

## Kapitel 8: Verfahren für Rangdaten

Der Mann-Whitney <i>U</i> -Test	1
Der Wilcoxon-Test	3
Der Kruskal-Wallis <i>H</i> -Test	4
Vergleich von Mann-Whitney <i>U</i> -Test und Kruskal-Wallis <i>H</i> -Test	5

### Der Mann-Whitney *U*-Test

In Kapitel 8.1 dient eine Klassenarbeit in einer Schule als Beispielanwendung für einen *U*-Test. Wir werden an dieser Stelle die Berechnung dieses Beispiels mit R nachvollziehen. Öffnen Sie dazu den Datensatz „Rangdaten.sav“.

```
library(foreign)
rang <- read.spss("Rangdaten.sav",
                 to.data.frame = TRUE)
```

Um einen Mann-Whitney-*U*-Test auszuführen, verwendet man die Funktion `wilcox.test()`. Geben Sie zunächst die abhängige Variable ein, die Sie auf Unterschiede zwischen den Gruppen testen möchten. Danach geben Sie die unabhängige Variable an, welche Gruppen miteinander verglichen werden sollen.

In manchen Fällen gibt es mehr als zwei unterschiedliche Gruppen (z.B. Erstklässler, Zweitklässler usw.). Wenn dieser Fall eintritt, müssen Sie R anweisen, nur die relevanten Gruppen zu berücksichtigen, um einen *U*-Test ausführen zu können. Verwenden Sie die Funktion `subset()` und tragen Sie die gewünschten Gruppen ein. Da es im vorliegenden Fall nur zwei Gruppen gibt, ist dieser Schritt hier nicht notwendig.

Mithilfe des Attributs `conf.level` können Sie das gewünschte Signifikanzniveau und die Größe des Konfidenzintervalls festlegen. Da R grundsätzlich zweiseitig testet, darf bei gerichteten Hypothesen das Signifikanzniveau verdoppelt werden. Tragen Sie daher für den vorliegenden Fall ein Signifikanzniveau von 10% ein.

```
rang.modell <- wilcox.test(rang$punkte ~ rang$sex,
                          conf.level = 0.90)
rang.modell
```

Sie erhalten folgenden Output.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

```
wilcoxon rank sum test
```

```
data: rang$punkte by rang$sex
```

```
w = 62, p-value = 0.05453
```

```
alternative hypothesis: true location shift is not equal to 0
```

In der letzten Zeile formuliert R die Alternativhypothese. Es wird das von anderen Outputs bekannte zweiseitige Signifikanzniveau ausgegeben. Während bei den meisten anderen Verfahren signifikante Kennwerte durch Sternchen (\*) gekennzeichnet wurden, gibt R hier lediglich den  $p$ -Wert aus.

Anders als in Kapitel 8.1 vorgestellt, vergibt R die Ränge so, dass die höchsten Werte die niedrigsten Ränge einnehmen. Der erste Rang geht also an die höchste Punktzahl. Mit anderen Worten: R hantiert mit den Rangplatzunterschreitungen, während sich die Darstellung in Kapitel 8.1 an den Rangplatzüberschreitungen orientiert. Diese Vorgehensweise führt bei der Berechnung der Rangsumme (und damit der mittleren Rängen) zu deskriptiv anderen Ergebnissen als in Kapitel 8.1.

Der von R ausgegebene  $W$ -Wert entspricht dem in Kapitel 8.1 vorgestellten Wert  $U' = 62$ .

Für große Stichproben lässt sich der  $U'$ -Wert zunächst in einen  $U$ -Wert und dann in einen  $z$ -Wert überführen (siehe Kapitel 8.1 für die genaue Vorgehensweise). Die unterschiedlichen Vorzeichen gehen auf die unterschiedliche Berechnung des  $U$ -Wertes zurück. Dieser Unterschied ist für die Signifikanzberechnung unerheblich, bei gerichteten Hypothesen ist es aber natürlich wichtig zu überprüfen, dass die Unterschiede in der vorhergesagten Richtung vorliegen.

