

Nichtparametrischer Vergleich von Testprofilen und Verlaufskurven bei unabhängigen Stichproben

VON HANS-PETER KRÜGER und HERBERT BUCHTA, Würzburg

Zusammenfassung, Summary, Résumé

Werden k Testwerte an zwei oder mehreren Stichproben von P_b n erhoben, stellt sich das Problem des Vergleichs von Testprofilen. Ähnliche Fragen treten auf, wenn statt k verschiedenen Testwerten ein Meßwert zu k Zeitpunkten an einem P_b erhoben wird. Zur Auswertung dieser Fragestellung wird ein Vorgehen nach dem WILCOXON-Rangsummentest vorgeschlagen, der es erlaubt (1) Unterschiede aller Art (2) Unterschiede in der Lage und (3) Unterschiede in der Form der Profile zu prüfen. Das Vorgehen wird an einem Beispiel aufgezeigt. Erweiterungen auf andere Fälle werden diskutiert.

A non-parametric comparison of testprofiles and trend curves in independent samples

If k -test scores are collected in two or more samples of subjects the problem of a comparison of test-profiles arises. Similar problems can be observed in those cases in which instead of k -different test-scores a test result of k -times is ascertained from a subject. For purposes of interpretation an approach (Wilcoxon rank sum test) is suggested allowing the examination of 1) differences of all kind, 2) differences of location and 3) differences of the forms of profiles. This approach is demonstrated by means of an example. Possibilities of applying this approach in other cases are discussed
(L. Candors)

Comparison non-paramétrique de profils de tests et de courbes d'évolution dans le cas de sondages indépendants

Si l'on obtient n valeurs sur deux ou plusieurs sondages opérés sur des sujets, le problème se pose de la comparaison des profils de tests. Semblables questions se posent également si au lieu de n valeurs de test différentes on prévêlve sur un sujet une valeur correspondant à n instants. Pour la mise en œuvre des résultats, on propose un procédé inspiré par le test de sommes de WILCOXON, permettant d'établir des différences de toute espèce, des différences de location et des différences de forme des profils. Un exemple illustre le procédé. On discute d'une extension possible à d'autres cas.
(J. Chanel)

CRONBACH und GLESER (1953) haben für die Untersuchung von Meßwertprofilen eine Liste von interessierenden Fragen zusammengestellt. Darin nimmt der Vergleich der Profile zweier Gruppen von Pbn, die sich in Behandlungs- oder Schichtungsvariablen unterscheiden, eine zentrale Stellung ein. Die Hinweise der Autoren zur Beantwortung solcher Fragestellungen basieren auf parametrischen Voraussetzungen, die in praxi oft schwer zu erfüllen sind. Für den Fall des Vergleichs paariger Stichproben (z. B. Prä-Post-Vergleiche) haben jüngst LEHMACHER und LIENERT (1980) eine robuste nonparametrische Auswertungsmethode über simultane WILCOXON-Tests vorgestellt. Im folgenden werden diese Überlegungen auf den Fall des Vergleichs von Profilen bei zwei und mehreren unabhängigen Stichproben angewendet.

Fragestellung und Vorgehen an einem Beispiel

BUCHTA (1980) untersuchte in einer psychometrischen Studie die Unterschiede zwischen Stotterern und Dysgrammatikern/Stammlern an Kindern einer Sprachheilschule zwischen 5 und 8 Jahren. Aus diesen Daten wurden 2 Stichproben mit je $N_1 = N_2 = 15$ Pbn entnommen, deren Ergebnisse in $k = 4$ Einzeltests in Tab. 1 dargestellt sind. Aus substanzwissenschaftlichen Gründen wurden folgende 4 Tests ausgewählt:

- a) BZ: Bilder zuordnen – visuell-motorische Assoziation zur Prüfung der nichtverbalen Repräsentation
- b) OF: Objekte finden – Wahrnehmungsgeschwindigkeit zur Prüfung der nichtverbalen Integration
- c) WV: Wortverständnis – Prüfung der verbalen Repräsentation
- d) ZF: Zahlenfolgegedächtnis – Akustisch-sprachliche Automatik der Sprache zur Prüfung der verbalen Integration

Wie aus der Testbeschreibung bereits hervorgeht, liegt der Auswahl ein zweidimensionales Modell der Sprachbehinderung zugrunde, wobei die Faktoren Repräsentation–Integration und nichtverbal–verbal zu unterscheiden sind. Die Ergebnisse der Pbn wurden nach den entsprechenden Testvorschriften in T-Werte ($T = 50 + 10z$) umgewandelt. Abbildung 1 zeigt die Mittelwertprofile der beiden Gruppen sowohl für die Einzeltests wie für die beiden Dimensionen.

