

Kapitel 8: Verfahren für Rangdaten

Anmerkung	1
Mann-Whitney <i>U</i> -Test	1
Wilcoxon-Test	3
Kruskal-Wallis <i>H</i> -Test	3
Literatur	6

Anmerkung

In Kapitel 8 der Bücher wird erwähnt, dass für nichtparametrische Daten keine eigenen Prozeduren zur Berechnung von Effekt- und Teststärken vorhanden sind, weshalb ein Rückgriff auf die Methoden für parametrische Daten notwendig ist. Dies ist mittlerweile allerdings nur noch beim Kruskal-Wallis *H*-Test notwendig, da mit der neuesten Version von G*Power (3.1) für den Mann-Whitney *U*-Test und den Wilcoxon-Test ein eigener Berechnungsweg umgesetzt wurde.

Mann-Whitney *U*-Test

Da die Bestimmung der Test- und Effektstärke für den allgemeinen Fall zu aufwändig ist, beschränken wir uns auf die Fälle, in denen der *U*-Test als Alternative zum *t*-Test verwendet wird. Der interessierte Leser sei für eine genaue Betrachtung auf Lehmann (1975) verwiesen.

Obwohl der Mann-Whitney *U*-Test ein verteilungsfreies Verfahren ist, muss für die Berechnung des Stichprobenumfangs und der Teststärke die Verteilungsform beachtet werden. Die Verteilungsform muss vor der Berechnung festgelegt werden, da die berechneten Werte von ihr abhängig sind.

G*Power bietet die drei Voreinstellungen „Normal“, „Laplace“ und „Logistic“. Die genannten Verteilungen sind alle kontinuierlich und symmetrisch, unterscheiden sich jedoch bezüglich der Kurtosis (Steilheit). Zusätzlich gibt es die Auswahlmöglichkeit „min ARE“. Die „asymptotic relative efficiency“ (ARE) beschreibt die Effizienz des *U*-Tests im Vergleich zum *t*-Test. Der kleinste Wert, den ARE annehmen kann, ist 0,864. Die Einstellung „minARE“ geht also vom schlechtesten Fall aus und schätzt damit die Teststärke konservativ. Der ARE-Wert kann auch per Hand berechnet werden. Für eine genauere Betrachtung sei auf die Darstellung im G*Power Handbuch verwiesen (<http://www.gpower.hhu.de>).

Stichprobenumfangsplanung bzw. Teststärkeanalyse a priori

Öffnen Sie G*Power und wählen Sie „Test family“: *t*-tests, „Statistical test“: Means: Wilcoxon-Mann-Whitney test (two groups), „Type of power analysis“: A priori: Compute required sample size. Sie sehen nun folgendes Fenster vor sich:

G*Power-Ergänzungen

Rasch, Frieze, Hofmann & Naumann (2014). *Quantitative Methoden. Band 2* (4. Auflage). Heidelberg: Springer.

The screenshot shows the G*Power software interface with the following settings:

- Test family: t tests
- Statistical test: Means: Wilcoxon-Mann-Whitney test (two groups)
- Type of power analysis: A priori: Compute required sample size - given α , power, and effect size
- Input Parameters:
 - Tail(s): One
 - Parent distribution: Normal
 - Effect size d: 0.5
 - α err prob: 0.05
 - Power ($1 - \beta$ err prob): 0.95
 - Allocation ratio N2/N1: 1
- Output Parameters:
 - Noncentrality parameter δ : ?
 - Critical t: ?
 - Df: ?
 - Sample size group 1: ?
 - Sample size group 2: ?
 - Total sample size: ?
 - Actual power: ?

Zum Beispiel möchte ein Forscher bei der Auswertung seiner Daten eine Teststärke von 90% bei einem Signifikanzniveau von $\alpha = 0,05$ erreichen. Er testet die Hypothese einseitig und erachtet einen Unterschied zwischen den zwei Gruppen ab einem großen Effekt als relevant. Des Weiteren geht der Forscher davon aus, dass seine Daten normalverteilt sind.

