

Simultane U-Tests zur exakten Prüfung von Haupt- und Wechselwirkungen an 2^2 -faktoriellen Versuchsplänen

VON HANS-PETER KRÜGER, Erlangen–Nürnberg

Zusammenfassung, Summary, Résumé

Ausgehend von einem Ansatz zur nichtparametrischen Prüfung von Wechselwirkungen (BREDENKAMP, 1974), der nur unter asymptotischen Bedingungen gilt, wird empfohlen, auf Wechselwirkung in dem praktisch wichtigen 2^2 -faktoriellen Versuchsplan exakt dadurch zu prüfen, daß man die Diagonalzellen überkreuzend zusammenfaßt und mittels U-Test (Cross-Over-U-Test) auf Lageunterschiede prüft. Zugleich kann durch Zeilen- und/oder Spaltenzusammenfassungen auf Hauptwirkungen geprüft werden. Drei simultane Tests können sodann eine parametrische ANOVA aussageinferentiell voll ersetzen! Die Durchführung der Tests wird an einem Beispiel illustriert. Zusätzlich wird auf die Möglichkeit der Prüfung auf Wechselwirkungen höherer Ordnung in 2^n -faktoriellen Versuchsplänen hingewiesen.

Simultaneous U-tests for the exact examination of main and interaction-effects of 2^2 factorial experimental designs

According to the results of non-parametric interaction testing (BREDENKAMP, 1974) an attempt is made to test exactly the interaction in a 2^2 factorial design in which a-symptotic testing sensu Bredenkamp is in appropriate. Assuming A+ and A- as modalities of factor A and B+ and B- as modalities of factor B, cross-over pooling A+B+ A-B- versus A+B- A-B+ results in a two sample U-test for differences in location which is equivalent to a test for interaction effects between A and B. In a similar way rows and columns may be pooled thus helping to develop exact tests for main effects of A and B. Thus ANOVA may be substituted by three simultaneous U-tests. The respective procedure is illustrated by a numerical example and may be generalized on 2^n factorial designs. (L. Canders)

Tests U simultanés assurant un contrôle exact des effets principaux et réciproques dans des systèmes d'expériences 2^2 factoriels

Partant d'une esquisse de contrôle non-paramétrique d'effets réciproques (BREDENKAMP, 1974) valable uniquement dans des conditions non asymptotiques, on recommande de contrôler avec exactitude les effets réciproques dans le système d'expériences 2^2 factoriel, de grande importance pratique, en unissant les cellules diagonales «en croix» et en utilisant le test U (Cross-Over-U-Test) pour contrôler les différences de situation. On peut en même temps contrôler les effets principaux en unissant

les lignes ou les colonnes. Trois tests simultanés peuvent alors remplacer parfaitement, par interférence d'informations, une ANOVA paramétrique. On illustre d'un exemple l'application des tests. On indique en outre la possibilité d'éprouver des effets réciproques d'un ordre plus élevé dans des systèmes d'expériences à 2^n facteurs. (J. Chanel)

Fragestellung und Begründung

BREDENKAMP (1974) hat jüngst in dieser Zeitschrift einen Ansatz zur nonparametrischen Prüfung von Wechselwirkungen vorgestellt, der dem längst fälligen Desiderat genügt, auf Wechselwirkungen zwischen zwei oder mehr als zwei Faktoren auch dann zu prüfen, wenn parametrische Voraussetzungen zur Prüfung mittels Varianzanalyse nicht erfüllt sind und/oder nicht überprüft werden können. Die empfohlenen Tests sind Verallgemeinerungen des H-Tests von KRUSKAL und WALLIS (1952) und des X_r^2 -Tests von FRIEDMAN (1937) und also asymptotische, nur auf größere Stichproben anzuwendende Rang-Tests. Trotz ihrer Asymptotik sind sie von hohem forschungspraktischen Nutzen, da nichtparametrische Schnell-Tests zur Prüfung von Wechselwirkungen in faktoriellen Versuchsplänen nur für diskretisierte bzw. dichotomierte Variablen vorliegen (LIENERT und KRAUTH, 1974 a und b) und wirksame, insbesondere exakte Tests für rangtransformierte Variablen noch fehlen!

