Verteilungsfreie Testverfahren

Stefan Götz
Universität Wien und Akademisches Gymnasium Wien

Zusammenfassung

Aus verschiedenen Gründen kann es passieren, daß die Testverfahren, die derzeit in den gängigen Schulbüchern vorgestellt werden, in der Praxis nicht angewandt werden können. Als — übrigens lehrplankonforme — Alternative zu diesen sogenannten „Parametertests" bieten sich „verteilungsfreie Testverfahren" an, die unter gewissen Bedingungen und Einschränkungen universeller einsetzbar sind. Es sollen verschiedene derartige Verfahren vorgestellt werden (mit Beispielen!). Dabei wird erkennbar werden, daß einige dieser Tests (wegen ihres elementaren Aufbaus) durchaus sinnvoll in die siebente Klasse passen würden. Dies umso mehr, da das „Testen mit der Binomialverteilung" in der siebenten Klasse ohnehin zum Teil unter vielen technischen Details leidet, die überdies oft mit der eigentlichen Idee des Testens im Grunde nur wenig zu tun haben. Ohne letzteres negativ zu beurteilen, soll einfach eine brauchbare Alternative besprochen werden.

1 Einführung

In der siebenten und achten Klasse der AHS steht *„Beurteilende Statistik"* im Lehrplan. Sie wird meist in die Themenbereiche *„Testen von Hypothesen"* und *„Schätzen von Parametern"* unterteilt. Ersterem wollen wir uns widmen. In den gängigen Schulbüchern wird in der siebenten Klasse mit der Binomialverteilung, in der achten Klasse mit der Normalverteilung getestet. Diese Testverfahren können verallgemeinert und ausgeweitet werden (t-Tests, χ^2-Tests, F-Tests), und bilden die große und wichtige Gruppe der *„Parametertests"*.

Wir wollen nun vier Gründe angeben, die das Anwenden dieser Parametertests nicht legitim erscheinen lassen: *Erstens* kann es passieren, daß die zugrundeliegende Verteilungsfunktion nicht allen Anforderungen genügt, die ein Parametertest fordert. So verlangt z.B. der t-Test zum Vergleich der Erwartungswerte zweier (unabhängiger) Stichproben, daß erstens letztere einer normalverteilten Grundgesamtheit entnommen wurden, und zweitens, daß die Varianzen, so sie nicht bekannt sind, wenigstens gleich sind. Es stellt sich die Frage nach der sog.

„Robustheit" eines Testverfahrens, d.h. wie empfindlich ist ein Test gegenüber Abweichungen von den eigentlich geforderten Voraussetzungen? (Z.B. ist der F-Test empfindlicher als der t-Test.)

Zweitens: Das erhobene Merkmal hat nicht kardinales Meßniveau. Wir unterscheiden im wesentlichen drei Merkmalstypen. Bei *nominalskalierten* Daten werden den verschiedenen Merkmalsausprägungen verschiedene Ziffern zugeordnet. Diese Zuordnung ist bijektiv, ansonsten beliebig. Beispiele wären Blutgruppen, Geschlechter,... An sinnvollen statistischen Größen können nur Modalwerte und relative Häufigkeiten berechnet werden.

Können die erhobenen Daten sinnvoll gereiht werden, so liegen *ordinalskalierte* Daten vor. Zu beachten ist dabei, daß die Abstände zwischen den einzelnen Ausprägungen nicht gleichbedeutend sein müssen. Eine Differenzenbildung (bei Zuordnung reeller Zahlen — ebenfalls geordnet) ist daher nicht unbedingt zulässig. Das klassische Beispiel hiefür sind die Schulnoten, aber auch Erdbebenstärken,... Zusätzlich zu den bei den nominalskalierten Daten angeführten Größen können jetzt auch schon Quantile (Spezialfall: Median) berechnet werden.

Das höchste Meßniveau stellen *kardinalskalierte* Daten dar, z.B. Längen, Massen,... Mit ihnen kann gerechnet werden. Daher können zusätzlich zu den bisher erwähnten statistischen Größen das arithmetische Mittel, die Varianz und bei zweidimensionalen Daten der Korrelationskoeffizient sinnvollerweise erhoben werden.

Drittens müssen die zugrundeliegenden Stichprobenvariablen X_i ($i = 1, \ldots, n$; wobei n der Stichprobenumfang sei), welche ihrer Natur nach Zufallsvariablen (ZV) sind, nicht alle *einer* Verteilungsfunktion genügen oder sie sind nicht unabhängig voneinander. Z.B. kann ein Merkmal bei einem Patienten vor (ZV X) und nach (ZV Y) einer bestimmten Behandlung erhoben werden. X und Y sind sicher nicht voneinander unabhängig.

Viertens werden oft zur Rechtfertigung der Anwendung von Parametertests Grenzwertsätze herangezogen, die naturgemäß nur für große Stichprobenumfänge n brauchbare Näherungslösungen liefern. Diese großen n sind oft aus Kosten- oder Zeitgründen nicht realisierbar.

Diese (und noch andere) Aspekte lassen nach Alternativen zu den Parametertests suchen, wie sie z.B. die sog. *„Verteilungsfreien Testverfahren"* darstellen. Ihr Name rührt daher, daß die Verteilung der Stichprobenvariablen nicht (vollständig) bekannt sein muß, die Testverteilung (zur Berechnung der kritischen Werte) natürlich schon. Diese Testverfahren sind daher universeller einsetzbar[1]. Einige sollen in dieser Arbeit (mit Beispielen!) vorgestellt werden. Es wird sich zeigen, daß ihr Aufbau *elementar* ist und sie daher auch bezüglich ihres Schwierigkeitsgrades durchaus in die siebente bzw. achte Klasse passen.

[1] Allerdings hat das auch seinen Preis, wie wir noch sehen werden.

2 Zur Stellung im Lehrplan

Im Lehrplan der siebenten Klasse findet man im Abschnitt „Testen und Schätzen" das Stichwort „nichtparametrisches Schätzen", welches im zugehörigen Lehrplankommentar von *Günther Malle* unter der Überschrift „Verteilungsunabhängige Testverfahren" (Nichtparametrische Testverfahren) an Hand eines Beispiels näher ausgeführt wird.

3 Spezielle Vorbemerkungen

Für soziologische oder psychologische Untersuchungen werden oft Tests herangezogen, deren Ergebnis eine Punkteanzahl ist. Fassen wir diese Ergebnisse (von n Versuchspersonen geliefert) als Daten auf, so erhalten wir (wenigstens) *ordinales* Meßniveau mit *stetigen* Merkmalsausprägungen. Letzteres soll bedeuten, daß die Wahrscheinlichkeit, zwei gleiche Daten zu erhalten, Null ist. Dies ist durch Verfeinerung der Auswertung stets zu erreichen.

Wir benötigen nun folgende

DEFINITION: Es liegen n paarweise verschiedene Daten vor: $x_1, \ldots, x_n$. Wir bilden den Vektor $(x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ mit

$$x_{(1)} < x_{(2)} < \ldots < x_{(n)} .$$

Die Abbildung

$$R: \{x_1, \ldots, x_n\} \mapsto \{1, \ldots, n\}$$

mit

$$R(x_i) = j ,$$

wobei x_i in obigem Vektor den j-ten Platz einnimmt, heißt *Rangabbildung*. Der Vektor $(R(x_1), \ldots, R(x_n))$ heißt *Rangstatistik*.

Bemerkungen:

- Der Übergang von der tatsächlichen Stichprobe zur Rangstatistik bedeutet natürlich einen *Informationsverlust*.
- Durch Rundungen der Ergebnisse kann es passieren, daß $x_i \neq x_j \; \forall i \neq j$ $(1 \leq i, j \leq n)$ verletzt wird. Es kommen also gleiche Daten vor, wir sprechen von *Bindungen*. Wie werden jetzt die Ränge vergeben? Diese Frage führt uns zur *Methode der Durchschnittsrangbildung*, die an einem Beispiel demonstriert werden soll:

Beispiel:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
5	7	7	2	3	5	6	4	5	8

Wir stellen eine Zweier-Bindung ($x_2 = x_3 = 7$) und eine Dreier-Bindung ($x_1 = x_6 = x_9 = 5$) fest. Durch Ordnen erhalten wir

$x_{(i)}$	$x_4 <$	$x_5 <$	$x_8 <$	$x_1 =$	$x_6 =$	$x_9 <$	$x_7 <$	$x_2 =$	$x_3 <$	x_{10}
$R(x_{(i)})$	1	2	3	?	?	?	7	?	?	10

Für x_1, x_6 und x_9 stehen die Ränge 4, 5 und 6 zur Verfügung. Da $x_1 = x_6 = x_9$ gilt, wird über diese Ränge gemittelt:

$$\frac{4+5+6}{3} = 5 ,$$

und dieser Durchschnittsrang wird sowohl x_1 als auch x_6 und x_9 zugeteilt. Ebenso für $x_2 = x_3$:

$$\frac{8+9}{2} = 8.5 ,$$

somit erhalten wir die Rangstatistik

$$(5, 8.5, 8.5, 1, 2, 5, 7, 3, 5, 10) .$$

Einen letzten Gesichtspunkt im Rahmen dieser Vorbemerkungen wollen wir zur Sprache bringen: Bei kardinalskalierten Daten kann unter gewissen Voraussetzungen sowohl ein Parametertest als auch ein verteilungsfreies Testverfahren angewendet werden. Das bedeutet, daß zwei unterschiedliche Verfahren miteinander in Konkurrenz treten. Wie können sie miteinander verglichen werden?

