

Vergleich Neu gegen Alt

Häufig kommt es vor, dass man zwei „Dinge“ z.B. ein neues Produkt gegen eine alte miteinander vergleichen will. Dabei geht es um ein bestimmtes Kriterium, das wenn möglich, mit einem Messwert zu beschreiben ist. Hat man eine Anzahl „Neue“ und „Alte“ Teile oder Systeme, so werden diese nach ihrer Bewertung der Reihe nach dargestellt. Gibt es zum Beispiel je 2 Teile, so könnte sich folgende Reihenfolge ergeben:

Teil	Bewertung
N	1,2
N	1,1
A	0,9
A	0,8

Es wird hier vorausgesetzt, dass die höhere Bewertung die bessere ist. Nach der sich dann ergebenden Reihenfolge erscheint das Ergebnis zunächst eindeutig. Die beiden Neu-Teile liegen vor den Alt-Teilen. Die Reihenfolge könnte aber auch zufällig die gleiche sein. Für 2 Neu und 2 Alt ergeben sich insgesamt 6 verschiedene Möglichkeiten:

N	N	N	A	A	A
N	A	A	N	N	A
A	N	A	N	A	N
A	A	N	A	N	N

Allgemein wird die Anzahl der möglichen Varianten durch:

$$\text{Varianten} = \frac{(n_{\text{neu}} + n_{\text{alt}})!}{n_{\text{neu}}! \cdot n_{\text{alt}}!}$$

bestimmt. Die Wahrscheinlichkeit des Ergebnisses in der ersten Spalte ist also 1/6 oder 16.667%.

Mit dieser Stichprobe soll eine Aussage über die Grundgesamtheit gemacht werden. Hierfür stellt man die Nullhypothese auf, dass Neu tatsächlich besser ist, als Alt. In gleicher Weise, wie für statistische Tests ein Signifikanzniveau festgelegt wird, soll hier ein Grenzwert von 5% gelten. Da dieser den Grenzwert mit 16.667% jedoch weit überschritten wird, ist das Ergebnis nicht signifikant. Grundsätzlich sollte das Signifikanzniveau vorher festgelegt werden. Bei einer Reihenfolge, bei der immer alle Neuen vor den Alten liegen, braucht man nur noch den Kehrwert der Anzahl der Varianten gegen das festgelegte Signifikanzniveau zu vergleichen. Die Nullhypothese, dass Neu besser ist als Alt, ist immer abzulehnen, wenn $100\% / \text{Varianten} > \text{Signifikanzniveau}$ ist.

Die folgende Tabelle zeigt die entsprechenden Stichprobenumfänge für verschiedene Signifikanzniveaus, bei der die Nullhypothese nicht abzulehnen ist:

Signifikanz-niveau	Umfang N	Umfang A
0,1%	2	43
0,1%	3	16
0,1%	4	10
0,1%	5	8
0,1%	6	6
1%	2	13
1%	3	7
1%	4	5
1%	5	4
1%	6	
5%	1	19
5%	2	5
5%	3	3
5%	4	3
5%	5	
10%	1	9
10%	2	3
10%	3	2
10%	4	
10%	5	

Ist ein Stichprobenumfang von mindestens je 10 vorhanden, so kann man die Hypothese auch prüfen, wenn sich die Rangfolge überlappt. Z.B.:

NNNANNAANNNNAAANAAAA

Unterscheiden sich die Stichprobenumfänge von N und A, dann muss der Stichprobenumfang für N größer sein. Dabei darf aber ein Verhältnis dieser Umfänge von 5:4 nicht überschritten werden. Im Falle einer Überlappung sind die jeweils gleichen Neu am Anfang zu zählen, bis Alt auftritt und vom Ende her Alt zu zählen, bis Neu auftritt. In unserem Beispiel ergibt sich die Anzahl gleicher Nennungen $3+4=7$. Diese Zahl ist gegen den folgenden Grenzwert zu vergleichen:

Signifikanz-niveau	Grenzwert
5%	6
1%	9
0,1%	12

Ist der Wert gleicher Nennungen größer, als der Grenzwert unter Berücksichtigung des vorher gewählten Signifikanzniveaus, so ist die Nullhypothese, dass Neu besser ist, als Alt zu bestätigen. In unserem Beispiel ist dies für 5% der Fall, da $7 > 6$ ist.

Ist eine messbare Bewertung nicht möglich, sondern nur eine vergleichende Prüfung, so kann der Test trotzdem angewendet werden, z.B. wenn das Design eines Autos miteinander verglichen werden soll. Das Prinzip ist aber immer das gleiche. Es wird eine

Rangfolge Neu gegen Alt aufgestellt. Die „Tester“ müssen die Autos unabhängig voneinander vergleichen, welches Auto schöner ist. Nennen von 5 alle ein und dasselbe Auto, dann ist dieses unter dem vorher festgelegten Signifikanzniveaus von 1% das schönere. Trifft jedoch nur einer eine andere Wahl, dann sind die Autos nicht unterscheidbar. Für diesen Vergleich gelten folgende Grenzwerte:

Signifikanzniveau	Grenzwert
5%	3
1%	5
0,1%	6

Auch hier ist bei einer Mindestanzahl von 10 Vergleiche der Test möglich, wenn unterschiedliche Nennungen auftreten. Aus den Nennungen wird der Vergleichswert durch

$$2 * |\text{NennungN} - \text{NennungA}|$$

berechnet. Hier gelten wiederum die Grenzwerte, wie bei der messbaren Bewertung (6-9-12).

Für die dargestellten Verfahren ist die Vorlage [Vergleich Neu gegen Alt.vxg](#) anzuwenden. Es wird immer davon ausgegangen, dass bei der Bewertung unabhängig von den Zahlenwerten Neu die „bessere“ Bewertung sein muss.

Siehe auch U-Test nach Wilcoxon, Mann und Whitney