i = 1, 2, \dots, n_j$  und  $j = 1, 2, \dots, g$ .

Im Output unten sind die Rangsummen zu sehen:

$$r_j = \sum_{i=1}^{n_j} \text{Rang}(x_{ij})$$

Die Liste der  $t_j$  (mit  $j = 1, 2, \dots, k$ ) enthält nur Einsen, bis auf eine 2 für den doppelt vorkommenden Wert. Im Beispiel ist  $k = 3 \cdot 6 - 1 = 17$ , da sich die Länge dieser Liste für jeden Wert, der sich wiederholt, um 1 reduziert (siehe Kapitel 1.5).

Nun kommen wir zur Berechnung der Prüfgröße  $h$ . Für deren Berechnung benötigen wir die Erwartungswerte der einzelnen  $R_j$  (deren Realisierungen die Rangsummen  $r_j$  sind):

$$E(R_j) = \frac{n_j(n+1)}{2}$$

Hiermit ergibt sich die Prüfgröße h:

$$h = \frac{12}{n(n+1)} \sum_{j=1}^g \frac{1}{n_j} (r_j - E(R_j))^2 = \frac{12}{n(n+1)} \sum_{j=1}^g \frac{r_j^2}{n_j} - 3(n+1)$$

Bei Bindungen sollte h korrigiert werden:

$$h^* = \frac{h}{1 - \frac{1}{n(n-1)(n+1)} \sum_{j=1}^k t_j(t_j-1)(t_j+1)}$$

h bzw.  $h^*$  ist eine Realisierung einer (wie immer unter  $H_0$ ) asymptotisch, Chi-Quadrat- verteilten zufälligen Größe mit  $g - 1$  Freiheitsgraden, womit wir den p-Wert berechnen können. In der Praxis sollten aber, da wir mit dem Test über die asymptotische Verteilung den p-Wert berechnen, die Teilstichprobenumfänge größer oder gleich 5 und mindestens 3 Teilstichproben vorhanden sein.

Der p-Wert berechnet sich dann über (treten keine Bindungen auf ist  $h = h^*$ ):

$$\text{p-Wert} = 1 - F_{\chi_{g-1}^2}(h^*)$$

Wie anhand des p-Wertes zu sehen ist, kann die Nullhypothese der Gleichheit der Verteilungen verworfen werden. Wir haben somit einen signifikanten Unterschied zwischen den Teilstichproben nachgewiesen.

Im Beispiel ist  $h = 6,5994\dots$ ,  $h^* = 6,6062\dots$  und p-Wert  $\approx 0,0368$ . Somit kann man auf einem Signifikanzniveau von 5% einen Unterschied zwischen mindestens zwei Teilstichproben bgl. deren Verteilung nachweisen. In Büchern wie z.B. in [3], [8] und [9] findet man Tabellen der exakten Verteilung für den Fall, dass keine Bindungen vorliegen.

### Umsetzung mit SAS:

```
data dat1;
input x y;
datalines;
1 8.7
1 1.1
1 4.9
1 3.8
1 7.5
1 16.7
2 17.1
2 15.3
2 14.5
2 12.0
2 5.8
2 9.3
3 17.5
3 9.0
3 11.3
3 20.2
3 16.3
3 16.7
run;

proc npar1way data = dat1 wilcoxon;
class x;
var y;
run;
```

**SAS-Output zur Prozedur NPAR1WAY:**

Das SAS System

Die Prozedur NPAR1WAY

**Wilcoxon-Scorewerte (Rangsummen) für Variable y  
Klassifiziert nach Variable x**

x	N	Summe der Scorewerte	Erwartet unter H0	Std.abw. unter H0	Mittelwert- Score
1	6	31.50	57.0	10.671568	5.250000
2	6	61.00	57.0	10.671568	10.166667
3	6	78.50	57.0	10.671568	13.083333

**Für gleiche Werte wurden durchschnittliche Scorewerte verwendet.**

**Kruskal-Wallis-Test**

**Chi-Quadrat** 6.6062

**DF** 2

**Pr > Chi-Quadrat** 0.0368