Dazu betrachten wir die sog. *Gütefunktion*, die in Abhängigkeit von dem Wert des Parameters θ, über den die zu testende Hypothese formuliert wurde, die Wahrscheinlichkeit angibt, daß die Testvariable T in den *Ablehnungsbereich* K fällt:

$$g(\theta) = P(T \in K)_\theta .$$

Zum Beispiel:

$$H_0 : \theta \leq \theta_0$$
$$H_1 : \theta > \theta_0$$

Es gilt: $g(\theta_0) = \alpha$, wobei α (wie üblich) die Irrtumswahrscheinlichkeit für einen Fehler 1. Art (i.e. H_0 zu Unrecht ablehnen) bedeutet[2]. β ist die (jeweilige) Irrtumswahrscheinlichkeit für einen Fehler 2. Art (i.e. H_0 zu Unrecht beibehalten):

$$g(\theta') = 1 - \beta' \qquad \forall \theta' \in (\theta_0, \infty)$$

[2] Diese Bezeichnung behalten wir die ganze Arbeit hindurch bei.

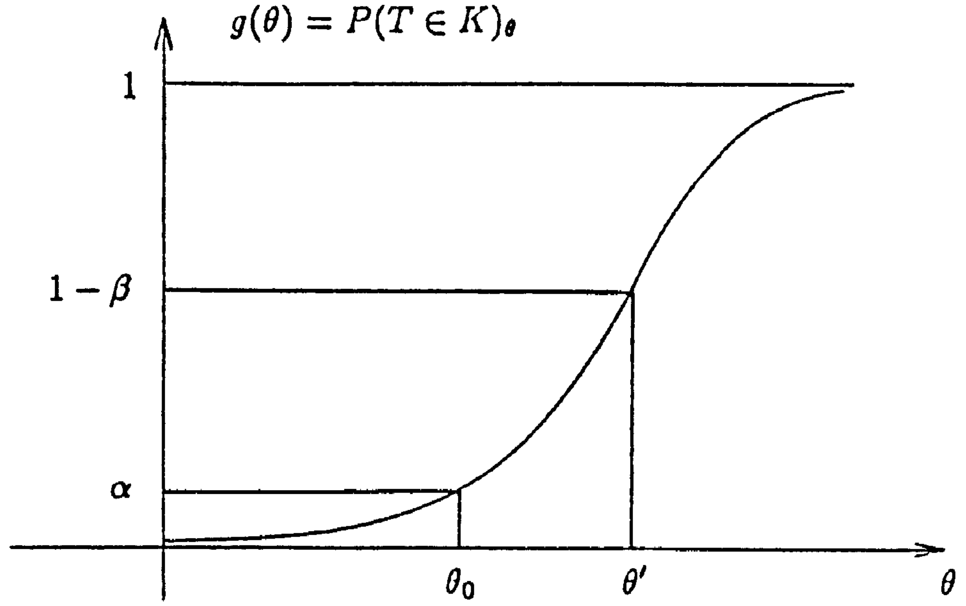

Abbildung 1: Skizze der Gütefunktion

in unserem Beispiel.
Man will nun, daß $g(\theta)$ im Bereich (θ_0, ∞) möglichst rasch nach θ_0 Werte nahe bei 1 erreicht (siehe auch Abbildung 1). Je steiler der Anstieg von $g(\theta)$ hier ausfällt, desto „schärfer" ist der Test, d.h. desto kleiner wird β bei vorgegebenem α für einen bestimmten Parameterwert $\theta' \in (\theta_0, \infty)$. Eine Verbesserung dieser Schärfe ist — bei gegebenem Problem und Testverfahren — i.a. nur durch eine Erhöhung des Stichprobenumfangs n erreichbar.

Damit wird folgende DEFINITION einsichtig: Für ein gegebenes Problem sei der Test A der mit der größten Güte. Der Test B mit dem Stichprobenumfang n_B besitze diesselbe Güte wie der Test A mit dem Stichprobenumfang n_A.
Dann heißt

$$\frac{n_A}{n_B} \cdot 100\%$$

relative Effizienz (RE) und

$$\lim_{\substack{n_A \to \infty \\ n_B \to \infty}} RE$$

asymptotische relative Effizienz (ARE) des Tests B.

Bemerkungen:

- $0 < RE \leq 1$ und $0 < ARE \leq 1$.
- Für kleine n zeigt sich, daß verteilungsfreie Testverfahren eine hohe Effizienz besitzen, daher liegt hier ein Hauptanwendungsgebiet dieser Verfahren.

4 Konkrete Testverfahren

4.1 Ein Test für nominalskalierte Daten: der Binomialtest

Beginnen wir mit einem

BEISPIEL: Eine bestimmte Behandlungsmethode A gegen Rheuma heilt erfahrungsgemäß mit einer Wahrscheinlichkeit p_0 von 0.3. Eine neuentwickelte Methode B wird an 10 Patienten ausprobiert mit dem Ergebnis, daß 4 von 10 geheilt wurden. Kann eine signifikante (d.h. $\alpha = 5\%$) Verbesserung festgestellt werden?

LÖSUNG: Als Hypothesen formulieren wir

$$H_0 : p \leq p_0 = 0.3$$
$$H_1 : p > 0.3 \; .$$

(Es ist günstig, als Nullhypothese jene Aussage auszuwählen, von welcher man annimmt, daß sie abgelehnt wird. Andernfalls wird man sich nie gegen H_0 aufgrund des Testergebnisses entscheiden können, was keinesfalls gleichbedeutend mit einer Annahme ist.)

Wir lehnen H_0 ab, wenn viele Patienten geheilt werden. Der *Ablehnungsbereich* K hat daher die Form $\{c, c+1, \ldots, 10\}$. Die Bestimmung des *kritischen Wertes* c erfolgt durch *Probieren:*
Es muß gelten:

$$P(X \geq c)_{H_0} \leq 0.05 \; , \tag{1}$$

wobei die ZV X die Anzahl der Geheilten anzeigt. X ist also unter H_0 (darauf deutet der Index in (1) hin) binomialverteilt mit den Parametern $n = 10$ und $p = p_0 = 0.3$. Wir schreiben (1) aus:

$$\binom{10}{c} 0.3^c 0.7^{10-c} + \binom{10}{c+1} 0.3^{c+1} 0.7^{10-c-1} + \ldots + \binom{10}{10} 0.3^{10} 0.7^0 \leq 0.05 \; .$$

Durch Einsetzen für c erkennen wir

$$P(X \geq 6)_{H_0} < 0.05$$

und

$$P(X \geq 5)_{H_0} > 0.05 \; ,$$

was

$$K = \{6, 7, 8, 9, 10\}$$

zur Folge hat. Da $4 \notin K$ ist, dürfen wir H_0 auf dem Signifikanzniveau $\gamma = 0.95$ nicht ablehnen ($\gamma \doteq 1 - \alpha$). $\diamond$

Bemerkung: Warum eigentlich bedeutet „unter der Annahme von H_0", daß $p = p_0 = 0.3$ zur Berechnung von K zu setzen ist? H_0 ließe auch andere Werte für p zu ($H_0 : p \leq p_0 = 0.3$)! Dazu überlege man sich, daß für $p < p_0 = 0.3$ die Wahrscheinlichkeit $P(X \geq c)_p$ sicher kleiner ist als für $p = p_0 = 0.3$. Somit ist unser Ablehnungsbereich der kleinste Ablehnungsbereich und hat für alle p-Werte mit $p \leq 0.3$ (entspricht H_0) Gültigkeit!

Allgemein beschrieben sieht der Binomialtest wie folgt aus: Es liege ein *dichotomes* Merkmal vor, von dem eine Stichprobe $(x_1, \ldots, x_n)$ erhoben wurde. Es handelt sich also um einen sog. *Einstichprobentest.* Die möglichen Ausprägungen der zugehörigen ZV X_i (beschreibt die i-te Stichprobenentnahme; $i = 1, \ldots, n$) seien „0" oder „1". Die X_i seien paarweise unabhängig voneinander.
Als *Hypothesen* formulieren wir $\forall i = 1, \ldots, n$

$$\begin{aligned} H_0 &: P(X_i = 1) \leq p_0 \\ H_1 &: P(X_i = 1) > p_0 \ , \end{aligned}$$

andere Hypothesenformulierungen können analog behandelt werden.
Die *Testvariable* X sei $\sum_{i=1}^{n} X_i$, sie zählt also die „1"-Ausprägungen. X ist unter H_0 binomialverteilt mit den Parametern n und p_0. Der *Ablehnungsbereich*

$$K = \{c, \ldots, n\}$$

zum Signifikanzniveau $\gamma = 1 - \alpha$ kann mit Hilfe von

$$P(X \geq c)_{H_0} \leq \alpha$$

bestimmt werden, wobei c minimal zu wählen ist.
Die *Testvorschrift* lautet: wenn $x \in K$, wird H_0 verworfen und umgekehrt.

Ein interessanter *Spezialfall* tritt dann ein, wenn $p_0 = \frac{1}{2}$ ist. Die Hypothesen lauten dann im Fall eines rechtsseitigen Tests

$$\begin{aligned} H_0 &: P(X_i = 1) \leq \frac{1}{2} \\ H_1 &: P(X_i = 1) > \frac{1}{2} \ , \end{aligned}$$

während die linksseitige Formulierung durch

$$\begin{aligned} H_0 &: P(X_i = 1) \geq \frac{1}{2} \\ H_1 &: P(X_i = 1) < \frac{1}{2} \end{aligned}$$

gegeben ist. Für gegebenes n und α werden die kritischen Werte $c_{1-\alpha}$ und c_α minimal bzw. maximal unter den Bedingungen

$$P(X \geq c_{1-\alpha}) = \sum_{i=c_{1-\alpha}}^{n} \binom{n}{i} \left(\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^n \sum_{i=c_{1-\alpha}}^{n} \binom{n}{i} \leq \alpha$$

bzw.

$$P(X \leq c_\alpha) = \sum_{j=0}^{c_\alpha} \binom{n}{j} \left(\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^n \sum_{j=0}^{c_\alpha} \binom{n}{j} \leq \alpha$$

gewählt.
Ersetzen wir das „$\leq$"-Zeichen durch das Gleichheitszeichen, so folgt

$$\sum_{i=c_{1-\alpha}}^{n} \binom{n}{i} = \sum_{j=0}^{c_\alpha} \binom{n}{j} = 2^n \cdot \alpha .$$

Wegen

$$\binom{n}{n} = \binom{n}{0}$$
$$\binom{n}{n-1} = \binom{n}{1}$$
$$\vdots$$

erkennen wir, daß in beiden Summen gleichviele Summanden stehen müssen, nämlich $n - c_{1-\alpha} + 1$ bzw. $c_\alpha + 1$, woraus wir

$$c_\alpha = n - c_{1-\alpha}$$

ableiten.