Die Effektgröße wird unter „Effect size d“ festgelegt. Hier müssen wir uns an den Konventionen des t -Tests orientieren, obwohl die Übertragung auf den U -Test fraglich ist. Das Signifikanzniveau wird unter „ α err prob“ eingestellt, die Teststärke unter „Power“ und ob einseitig bzw. zweiseitig getestet wird unter „Tail(s)“. Das „Allocation ratio“ gibt das Verhältnis der Größe der beiden Gruppen wieder, sind diese gleich ergibt sich ein Wert von 1. Die Verteilungsform wird unter „Parent distribution“ festgelegt. Nach Eingabe und Berechnung („Calculate“) ergibt sich folgendes Bild:

The screenshot shows the G*Power software interface with the following settings and results:

- Test family: t tests
- Statistical test: Means: Wilcoxon-Mann-Whitney test (two groups)
- Type of power analysis: A priori: Compute required sample size - given α , power, and effect size
- Input Parameters:
 - Tail(s): One
 - Parent distribution: Normal
 - Effect size d: 0.8
 - α err prob: 0.05
 - Power ($1 - \beta$ err prob): 0.90
 - Allocation ratio N2/N1: 1
- Output Parameters:
 - Noncentrality parameter δ : 2.9768687
 - Critical t: 1.6739010
 - Df: 53.3859202
 - Sample size group 1: 29
 - Sample size group 2: 29
 - Total sample size: 58
 - Actual power: 0.9021592

The values 29, 29, and 58 in the Output Parameters section are circled in red.

Um den erwarteten Effekt mit einer Wahrscheinlichkeit von 90% zu finden sollte der Forscher insgesamt 58 Versuchspersonen, also 29 pro Bedingung untersuchen.

Berechnen der Teststärke a posteriori

Die Teststärkebestimmung a posteriori gibt Aufschluss darüber, wie hoch die Chance ist unter den bestehenden Bedingungen den vermuteten Effekt überhaupt zu entdecken. Auch hier muss die Verteilungsform vorher festgelegt werden.

Nehmen wir an der Forscher hat insgesamt 18 Versuchspersonen befragt, das Signifikanzniveau betrug $\alpha = 0,05$ und es wurde ein großer Effekt angenommen. Die Hypothese wurde einseitig getestet und die Daten sind normalverteilt. Um die Teststärke einer Studie a posteriori zu schätzen, wechseln Sie bei „Type of power analysis“ auf „Post hoc: Compute achieved power“. Nach Eingabe der oben genannten Parameter und Berechnung ergibt sich folgendes Bild:

The screenshot shows the G*Power software interface. The 'Test family' is set to 't tests' and the 'Statistical test' is 'Means: Wilcoxon-Mann-Whitney test (two groups)'. The 'Type of power analysis' is 'Post hoc: Compute achieved power - given α , sample size, and effect size'. Under 'Input Parameters', 'Tail(s)' is 'One', 'Parent distribution' is 'Normal', 'Effect size d' is 0.8, ' α err prob' is 0.05, and sample sizes for both groups are 9. Under 'Output Parameters', the 'Power (1 - β err prob)' is 0.4757464, which is circled in red. Other output parameters include Noncentrality parameter δ (1.6583719), Critical t (1.7516210), and Df (15.1887339). Buttons for 'Options', 'X-Y plot for a range of values', and 'Calculate' are visible at the bottom.

Unter diesen Bedingungen beträgt die Teststärke knapp 47,57%. Der β -Fehler liegt also bei knapp 50%. Sollte der Forscher seine Hypothese auf Grund dieser Daten verwerfen und einen Nullunterschied zwischen den Gruppen annehmen, würde er mit 50%iger Wahrscheinlichkeit einen Fehler machen. Er könnte also ebenso gut eine Münze werfen.

Wilcoxon-Test

Die Berechnung erfolgt analog zum bereits erläuterten Mann-Whitney U -Test. Bitte beachten Sie, dass auch hier die Verteilungsform festgelegt werden muss. Wählen Sie in G*Power „Test family“: t-tests und „Statistical test“: Means: Wilcoxon signed-rank test (matched pairs) und folgen Sie den Berechnungsschritten des U -Tests.

Kruskal-Wallis H -Test

Dieser Test ist die nichtparametrische Alternative zur einfaktoriellen Varianzanalyse. Zur Stichprobenumfangsplanung und Teststärkenberechnung greifen wir daher auf die Methoden der einfaktoriellen Varianzanalyse ohne Messwiederholung zurück.