Einer der wichtigsten faktoriellen Versuchspläne ist der 2×2 oder 2^2 -faktorielle Plan, in welchem 2 Faktoren zu je w Stufen (Modalitäten, Graduationen) auf Haupt- und Wechselwirkungen geprüft werden sollen, z. B. zwei Behandlungen (Lehrprogramme A_1 und A_2) in 2 Schichten von V_{pn} (gute, B_1 und schlechte Schüler, B_2), oder die Wirkung zweier Chargen (A_1 = Tranquilizer gegen A_2 = Placebo) in zwei Situationen (B_1 mit, B_2 ohne Stress). Der 2×2 -Plan ist nicht nur der einfachste, sondern auch der ökonomischste multifaktorielle Versuchsplan, da er ceteris paribus die wenigsten V_{pn} (V_{te} , P_{tn}) erfordert, um die Frage, ob eine Wechselwirkung zwischen den beiden Faktoren A und B besteht, nachzuweisen (welche Frage oft wichtiger ist als die nach eventuell bereits bekannten Hauptwirkungen der einzelnen Faktoren).

Gerade der in praxi so bedeutsame 2×2 -Versuchsplan hat nun die geringsten Chancen, asymptotische Bedingungen zu erfüllen, wenn man davon ausgeht, daß der H-Test in diesem Fall nach BREDENKAMP erst zu vertreten ist, sofern die 2×2 Zellen mit je $n \geq 10$ Individuen besetzt sind. Nicht einmal ein Schulklassenversuch kann heutzutage diese Bedingung mehr erfüllen, da kaum noch Klassen mit 40 oder mehr Schülern existie-

ren. Selbst wenn man von der oft unvermeidlichen „experimentellen Mortalität“ absieht, die bedingt, daß man bei Ausfall von beispielsweise 2 Vpn in einer einzigen Zelle alle übrigen 3 Zellen ebenfalls um 2 – nach Los auszuwählenden – Vpn zu reduzieren hat, um die Orthogonalität des Versuchsplans zu wahren, sind $n \geq 10$ Vpn pro Zelle zu gewährleisten.

Wie prüft man nun exakt in einem 2×2 -Design auf Wechselwirkungen, wenn die Zellen schwächer als gefordert besetzt sind, oder wenn Versuchspersonen oder deren Ergebnisse planwidrig ausfallen? Im folgenden wird versucht, auf intuitiver Basis einen Zugang zu einer exakten Prüfung auf Wechselwirkung zu eröffnen. Dabei gehen wir davon aus, daß in dem Spezialfall je zweier Stufen eines Faktors der H-Test durch den U-Test von MANN und WHITNEY (1947) ersetzt werden kann, wenn die Voraussetzungen für eine ANOVA (Intervallskalierung, Normalverteilung und Varianzhomogenität der Meßwerte in den 2×2 -Zellen) nicht erfüllt sind oder nicht abgeklärt werden können.

Rationale und Testansatz

Wir gehen bei unserem Ansatz von der Nullhypothese aus, wonach in einem 2×2 -Versuchsplan weder Hauptwirkungen der beiden Faktoren A und B noch eine Wechselwirkung zwischen ihnen vorliegt.

- H_1 : Unter dieser Alternative zu H_0 prüfen wir auf Hauptwirkung des Faktors A mit $a=2$ Stufen mittels des U-Tests für zwei unabhängige Stichproben, um etwaige Lageunterschiede zwischen A_1 und A_2 entgegen H_0 zu entdecken; wir poolen hierbei über die 2 Stufen des Faktors B, d. h. wir werfen die B-Stufen zusammen.
- H_2 : Sodann prüfen wir auf Hauptwirkung des Faktors B, indem wir über die Stufen von A poolen, ebenfalls mittels des U-Tests.
- H_3 : Schließlich prüfen wir – und das entspricht der Auswertung eines Cross-Over-Designs (COCHRAN und COX, 1957, Kap. 4) auf Wechselwirkungen zwischen A und B, indem wir zellenüberkreuzend poolen, d. h. die homonymen Zellen $A_1 B_1$ mit $A_2 B_2$ zusammenwerfen und die heteronymen Zellen $A_1 B_2$ mit $A_2 B_1$ zusammennehmen (Cross-Over-U-Test).