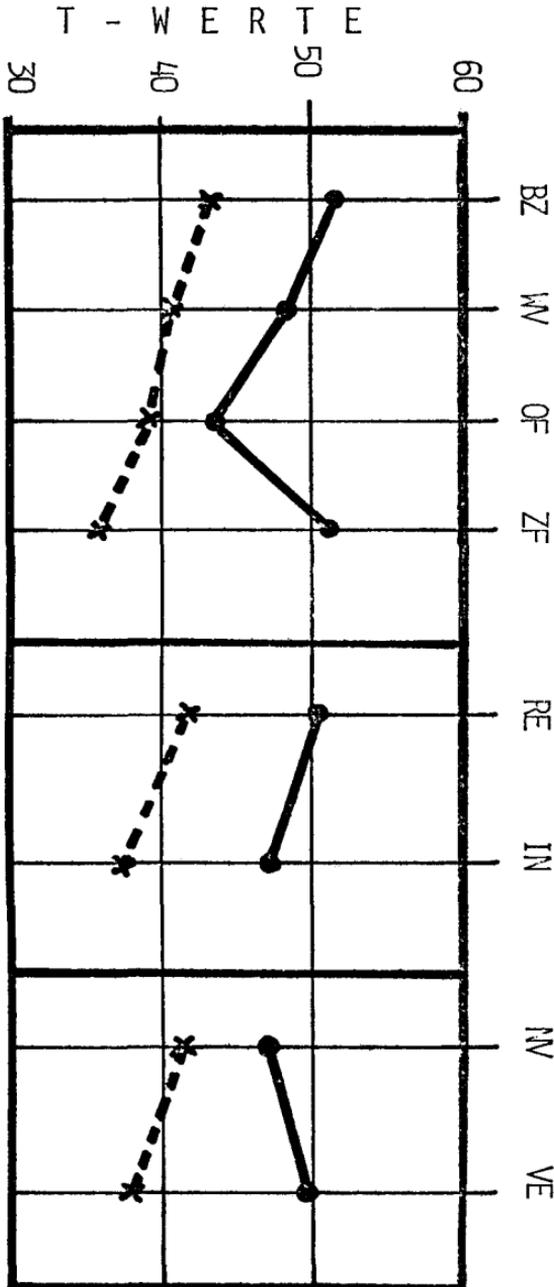


Abb. 1: Die durchgezogene Linie gibt das Profil der Stotterer, die gestrichelte die der Dysgrammatiker/Stammer wieder. BZ = Bilder zuordnen, WV = Wortverständnis, OF = Objekte finden, ZF = Zahlenfolgen, RE = Repräsentation, IN = Integration, NV = nonverbal, VE = verbal. Eingetragen sind die T-Werte.

Beim Vergleich dieser Profile sind folgende Fragestellungen möglich:

1. Omnibus-Alternative: unterscheiden sich die beiden Gruppen in einem oder mehreren Subtests? Hier werden über simultane Einzeltests sowohl Lage- wie Formunterschiede der Profile abgeprüft.
2. Lage-Alternative: unterscheiden sich die beiden Gruppen im Gesamtniveau des Profils?
3. Form-Alternative: Ist die Form der Profile unabhängig von der Lage der Profile zwischen den Gruppen unterschiedlich?

Alternativen 2 und 3 verlangen, daß

- die Einzeltests gleich skaliert sind und daß
- eine inhaltlich zu vertretende physio- oder pathotrope Polung der Einzeltests vorgenommen wurde (etwa: hoher Score = hohe Ausprägung des Merkmals).

