1.5 Berechnung von Rangzahlen

Bei vielen nichtparametrischen Verfahren spielen die so genannten Rangzahlen eine wesentliche Rolle, denn über diese werden hier die Prüfgrößen berechnet. Dies steht im Gegensatz zu den parametrischen Verfahren, bei denen die Prüfgrößen über die eigentlichen Beobachtungen berechnet werden und die oft voraussetzten, dass die Daten aus einer normalverteilten Grundgesamtheit stammen. Bei einigen nichtparametrischen Verfahren, bei denen die Prüfgrößen über die Rangzahlen berechnet werden, muss das Datenniveau (u.a. falls keine Berechnungen wie Subtraktionen mit den Daten durchgeführt werden, wie beim Vorzeichentest oder Vorzeichenrangtest) nur einer Ordinalskala genüge. Allerdings können bei zu geringem Datenniveau viele Werte mehrfach vorkommen, was problematisch sein könnte.

Die zugrunde liegenden Daten werden bei der Rangzahlenvergabe durch eine monotone Transformation auf die Rangzahlen und somit auf die rationalen Zahlen abgebildet. Die Information über die Abstände der originalen Werte gehen dabei verloren. Ist das zugrunde liegende Datenniveau metrisch, so stellt dies natürlich einen Informationsverlust dar, der aber in Bezug auf die Effizienz des Verfahrens in vielen Fällen unerheblich ist.

Die Daten im Beispiel sind:

v1
150
155
140
156
140
156
180
156
150

Wenn man diese Daten in der Spalte v1 eingibt und dann **→Univariate Statistik →Rangzahlen berechnen** wählen, dann sehen Sie die sortierte Stichprobe mit den Rangzahlen (es wurden noch Hinweise zur Berechnung der Rangzahlen eingefügt):

i	sortierte Stichprobe (y_i)	Rang(y_i)
1	140	1.5 = (1+2)/2
2	140	1.5 = (1+2)/2
3	150	3.5 = (3+4)/2
4	150	3.5 = (3+4)/2
5	155	5
6	156	7 = (6+7+8)/3
7	156	7 = (6+7+8)/3
8	156	7 = (6+7+8)/3
9	180	9

Bei der Vergabe der Rangzahlen wird so vorgegangen, dass der kleinsten Beobachtung der Rang 1 und der größten Beobachtung der Rang n (= Stichprobenumfang) zugewiesen wird. Kommen Beobachtungen doppelt vor (so genannte Bindungen), so wird diesen das arithmetische Mittel der entsprechenden Rangzahlen zugewiesen. In unserem Beispiel ist der kleinste Wert 140. Es kommen zwei Beobachtungen mit diesem Wert vor. Man müsste eigentlich für diese beiden kleinsten Beobachtungen die Ränge 1 und 2 vergeben. Da diese aber doppelt vorkommen, erhalten beide Beobachtungen den Rang 1,5, also den Mittelwert aus den Rängen 1 und 2.

Würden oben keine Bindungen vorkommen, dann würden alle Rangzahlen der Beobachtungen mit den Nummern der Beobachtung in der sortierten Stichprobe übereinstimmen. Also wenn y_i die Werte der sortierten Stichprobe sind, dann würde Rang(y_i) = i gelten.

Dadurch, dass wir die Rangzahlen derart definiert haben, gilt

$$\sum_{i=1}^{n} \text{Rang}(x_i) = \frac{n(n+1)}{2}$$

und somit ergibt sich der Mittelwert der Rangzahlen:

$$\bar{r} = \frac{n+1}{2}$$

Bei einigen nichtparametrischen Testverfahren ändert sich u.a. die Varianz der Prüfgröße, wenn Bindungen vorkommen. Für deren Berechnung müssen deshalb zweifach, dreifach, … vorkommende Werte berücksichtigt werden. Aus diesem Grund benötigen wir noch eine Folge $(t_i)_{i=1,2,\ldots k}$ der absoluten Häufigkeiten. Um diese zu bestimmen wurde die Tabelle der absoluten Häufigkeiten in unserem Beispiel bestimmt:

Beobachtung (Ausprägungen)	absolute Häufigkeit
140	2
150	2
155	1
156	3
180	1

In welcher Reihenfolge die absoluten Häufigkeiten in einer Folge festgelegt werden, ist im Prinzip egal. Wir definieren nun diese Folge für unser Beispiel über die Häufigkeiten, die nach den Größen der zugehörigen Ausprägungen aufsteigend sortiert wurden (wie in obiger Tabelle):

Ausprägung	absolute Häufigkeit
140	$t_1 = 2$
150	$t_2 = 2$
155	$t_3 = 1$
156	$t_4 = 3$
180	$t_5 = 1$

Damit ist k = 5, da nur 5 Ausprägungen vorkommen. Das Wort Ausprägung haben wir oben für die vorkommenden Werte der Variable bzw. Spalte v1 verwendet, d.h. die voneinander verschiedenen Beobachtungen. Kommen keine Bindungen vor, dann ist k = n. Ansonsten ist k genau die Anzahl der verschiedenen Werte, bzw. der Anzahl der Ausprägungen.

Umsetzung mit SAS:

```
data dat3;
input x;
cards;
150
155
140
156
140
156
180
156
150
run;


proc rank data = dat3
out=datout;
var x;
ranks rang;
run;
```

```
proc sort data=datout;
by x;
run;

proc print data=datout;
run;
```

SAS-Output zur Prozedur RANK:

Beob.	x	rang
1	140	1.5
2	140	1.5
3	150	3.5
4	150	3.5
5	155	5.0
6	156	7.0
7	156	7.0
8	156	7.0
9	180	9.0