Damit haben wir $r=3$ simultane U-Tests am gleichen Datenmaterial durchzuführen. Das bedeutet, daß die 3 Prüfgrößen U nach einem adjustierten Alpha-Risiko α^* zu beurteilen sind. Da die 3 simultanen U-Tests nun aber nicht unabhängig sind (wegen der Zellenüberkreuzung), ist $\alpha^* = \alpha/3$ zu setzen (KRAUTH und LIENERT, 1973) statt $\alpha^* = 1-(1-\alpha)^3$ nach BREDEKAMP (1974). Die adjustierte Schranke U_{α^*} entnehmen wir der kumulierten Prüfverteilung von U, wie sie etwa bei OWEN (1962,

Tab. 11.3), JACOBSON (1963) oder KRÜGER & WALL (1977) vertafelt sind. (Falls a^+ dort nicht angeführt ist, nehme man die zu dem nächstniedrigen aufgeführten a^+ gehörige Schranke U_{a^+} !).

Je nachdem, ob unter den 3 Alternativen zu H_0 die Richtung der Hauptwirkungen und/oder der Wechselwirkung spezifiziert wurde oder nicht, ist einseitig mit U_{a^+} oder zweiseitig mit $U_{a^+ / 2}$ zu prüfen¹!

Testentscheidung

1. Sind Test 1 und/oder Test 2 signifikant, nicht aber Test 3, dann bestehen nur Hauptwirkungen in A und/oder B, die stets als Lageunterschiede zwischen den korrespondierenden Faktorstufen zu interpretieren sind.
2. Ist weder Test 1 noch Test 2 signifikant, wohl aber Test 3, dann besteht eine von Hauptwirkungen freie Wechselwirkung, die einer doppelten disordinalen Wechselwirkung i.S. BREDEKAMPS entspricht.
3. Ist der Test 1 und zugleich der Test 3 signifikant, nicht aber der Test 2, so besteht eine von der Hauptwirkung A überlagerte Wechselwirkung, die meist als disordinale Wechselwirkung interpretiert wird. Analoges gilt für eine durch B überlagerte Wechselwirkung, wenn Test 2 und 3 signifikant sind, nicht aber Test 1.
4. Sind alle drei Tests signifikant, dann besteht eine von beiden Hauptwirkungen A und B überlagerte Wechselwirkung, die meist als ordinale Wechselwirkung i. S. BREDEKAMPS zu interpretieren sein wird.

Abbildung 1 gibt die Idealfälle der vier möglichen Entscheidungen in bekannter Weise anschaulich wieder:

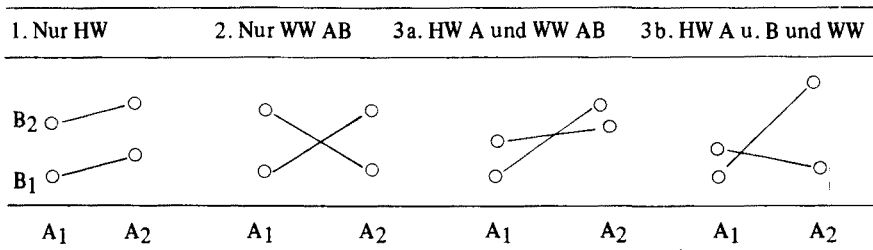


Abb. 1: Mögliche Entscheidungen des Cross-Over-U-Tests im 2x2-Design
(HW = Hauptwirkung, WW = Wechselwirkung)

1 Die Möglichkeit, einseitig zu prüfen, ist übrigens in Bredenkamps H-Test nicht gegeben, und so bedeutet die Anwendung (simultaner) U-Tests anstelle (simultaner) H-Tests in jedem Fall eine Verschärfung, und zwar auch dann, wenn die Zellenbesetzung (mit $n > 10$) für den H-Test ausreichen.