Eine *andere Lösungsmöglichkeit* unseres Beispiels besteht darin, die Wahrscheinlichkeit für das beobachtete und noch extremere Ereignisse (unter H_0) auszurechnen:

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - \binom{10}{0} 0.3^0 \cdot 0.7^{10} - \binom{10}{1} 0.3 \cdot 0.7^9 -$$
$$- \binom{10}{2} 0.3^2 \cdot 0.7^8 - \binom{10}{3} 0.3^3 \cdot 0.7^7 \approx 0.35 > \alpha ,$$

d.h. diese Ergebnisse gehören nocht nicht zu den extremsten (die Wahrscheinlichkeit ihres Eintretens liegt über 5%), wir lehnen H_0 daher nicht ab.

In der klassischen Testtheorie ist es allerdings üblich (und gehört dort sozusagen zum guten Ton), das Stichprobenergebnis nicht von vornherein in den Entscheidungsprozeß einfließen zu lassen (so wie das hier der Fall ist), sondern erst den Ablehnungsbereich zu bestimmen und dann zu schauen, ob das Stichprobenergebnis darin enthalten ist oder nicht. Damit soll Manipulationsmöglichkeiten vorgebaut werden.

Im wesentlichen entspricht der Binomialtest dem Testen mit der Binomialverteilung, wie es in den meisten Büchern für die siebente Klasse angeboten wird, allerdings kommt hier die Binomialverteilung in natürlicher Weise ins Spiel und wird nicht apriori vorausgesetzt.

4.2 Tests für ordinalskalierte Daten

4.2.1 Der Iterationstest

Es handelt sich auch hierbei um einen *Einstichprobentest*, der die *Zufälligkeit* des Zustandekommens von Stichproben überprüft. Die Problematik soll an einem einführenden *Beispiel* demonstriert werden:

Ein Experiment werde zehnmal durchgeführt mit folgendem Ergebnis (E...Erfolg, M...Mißerfolg):

$$\underbrace{M}_{1}\underbrace{EEE}_{2}\underbrace{MM}_{3}\underbrace{E}_{4}\underbrace{MM}_{5}\underbrace{E}_{6}\ .$$

DEFINITION: Unter einer *„Iteration"* verstehen wir eine Folge von identen Symbolen, der entweder davon verschiedene Symbole voraus- und nachfolgen oder kein Symbol voraus- oder nachfolgt.

In unserem Beispiel zählen wir also sechs Iterationen.
Zwei Extremfälle können dabei auftreten:

$$\underbrace{MMMMM}_{1}\underbrace{EEEEE}_{2} \quad \text{und vice versa: zwei Iterationen}$$

oder

$$\underbrace{M}_{1}\underbrace{E}_{2}\underbrace{M}_{3}\underbrace{E}_{4}\underbrace{M}_{5}\underbrace{E}_{6}\underbrace{M}_{7}\underbrace{E}_{8}\underbrace{M}_{9}\underbrace{E}_{10}$$

bzw.

$$\underbrace{E}_{1}\underbrace{M}_{2}\underbrace{E}_{3}\underbrace{M}_{4}\underbrace{E}_{5}\underbrace{M}_{6}\underbrace{E}_{7}\underbrace{M}_{8}\underbrace{E}_{9}\underbrace{M}_{10} \quad : \text{zehn Iterationen.}$$

Als *Nullhypothese* formulieren wir

$$H_0\text{: Die Stichprobe ist zufällig entstanden,}$$

was bedeuten soll, daß nicht zu viele und nicht zu wenige Iterationen in der Stichprobe vorkommen.

Die Frage stellt sich nun nach der Verteilung der Anzahl der Iterationen. Auskunft darüber gibt folgender

SATZ. Es liege eine Stichprobe vom Umfang n vor, wobei die Ausprägungen A und B eines dichotomen Merkmals festgestellt wurden: n_1-mal wurde A gezählt, n_2-mal B $(n_1 + n_2 = n)$. Dann besitzt unter H_0 die Anzahl R der Iterationen die Verteilung

$$P(R = 2r) = \frac{2 \cdot \binom{n_1-1}{r-1} \cdot \binom{n_2-1}{r-1}}{\binom{n}{n_1}}$$

bzw.

$$P(R = 2r + 1) = \frac{\binom{n_1-1}{r-1} \cdot \binom{n_2-1}{r} + \binom{n_1-1}{r} \cdot \binom{n_2-1}{r-1}}{\binom{n}{n_1}} .$$

BEWEIS: Die Anzahl der möglichen Fälle beträgt

$$\frac{n!}{n_1!n_2!} ,$$

da wir die möglichen Anordnungen von n_1 A-Zeichen und n_2 B-Zeichen zählen: *Permutation mit Wiederholung.*
Wegen

$$\frac{n!}{n_1!n_2!} = \frac{n!}{n_1!(n-n_1)!} = \binom{n}{n_1}$$

ergibt sich eine alternative Interpretation: Suche n_1 Platznummern aus n Platznummern für die A-Zeichen heraus: *Kombination ohne Wiederholung.*

Sei jetzt $R = 2r$, es liege also eine gerade Anzahl von Iterationen vor. Dann gibt es r A-Iterationen und r B-Iterationen, da jede A-Iteration von einer B-Iteration begrenzt wird und umgekehrt.

Wir fassen nun jede A-Iteration als Kästchen auf (es gibt r davon) und belegen jedes Kästchen mit einem A. Es bleiben also $n_1 - r$ A-Zeichen über. Diese müssen aufgeteilt werden:

	1	2	$\dots$	$n_1 - r - 1$	$n_1 - r$
	A	A	$\dots$	A	A
Kästchennummer (z.B.)	$r - 1$	17		3	4

Kommt in der untersten Zeile eine Nummer k-mal vor, so bedeutet das: das Kästchen mit dieser Nummer enthält $k + 1$ A-Zeichen.

Die zu beantwortende *Frage* lautet nun: wieviele Möglichkeiten der Zuteilung von Kästchennummern gibt es?

Zur *Beantwortung* dieser Frage bemühen wir das *Urnenmodell:* Wir legen die r Kästchennummern in eine Urne und ziehen $n_1 - r$ daraus mit Zurücklegen. Auf die Reihenfolge kommt es dabei nicht an: siehe oben bzw. woher sich eine Iteration ihre A-Zeichen holt, ist gleichgültig. Es liegt also eine *Kombination mit Wiederholung* vor, das sind

$$\binom{(n_1 - r) + r - 1}{n_1 - r} = \binom{n_1 - 1}{n_1 - r}$$

Möglichkeiten.
Wegen

$$\binom{n_1 - 1}{n_1 - r} = \binom{n_1 - 1}{r - 1}$$

gibt es also $\binom{n_1-1}{r-1}$ Möglichkeiten der Anordnung der A-Zeichen, um r Iterationen zustande zu bringen.
Analog gibt es

$$\binom{n_2 - 1}{r - 1}$$

Möglichkeiten für die B-Zeichen.
Insgesamt erhalten wir

$$2 \cdot \binom{n_1 - 1}{r - 1}\binom{n_2 - 1}{r - 1}$$

günstige Möglichkeiten, wobei der Faktor 2 daher rührt, daß jede Anordnung von vorne oder von hinten gelesen werden kann, d.h. mit einem A- oder B-Zeichen beginnend.

Sei jetzt $R = 2r + 1$. Wir unterscheiden zwei Fälle:
1. Fall: Die Anzahl der A-Iterationen sei $r + 1$, die der B-Iterationen ist dann also gleich r. Jede Anordnung beginnt und endet also mit einem A.

Für die A-Zeichen stehen $r+1$ Kästchen bereit, d.h. $n_1 - r - 1$ A-Zeichen können beliebig auf sie verteilt werden. Nach oben ergibt das

$$\binom{(n_1 - r - 1) + (r + 1) - 1}{n_1 - r - 1} = \binom{n_1 - 1}{n_1 - r - 1} = \binom{n_1 - 1}{r}$$

Möglichkeiten.

Für die B-Zeichen gilt analoges, allerdings treten hier nur r Kästchen auf; was

$$\binom{n_2 - 1}{r - 1}$$

Möglichkeiten der Aufteilung (siehe oben) zur Folge hat.

Die Gesamtanzahl der Möglichkeiten für den 1. Fall beträgt daher

$$\binom{n_1-1}{r}\cdot\binom{n_2-1}{r-1}.$$

2. Fall: Die Anzahl der A-Iterationen sei nun r und die der B-Iterationen folglich $r+1$. In der mathematischen Behandlung dieses 2. Falles tauschen im Vergleich zum 1. Fall n_1 und n_2 nur Platz.

Damit ergeben sich *insgesamt* (1. Fall und 2. Fall)

$$\binom{n_1-1}{r}\cdot\binom{n_2-1}{r-1}+\binom{n_1-1}{r-1}\cdot\binom{n_2-1}{r}$$

Möglichkeiten. ∎

Auf Grund dieser Verteilung entsteht eine *Tabelle* der *kritischen Werte* r_α und $r_{1-\alpha}$ für den links- bzw. rechtsseitigen Test:

$$P(R\leq r_\alpha)\leq\alpha \qquad \text{und} \qquad P(R\leq r_\alpha+1)>\alpha$$

bzw.

$$P(R\geq r_{1-\alpha})\leq\alpha \qquad \text{und} \qquad P(R\geq r_{1-\alpha}-1)>\alpha$$

(siehe Tabelle 1 im Anhang). Ist also $r\leq r_\alpha$ bzw. $r\geq r_{1-\alpha}$, so wird H_0 abgelehnt, andernfalls nicht.

