

034 Nichtparametrische Test

Hin und wieder sind bei parametrischen Tests die Voraussetzungen nicht erfüllt. Es können etwa die Daten nicht normalverteilt sein, die Stichprobe kann zu klein sein oder es kommen Ausreisser vor. In solchen Fällen können nichtparametrische Tests verwendet werden. Häufig wird der Test nicht mit den Daten, sondern mit ihren Rängen der Daten durchgeführt.

Im Folgenden werden zwei gepaarte (abhängige, verbundene) Stichproben verglichen. Dabei bedeutet gepaart, dass zum Beispiel für jede Person Messwerte zu zwei verschiedenen Zeitpunkten, ohne oder mit Behandlung vorliegen oder auch in der Landwirtschaft Ernteerträge mit zwei verschiedenen Düngemitteln. Der Test heisst Wilcoxon-Vorzeichen-Rangtest. Alternativ kann auch der **Vorzeichentest** verwendet werden.

Der Wilcoxon-Test ist das parameterfreie Gegenstück zum t-Test für 2 abhängige Stichproben. Einen entsprechenden Test gibt es auch für zwei unabhängige Stichproben: Wilcoxon-Mann-Whitney-Test oder u-Test. Als Alternative steht auch der **Welchtest** mit Rangzahlen zur Verfügung.

Wilcoxon-Vorzeichentest

Die Grundidee wird am folgenden Beispiel erläutert:

Beispiel:

Die Wirkung eines neuen Schlafmittels wird an 10 weiblichen Probanden getestet. Für jede Probandin wird die Zeit in Minuten zwischen dem Zubettgehen und dem Einschlafen gemessen, und zwar einmal mit dem neuen Schlafmittel und einmal mit einem Placebo (ohne Wirkstoff). Die Reihenfolge Schlafmittel/Placebo wird für jede Probandin zufällig bestimmt.

Probandin	mit Placebo	mit Schlafmittel	Differenz	Betrag der Differenz	Rang des Betrags	R+	R-
1	18	10	8	8	7	7	
2	24	11	13	13	9	9	
3	22	12	10	10	8	8	
4	6	11	-5	5	4		4
5	10	10	0	0			
6	7	5	2	2	1.5	1.5	
7	8	10	-2	2	1.5		1.5
8	10	17	-7	7	6		6
9	14	10	4	4	3	3	
10	9	15	-6	6	5		5
Mittelwert	12.8	11.1				28.5	16.5
						R+	R-
					Kontrolle:	45	45

Es soll geklärt werden, ob es einen Unterschied zwischen den mittleren Werten (Medianen) bei den zwei gepaarten Messungen gibt. Dazu bildet man die Differenzen der Messpaare. Wertepaare mit der Differenz 0 werden weggelassen (gelb)

Als Hypothesen formuliert:

H0: Nullhypothese:

positive und negative Werte sind gleichwahrscheinlich, Die beiden Werte des Paares unterscheiden sich nicht wesentlich. Der Median der Differenzen ist 0.

H1: Alternativhypothese:

Es treten auffällig viele kleine oder grosse Differenzen auf. Es besteht ein Unterschied zwischen den Werten (vorher oder nachher, mit oder ohne Behandlung). Der Median der Differenzen ist von 0 verschieden.

H0: Nullhypothese

positive und negative Werte sind gleichwahrscheinlich. Die beiden Werte des Paares unterscheiden sich nicht wesentlich. Der Median der Differenzen ist 0.

HA: Alternativhypothese

Es treten auffällig viele kleine oder grosse Differenzen auf. Es besteht ein Unterschied zwischen den Werten (vorher/nachher, mit/ohne Behandlung).

1.

Es wird eine Rangliste für die Beträge der Differenzen erstellt. Kommen Messwerte mehrfach (engl. ties) vor, so erhalten alle den Mittelwert der Ränge zugeteilt. Kommt z.B eine Differenz viermal vor und sind die möglichen Ränge 1, 2, 3, 4, dann wird ihnen der Mittelwert der Ränge 2.5 zugewiesen.

2.

Es werden in zwei Spalten die positiven und die negativen Differenzen und die entsprechenden Rangsummen R^+ und R^- gebildet. Die Summe von R^+ und R^- muss mit der Summe der n ersten natürlichen Zahlen $\frac{1}{2} \cdot n \cdot (n + 1)$ übereinstimmen.

Frage:

Gibt es einen Unterschied zwischen den mittleren Werten (Medianen) zwischen den zwei gepaarten Messungen? Oder als Hypothesen formuliert:

Die zu Grunde liegende Idee ist, dass unter der Nullhypothese H0 die Werte R^+ und R^- ungefähr gleich sein sollten, also nahe bei $\mu = \frac{1}{4} \cdot n \cdot (n + 1)$, der Hälfte der Gesamtangsumme.