Da die Stichprobe in unserem Beispiel sehr klein war, orientieren wir uns bei der Beurteilung der Signifikanz an der „exakten Signifikanz“. Bei zweiseitiger Testung ist das Resultat knapp nicht signifikant. Da die Hypothese des Lehrers aber gerichtet war („Mädchen schneiden in der Klassenarbeit besser ab“), dürfen wir einseitig auf Signifikanz prüfen, was zu einer Wahrscheinlichkeit von weniger als 5% für diesen Unterschied führt.

Im Datensatz ist leicht nachzuvollziehen, wie die Rangsummen bzw. mittleren Rangplätze berechnet werden. Sortieren Sie dafür den Datensatz nach aufsteigenden Punkten.

```
rang <- rang[order(rang$punkte),]
```

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

	sex Geschlecht	punkte Punkte
9	Jungen	3
10	Jungen	5
11	Jungen	8
1	Mädchen	9
12	Jungen	11
13	Jungen	12
14	Jungen	14
15	Jungen	16
2	Mädchen	17
3	Mädchen	18
16	Jungen	19
4	Mädchen	20
5	Mädchen	21
17	Jungen	22
6	Mädchen	24
18	Jungen	25
7	Mädchen	26
8	Mädchen	28

Für Mädchen, die mit „1“ kodiert sind:

$$\text{Mittlerer Rangplatz}_{\text{Mädchen}} = \frac{4 + 9 + 10 + 12 + 13 + 15 + 17 + 18}{N_{\text{Mädchen}}} = \frac{98}{8} = 12,25$$

Für Jungen, die mit „2“ kodiert sind:

$$\text{Mittlerer Rangplatz}_{\text{Jungen}} = \frac{1 + 2 + 3 + 4 + 5 + 6 + 7 + 11 + 14 + 16}{N_{\text{Jungen}}} = \frac{73}{10} = 7,3$$

Hier wird deutlich, dass die Rangunterschiede in der vorhergesagten Richtung vorliegen. Die einseitige Signifikanztestung ist also zulässig.

## Der Wilcoxon-Test

Der Wilcoxon-Test für abhängige Stichproben ist innerhalb der nichtparametrischen Verfahren das Pendant zum *U*-Test für unabhängige Stichproben. Damit ist er andererseits auch das nichtparametrische Pendant zum *t*-Test für abhängige Stichproben für den Fall, dass dessen Voraussetzungen verletzt sind.

Im Datensatz „Rangdaten.sav“ finden Sie die Werte des in Kapitel 8.2 diskutierten Beispiels über ein Entspannungsprogramm abgetragen. Die Hypothese lautete „Die Teilnehmer sollten sich nach dem Training entspannter fühlen als davor“.

Um mit R einen Wilcoxon-Test durchzuführen, verwenden Sie wieder die Funktion `wilcox.test()`. Die relevanten Variablen lauten – analog zum Buch – „vorher“ und „nachher“.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

## R-Ergänzungen

Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

Tragen Sie diese beiden Variablen in die Klammern ein. Ergänzen Sie noch `paired = TRUE`, weil es sich um abhängige Daten handelt und `correct = FALSE`. Da auch hier die Hypothese gerichtet ist, muss das Signifikanzniveau verdoppelt werden.

```
rang.wil <- wilcox.test(rang$vorher, rang$nachher,  
                        paired = TRUE,  
                        correct = FALSE,  
                        conf.level = 0.90)  
rang.wil
```

### wilcoxon signed rank test

```
data: rang$vorher and rang$nachher  
V = 1.5, p-value = 0.02009  
alternative hypothesis: true location shift is not equal to 0
```

Genauso wie im Buch verwendet R den kleineren der beiden „gerichteten Ränge“ als Kennwert, was mit der Reihenfolge der eingetragenen Variablen zusammenhängt. Dieser stimmt genau mit dem in Kapitel 8.2 per Hand ermittelten überein. Der Wilcoxon-Test liefert ein signifikantes Ergebnis,  $p = 0,02$  bei zweiseitiger Testung. Da die Hypothese des Forschers gerichtet war, dürfen wir einseitig testen. Das Ergebnis korrespondiert exakt mit dem aus Tabelle G (Band 1) ermittelten überein.