Abschließend wollen wir ein BEISPIEL zum Iterationstest vorführen:
Bei einer Verkehrszählung wird von 8 bis 16^{30} Uhr für jede halbe Stunde die Anzahl der vorbeifahrenden Autos festgehalten:
176, 183, 151, 149, 192, 203, 197, 180, 175, 214, 226, 231, 237, 243, 249, 250, 258.
Ist die Reihenfolge der Anzahlen zufällig oder gibt es Tendenzen ($\alpha=10\%$)?

LÖSUNG: Zuerst dichotomisieren wir die kardinalskalierten Daten: wir setzen ein „+", wenn ein Wert kleiner ist als sein Nachfolger, ein „−" im umgekehrten Fall:

$$\underbrace{+}_{1}\underbrace{--}_{2}\underbrace{++}_{3}\underbrace{---}_{4}\underbrace{++++++++}_{5}.$$

Wir stellen nur $r=5$ Iterationen fest, was uns zur Formulierung von

$$H_1\text{: Es sind zuwenige Iterationen}$$

verleitet. Tatsächlich ergibt sich aus Tabelle 1 der kritische Wert

$$r_{0.1}=5 \qquad (n_1=5,\ n_2=11)\ ,$$

und wegen $r\leq r_{0.1}$ lehnen wir H_0 zugunsten von H_1 ab! ◊

4.2.2 Der Vorzeichentest

Dies ist ein *Zweistichprobentest* für *abhängige* Stichproben. Eine typische Situation für seine Anwendung ist jene, in der vor und nach einer „Behandlung" ein Merkmal an n voneinander unabhängigen Personen erhoben wird („vorher"–„nachher"): Sei $(x_1, y_1), \ldots, (x_n, y_n)$ die Realisation der (verbundenen) Stichproben. Wir setzen ein „+", wenn $x_i > y_i$ ist, und ein „−", wenn $x_i < y_i$ ist $(i = 1, \ldots, n)$. Wenn die Behandlung keine Auswirkung zeigt, erwarten wir genausoviele „+" wie „−".

Die *Hypothesen* im Falle eines zweiseitigen Tests lauten:

H_0 : Die Behandlung hat keine Auswirkung.
H_1 : Die Behandlung hat eine Auswirkung.

Bei Vorliegen eines einseitigen Tests ist

H_0: wie oben,
H_1: Die Auswirkung geht in die eine oder andere Richtung.

Sei nun P die ZV, welche die „+"-Zeichen zählt. Unter H_0 ist P dann binomialverteilt mit den Parametern n und $\frac{1}{2}$, was eine Zurückführung auf den Spezialfall des Binomialtests, den wir auch dort besprochen haben, bedeutet. Die *kritischen Werte* sind in *Tabelle 2* angeführt (es ist $P(P \leq c_\alpha) \leq \alpha$ und $P(P \leq c_\alpha + 1) > \alpha$), wobei ja — wie bereits besprochen — $c_{1-\alpha} = n - c_\alpha$ gilt.

Beim zweiseitigen Test lautet dann die *Testentscheidung*: H_0 ablehnen, wenn $p \leq c_{\frac{\alpha}{2}}$ oder $p \geq c_{1-\frac{\alpha}{2}}$ ist; im einseitigen Fall verwerfen wir H_0, wenn $p \leq c_\alpha$ bzw. $p \geq c_{1-\alpha}$ ist.

Bemerkung: Wenn für ein oder mehrere i $x_i = y_i$ ist, so werden diese Bindungen weggelassen, da ihnen weder ein „+"- noch ein „−"-Zeichen zugeordnet werden kann und sie so im Rahmen des Vorzeichentests nicht verwertbar sind. Dies bedeutet natürlich einen Informationsverlust, da n verringert wird.

BEISPIEL: Die folgende Tabelle zeigt die Noten einer bestimmten Klasse auf eine Mathematik- bzw. unmittelbar darauffolgende Physikschularbeit:

Schüler	M(x_i)	Ph(y_i)	Vorzeichen	S	M(x_i)	Ph(y_i)	VZ
1	1	2	−	9	3	4	−
2	3	4	−	10	2	1	+
3	2	4	−	11	3	4	−
4	4	3	+	12	4	5	−
5	5	5	0	13	3	1	+
6	3	5	−	14	2	2	0
7	2	3	−	15	2	3	−
8	1	3	−				

Frage: Ist die Physikschularbeit signifikant schlechter ausgefallen als die Mathematikschularbeit (etwa wegen der unmittelbar vorangegangenen Mathematikschularbeit)?

LÖSUNG: Die Vorzeichenspalte der obigen Tabelle gehört bereits zur Lösung. Wir zählen also 3 Pluszeichen, 2 Bindungen und 10 Minuszeichen ($n = 15$). Wir testen einseitig, der neue Stichprobenumfang n' beträgt $n - 2 = 15 - 2 = 13$. Aus Tabelle 2 lesen wir den kritischen Wert

$$c_{0.05} = 3$$

ab, wegen $p = 3 \leq c_{0.05} = 3$ lehnen wir H_0 zugunsten von H_1: „Die Physikschularbeit ist signifikant schlechter ausgefallen als die Mathematikschularbeit" ab. ◊

Bemerkung: $RE = 95\%$ für $n = 6$, $ARE = 63\%$. Die Zahlen beziehen sich auf den t-Test für verbundene Stichproben. Die geringe ARE ist durch den großen Informationsverlust, der durch die Ignoranz der Größe der Differenzen entsteht, erklärbar.

4.2.3 Der Mediantest

Sei X eine (im herkömmlichen Sinne) stetige ZV mit der Verteilungsfunktion F. Dann heißt jede reelle Zahl

$$\mathcal{M} \qquad \text{mit} \qquad F(\mathcal{M}) = \frac{1}{2}$$

Median von X. (Bei streng monotonem F gibt es nur *einen* Median.)
Daraus folgt

$$P(X < \mathcal{M}) = P(X > \mathcal{M}) = \frac{1}{2} \, .$$

Sei nun $x = (x_1, \ldots, x_n)$ eine Stichprobe vom Umfang n, wobei die zugehörigen (voneinander unabhängigen) ZV X_i $(i = 1, \ldots, n)$ alle so wie X verteilt sind. (Das erhobene Merkmal habe also ausnahmsweise *kardinales* Meßniveau.)

Die *Hypothesen* lauten

$$\begin{aligned} H_0 : \mathcal{M} = \mathcal{M}_0 \quad &\text{bzw.} \quad \mathcal{M} \geq \mathcal{M}_0 \quad \text{oder} \quad \mathcal{M} \leq \mathcal{M}_0 \\ H_1 : \mathcal{M} \neq \mathcal{M}_0 \quad &\text{bzw.} \quad \mathcal{M} < \mathcal{M}_0 \quad \text{oder} \quad \mathcal{M} > \mathcal{M}_0 \, . \end{aligned}$$

Betrachten wir

$$Y \doteq X - \mathcal{M}_0 \, ,$$

dieser neuen ZV entspricht die Stichprobe

$$y = (x_1 - \mathcal{M}_0, \ldots, x_n - \mathcal{M}_0) \, .$$

Auf diese Stichprobe kann der *Vorzeichentest* angewendet werden: Sei Z die ZV, die die positiven Werte in Y zählt. Dann ist Z unter H_0 binomialverteilt mit den Parametern n und $\frac{1}{2}$.

Wir erhalten folgende *Testentscheidungen:*

a) $H_0 : \mathcal{M} = \mathcal{M}_0$ wenn $z \le c_{\frac{\alpha}{2}}$ oder $z \ge n - c_{\frac{\alpha}{2}} = c_{1-\frac{\alpha}{2}}$

$\Longrightarrow$ Entscheidung für $H_1 : \mathcal{M} \neq \mathcal{M}_0$.

b) $H_0 : \mathcal{M} \le \mathcal{M}_0$ wenn $z \ge n - c_\alpha = c_{1-\alpha}$

$\Longrightarrow$ Entscheidung für $H_1 : \mathcal{M} > \mathcal{M}_0$.

c) $H_0 : \mathcal{M} \ge \mathcal{M}_0$ wenn $z \le c_\alpha$

$\Longrightarrow$ Entscheidung für $H_1 : \mathcal{M} < \mathcal{M}_0$.

4.2.4 Der Mann-Whitney-U-Test

Nun haben wir es mit einem *Zweistichprobentest* für *unabhängige* Stichproben X und Y mit dem Umfang n bzw. m zu tun.

Die *Hypothesen* betreffen die *Mediane* $\mathcal{M}_X$ bzw. $\mathcal{M}_Y$ der beiden Stichproben: Für einen zweiseitigen Test lauten sie

$$H_0 : \mathcal{M}_X = \mathcal{M}_Y$$
$$H_1 : \mathcal{M}_X \neq \mathcal{M}_Y ,$$

während beim einseitigen Test entweder zwischen

$$\begin{array}{l} H_0 : \mathcal{M}_X = \mathcal{M}_Y \\ H_1 : \mathcal{M}_X < \mathcal{M}_Y \end{array} \qquad \text{oder} \qquad \begin{array}{l} H_0 : \mathcal{M}_X = \mathcal{M}_Y \\ H_1 : \mathcal{M}_X > \mathcal{M}_Y \end{array}$$

entschieden wird.

Trifft H_0 zu, so werden in der *vereinigten, geordneten* Stichprobe die Elemente von X und Y *gut durchmischt* sein. Sind die Elemente der einen Stichprobe nur in der oberen (unteren) Hälfte der Gesamtstichprobe zu finden, so wird das nicht für H_0 sprechen!