Sind diese Voraussetzungen nicht zu erfüllen, ist nur Alternative 1 abzuprüfen.

Alle 3 Fragestellungen sind multivariater Natur und könnten durch entsprechende multivariate Verfahren parametrischer oder nicht-parametrischer Art angegangen werden (für die multivariate ANOVA siehe etwa AHRENS und LÄUTER, 1974, für multivariate Rangverfahren PURI und SEN, 1971). In der Regel sind die Ergebnisse aus diesen Analysen jedoch schwierig zu interpretieren. Praktikabler sind meist mehrere, auf die spezielle Fragestellung abgestimmte simultane Tests, die eine direkte Lokalisierung der Unterschiede erlauben und damit die Interpretation wesentlich erleichtern.

Alle 3 Alternativen können bei Vorliegen dieser Voraussetzungen durch die Bildung entsprechender Summen und Differenzen aus den Daten geprüft werden. Jeweils zugrundegelegt wird bei den simultanen Tests das Rationale des WILCOXON-Rangsummen-Tests (1945) für unabhängige Stichproben (für eine Beschreibung siehe LIENERT, 1973, 230 ff.). Dazu werden die Meßwerte beider Gruppen in eine gemeinsame Rangreihe gebracht und pro Gruppe die Rangsumme bestimmt. Der Erwartungswert für diese Rangsummen T_1 und T_2 ergibt sich aus

$$\mu_T = N_1 (N + 1)/2$$

N_1 bezeichnet dabei den Umfang der ersten, N_2 den Umfang der zweiten Stichprobe. N ergibt sich aus $N_1 + N_2$.

Die Standardabweichung beträgt

$$\sigma_T = \sqrt{N_1 N_2 (N + 1)/12}$$

Asymptotisch ist T normalverteilt, so daß für genügend große Stichprobenumfänge mit

$$z = \frac{T - \mu_T}{\sigma_T}$$

bzw. mit Stetigkeitskorrektur für $N \leq 60$ mit

$$z = \frac{|T - \mu_T| - 1/2}{\sigma_T}$$

in die Tafel der Normalverteilung (0,1) eingegangen werden kann. Haben die beiden Stichproben unterschiedliche Umfänge, wird $T = T_1$ (Rangsumme der kleineren Stichprobe) gesetzt.

Die Alternativhypothese ist substanzwissenschaftlich begründet ein- oder zweiseitig abzufassen. Da mehrere simultane Tests am gleichen Material durchgeführt werden, ist die Irrtumswahrscheinlichkeit α nach dem BONFERONI-Prinzip entsprechend zu adjustieren (siehe KRAUTH und LIENERT, 1973). Bei r simultanen Tests ergibt sich ein adjustiertes α° von

$$\alpha^\circ = \alpha/r$$

Dieses Vorgehen ist konservativ: das adjustierte α° als nominales Niveau ist immer kleiner, höchstens gleich dem faktischen α .

Treten in den Meßwertreihen Bindungen auf (ties), sind diese durch Bildung von mittleren Rängen aufzufangen. Entsprechend verringert sich die Varianz der Prüfgröße. Nach dem Vorschlag von RAATZ (1966a, 1966b) bildet man mit f_i die Häufigkeit, in der jeder einzelne Rang auftritt. Liegt eine Zweierbindung vor, wird $f_i = 2$ usf. Die reduzierte Standardabweichung von T ergibt sich dann aus

$$\sigma_{T \text{ corr}} = \sqrt{\frac{N_1 N_2}{12 N(N-1)} (N^3 - \sum f_i^3)}$$

Liegen keine Bindungen vor, werden alle $f_i = 1$ und die Formel verkürzt sich zur oben angegebenen. Das Verfahren bei vielen Bindungen wird zum Abschluß diskutiert.

Die Überprüfung der Omnibus-Hypothese

Für jeden der $k = 4$ Untertests werden die Ränge über alle $N = N_1 + N_2 = 30$ Pbn gebildet und anschließend die Rangsummen T_1 und T_2 berechnet. Die Ränge für den ersten Untertest BZ sind der Spalte 7 der Tabelle 1 zu entnehmen. Die Rangsumme der Stotterer ist $T_1 = 292$, die der Dysgrammatiker ist $T_2 = 173$.