Entscheidungsinterpretation

Wie sind nun die vier möglichen Testergebnisse zu interpretieren? Die Entscheidung 1 bietet keine Interpretationsschwierigkeiten, da hier nur die beiden Hauptwirkungen (oder eine davon) als generelle Wirkungen (behandlungsbedingt bei A, schichtungsbedingt oder situationsbedingt bei B) zu deuten sind.

Die Entscheidung 2 bietet (paradoxerweise) ebenfalls keine Interpretationsschwierigkeiten, da hier eine hauptwirkungsfreie Wechselwirkung vorliegt: Die Behandlung A wirkt in den beiden Schichten von B in gleichem Ausmaß genau gegensinnig. Die Behandlung wirkt nur differentiell, nicht generell!

Die Entscheidung 3 verbindet eine Hauptwirkung A mit einer Wechselwirkung. a) Ist diese Wechselwirkung vom ordinalen Typ, wie in Abb. 1, dann darf die Hauptwirkung trotz der sie überlagernden Wechselwirkung als generelle interpretiert werden. b) Ist die Wechselwirkung vom disordinalen Typ (B_2 von A_1 nach A_2 fallend statt steigend wie B_1), dann darf die Hauptwirkung von A nicht generell, sondern nur differentiell (schichtspezifisch) interpretiert werden².

Entscheidung 4 ist u. U. am schwierigsten zu interpretieren: Wir müssen hier – unter Benutzung von BREDENKAMPS Wechselwirkungs-Terminologie – vier Möglichkeiten unterscheiden:

a) Eine ordinale Wechselwirkung, wie sie in Abb. 1 dargestellt worden ist, erlaubt es, beide Hauptwirkungen (A und B) zu interpretieren und zwar als generelle Wirkungen.

b) Eine einfach-disordinale Wechselwirkung bezüglich A, wie sie entsteht, wenn B_1 von A_1 nach A_2 flach abfiele (statt leicht anstiege) und

c) eine einfach-disordinale Wechselwirkung bezüglich B, wie sie entsteht, wenn die flach steigende B_1 -Verbindung (zwischen A_1 und A_2) so weit nach oben parallel verschoben würde, daß sich beide Verbindungen kreuzen. Bei b) wäre die Hauptwirkung von A, bei c) die Hauptwirkung von B, obschon jeweils nachgewiesen, nicht (generell) zu interpretieren.

d) Würde eine flach fallende (statt steigende) B_1 -Verbindung in gleicher Weise nach oben parallel verschoben, dann entsteht eine doppelt-disordinale Wechselwirkung i.S. BREDENKAMPS. Hier darf weder die Hauptwirkung von A noch die von B (generell) interpretiert werden, sondern – wie auch in den Fällen b) und c) – nur die differentielle Wirkung von A und die differentielle Wirkung von B.

2 Ob der Fall a) oder der Fall b) vorliegt, sollte nicht nur inspektiv-deskriptiv, sondern inferentiell entschieden werden: Dabei wird sich allerdings ein „Weder-Noch“-Fall c) ergeben, auf den hier nicht einzugehen ist.

Abbildung 2 macht die 4 Fälle Wechselwirkungs-überlagerter Hauptwirkungen anschaulich sichtbar und begründet intuitiv den Verzicht auf die Interpretation statistisch nachgewiesener Hauptwirkungen.