Zahlenbeispiel: Stichprobe X

i	1	2	3	4
x_i	5	7	6	9

Stichprobe Y

j	1	2	3
y_j	2	8	11

Die vereinigte, geordnete Stichprobe ergibt sich daraus zu

$$y_1 < x_1 < x_3 < x_2 < y_2 < x_4 < y_3 \; .$$

Als *Maß* für die *Durchmischung* in dieser Stichprobe definieren wir die ZV U_1 und U_2:
U_1 zeige an, wie oft ein Y-Wert einem X-Wert vorausgeht, und
U_2 dementsprechend, wie oft ein X-Wert einem Y-Wert vorausgeht.
In unserem Beispiel sind die Realisationen u_1 und u_2 von U_1 und U_2:

$$u_1 = 1 + 1 + 1 + 2 = 5$$

$$u_2 = 0 + 3 + 4 = 7 \; .$$

Zwei *Extremfälle* können auftreten:

1. $y_1 < y_2 < y_3 < x_1 < x_2 < x_3 < x_4$.
   Hier ist $u_1 = 3 + 3 + 3 + 3 = 12$ und $u_2 = 0 + 0 + 0 = 0$.

2. $x_1 < x_2 < x_3 < x_4 < y_1 < y_2 < y_3$.
   Nun ist $u_1 = 0 + 0 + 0 + 0 = 0$ und $u_2 = 4 + 4 + 4 = 12$.

In jedem Fall ist ein $u_i = 0$ $(i = 1, 2)$.
Bemerkung: Die x_i $(i = 1, \ldots, n)$ und die y_j $(j = 1, \ldots, m)$ können jeweils untereinander vertauscht werden, ohne die Werte u_k $(k = 1, 2)$ zu verändern.

Als *Testvariable* drängt sich nach diesen Überlegungen

$$U \doteq \min\{U_1, U_2\}$$

auf. Ist sie kleiner oder gleich einem *kritischen Wert* u_α (siehe *Tabelle 3*), dann ist H_0 zu verwerfen. [Dabei ist $P(U \leq u_\alpha) \leq \alpha$ und $P(U \leq u_\alpha + 1) > \alpha$.]

Zur *praktischen Berechnung* können folgende Formeln bewiesen werden:

$$U_2 = n \cdot m + \frac{n \cdot (n+1)}{2} - R_1$$

und

$$U_1 = n \cdot m + \frac{m \cdot (m+1)}{2} - R_2 \; ;$$

wobei $R_1 = \sum_{i=1}^{n} R(X_i)$ und $R_2 = \sum_{j=1}^{m} R(Y_j)$ ist.
Der BEWEIS kann durch vollständige Induktion nach n und m erfolgen, er ist aber mühsam und entfällt daher.

KOROLLAR. $U_1 + U_2 = n \cdot m$.
BEWEIS: durch Einsetzen der eben angeführten Formeln.

Jetzt können wir auch (zumindest prinzipiell) das *Zustandekommen* von *Tabelle 3* erklären:
Betrachten wir den einfachsten Fall: $P(U = 0) = ?$, wobei wir H_0 — wie immer — zugrunde legen wollen.
Mögliche Fälle gibt es

$$(n+m)!\ .$$

Die *günstigen* Fälle kommen auf zwei Arten zu Stande: entweder liegen alle X-Werte vor den Y-Werten oder umgekehrt. D.h. es gibt

$$2 \cdot n! \cdot m!$$

solche Fälle, woraus insgesamt

$$P(U = 0)_{H_0} = \frac{2 \cdot n! \cdot m!}{(n+m)!}$$

folgt.
Die übrigen Wahrscheinlichkeiten für $U = 1, 2, \ldots$ können analog berechnet werden.

BEISPIEL: Die Anzahl der Schiffe, die in einen Hafen pro Zeiteinheit einlaufen, ist Poisson-verteilt, im Hafen A mit dem Parameter λ, im Hafen B mit dem Parameter μ. (λ und μ sind beide unbekannt.) Um die Frage zu klären, ob gleichviele Schiffe im Mittel in A und B einlaufen, wird zwölf Stunden lang jede Stunde die Anzahl der einlaufenden Schiffe notiert:

Stunde	1	2	3	4	5	6	7	8	9	10	11	12
Hafen A	15	5	7	13	12	10	16	14	9	4	11	8
Hafen B	23	19	6	17	18	17	13	15	12	5	20	11

Wie ist zu entscheiden ($\alpha = 5\%$)?

LÖSUNG: Der t-Test kann nicht angewendet werden, da die Grundgesamtheit nicht normalverteilt ist. Wir testen daher zweiseitig mit dem Mann-Whitney-U-Test:

$$H_0 : \mathcal{M}_A = \mathcal{M}_B$$
$$H_1 : \mathcal{M}_A \neq \mathcal{M}_B\ .$$

Wir erstellen dazu die vereinigte, geordnete Stichprobe, wobei gegebenenfalls die Methode der Durchschnittsrangbildung angewendet wird:

Hafen	A	A	B	B	A	A	A	A	A	B	A	B
Wert	4	5	5	6	7	8	9	10	11	11	12	12
Rang	1	2.5	2.5	4	5	6	7	8	9.5	9.5	11.5	11.5
Hafen	A	B	A	A	B	A	B	B	B	B	B	B
Wert	13	13	14	15	15	16	17	17	18	19	20	23
Rang	13.5	13.5	15	16.5	16.5	18	19.5	19.5	21	22	23	24

Die Rangsummen sind

$r_A = 1 + 2.5 + 5 + 6 + 7 + 8 + 9.5 + 11.5 + 13.5 + 15 + 16.5 + 18 = 113.5$

$r_B = 2.5 + 4 + 9.5 + 11.5 + 13.5 + 16.5 + 19.5 + 19.5 + 21 + 22 + 23 + 24 = 186.5$,

woraus

$$u_A = 12 \cdot 12 + \frac{12 \cdot 13}{2} - 186.5 = 35.5$$

und

$$u_B = 12 \cdot 12 + \frac{12 \cdot 13}{2} - 113.5 = 108.5$$

folgt.
Damit ist

$$u = \min\{u_A, u_B\} = 35.5 \ .$$

Aus Tabelle 3 (es ist $n_1 = n$ und $n_2 = m$) lesen wir den kritischen Wert

$$u_{0.05} = 37$$

ab (Achtung, wir testen zweiseitig!), wegen $u < u_{0.05}$ verwerfen wir H_0! ◇

Bemerkungen:

- $RE = 95\%$ für kleine Stichprobenumfänge, $ARE = 95.5\%$. (Beide Zahlen beziehen sich auf den t-Test für unverbundene Stichproben.)
- Man kann zeigen: unter H_0 ist die Verteilung von U symmetrisch mit $E(U) = \frac{n \cdot m}{2}$ und $\sigma^2(U) = \frac{n \cdot m \cdot (n+m+1)}{12}$. ($E$ bedeutet dabei der Erwartungswert und σ^2 die Varianz.)

4.3 Tests für kardinalskalierte Daten

4.3.1 Der Randomisierungstest für Einstichproben

Randomisierungstests finden ihre Anwendung immer dann, wenn die Normalverteilungsannahme *nicht* getroffen werden kann, und gehen auf einen Vorschlag von *Fisher* und *Pitman* zurück. Ihr Nachteil liegt darin, daß ihre Durchführung sehr aufwendig ist, da die zugrundeliegende Verteilung immer neu berechnet werden muß.

Zum eigentlichen Test:
Gegeben sei eine Stichprobe $(x_1, \ldots, x_n)$, die zugehörigen (voneinander unabhängigen) ZV X_i seien alle in gleicher Weise symmetrisch um ihren Erwartungswert μ verteilt ($i = 1, \ldots, n$):
$E(X_i) = \mu \quad \forall i = 1, \ldots, n$.

Die *Hypothesen* lauten (zweiseitig)

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

und (einseitig)

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu > \mu_0 \qquad \text{oder} \qquad H_1 : \mu < \mu_0 \ .$$

Wir bilden folgende *Differenzen:*

$$d_i = x_i - \mu_0 \qquad \forall i = 1, \ldots, n \ .$$

Die zugehörige ZV D_i nimmt unter H_0 mit *gleicher Wahrscheinlichkeit positive* und *negative* Werte an. Insgesamt ergeben sich daher 2^n gleichwahrscheinliche Vorzeichenbelegungen der n Differenzen $|D_i|$.

Als *Testvariable* legen wir

$$D \doteq \sum_{i=1}^{n} D_i$$

fest. Wenn $|D|$ zu groß wird, müssen wir H_0 verwerfen.

Ein *Zahlenbeispiel* soll das eben Gesagte illustrieren:

x_i	μ_0	d_i
5	4	1
2	4	−2
7	4	3

Wir rechnen

$$d = 1 - 2 + 3 = 2$$

aus. Insgesamt gibt es 2^3 unter H_0 gleichwahrscheinliche Vorzeichenbelegungen der einzelnen Differenzen:

d_1	1	−1	1	1	−1	−1	1	−1
d_2	2	2	−2	2	−2	2	−2	−2
d_3	3	3	3	−3	3	−3	−3	−3
d	6	4	2	0	0	−2	−4	−6

Daraus erhalten wir folgende Verteilung von D:

d	6	4	2	0	−2	−4	−6
$P(D=d)$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Beachten wir, daß die Verteilung von D erst *nach* der Realisation der Stichprobe feststeht.

Wir *entscheiden* nun so: Wir lehnen H_0 genau dann ab, wenn D einen Randwert (Name des Tests!) annimmt. Darunter verstehen wir die kleinsten und größten Werte, die D überhaupt annehmen kann. Sie werden durch die $[\alpha \cdot 2^n]$ extremsten Vorzeichenbelegungen der d_i festgelegt.
[Bei unserem Beispiel: „6", der größte Wert, den D annehmen kann, wird durch die extreme Vorzeichenbelegung „+ + +" geliefert, der kleinste Wert (= −6) durch „− − −".]