Tabelle 1

T-Werte der Pbn in den einzelnen Untertests und deren Ränge. BZ = Bilder zuordnen, OF = Objekte finden, WV = Wortverständnis, ZF = Zahlenfolgen. Gesamt = Summe aus allen Untertests, Diff = Differenz.

Pb	BZ	OF	WV	ZF	BZ		GESAMT		BZ-OF		(BZ+WV) - (OF+ZF)		
					x	Rang	Σ	Rang	Diff	Rang	Diff	Rang	
01	38	20	43	25	38	4.5	126	1	+18	26	+36	30	
02	43	38	30	43	43	8.5	154	10	+ 5	15	- 8	4	
03	60	37	47	52	60	29	196	21	+23	28.5	+18	22.5	
04	45	59	57	54	45	12	215	29	-14	1	-11	2	
05	50	42	53	66	50	18.5	211	26.5	+ 8	17.5	- 5	7.5	R
06	43	54	51	44	43	8.5	192	20	-11	2	- 4	9.5	ER
07	53	49	44	55	53	24.5	201	22	+ 4	12.5	- 7	5.5	RR
08	51	30	58	65	51	21	204	23	+21	27	+14	21	ER
09	53	45	41	36	53	24.5	175	12	+ 8	17.5	+13	19.5	ER
10	51	51	38	48	51	21	188	18.5	0	9	-10	3	ER
11	71	54	51	50	71	30	226	30	+17	24.5	+18	22.5	OR
12	59	45	53	48	59	28	205	24	+14	23	+19	24.5	ER
13	49	53	55	55	49	16.5	212	28	- 4	6	- 4	9.5	S
14	57	33	57	60	57	27	207	25	+24	30	+21	26	
15	50	42	53	56	50	18.5	211	26.5	+ 8	17.5	- 5	7.5	
16	38	33	52	23	38	4.5	146	8	+ 5	15	+34	29	
17	51	57	42	31	51	21	181	15	- 6	5	+ 5	15.5	
18	34	36	45	24	34	2	139	4	- 2	7	+19	24.5	
19	49	41	51	46	49	16.5	187	17	+ 8	17.5	+13	19.5	
20	40	48	49	44	40	7	181	15	- 8	4	- 3	11	
21	45	46	25	24	45	12	140	5	- 1	8	0	12	
22	37	34	45	21	37	3	137	3	+ 3	10.5	+27	28	
23	39	48	48	46	39	6	181	15	- 9	3	- 7	5.5	
24	53	36	33	26	53	24.5	148	9	+17	24.5	+24	27	
25	45	34	30	36	45	12	145	6.5	+11	22	+ 5	15.5	
26	32	28	28	44	32	1	132	2	+ 4	12.5	-12	1	
27	53	50	42	43	53	24.5	188	18.5	+ 3	10.5	+ 2	13	
28	47	24	28	46	47	14	145	6.5	+23	28.5	+ 5	15.5	
29	44	39	45	39	44	10	167	11	+ 5	15	+11	18	
30	48	39	44	48	48	15	179	13	+ 9	21	+ 5	15.5	D
Spalte	2	3	4	5	6	7	8	9	10	11	12	13	

Der Erwartungswert beträgt

$$\mu_T = 15(30 + 1)/2 = 232.5$$

Es liegen $f_1 = 12$ Ränge ohne Bindungen vor, $f_2 = 4$ Zweierbindungen (bei 4.5, 8.5, 16.5 und 18.5), $f_3 = 2$ Dreierbindungen (bei 12 und 21) und $f_4 = 1$ Viererbindung (bei 24.5). Die $\sum f_i^3$ beträgt damit

$$\sum f_i^3 = 12 \cdot 1^3 + 4 \cdot 2^3 + 2 \cdot 3^3 + 1 \cdot 4^3 = 162$$

Die korrigierte Standardabweichung ist

$$\sigma_{T_{\text{corr}}} = \sqrt{\frac{15 \cdot 15}{12 \cdot 30 \cdot (30 - 1)} (30^3 - 162)} = 24.050$$