Bei auffällig nach unten oder oben abweichenden Teilsummen

R^+ bzw. R^- wird man die Nullhypothese ablehnen.

Testgrösse ist die kleinere Rangsumme, im Beispiel $R^- = 15$. - Ist die Anzahl der Messwertpaare klein, dann kann die Entscheidungsregel kombinatorisch ermittelt werden. Es gibt insgesamt $2^9 = 512$ Vorzeichenfolgen. In der Tabelle ist ein Ausschnitt aufgeführt. Die Liste für R^+ beginnt mit 0 (lauter Pluszeichen) und endet mit 45 (lauter Minuszeichen). Gibt man etwa das Signifikanzniveau 5% (zweiseitig) vor, so wird man die Nullhypothese H_0 (kein Unterschied zwischen den beiden Methoden) ablehnen, wenn die kleinere Rangsumme höchstens in 2.5% aller Fälle auftritt. Wegen 2,5% von 512 also 12.8 gehören maximal 12 Folgen am Anfang und am Schluss zum Verwerfungsbereich. Der kritische Wert ist also 5. Für die Ränge 1 – 5 wird man also die Nullhypothese verwerfen. Da im Beispiel $R^- = 15$ ist, wird man die Hypothese, dass beide Methoden gleich gut sind, also vorläufig beibehalten.

Quelle: Strick)

Rang									Summe der neg. Rangzahlen	
1	2	3	4	5	6	7	8	9		
+	+	+	+	+	+	+	+	+	0	Verwerfungsbereich
-	+	+	+	+	+	+	+	+	1	
+	-	+	+	+	+	+	+	+	2	
+	+	-	+	+	+	+	+	+	3	
-	-	+	+	+	+	+	+	+	4	
+	+	+	-	+	+	+	+	+	4	
-	+	-	+	+	+	+	+	+	5	
+	+	+	+	-	+	+	+	+	5	
-	+	+	-	+	+	+	+	+	5	
+	-	-	+	+	+	+	+	+	5	
+	+	+	+	+	-	+	+	+	Annahmehbereich	
-	+	+	+	-	+	+	+	+		
+	-	+	-	+	+	+	+	+		
-	-	-	+	+	+	+	+	+		
+	+	+	+	+	+	-	+	+		
-	+	+	+	+	-	+	+	+		
+	-	+	+	-	+	+	+	+		
.		
.		
.		
-	+	-	-	+	-	-	-	-		
+	-	-	-	-	+	-	-	-		
-	-	-	-	-	-	+	-	-		
+	+	+	-	-	-	-	-	-		
-	+	-	+	-	-	-	-	-		
+	-	-	-	+	-	-	-	-		
-	-	-	-	-	+	-	-	-		
-	+	+	-	-	-	-	-	-	40	Verwerfungsbereich
+	-	-	+	-	-	-	-	-	41	
-	-	-	-	+	-	-	-	-	42	
+	-	+	-	-	-	-	-	-	43	
-	-	+	-	-	-	-	-	-	44	
-	+	-	-	-	-	-	-	-	45	
+	-	-	-	-	-	-	-	-	45	
-	-	-	-	-	-	-	-	-	45	

In der Praxis verwendet man Tabellen oder eine Statistiksoftware z. B. R.

Kritische Grenzen für den Test von Wilcoxon

Kritische Grösse $r_{0.025}(n)$ zur Durchführung des Tests von WILCOXON auf dem 5%-Niveau. Die Nullhypothese wird verworfen, falls

$$r_+ \leq r_{0.025}(n) \text{ oder } r_+ \geq \frac{n(n+1)}{2} - r_{0.025}(n)$$

Für $n < 6$ sind die Grenzen nicht definiert, auf dem Niveau 5% wird die Nullhypothese immer akzeptiert.

n	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
$r_{0.025}$	0	2	3	5	8	10	13	17	21	25	29	34	40	46	52	58	65	73

Für grössere Werte von n kann man die Näherung

$$r_{0.025}(n) \approx \frac{n(n+1)}{4} - 1.96 \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

verwenden.

Für $n > 25$ kann die Testgrösse durch die Normalverteilung approximiert werden:

Es gilt für den Erwartungswert $\mu = \frac{1}{4} \cdot n \cdot (n + 1)$

und die Standardabweichung $\sigma = \sqrt{\frac{1}{24} \cdot n \cdot (n + 1) \cdot (2n + 1)}$