Bemerkungen:

- $[a]$ ist die nächstkleinere ganze Zahl zu $a \in R/Z$ bzw. wenn $a \in Z$, dann ist $[a] = a$ (Gauß-Klammer).
- α — wie immer die Irrtumswahrscheinlichkeit für einen Fehler 1. Art — ist nun nicht mehr beliebig vorgebbar:
  $\alpha = 0.01$ würde in unserem Beispiel bedeuten:
  $$[\alpha \cdot 2^n] = [\alpha \cdot 2^3] = [0.08] = 0 \ ,$$
  d.h. *kein* D-Wert würde zur Ablehnung von H_0 führen.
  Beim einseitigen Testen erreichen wir $\alpha_{min} = \frac{1}{8}$, beim zweiseitigen Testen gar nur $\alpha_{min} = \frac{1}{4}$.
- Der Rechenaufwand zur Bestimmung der möglichen Werte von D steigt exponentiell, allerdings ist es oft nicht notwendig, die ganze Verteilung von D zu kennen, wie das folgende Beispiel zeigt.

BEISPIEL: Das Füllgewicht bestimmter Dosen soll 280 g betragen. Eine Konsumentenschutzorganisation zieht zur Überprüfung eine Stichprobe vom Umfang $n = 10$ mit folgendem Ergebnis (in g):

$$262,\ 275,\ 279,\ 281,\ 269,\ 273,\ 278,\ 287,\ 259,\ 277.$$

Das arithmetische Mittel dieser Werte ist

$$\bar{x} = 274 \ g \ ,$$

was Anlaß zu folgender Hypothesenformulierung gibt:

$$H_0 : \mu = \mu_0 = 280$$
$$H_1 : \mu < 280 \ .$$

Kann H_0 auf dem Signifikanzniveau $\gamma = 99\%$ aufrecht erhalten werden?

LÖSUNG: 1.) Vorerst treffen wir keine Normalverteilungsannahme, was den Randomisierungstest zur Anwendung bringt:

x_i	d_i	x_i	d_i
262	−18	273	−7
275	−5	278	−2
279	−1	287	7
281	1	259	−21
269	−11	277	−3

Aus der Tabelle lesen wir

$$d = \sum_{i=1}^{10} d_i = -60$$

ab. Wegen

$$[\alpha \cdot 2^n] = [0.01 \cdot 2^{10}] = [10.24] = 10$$

rekrutiert sich der *Ablehnungsbereich* K aus den zehn kleinsten Werten für D, welche durch geeignete Vorzeichenbelegungen der $|d_i|$ berechnet werden können:

$\|d_i\|$	18	5	1	1	11	7	2	7	21	3	$d = \sum_{i=1}^{10} d_i$
1.	−	−	−	−	−	−	−	−	−	−	−76
2.	−	−	+	−	−	−	−	−	−	−	−74
3.	−	−	−	+	−	−	−	−	−	−	−74
4.	−	−	+	+	−	−	−	−	−	−	−72
5.	−	−	−	−	−	−	+	−	−	−	−72
6.	−	−	+	−	−	−	+	−	−	−	−70
7.	−	−	−	+	−	−	+	−	−	−	−70
8.	−	−	−	−	−	−	−	−	−	+	−70
9.	−	−	+	−	−	−	−	−	−	+	−68
10.	−	−	−	+	−	−	−	−	−	+	−68

Der kleinste Wert von D wird erreicht, wenn alle $|d_i|$ mit negativem Vorzeichen belegt werden, der zweitkleinste Wert dadurch, daß das kleinste $|d_i|$ mit einem Pluszeichen belegt wird, während die übrigen $|d_i|$ weiterhin negativ belassen werden, usw.

Der Ablehnungsbereich K ist

$$\{-76, -74, -72, -70, -68\} ,$$

was wegen $d = -60 \notin K$ zur Beibehaltung von H_0 führt. (Das Beispiel wird fortgesetzt.)

Bemerkungen:

- Die tatsächliche Irrtumswahrscheinlichkeit $\alpha' = \frac{10}{2^{10}}$ ist kleiner als die vorgegebene ($\alpha = 0.01$), was wegen der Benutzung der Gauß-Klammer zur Festlegung von K klar ist.
- K besteht unter Umständen aus weniger als $[\alpha \cdot 2^n]$ Werten, wenn sich nämlich ein und derselbe Wert für $m > 1$ verschiedene Vorzeichenbelegungen ergibt. Dieser Wert ist dann auch m-mal zu zählen.
- Nulldifferenzen werden gestrichen.

2.) Wenn wir die Normalverteilungsannahme treffen, so liegt ein t-Test vor. Die Entscheidungsregel lautet jetzt:

$$H_0 \quad \text{ablehnen} \Longleftrightarrow \bar{x} \leq \mu_0 - t_{n-1,1-\alpha} \cdot \frac{s}{\sqrt{n}} \;,$$

wobei $t_{n,\alpha}$ das α-Quantil der Student-Verteilung mit n Freiheitsgraden bedeutet und $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ die empirische, $(n-1)$-gewichtete Standardabweichung.
Das Einsetzen der entsprechenden Werte ergibt:

$$280 - 2.82 \cdot \frac{8.59}{\sqrt{10}} \approx 272\, g \;,$$

was eine — knappe — Nichtverwerfung von H_0 zuläßt. ◇

Wir sehen: Ein Mehr an Information läßt eher eine Ablehnung von H_0 zu, weniger Information läßt das Verfahren konservativ werden!

4.3.2 Der Wilcoxon-Test

Dieser *Zweistichprobentest* für *abhängige* Stichproben steht in Konkurrenz zum Vorzeichentest, letzterer kann ja schon für ordinalskalierte Daten eingesetzt werden.

Also: Gegeben sei eine Stichprobe $(x_1, y_1), \ldots, (x_n, y_n)$, wobei wohl die zugehörigen ZV X_i und Y_i voneinander abhängig sind $(i = 1, \ldots, n)$, *nicht* aber die *Paare* (X_i, Y_i) $(i = 1, \ldots, n)$ voneinander.
Daher sind auch die ZV

$$D_i \doteq X_i - Y_i \qquad (i = 1, \ldots, n)$$

voneinander unabhängig. Zusätzlich wollen wir voraussetzen, daß die D_i symmetrisch um ihren Median $\mathcal{M}$ verteilt sind.

Unsere *Strategie* ist es nun, die Ränge der Beträge der Differenzen D_i zu berücksichtigen; wenn die „Behandlung" (siehe Vorzeichentest) ohne Wirkung war, so ist zu erwarten, daß die Rangsumme der positiven Differenzen D_i ungefähr gleich der der negativen ist. Wir nützen also (im Gegensatz zum Vorzeichentest) zusätzlich die Ranginformation.

Die *Hypothesen* werden wie folgt formuliert:

$$H_0 : \mathcal{M} = 0$$
$$H_1 : \mathcal{M} \neq 0$$

für das zweiseitige Testen. Bei einseitigen Tests liegen die Hypothesen

$$H_0 : \mathcal{M} = 0$$
$$H_1 : \mathcal{M} < 0 \qquad \text{bzw.} \qquad H_1 : \mathcal{M} > 0$$

zugrunde.

Die *Testvariable* W ist durch

$$W \doteq \sum_{i=1}^{n} \varphi(D_i) \cdot R(|D_i|)$$

definiert, wobei

$$\varphi(D_i) = \begin{cases} 1 \text{ falls } D_i > 0 \\ 0 \text{ falls } D_i < 0 \end{cases}$$

und $R(|D_i|)$ der Rang von $|D_i|$ ist.
W ist also die Rangsumme der positiven Differenzen.

Wenn $x_i < y_i$ ist $\forall i = 1, \ldots, n$, dann ist $w = 0$,
wenn $x_i > y_i$ ist $\forall i = 1, \ldots, n$, dann ist $w = \frac{n \cdot (n+1)}{2}$.
Wenn $w \approx \frac{n \cdot (n+1)}{4}$ ist, dann entspricht die Rangsumme der positiven Differenzen ungefähr jenen der negativen. Wie kommt es dazu?

Erstens müssen *links* und *rechts von* $\mathcal{M} = 0$ *gleichviele Differenzen* zu finden sein, was die Definition des Medians fordert, und zweitens muß die *Summe* der *positiven* Differenzen *gleich* der der *negativen* sein, was aus der geforderten Symmetrie der Verteilung um ihren Median hervorgeht.
Sind diese *beiden* Punkte im Großen und Ganzen erfüllt, so ist nichts gegen H_0 einzuwenden, das Eintreten nur eines Punktes allein genügt i.a. nicht.

Die *kritischen Werte* w_α können aus den 2^n unter H_0 gleichwahrscheinlichen Vorzeichenbelegungen der n Beträge $|D_i|$ $(i = 1, \ldots, n)$ gewonnen werden: für jede Vorzeichenbelegung ist die jeweilige Rangsumme der nun positiven Differenzen auszurechnen, woraus die *Verteilung* von W und damit auch die kritischen Werte von W folgen. In *Tabelle 4* sind letztere vertafelt. Es ist

$$P(W \leq w_\alpha) \leq \alpha \qquad \text{und} \qquad P(W \leq w_\alpha + 1) > \alpha$$

bzw.

$$P(W \geq w_{1-\alpha}) \leq \alpha \qquad \text{und} \qquad P(W \geq w_{1-\alpha} - 1) > \alpha \, .$$

Die *Testentscheidung* lautet: Beim zweiseitigen Testen verwerfen wir H_0, wenn $w \leq w_{\frac{\alpha}{2}}$ oder $w \geq w_{1-\frac{\alpha}{2}}$ ist, im einseitigen Fall lehnen wir H_0 ab, wenn $w \leq w_\alpha$ bzw. $w \geq w_{1-\alpha}$ ist.

Bemerkungen:

- Nulldifferenzen werden gestrichen.
- Beim Auftreten betragsgleicher Differenzen wird wiederum die Methode der Durchschnittsrangbildung herangezogen.
- Man kann zeigen:
  Unter H_0 ist $E(W) = \frac{n\cdot(n+1)}{4}$ und $\sigma^2(W) = \frac{n\cdot(n+1)\cdot(2n+1)}{24}$.