Da $N < 60$ ist das stetigkeitskorrigierte

$$z = \frac{|292 - 232.5| - 0.5}{24.050} = 2.453$$

Insgesamt werden $r = 4$ Einzeltests durchgeführt. Legt man ein Signifikanzniveau von $\alpha = 0,05$ zugrunde, ergibt sich ein adjustiertes $\alpha^\circ = 0,05/4 = 0,0125$. Aus theoretischen Erwägungen heraus wird hier eine einseitige Fragestellung formuliert, wonach Stammler/Dysgrammatiker schlechter abschneiden als Stotterer. Der entsprechende z-Wert liegt damit bei $z = 2.24$. Die Nullhypothese gleichen Abschneidens ist damit zu verwerfen.

Entsprechend diesem Vorgehen wurden auch die übrigen Einzeltests abgeprüft. Es ergeben sich folgende z-Werte:

für BZ: $z = 2.453$

für OF: $z = 1.287$

für WV: $z = 2.410$

für ZF: $z = 3.325$

Außer für den Untertest OF ist damit H_0 in allen Fällen zurückzuweisen.

Ist der Untersucher weniger am Ergebnis der Einzeltests interessiert, sondern mehr an den Leistungen in bestimmten Bereichen, kann bei Vorliegen der entsprechenden Skalierung die Summe aus Einzeltests gebildet und auf gleiche Weise abgeprüft werden. Im Beispiel lassen sich für die einzelnen Dimensionen der Sprachbehinderung folgende Summen bilden:

BZ + WV als Repräsentationsleistungen

($z = 3.328$)

OF + ZF als Integrationsleistungen	(z = 3.282)
BZ + OF als nichtverbale Leistungen	(z = 2.057)
WV + ZF als verbale Leistungen	(z = 3.382)

Auch hier ist eine α -Adjustierung für $r = 4$ simultane Tests durchzuführen. H_0 ist lediglich für die nichtverbalen Leistungen beizubehalten.

Die Überprüfung der Lagehypothese

Für die Entscheidung, ob Stotterer insgesamt besser abschneiden als Dysgrammatiker, bildet man die Σx_i über alle k Einzeltests. Diese werden wieder über den WILCOXON-Test geprüft.

Daten und Ränge gehen aus Spalten 8 und 9 der Tabelle 1 hervor. Wie bei den Einzeltests wird eine einseitige Alternative formuliert, wonach Stotterer besser abschneiden als Dysgrammatiker. Es ergibt sich ein z -Wert von $z = 3.466$, der zur Ablehnung von H_0 führt.

Die Überprüfung der Formhypothese

Die unterschiedliche Form eines Profils drückt sich unabhängig von der Lage der Profilpunkte darin aus, daß die Differenzen zwischen aufeinanderfolgenden Punkten sich ändern. Zwei Profile sind damit dann unterschiedlich, wenn zwischen zwei entsprechenden Profilpunkten sich die Differenzen unterscheiden. Da bei Profilen in der Regel die Anordnung der Profilpunkte beliebig ist (anders als bei Verlaufskurven in der Zeit), bildet man bei k Profilpunkten alle $\binom{k}{2}$ Differenzen pro Pb. Diese Differenzen werden dann als Grundlage für die WILCOXON-Testprozedur genommen.

Für den Vergleich der Subtests BZ–OF wurden in Spalte 10 der Tabelle 1 die Differenzen gebildet, denen in Spalte 11 die Ränge zugeordnet wurden. Ein nach dem obigen Verfahren berechneter z -Wert von $z = 0.997$ zeigt an, daß beide Gruppen sich in der Form des Profils in diesen Subtests nicht unterscheiden.