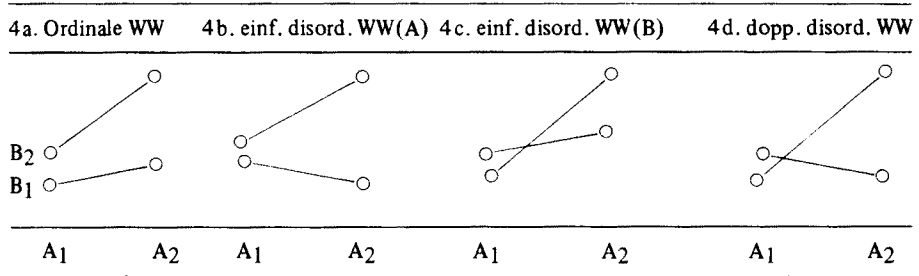


Abb. 2: Möglichkeiten der Wechselwirkungs-überlagerter Hauptwirkungen zweier 2-stufiger Faktoren – WW = Wechselwirkung

Im folgenden wird das Vorgehen an einem Beispiel eines 2x2-Versuchsplans mit einer Behandlungs- und einem Schichtungsfaktor illustriert und theoriebezogen (einseitig) geprüft.

Beispiel aus der Psychopharmakologie

N=20 Vpn wurden einem Vortest (KLT) unterworfen und nach ihrer Vortestleistung in eine leistungsstarke (L+) und eine leistungsschwache Gruppe (L-) mediandichotomiert. Jede der beiden Schichtungsgruppen wurde nach Los halbiert und in eine Versuchsgruppe (B+) und eine Kontrollgruppe (B-) unterteilt. Die Versuchsgruppe erhielt vor dem Schlafengehen ein mildes Barbiturat (0,2 g Cyclohexylbarbitursäure = Medomin^R), die Kontrollgruppe ein Placebo. Am Morgen des nächsten Tages wurde ein Haupttest mit allen 20 Vpn durchgeführt und die Leistungssteigerung im Haupttest X_h in Prozent der Vortestleistung X_v bestimmt:

$$X = 100(X_h - X_v) / X_v .$$

Tabelle 1 führt die n=5 X-Werte der 2 x 2 = 4 Zellen des Versuchsplanes und die dazugehörigen Rangwerte R (von 1 bis 20), sowie die Rangsummen R_{11} , R_{12} , R_{21} und R_{22} auf (adaptiert aus LIENERT, 1954).

Tabelle 1

Faktorstufen	B+			B-			
	X	R		X	R		
L+	+53	20		+17	9		T ₁
	+36	17		+24	13		
	+43	19		+29	14		
	+31	15	R ₁₁	+ 9	7	R ₁₂	
	+39	18		-12	1		
		89			44	133	
L-	+20	10		+34	16		T ₂
	- 3	2		+12	8		
	+ 7	5	R ₂₁	+ 9	6	R ₂₂	
	+22	11		+ 5	4		
	+ 4	3	31	+25	12	46	
		T ₁ = 120			T ₂ = 90	210	

Die zu prüfenden Hypothesen folgen aus DÜKERS (1956) Theorie der reaktiven Leistungssteigerung: Danach sollen unmerkliche Beeinträchtigungen, wie sie durch ein Hang-Over der Schlafmittelwirkung vermutet werden, zu einer kompensatorischen Leistungssteigerung führen. Die theoriebedingte erste Alternative H_1 zur Nullhypothese (keine Behandlungswirkung) impliziert $R(B+) > R(B-)$ oder verbalisierter: Die Barbituratgruppe zeigt höhere Leistungsrangwerte als die Placebogruppe. Da aufgrund vorgängiger Erfahrungen zu bezweifeln ist, daß DÜKERS Theorie generell gilt und vermutet wurde, daß sie nur für leistungsstarke Vpn gilt, wurde im Versuchsplan eine entsprechende Schichtung eingeführt. Die Wirkung des Schichtungsfaktors L selbst interessiert nicht, wohl aber seine Wechselwirkung mit dem Behandlungsfaktor B. Daher lautet die empiriebedingte zweite Alternative $H_2: R(B+L+, B-L-) > R(B+L-, B-L+)$.

Beide Alternativen sind einseitig und sollen wegen (bekanntermaßen) nicht-normalverteilter Zuwachsraten mittels BREDENKAMPS H-Test geprüft werden. Wegen zu kleiner Besetzungszahlen ($n=5$) wird jedoch die Methode der simultanen und exakten U-Tests (Behandlungs-U-Test und Wechselwirkungs-U-Test) vorgezogen.