BEISPIEL: Im Zuge der Aktion „Minus 10%“ erfolgen an neuralgischen Stellen des Straßenverkehrs bauliche Maßnahmen (wie z.B. Fahrbahnverengungen, Schwellen,...), wodurch eine Senkung der Unfallszahlen erhofft wird:

Straße i	vorher x_i	nachher y_i	d_i	$R(\lvert d_i\rvert)$	$\varphi(d_i)\cdot R(\lvert d_i\rvert)$
1	120	96	24	8	8
2	45	35	10	5	5
3	52	59	−7	3	0
4	79	64	15	7	7
5	28	22	6	2	2
6	16	8	8	4	4
7	93	81	12	6	6
8	31	34	−3	1	0

Ist eine signifikante Senkung der Unfallszahlen feststellbar?

LÖSUNG: Die letzten drei Spalten in obiger Tabelle gehören bereits zur Lösung. Es ergibt sich daraus

$$w = \sum_{i=1}^{8} \varphi(d_i)\cdot R(\lvert d_i\rvert) = 32 \, .$$

Zwischen den Hypothesen (einseitiger Test)

$$H_0 : \mathcal{M} = 0$$
$$H_1 : \mathcal{M} > 0$$

wird zugunsten von H_1 entschieden, da aus Tabelle 4 der kritische Wert $w_{0.95} = = 31 < 32$ abzulesen ist.

Auf dieses Problem hätten wir auch den *Vorzeichentest* anwenden können. Wir zählen 6 Pluszeichen. Aus Tabelle 2 lesen wir $c_{0.95} = 7$ ab, was eine Beibehaltung der Nullhypothese zur Folge hat. ◇

Wir sehen also, daß durch den Informationsverlust, den der Vorzeichentest bei Anwendung auf kardinalskalierte Daten mit sich bringt, das Testverfahren konservativer wird, also die Ergebnisse der Stichprobe noch extremer zuungunsten von H_0 ausfallen müssen, damit eine Ablehnung derselben erfolgen kann.

Bemerkungen:

- $RE \approx 95\%$ für kleine Stichprobenumfänge n, $ARE = 95.5\%$. (Beide Zahlen beziehen sich auf den t-Test für verbundene Stichproben.)
- Auch für dieses Problem existiert ein *Randomisierungstest*, welcher an Hand des BEISPIELS vorgeführt werden soll:

  Wir definieren nun
  $$D \doteq \sum_{i=1}^{n} D_i$$
  mit
  $$D_i = X_i - Y_i \qquad (i = 1, \ldots, n) .$$
  Die Hypothesenformulierung bleibt dieselbe.
  Gesucht sind die $[0.05 \cdot 2^8] = [12.8] = 12$ Vorzeichenbelegungen, die die größten Werte von $d = \sum_{i=1}^{8} d_i$ liefern, letztere bilden den *Ablehnungsbereich* K:

$\lvert d_i \rvert$	24	10	7	15	6	8	12	3	d
1.	+	+	+	+	+	+	+	+	85
2.	+	+	+	+	+	+	+	−	79
3.	+	+	+	+	−	+	+	+	73
4.	+	+	−	+	+	+	+	+	71
5.	+	+	+	+	+	−	+	+	69
6.	+	+	+	+	−	+	+	−	67
7.	+	+	−	+	+	+	+	−	65

8.	+	−	+	+	+	+	+	+	65
9.	+	+	+	+	+	−	+	−	63
10.	+	+	+	+	+	+	−	+	61
11.	+	+	−	+	−	+	+	+	59
12.	+	−	+	+	+	+	+	−	59

Wegen

$$d = 65 \in K = \{59, 61, 63, 65, 67, 69, 71, 73, 79, 85\}$$

können wir H_0 zugunsten von H_1 verwerfen. ◇

Beachte:

- Wie in 4.3.1 ist unter H_0 jede Vorzeichenbelegung gleich wahrscheinlich, woraus der Wertebereich und die Verteilung von D resultieren.
- $P(D \in K)_{H_0} = \frac{12}{2^8} \approx 0.047 < 0.05$.
- $RE = 100\%$, da die gesamte Information genutzt wird.
- Nulldifferenzen werden gestrichen.
- Für Stichprobenumfänge $n \geq 12$ wird der Rechenaufwand sehr groß ($[0.05 \cdot 2^{12}] = 204$), man wird dann auf den Wilcoxon-Test zurückgreifen. (Die entsprechenden Werte sind bis $n = 20$ in Tabelle 4 vertafelt.)

4.3.3 Der Randomisierungstest für zwei unabhängige Stichproben

Für die beiden unabhängigen Stichproben $(x_1, \ldots, x_n)$ bzw. $(y_1, \ldots, y_m)$ soll gelten: Die zugehörigen ZV X_i $(i = 1, \ldots, n)$, welche auch untereinander voneinander unabhängig seien, sollen alle wie eine ZV X verteilt sein, das gleiche fordern wir für die ZV Y_j $(j = 1, \ldots, m)$ bezüglich einer ZV Y. X und Y besitzen — höchstens bis auf die Erwartungswerte — dieselbe Verteilung:

$$E(X_i) = \mu_X \quad \forall i = 1, \ldots, n$$

und

$$E(Y_j) = \mu_Y \quad \forall j = 1, \ldots, m\ .$$

Beim zweiseitigen Test entscheiden wir zwischen

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y\ ,$$

im einseitigen Fall stehen

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X < \mu_Y \qquad \text{bzw.} \qquad H_1 : \mu_X > \mu_Y$$

zur Debatte.

Als *Testvariable* definieren wir

$$U \doteq \sum_{i=1}^{n} X_i - \sum_{j=1}^{m} Y_j \ .$$

Unter H_0 stammen beide Stichproben aus der selben Grundgesamtheit. Wir bilden die *vereinigte Stichprobe*, welche aus $n + m$ Elementen besteht. Diese kann auf

$$\binom{n+m}{n} = \binom{n+m}{m}$$

Arten auf zwei Teilstichproben vom Umfang n bzw. m aufgeteilt werden. Daraus resultiert die *Verteilung* und der *Wertebereich* von U. Ein *Zahlenbeispiel* soll dies verdeutlichen:

$$x_1 = 2$$
$$y_1 = 3 \qquad y_2 = 4$$

Unter H_0 gibt es $\binom{1+2}{1} = 3$ gleichwahrscheinliche Möglichkeiten der *Aufteilung*, d.h. aus $\{2, 3, 4\}$ eine Teilstichprobe vom Umfang $n = 1$ und eine vom Umfang $m = 2$ zu bilden:

x	y		u
2	3	4	−5
3	2	4	−3
4	2	3	−1

Die Verteilung von U unter H_0 ist dann:

$$P(U = -5) = \frac{1}{3} = P(U = -3) = P(U = -1) \ .$$

Die „extremen" Werte des Wertebereichs sind -5 und -1, im Fall von $u = -5$ kann H_0 zugunsten von $H_1 : \mu_X < \mu_Y$ mit $\alpha = \frac{1}{3}$ abgelehnt werden; lautet $H_1 : \mu_X \neq \mu_Y$, so kann H_0 (wenn $u = -5$ oder $u = -1$ ist) gar nur mit $\alpha = \frac{2}{3}$ verworfen werden.

Für großes n und m wird der Rechenaufwand sehr hoch!

Die Berechnung des *Ablehnungsbereiches* K erfolgt *„von außen nach innen"*: U wird jedenfalls maximal, wenn wir die größten Werte alle in die erste Stichprobe X und die kleinsten alle in die zweite Stichprobe Y geben. Im umgekehrten Fall wird U minimal. So bekommen wir die extremsten Ausprägungen von U. Weniger extreme Werte erreichen wir durch *schrittweises Austauschen* der jeweils kleinsten Werte in X und Y. Dies tun wir solange, bis wir die $\left[\alpha \cdot \binom{n+m}{n}\right]$ extremsten Möglichkeiten berechnet haben. Diese müssen durchaus nicht paarweise verschiedene Werte für U liefern, sodaß K unter Umständen auch aus weniger als $\left[\alpha \cdot \binom{n+m}{n}\right]$ Werten besteht.

BEISPIEL: Die Lebensdauer von zwei Typen (A und B) elektronischer Bausteine wird getestet. Da die Elemente sehr teuer sind, kann nur ein geringer Stichprobenumfang in Kauf genommen werden. Das Ergebnis lautet in Stunden:

A	x_i	136.7	122.3	153.2	107.8	144.5
B	y_j	161.4	149.1	158.9	155.6	

Unterscheiden sich die beiden Erwartungswerte signifikant?

LÖSUNG: Wir testen zweiseitig:

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y \; .$$

K umfaßt höchstens $\left[0.05 \cdot \binom{9}{4}\right] = 6$ Werte, das sind die drei kleinsten und die drei größten Ausprägungen. Wir beachten dabei, daß $\alpha' = \frac{6}{\binom{9}{4}} \approx 0.0476 < \alpha$ ist.

x					y				u
161.4	158.9	155.6	153.2	149.1	107.8	122.3	136.7	144.5	266.9
161.4	158.9	155.6	153.2	144.5	107.8	122.3	136.7	149.1	257.7
161.4	158.9	155.6	149.1	144.5	107.8	122.3	136.7	153.2	249.5
107.8	122.3	136.7	144.5	149.1	153.2	155.6	158.9	161.4	31.3
107.8	122.3	136.7	144.5	153.2	149.1	155.6	158.9	161.4	39.5
107.8	122.3	136.7	144.5	155.6	149.1	153.2	158.9	161.4	44.3

Die oberen drei Zeilen (Aufteilungen) ergeben die drei größten Werte für U, die unteren drei die drei kleinsten.
Wegen

$$u = 39.5 \in K = \{31.3, 39.5, 44.3, 249.5, 257.7, 266.9\}$$

lehnen wir H_0 ab! ◇

Beachte:

- Wenn wir $\alpha = 0.01$ wählen, so können wir wegen $\left[0.01 \cdot \binom{9}{4}\right] = 1$ nur mehr einseitig testen.
- $RE = 100\%$, die gesamte Information wird also genutzt.
- Der *Mann-Whitney-U-Test* (siehe 4.2.4), der dasselbe Testproblem für ordinalskalierte Daten behandelt, erscheint uns nun im Lichte des eben Besprochenen als *Randomisierungstest angewandt auf Ränge.*

5 Didaktischer Kommentar

Was *für* die Behandlung von verteilungsfreien Testverfahren im Unterricht spricht:

+ Es sind nur *elementare Techniken* zu ihrer Behandlung notwendig.