Da $k = 4$ Tests vorliegen, können insgesamt $\binom{4}{2} = 6$ Einzeltests angesetzt werden. Hypothesen für die Richtung des Unterschieds liegen nicht vor. Das adjustierte $\alpha^\circ = \alpha/6 = 0.05/6 = 0.0083$ ergibt bei zweiseitiger Fragestellung ein $z = 2.65$. Die Einzel-Formvergleiche erbringen

BZ–OF: nonverbal Repräsentation-Integration	$z = 0.997$
BZ–WV: Repräsentation verbal-nonverbal	$z = 0.145$

BZ–ZF: Repräsentation nonverbal-Integration verbal	$z = 1.702$
OF–WV: Integration nonverbal-Repräsentation verbal	$z = 0.644$
OF–ZF: Integration nonverbal-verbal	$z = 2.076$
WV–ZF: verbal Integration-Repräsentation	$z = 1.558$

Keiner der z -Werte übersteigt die gesetzte Schranke, so daß die Hypothese gleicher Form der Profile beibehalten werden kann. Allerdings zeigt sich unter der heuristischen Betrachtungsweise ein deutlicher Unterschied bei der Differenz OF–ZF. Er ist so zu interpretieren, daß bei den Dysgrammatikern auf der Integrationsstufe der Unterschied zwischen nonverbaler und verbaler Leistung wesentlich größer ist als bei Stotterern. Hier scheint der eigentliche Unterschied zwischen beiden Gruppen zu liegen.

Ähnlich wie bei den Lagetests können diese Formvergleiche auch für die beiden Dimensionen Repräsentation–Integration und nichtverbal–verbal durchgeführt werden. In Spalte 12 der Tabelle 1 wurde (BZ + WV)–(OF + ZF) gebildet. Damit wird verglichen, ob der Unterschied zwischen Repräsentation und Integration in den beiden Gruppen verschieden ist. Es ergibt sich ein z -Wert von $z = 0.727$. Die Prüfung (BZ + OF)–(WV + ZF) zur Bestimmung des Unterschieds zwischen verbaler und nichtverbaler Leistung ergibt ein $z = 1.287$. Beide Werte sind zu einem adjustierten $\alpha^\circ = \alpha/2 = 0.025$ nicht signifikant.

Diskussion und Erweiterungen

Der Vergleich von unabhängigen Profilen über simultane WILCOXON-Tests erweist sich in der Praxis als handliche Methode, die wenig Anforderungen an die Dignität der Daten stellt. Der Nachteil ist darin zu sehen, daß über die α -Adjustierung bei vielen simultanen Tests schnell extreme α° resultieren, die nur über große Stichprobenumfänge überhaupt erreichbar werden. Zwei Vorgehensweisen sind dabei zu empfehlen:

- a) es werden nur jene Profildifferenzen abgeprüft, für die substanzwissenschaftliche Hypothesen bestehen. Der Einzeltest wird aus einem inhaltlichen Modell heraus begründet. Die Erfahrung zeigt, daß damit die Zahl der Tests drastisch gesenkt wird.
- b) LEHMACHER und LIENERT (1980) empfehlen ein zweistufiges Vorgehen. Die Stichproben werden in 2 Zufallshälften geteilt. An der ersten Hälfte werden heuristisch alle Tests durchgeführt. Inferentiell werden an der 2. Stichprobe nur jene Tests durchgeführt, die bereits in der ersten deutliche Effekte gezeigt haben. Durch die Möglichkeit der einseitigen Fragestellung können maximal scharfe Testprozeduren eingebracht werden.

Liegen mehr als 2 Stichproben vor, ist der H-Test nach KRUSKAL-WALLIS (1952, siehe LIENERT, 1973, 263 ff.) indiziert. Liegen Hypothesen vor, daß die Stichproben ordinal sind (in unserem Beispiel etwa Dysgrammatiker – Stotterer – Normale) kann schärfer über den JONCKHEERE-Test (1954) geprüft werden. Treten sehr viel Bindungen auf, wird die Berechnung der Standardabweichung nach dem RAATZ-Verfahren sehr aufwendig. Einfacher ist hier die Testgröße S nach KENDALL (1970) zu handhaben, die von SILLITTO (1947) für einreihige, von BURR (1960) für zweireihige Rangbindungen beschrieben wurde (siehe zu den Verfahren LIENERT, aaO., 608 ff.). Üblicherweise werden diese Techniken zu den Korrelationsverfahren gezählt. Sie führen aber auch bei Homogenitätsprüfungen zum gleichen Ergebnis.