## Der Kruskal-Wallis *H*-Test

Der Kruskal-Wallis *H*-Test ist ein Verfahren für ordinalskalierte Daten für den Vergleich von mehr als zwei unabhängigen Gruppen. Ebenso kommt es bei intervallskalierten Daten zum Einsatz, wenn die Voraussetzungen der einfaktoriellen Varianzanalyse ohne Messwiederholung verletzt sind. Deshalb heißt dieses Verfahren auch Rangvarianzanalyse.

Wir werden an dieser Stelle das Beispiel aus Kapitel 8.3 mit R nachvollziehen. Dort ging es um den Vergleich von drei unabhängigen Gruppen mit jeweils fünf Versuchspersonen.

Um in R einen Kruskal-Wallis *H*-Test durchzuführen, verwenden Sie die Funktion `kruskal.test()`. Im Datensatz „Rangdaten.sav“ finden Sie die Variable „bedingung“, die als Gruppenvariable für dieses Beispiel dient. Die abhängige Variable ist als „av“ gekennzeichnet.

```
kruskal.test(rang$av ~ rang$bedingung)
```

### Kruskal-wallis rank sum test

```
data: rang$av by rang$bedingung  
Kruskal-wallis chi-squared = 6.5501, df = 2, p-value = 0.03782
```

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

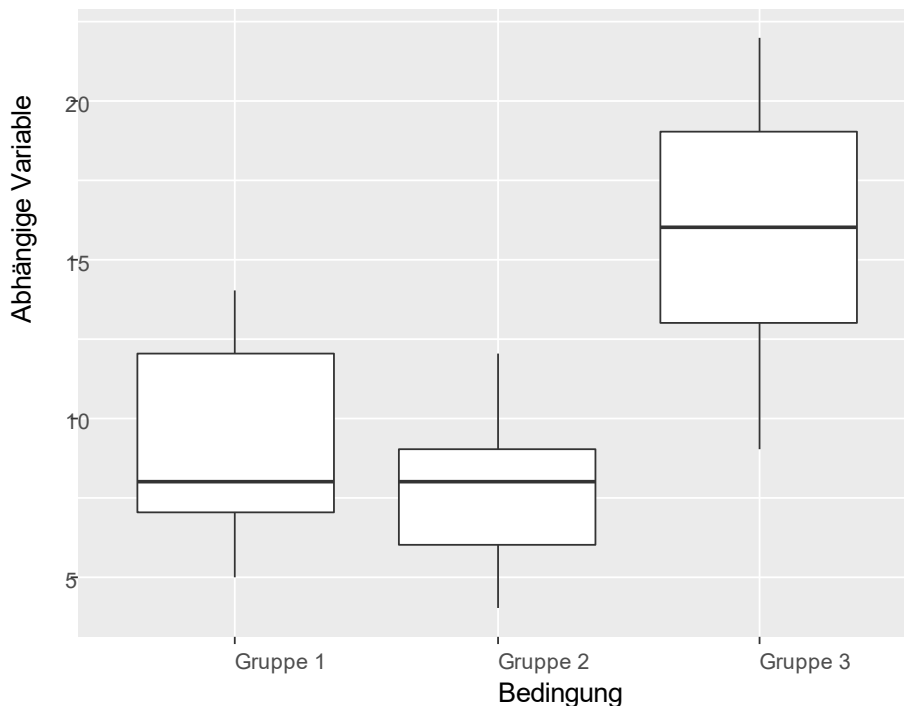
Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature

Der ausgegebene Wert „Kruskal-Wallis chi-squared“ entspricht dem per Hand ermittelten  $\chi^2$ -Wert. Die geringen Abweichungen sind auf einen etwas anderen Berechnungsalgorithmus von R zurückzuführen. Bei drei Gruppen hat die Verteilung zwei Freiheitsgrade. Das Ergebnis ist signifikant auf dem 5%-Niveau. Die mittleren Ränge der Gruppen unterscheiden sich voneinander.