Vor dem Versuch wurde das Signifikanzniveau $\alpha = 0,05$ vereinbart, und weil $r = 2$ Hypothesen zu prüfen sind, gilt $\alpha^* = 0,05/2 = 0,025$. Die einseitige U-Schranke für $n_1 = n_2 = 2n = 10$ lesen wir bei SIEGEL (1956, Tab. K, S. 276) zu $U = 23$ ab. Wird diese Schranke erreicht oder unterschritten, ist H_0 zugunsten von H_1 bzw. H_2 abzulehnen! Die U-Größen für unsere 2

Tests erhalten wir aus den Rangsummen: Bei gleichen Stichprobenumfängen von $n_1 = n_2 = 2n$ gilt (SIEGEL, 1956, Formel 6.7 a,b)

$$U = \text{Min. } (U_1, U_2),$$

wobei $U_1 = 4n^2 - n(2n+1) - R_1$ und $U_2 = 4n^2 - n(2n+1) - R_2$.

Für den Behandlungs-U-Test resultiert daher $U = U_1 = 4 \cdot 5^2 + 5(11) - 120$ (vgl. Tabelle 1) $= 35 > 23 = U_{a^+}$, so daß ein genereller Effekt im Sinne von DÜKERS Theorie (H_1) nicht nachzuweisen ist.

Für den Wechselwirkung-U-Test ergibt sich ein $U = U_1 = 4 \cdot 5^2 + 5(11) - (89+46) = 20 < 23 = U_{a^+}$, so daß ein differentieller Effekt i. S. der Theorie (H_2) nachzuweisen ist: Die reaktive Leistungssteigerung ist gemäß der Vorhersage unter H_2 nur bei den leistungsstarken (L^+), nicht bei den leistungsschwachen V_{pn} (L^-) anzutreffen³.

Testvarianten

(1) Im vorstehenden Beispiel war ein randomisierbarer Behandlungsfaktor A mit einem nicht-randomisierbaren Schichtungsfaktor L (organismischer Faktor nach EDWARDS, 1971, S. 8) kreuzklassifiziert, in welchem Fall nur die beiden genannten Alternativen (Hauptwirkung von A und WW von AL) in Frage kamen. Sind beide Faktoren (A und B) randomisiert, dann können $r = 3$ Alternativen mittels simultaner U-Tests geprüft werden (Hauptwirkung von A, Hauptwirkung von B und Wechselwirkung von AB).

Solches wäre in unserem Beispiel sinnvoll, wenn die Gültigkeit von DÜKERS Theorie von dem Intervall zwischen Verabreichung des Barbiturates und dem Zeitpunkt des kritischen Tests (Haupttests) abhängig angenommen worden wäre. In diesem Fall hätte man einen 2-stufigen Intervallfaktor (morgens testen, mittags testen) einzuführen und auszuwerten gehabt.

(2) Interessiert im Fall zweier randomisierter Faktoren nur die globale Frage, ob überhaupt irgendeine Behandlungswirkung nachzuweisen ist, gleichgültig aus welchen Wirkungskomponenten, Hauptwirkung und/oder Wechselwirkung sie sich zusammensetzt, dann prüft man mittels des exakten H-Tests für $k = 4$ unabhängige Stichproben am schärfsten (SIE-

3 Die differentielle Interpretation des Dükerschen Effektes hätte aber auch dann gegolten, wenn außer dem Wechselwirkungs-U-Test auch der Hauptwirkungs-U-Test signifikant ausgefallen wäre (was bei größerem n je Zelle hätte erreicht werden können): Denn die grafische Darstellung der R_{ij} gemäß Abbildung 2 ergibt eine einfache disordinale Wechselwirkung bezüglich des Behandlungsfaktors.

GEL, 1956, Table 0), schärfer jedenfalls als mittels dreier simultaner U-Tests wie oben unter (1)⁴.