Für Schnelltests empfiehlt sich die Mediandichotomierung der Meßwerte bzw. deren Differenzen. Die angegebenen Prüfungen reduzieren sich dann auf Mediantests, die einfach über die Tafeln von KRÜGER, LEHMACHER und WALL (1981) auszuwerten sind.

Soll untersucht werden, ob sich die Streuungen der Meßwerte in den einzelnen Gruppen unterscheiden, kann dies ebenfalls auf Rangbasis geschehen, wenn das Vorgehen von WESTENBERG (siehe dazu LIENERT, 1973) angewendet wird.

Selbstverständlich sind die Überlegungen dieses Artikels auch auf die Auswertungen von Verlaufskurven anzuwenden (k Messungen eines Meßobjekts pro Pb). Die Omnibus-Prüfung verläuft völlig analog. Bei der Formverschiedenheit ist es bei solchen Zeitreihen jedoch angebracht, nicht mehr alle $\binom{k}{2}$ Prüfungen durchzuführen, sondern lediglich die Differenzen der aufeinanderfolgenden Zeitpunkte zu bilden (Folgedifferenzen).

Eine ausführliche Diskussion des Vorgehens bei der Auswertung von Verlaufskurven findet der Leser in LEHMACHER (1980).

Literatur

- AHRENS, H. und LÄUTER, J.: Mehrdimensionale Varianzanalyse, Berlin: Akademie-Verlag, 1974
- BUCHTA, H.: Teilleistungsschwächen bei sprachbehinderten Kindern. Dissertation Würzburg 1981.
- BURR, E. J.: The distribution of Kendall's score S for a pair of tied rankings. *Biometrika*, 47, 1960, 151–171
- CRONBACH, J. L. und GLEESER, Goldine C.: Assessing similarity between profiles. *Psych. Bull.*, 50, 1963, 456–473
- JONCKHEERE, A. R.: A distribution-free k -sample test against ordered alternatives. *Biometrika*, 41, 1954, 133–145

- KENDALL, M. G.: Rank correlation methods. 4th ed. London: Griffin, 1970
- KRAUTH, J. und LIENERT, G. A.: Nichtparametrischer Nachweis von Syndromen durch simultane Binomialtest. *Biometrische Zeitschrift*, 15, 1973, 13–20
- KRÜGER, H.-P., LEHMACHER, W. und WALL, K.-D.: Statistische Tafeln für Sozial- und Biowissenschaftler. Band I: Die Vierfeldertafel. Stuttgart: Fischer, 1981
- KRUSKAL, W. H. und WALLIS, W. A.: Use of ranks in one-criterion variance analysis. *J. Am. Stat. Ass.*, 47, 1952, 583–621
- LEHMACHER, W.: Nonparametrische Auswertung von Verlaufskurven. Erscheint im Tagungsband des Kongresses für „Klinische Psychologie und Psychotherapie“ 1980 in Berlin.
- LEHMACHER, W. und LIENERT, G. A.: Nichtparametrischer Vergleich von Testprofilen und Verlaufskurven vor und nach einer Behandlung. Erscheint in „Psychologische Beiträge“
- LIENERT, G. A.: Verteilungsfreie Methoden in der Biostatistik. 2. Aufl. Band 1. Meisenheim: Hain, 1973
- PURI, M. L. und SEN, P. K.: Nonparametric methods in multivariate analysis. New York: Wiley, 1971
- RAATZ, U.: Eine Modifikation des White-Tests bei großen Stichproben. *Biometrische Zeitschrift*, 8, 1966a, 42–54
- RAATZ, U.: Wie man den White-Test bei großen Stichproben ohne die Verwendung von Rängen durchführen kann. *Archiv für die gesamte Psychologie*, 118, 1966b, 86–92
- SILLITTO, G. P.: The distribution of Kendall's τ coefficient in rankings containing ties. *Biometrika*, 34, 1947, 36–44
- WILCOXON, F.: Individual comparisons by ranking methods. *Biometrics* 1, 1945, 80–83

Anschrift der Verfasser:

Prof. Dr. H.-P. Krüger
und Dipl. Psych. H. Buchta
Institut für Psychologie
der Universität Würzburg
Domerschulstraße 13
8700 Würzburg