Wenn Sie möchten, können Sie sich mit `ggplot` ein Boxplot-Diagramm ausgeben lassen. Da es fehlende Werte auf der Variable „bedingung“ gibt, wurde der Datensatz zunächst gefiltert.

```
library(ggplot2)
rang.kru <- rang[complete.cases(rang$bedingung),]
rang.kru.box <- ggplot(rang.kru, aes(rang.kru$bedingung,
                                     rang.kru$av)) +
  geom_boxplot() +
  labs(x = "Bedingung",
       y = "Abhängige Variable")
rang.kru.box
```



### Vergleich von Mann-Whitney *U*-Test und Kruskal-Wallis *H*-Test

Der *U*-Test überprüft zwei unabhängige Gruppen mit ordinalskalierten Daten auf Unterschiedlichkeit. Damit entspricht er dem *t*-Test für unabhängige Stichproben, der dies für intervallskalierte Daten tut. Analog dazu vergleicht der Kruskal-Wallis *H*-Test mehr als zwei unabhängige Gruppen, so wie dies die einfaktorielle Varianzanalyse ohne Messwiederholung tut.

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Friese, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

## R-Ergänzungen

Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2* (5. Auflage). Heidelberg: Springer.

---

Diese Parallelen lassen die Vermutung zu, dass das Verhältnis zwischen  $U$ -Test und  $H$ -Test dem zwischen  $t$ -Test und einfaktorieller Varianzanalyse entspricht. In den R-Ergänzungen zu Kapitel 5 haben wir gesehen, dass ein  $t$ -Test identisch ist mit einer Varianzanalyse zweier Gruppen. Gilt diese Beziehung auch zwischen  $U$ -Test und Rangvarianzanalyse?

Gehen wir zurück zu dem im Abschnitt über den  $U$ -Test verwendeten Beispiel. Dort ging es um unterschiedliche Ergebnisse von Mädchen und Jungen in einer Klassenarbeit. Um den Vergleich zu verdeutlichen, wurde diesmal der von R per Standardeinstellung durchgeführte exakte Test und die Kontinuitätskorrektur deaktiviert.

```
rang.modell <- wilcox.test(rang$punkte ~ rang$sex,  
                           exact = FALSE,  
                           correct = FALSE,  
                           conf.level = 0.90)  
rang.modell
```

Wir erhalten folgenden Output:

```
      wilcoxon rank sum test  
  
data:  rang$punkte by rang$sex  
w = 62, p-value = 0.05061  
alternative hypothesis: true location shift is not equal to 0
```

Wenn die Rangvarianzanalyse wirklich eine Erweiterung des  $U$ -Tests von zwei auf mehrere unabhängige Gruppen ist (so wie die Varianzanalyse eine Erweiterung des  $t$ -Tests ist), dann sollten wir identische Ergebnisse erhalten, wenn wir dieselben Daten mit der Rangvarianzanalyse auswerten.

Verwenden Sie die Funktion `kruskal.test()` und wählen Sie die Variable „sex“ als Gruppenvariable und „punkte“ als abhängige Variable.

```
kruskal.test(rang$punkte ~ rang$sex)
```

Sie erhalten folgende Werte:

```
      Kruskal-wallis rank sum test  
  
data:  rang$punkte by rang$sex  
Kruskal-wallis chi-squared = 3.8211, df = 1, p-value = 0.05061
```

Wie erwartet entspricht der  $p$ -Wert für die asymptotische Berechnung der Signifikanz der Angabe, die R bei Verwendung des  $U$ -Tests für dieses Datum ausgibt. Obwohl sich die vorgestellten Berechnungen der Kennwerte zwischen beiden Verfahren unterscheiden, gilt also:  $U$ -Test und Rangvarianzanalyse liefern identische Ergebnisse im Fall von zwei unabhängigen Gruppen, so wie dies auch für den  $t$ -Test und die einfaktorielle Varianzanalyse gilt (siehe dazu Kapitel 5, insbesondere die R-Erläuterungen zu Kapitel 5).

Quelle: <https://lehrbuch-psychologie.springer.com/content/zu-den-spss-r-und-gpower-aufgaben-und-ergaenzungen>

Aus: Rasch, Frieze, Hofmann & Naumann (2021). *Quantitative Methoden. Band 2*, 5. Auflage. Heidelberg: Springer.

© Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature