1.6 Der Vorzeichentest

In diesem Kapitel soll der Vorzeichentest bzw. Zeichentest vorgestellt werden, mit dem man Hypothesen bezüglich des Medians der unabhängig und identisch stetig verteilten Zufallsvariablen $X_1, X_2, \ldots, X_n$ testen kann, deren Realisierungen als eine Stichprobe $x_1, x_2, \ldots, x_n$ vorliegt. Das Datenniveau sollte hier sogar metrisch sein. Es wird, wie bei allen hier vorgestellten nichtparametrischen Verfahren, keine Normalverteilung der X_i vorausgesetzt. Nun können wir die folgenden Hypothesen testen:

H_0: Median = Median_0
gegen
H_1: Median $\neq$ Median_0

Im folgenden Beispiel wollen wir nun

H_0: Median = 2
gegen
H_1: Median $\neq$ 2

testen.

Wir verwenden folgende Daten:

v1
1
2
5
5
4
3
2
5

Wenn Sie dann →**Univariate Statistik** wählen, können Sie neben dem Button Vorzeichentest den Wert für Median_0 eintragen, also hier 2. Danach können Sie auf →**Vorzeichentest** klicken und erhalten den folgenden Output:

Vorzeichentest

Anzahl Werte größer als 2	5
Anzahl Werte kleiner als 2	1

	H0: Median = 2 gegen H1: Median <> 2	H0: Median >= 2 gegen H1: Median < 2	H0: Median <= 2 gegen H1: Median > 2
p-Werte	0.21875	0.984375	0.109375

Kommen wir nun zur Beschreibung des Tests bzw. des Outputs. Wir bestimmten die transformierte Stichprobe $y_i = x_i - \text{Median}_0 = x_i - 2$:

-1, 0, 3, 3, 2, 1, 0, 3

Die Zufallsvariablen Y_i (deren Realisierungen die y_i sind) hätten unter H_0 den Median 0. Aus der transformierten Stichprobe werden nun alle Nullen entfernt:

-1, 3, 3, 2, 1, 3

Unser (neuer) Stichprobenumfang ist nun n = 6. Eine Beobachtung ist negativ und k = 5 sind positiv. Unter H_0 ist die Anzahl k der positiven Werte y_i eine Realisierung einer mit den Parametern n und $p = ½ (= P(Y_i > 0))$ binomialverteilten Zufallsvariablen K, da bei einer stetigen Zufallsvariablen die Wahrscheinlichkeit 50% beträgt, dass diese einen Wert größer dem Median annimmt. Bei einer stetigen Zufallsvariablen Y ist $P(Y = 0) = 0$.

H_0 kann verworfen werden, wenn

$$P_{B(n,1/2)}(X \le k) \le \alpha / 2 \text{ oder}$$
$$P_{B(n,1/2)}(X \ge k) = 1 - P_{B(n,1/2)}(X \le k-1) \le \alpha / 2 \ .$$

Oben ist zu beachten, dass bei einer diskreten Verteilung, wie hier der Binomialverteilung, P(X = k) größer als Null ist für wenn k = 0, 1, …, n ist. Aus diesem Grund ist

$$P(X < k) = P(X \le k - 1)$$

ist, was oben verwendet wurde.

Somit kann die Nullhypothese verworfen werden, wenn

$$2 \cdot P_{B(n,1/2)}(X \le k) \le \alpha \text{ oder } 2 \cdot (1 - P_{B(n,1/2)}(X \le k-1)) \le \alpha \ .$$

Somit ergibt sich der zweiseitige p-Wert:

$$\text{p-Wert} = \min \{2 \cdot P_{B(n,1/2)}(X \le k),\ 2 \cdot (1 - P_{B(n,1/2)}(X \le k-1)),\ 1\}$$

Im Beispiel gilt:

$$\begin{aligned}\text{p-Wert} &= \min \{2 \cdot P_{B(6,1/2)}(X \le 5),\ 2 \cdot (1 - P_{B(5,1/2)}(X \le 4)),\ 1\} \\ &= 0.21875\end{aligned}$$

Diesen p-Wert finden Sie auch im Output (in der ersten Spalte für den zweiseitigen Test). Die Nullhypothese kann also beispielsweise auf einem Signifikanzniveau von 5% nicht verworfen werden (0,21875 > 0,05). Damit weicht der Median nicht signifikant vom Wert 2 ab. Daneben befinden sich im Output die p-Werte für die beiden einseitigen Tests.

Für „H_0: Median ≥ $Median_0$” gegen „H_1: Median < $Median_0$“ wird H_0

verworfen, wenn k „zu klein ist“, also wenn

$$\text{p-Wert} = P_{B(n,1/2)}(X \le k) \le \alpha.$$

Im Beispiel gilt:

$$\text{p-Wert} = P_{B(6,1/2)}(X \le 5) = 0{,}984375 > 0{,}05$$

Für „H_0: Median ≤ Median_0” gegen „H_1: Median > Median_0“ wird H_0 verworfen, wenn k „zu groß ist“, also wenn

$$\text{p-Wert} = 1 - P_{B(n,1/2)}(X \le k-1) \le \alpha.$$

Im Beispiel gilt:

$$\text{p-Wert} = 1 - P_{B(6,1/2)}(X \le 4) = 0{,}109375 > 0{,}05$$

Somit könnten in den beiden einseitigen Tests die Nullhypothesen ebenfalls nicht auf einem Signifikanzniveau von 5% verworfen werden.

Bemerkung: Für $n \ge 200$ und $np(1-p) > 9$ erscheinen im Output approximative p-Werte. Da

$$z = \frac{k - E(X)}{\sqrt{Var(X)}} = \frac{k - np}{\sqrt{np(1-p)}} \text{ mit } p = ½$$

(unter H_0) asymptotisch standardnormalverteilt ist, werden als (approximative) p-Werte $2(1-F_{N(0,1)}(|z|))$, $F_{N(0,1)}(z)$ und $1-F_{N(0,1)}(z)$ ausgegeben, wobei eine Stetigkeitskorrektur verwendet wird.

Umsetzung mit SAS:

```
data dat1;
input x;
cards;
1
2
5
5
4
3
2
5
run;

proc univariate data = dat1 mu0=2;
   var x;
run;
```

SAS-Output zur Prozedur UNIVARIATE:

Die Prozedur UNIVARIATE
Variable: x

Momente			
N	8	**Summe Gewichte**	8
Mittelwert	3.375	**Summe Beobacht.**	27
Std.abweichung	1.59798981	**Varianz**	2.55357143
Schiefe	-0.2581915	**Kurtosis**	-1.7422466
Unkorr. Qu.summe	109	**Korr. Quad.summe**	17.875
Variationskoeff.	47.3478462	**Stdfeh. Mittelw.**	0.56497471

Grundlegende Statistikmaße			
Lage		**Streuung**	
Mittelwert	3.375000	**Std.abweichung**	1.59799

Grundlegende Statistikmaße			
Lage		Streuung	
Median	3.500000	**Varianz**	2.55357
Modalwert	5.000000	**Spannweite**	4.00000
		Interquartilsabstand	3.00000

Tests auf Lageparameter: Mu0=2				
Test	**Statistik**		**p-Wert**	
Studentsches t	**t**	2.433737	**Pr > |t|**	0.0452
Vorzeichen	**M**	2	**Pr >= |M|**	0.2188
Vorzeichen-Rang	**S**	9	**Pr >= |S|**	0.0938