(3) Hat der Untersucher eine theoretisch begründete Hypothese darüber, in welcher Steigerungsreihe die $k = 4$ unabhängigen Stichproben aufeinanderfolgen (Trendhypothese bzgl. Lokation), dann prüft er noch schärfer mittels des exakten S_j -Tests von JONCKHEERE (vgl. LIENERT, 1973, Kap. 6.2.4.): In unserem Beispiel hätte etwa theoretisch postuliert werden können, daß unter H_1 folgende Steigerungsreihe zu erwarten sei: $M+L+ > M+L- > M-L+ > M-L-$. Gegen diese Alternative hätte der S_j -test relativ „versagt“, da die tatsächlich beobachtete Rangordnung eine andere ist als die unter H_1 postulierte Rangordnung der Rangsummen je Zelle!

(4) Falls die Observablen grob gruppiert sind (z. B. binäre Variablen) kann ohne Informationsverlust statt des U-Tests bei großem Stichproben- n nach SUTCLIFFE (1957) über eine asymptotische X^2 -Zerlegung oder bei kleinem n über exakte Wahrscheinlichkeiten nach MYERS (1958) geprüft werden.

(5) Wie unser Beispiel schon impliziert, ist es möglich, den U-Test auch auf Differenzwerte anzuwenden und damit auf die Originalmeßwerte zu verzichten. Das kann der Fall sein z. B. bei Untersuchungen über Behandlungswirkungen, wo dann als Observablen die Differenzen nachher-vorher eingehen. Explizit vorgeschlagen wurde ein solches Vorgehen von BUCK (1975).

Verallgemeinerung auf 2^n -faktorielle Pläne

In 2^n -faktoriellen Versuchsplänen ergeben sich als prüfbare Wirkungen

- ($\overset{n}{1}$) Hauptwirkungen
- ($\overset{n}{2}$) Wechselwirkungen 1. Ordnung
- ($\overset{n}{3}$) Wechselwirkungen 2. Ordnung
- \vdots
- ($\overset{n}{n}$) Wechselwirkungen ($n-1$)ter Ordnung

4 Merkwürdigerweise wird auf die Möglichkeit, kreuzklassifizierte Faktoren ebenso wie singuläre k -stufige Faktoren mittels des H-Tests zu prüfen, in der nonparametrischen Testliteratur nicht explizit hingewiesen, obschon solch ein Hinweis wegen des Mangels faktorieller Auswertungsmethoden sehr nützlich wäre! Selbstverständlich können auch trifaktorielle Tests, z. B. ein 2^3 -faktorieller Versuchsplan als unifaktorieller mit $k = 2^3$ nach dem H-Test ausgewertet werden, wenn man auf die „Varianzquellenidentifikation“ verzichtet bzw. sie nur heuristisch aufweist.

Diese Wirkungen können völlig analog dem beschriebenen Verfahren über das Poolen der entsprechenden Gruppen geprüft werden. In einem 2^3 -Versuchsplan sind das

$\binom{3}{1}$ = 3 Hauptwirkungen, nämlich A, B, C

$\binom{3}{2}$ = 3 Wechselwirkungen 1. Ordnung, nämlich AxB, AxC und BxC und

$\binom{3}{3}$ = 1 Wechselwirkung 2. Ordnung, nämlich AxBxC

Zur Verdeutlichung des Vorgehens führen wir als Notation für die je zwei Faktorstufen die Vorzeichen „+“ und „-“ ein. Im 2^3 -Fall entstehen so 8 Faktorkombinationen (und damit Versuchsgruppen), die in Tabelle 2 aufgeführt sind.