Stichprobenumfangsplanung bzw. Teststärkeanalyse a priori

Öffnen Sie G*Power und wählen Sie „Test family“: F-tests, „Statistical test“: ANOVA: Fixed Effects, omnibus, one-way. Sowie „Type of power analysis“: A priori: Compute required sample size. Sie sehen nun folgendes Fenster vor sich:

The screenshot shows the G*Power software interface. The 'Test family' is set to 'F tests' and the 'Statistical test' is 'ANOVA: Fixed effects, omnibus, one-way'. The 'Type of power analysis' is 'A priori: Compute required sample size - given α , power, and effect size'. Under 'Input Parameters', 'Effect size f' is 0.25, ' α err prob' is 0.05, 'Power (1- β err prob)' is 0.95, and 'Number of groups' is 5. Under 'Output Parameters', all values are marked with a question mark. A 'Calculate' button is visible at the bottom right.

Als Beispiel soll ein Vergleich von drei Gruppen dienen. Der Forscher möchte mit einer Teststärke von 80% und einem Signifikanzniveau von $\alpha = 0,05$ einen mittleren Effekt von $f = 0,25$ finden. Nach Eingabe der Parameter und Berechnung („Calculate“) ergibt sich folgendes Bild:

The screenshot shows the G*Power software interface after calculation. The 'Input Parameters' are: 'Effect size f' is 0.25, ' α err prob' is 0.05, 'Power (1- β err prob)' is 0.80, and 'Number of groups' is 3. The 'Output Parameters' are: 'Noncentrality parameter λ ' is 9.9375000, 'Critical F' is 3.0540042, 'Numerator df' is 2, 'Denominator df' is 156, 'Total sample size' is 159 (circled in red), and 'Actual power' is 0.8048873.

Um einen tatsächlich existierenden Effekt unter diesen Bedingungen zu finden, sollten insgesamt 159 Versuchspersonen, bzw. 53 pro Gruppe untersucht werden. Ist keine Effektgröße aus Vorstudien oder anderen Untersuchungen aus der Literatur bekannt, so empfiehlt sich eine Orientierung an den von Cohen (1988) vorgeschlagenen Konventionen.

Berechnen der Teststärke a posteriori

Da die Zusammenhänge beim Kruskal-Wallis-Test sehr kompliziert sind, kann es bei der Teststärkenberechnung mittels der Methoden der einfaktoriellen Varianzanalyse zu erheblichen Abweichungen kommen. Genauer dazu findet der interessierte Leser in Lehmann (1975). Nehmen wir an, der Forscher habe insgesamt 60 Versuchspersonen in drei Gruppen untersucht. Das Signifikanzniveau betrug $\alpha = 0,05$ und es wurde ein Effekt mittlerer Größe angenommen.

Öffnen Sie G*Power und wählen Sie „Test family“: F- tests, „Statistical test“: ANOVA: Fixed Effects, omnibus, one-way. Sowie „Type of power analysis“: Post hoc: Compute achieved power. Nach Eingabe der Parameter und Berechnung ergibt sich folgendes Bild:

The screenshot shows the G*Power software interface with the following settings:

- Test family: F tests
- Statistical test: ANOVA: Fixed effects, omnibus, one-way
- Type of power analysis: Post hoc: Compute achieved power – given α , sample size, and effect size
- Input Parameters:
 - Effect size f: 0.25
 - α err prob: 0.05
 - Total sample size: 60
 - Number of groups: 3
- Output Parameters:
 - Noncentrality parameter λ : 3.7500000
 - Critical F: 3.1588427
 - Numerator df: 2
 - Denominator df: 57
 - Power (1- β err prob): 0.3744311 (circled in red)

Buttons at the bottom include "X-Y plot for a range of values" and "Calculate".

Die Wahrscheinlichkeit unter diesen Bedingungen einen Effekt zu finden beträgt 37,44%. Der Forscher konnte also einen Effekt mittlerer Größe mit seiner Anzahl Versuchspersonen nur mit einer sehr geringen Wahrscheinlichkeit entdecken.

Literatur

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NY: Erlbaum.

Lehmann, E. L. (1975): *Nonparametrics: Statistical methods based on ranks*. San Francisco, CA: Holden-Day.