Tabelle 2

Gruppe Nr.	A	B	C	WW	A x B x C
1	+	+	+	+	X + X + = +
2	+	+	-	+	X + X - = -
3	+	-	+	+	X - X + = -
4	+	-	-	+	X - X - = +
5	-	+	+	-	X + X + = -
6	-	+	-	-	X + X - = +
7	-	-	+	-	X - X + = +
8	-	-	-	-	X - X - = -

Zur Berechnung der Haupt- und Wechselwirkungen müssen folgende Gruppen gepoolt werden:

HW A	= 1+2+3+4	gegen	5+6+7+8
HW B	= 1+2+5+6	gegen	3+4+7+8
HW C	= 1+3+5+7	gegen	2+4+6+8
WW AxB	= 1+2+7+8	gegen	3+4+5+6
WW AxC	= 1+3+6+8	gegen	2+4+5+7
WW BxC	= 1+4+5+8	gegen	2+3+6+7
WW AxBxC	= 1+4+6+7	gegen	2+3+5+8

Die Frage, welche Gruppen bei den Wechselwirkungen gepoolt werden müssen, ist einfach über das Durchmultiplizieren der Vorzeichen zu entscheiden, wie es in Tabelle 2 für die Wechselwirkung AxBxC angegeben ist. Dabei sind die Gruppen mit dem Multiplikationsergebnis „+“ den anderen mit dem Ergebnis „-“ gegenüberzustellen. Selbstverständlich müssen alle Zellen (= Faktorkombinationen) mit der gleichen Zahl von Beobachtungen besetzt sein. Dieses n kann aber beim U-Test auch sehr klein sein.

Die α -Adjustierung hat, wie beschrieben, nach der Zahl der durchgeführten simultanen Tests zu geschehen. Werden im 2^3 -Fall alle Wirkungen geprüft, ergibt sich ein $\alpha = \alpha/7$.

Literatur

- BREDEKAMP, J.: Nonparametrische Prüfung von Wechselwirkungen. Psychol. Beitr. 16 (1974), 398–416.
- BUCK, W.: Paardifferenzen-U-Test, ein verteilungsfreier Wirkungs-Vergleich zweier Behandlungen. Arzneimittelforschung 25 (1975), 825–827.
- COCHRAN, W. G., und Gertrude M. COX: Experimental designs. 2. Auflage. New York: Wiley 1957.
- DÜKER, H.: Über reaktive Anspannungssteigerung. Z. exp. u. ang. Psych. 10 (1963), 46–72.
- EDWARDS, A. L.: Versuchsplanung in der psychologischen Forschung. (Deutsch von H. Huber und W. Strehse). Weinheim: Beltz 1971.
- FRIEDMAN, M.: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J. Amer. Statist. Assoc. 32 (1937), 675–701.
- JACOBSON, J. E.: The Wilcoxon Two sample statistic: Tables and Bibliography. J. Amer. Statist. Assoc. 58 (1963), 1086–1103.
- JONCKHEERE, A. R.: A distribution free k sample test against ordered alternatives. Biometrika 41 (1954), 133–145.
- KRAUTH, J., und G. A. LIENERT: Nichtparametrischer Nachweis von Syndromen durch simultane Binomialtests. Biom. Zeitschr. 15 (1973), 13–20.
- KRÜGER, H. P. und K. D. WALL: Statistische Tafeln für Sozial- und Biowissenschaftler. 5 Bde. Weinheim: Beltz, 1977
- KRUSKAL, W. H., und W. A. WALLIS: Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. 47 (1952), 583–621.
- LIENERT, G. A.: Verteilungsfreie Methoden in der Biostatistik. Bd. 1. 2. Auflage. Meisenheim: Hain 1973.
- , und J. KRAUTH: Konfigurationsfrequenzanalyse: IX. Auswertung multivariater klinischer Untersuchungspläne. Z. Klin. Psychol. u. Psychother. 22 (1974), 3–17.
- MYERS, J. L.: Exact probability treatments of factorial designs. Psych. Bull. 55 (1958), 59–61.
- OWEN, D. B.: Handbook of Statistical Tables. Reading/Mass.: Addison Wesley 1962.
- SIEGEL, S.: Nonparametric Statistic for the behavioral Sciences. New York: McGraw-Hill 1956.
- SUTCLIFFE, J. P.: A general method of analysis of frequency data for multiple classification designs. Psych. Bull. 54 (1957), 134–137.

Dr. Hans-Peter Krüger
 Fachbereich Erziehungs- und Kulturwissenschaften
 der Universität Erlangen-Nürnberg
 Regensburger Straße 160
 D 8500 Nürnberg