+ Beim Vorliegen kardinalskalierter Daten kann einerseits ein *Vergleich* mit Parametertests stattfinden; andererseits können auch verteilungsfreie Testverfahren miteinander verglichen werden, nämlich solche, die schon bei ordinalskalierten Daten anwendbar sind, mit jenen, die nur für kardinalskalierte Daten geeignet sind. Zweierlei wird daraus erkennbar: Erstens sieht man, wie die verschiedenen Testverfahren die ihnen zur Verfügung stehende Information [aus der(n) Stichprobe(n)] unterschiedlich nützen und daraus folgt zweitens, welcher Test welchem Problem angepaßt ist.

+ Ihre *universelle Anwendbarkeit.*

+ Sie bieten viele numerische (i.e. kombinatorische) Probleme, die in den *Informatikunterricht* einfließen könnten.

+ Sie stellen eine Bereicherung (Ergänzung) des Stochastikunterrichts insofern dar, als daß sie als Themen für *Fachbereichsarbeiten* bzw. für den *Wahlpflichtgegenstand* verwendet werden können.

Was *gegen* die Behandlung von verteilungsfreien Testverfahren im Unterricht spricht:

- Ihre *spröde* Handhabbarkeit.

- Die *Voraussetzungen*, die das Testproblem erfüllen muß, richten sich nach dem Testverfahren und nicht umgekehrt (im Gegensatz zu den Parametertests!).

- Ebenso scheinen die *Testvariablen* bis zu einem gewissen Grad konstruiert zu sein im Sinne von „sich nicht aufdrängend" (im Gegensatz zu den Parametertests!).

- Ein wesentlicher Punkt ist der *Vergleich* mit *Parametertests*, letztere müssen also wenigstens exemplarisch unterrichtet werden.

- Das Prinzip von Parametertests kann durch *Zeichnungen* (etwa der „Gaußschen Glockenkurve") in didaktisch sehr sinnvoller Weise erläutert und unterstrichen werden, die Grundideen verteilungsfreier Testverfahren dagegen können nur *verbal* erklärt werden (siehe die vorliegende Arbeit), was im schlechtesten Fall hängenbleibt, sind die *Tabellen* mit den vielen kritischen Werten. Daher sollte das *Prinzip* des *Testens* selbst *jedenfalls am Beispiel eines Parametertests* demonstriert werden.

6 Literatur

Billeter, E.: Grundlagen der erforschenden Statistik: statistische Testtheorie. Springer, Wien-New York 1972.

Bosch, K.: Elementare Einführung in die angewandte Statistik. Vieweg, Braunschweig/Wiesbaden 1987.

Büning, H. u. Trenkler, G.: Nichtparametrische statistische Methoden: mit Aufgaben und Lösungen und einem Tabellenanhang. De Gruyter, Berlin-New York 1978.

Bürger, H. u.a.: Mathematik AHS-Oberstufe Kommentar. Österreichischer Bundesverlag, Wien 1991.

Bürger, H., Fischer, R. u. Malle, G.: Mathematik Oberstufe. Arbeitsbuch mit Lösungen für die AHS. Band 3. Hölder-Pichler-Tempsky, Wien 1991.

Bürger, H., Fischer, R. u. Malle, G.: Mathematik Oberstufe. Arbeitsbuch mit Lösungen für die AHS. Band 4. Hölder-Pichler-Tempsky, Wien 1992.

Fisz, M.: Wahrscheinlichkeitsrechnung und mathematische Statistik. VEB Deutscher Verlag der Wissenschaften, Berlin 1962.

Götz, S.: Einführung in die Schätz- und Testtheorie und Gedanken über den Einsatz im Mathematikunterricht. Diplomarbeit. Universität Wien 1990.

Hasibeder, G. u.a.: Grundkurs Statistik: Eine Anwendungshilfe für Biologen und Mediziner. Prugg, Eisenstadt 1986.

Heller, W.D., Lindenberg, H., Nuske, M. u. Schriever, K.H.: Studien- und Unterrichtsmaterial zur Lehrerfortbildung. Band 4: Schließende Statistik; mit vollständig gelösten Aufgaben. Birkhäuser, Basel 1980.

Reichel, H.-C., Hanisch, G. u. Müller, R.: Wahrscheinlichkeitsrechnung und Statistik. Hölder-Pichler-Tempsky, Wien 1987.

Reichel, H.-C., Müller, R., Hanisch, G. u. Laub, J.: Mathematik Oberstufe. Arbeitsbuch mit Lösungen. Band 7. Hölder-Pichler-Tempsky, Wien 1991.

Reichel, H.-C., Müller, R. u. Hanisch, G.: Mathematik Oberstufe. Arbeitsbuch mit Lösungen. Band 8. Hölder-Pichler-Tempsky, Wien 1992.

Witting, H. u. Nölle, G.: Angewandte Mathematische Statistik: optimale finite und asymptotische Verfahren. B.G. Teubner, Stuttgart 1970.

Anschrift des Verfassers:

Mag. Stefan Götz
Institut für Mathematik der Universität Wien und AKG Wien
1090 Wien, Strudlhofgasse 4 bzw. 1010 Wien, Beethovenplatz 1

A Tabellen der kritischen Werte

A.1 Tabelle 1: Kritische Werte r_α zum *Iterationstest*

n_1	n_2			α					$1-\alpha$		
		0.005	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99	0.995
2	5	-	-	-	-	2	-	-	-	-	-
	8	-	-	-	2	2	-	-	-	-	-
	11	-	-	-	2	2	-	-	-	-	-
	14	-	-	2	2	2	-	-	-	-	-
	17	-	-	2	2	2	-	-	-	-	-
	20	-	2	2	2	3	-	-	-	-	-
5	5	-	2	2	3	3	9	9	10	10	-
	8	2	2	3	3	4	10	11	11	-	-
	11	3	3	4	4	5	11	-	-	-	-
	14	3	3	4	5	5	-	-	-	-	-
	17	3	4	4	5	6	-	-	-	-	-
	20	4	4	5	5	6	-	-	-	-	-
8	8	3	4	4	5	5	13	13	14	14	15
	11	4	5	5	5	7	14	15	15	15	16
	14	5	5	6	7	7	15	15	15	17	17
	17	5	6	7	7	8	15	16	17	-	-
	20	5	6	7	8	9	16	17	17	-	-
11	11	5	6	7	7	8	15	17	17	18	19
	14	6	7	8	8	9	17	18	19	20	20
	17	7	8	9	9	10	18	19	20	21	22
	20	8	8	9	10	11	19	20	21	22	22
14	14	7	8	9	10	11	19	20	21	22	23
	17	8	9	10	11	12	21	22	23	24	24
	20	9	10	11	12	13	22	23	24	25	25
17	17	10	10	11	12	13	23	24	25	25	25
	20	11	11	13	13	15	24	25	25	27	28
20	20	12	13	14	15	16	25	27	28	29	30

A.2 Tabelle 2: Kritische Werte c_α zum *Vorzeichentest*

n	α = 0.005	0.01	0.025	0.05	0.10
1	-	-	-	-	-
2	-	-	-	-	-
3	-	-	-	-	-
4	-	-	-	-	0
5	-	-	-	0	0
6	-	-	0	0	0
7	-	0	0	0	1
8	0	0	0	1	1
9	0	0	1	1	2
10	0	0	1	1	2
11	0	1	1	2	2
12	1	1	2	2	3
13	1	1	2	3	3
14	1	2	2	3	4
15	2	2	3	3	4
16	2	2	3	4	4
17	2	3	4	4	5
18	3	3	4	5	5
19	3	4	4	5	6
20	3	4	5	5	6
21	4	4	5	6	7
22	4	5	5	6	7
23	4	5	6	7	7
24	5	5	6	7	8
25	5	6	7	7	8
26	6	6	7	8	9
27	6	7	7	8	9
28	6	7	8	9	10
29	7	7	8	9	10
30	7	8	9	10	10
31	7	8	9	10	11
32	8	8	9	10	11
33	8	9	10	11	12
34	9	9	10	11	12
35	9	10	11	12	13
36	9	10	11	12	13
37	10	10	12	13	14
38	10	11	12	13	14
39	11	11	12	13	15
40	11	12	13	14	15
41	11	12	13	14	15
42	12	13	14	15	16
43	12	13	14	15	15
44	13	13	15	16	17
45	13	14	15	16	17
46	13	14	15	16	18
47	14	15	15	17	18
48	14	15	16	17	19
49	15	15	17	18	19
50	15	16	17	18	19

A.3 Tabelle 3: Kritische Werte u_α zum *Mann-Whitney-U-Test*

Die obere Zahl gilt für $\alpha = 0.025$, die untere für $\alpha = 0.05$.

n_2 \ n_1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	0	0
2	--	--	--	--	--	--	--	0	0	0	0	1	1	1	1	1	2	2	2	2
	--	--	--	--	0	0	0	1	1	1	1	2	2	2	3	3	3	4	4	4
3	--	--	--	--	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
	--	--	0	0	1	2	2	3	3	4	5	5	6	7	7	8	9	9	10	11
4	--	--	--	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
	--	--	0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
5	--	--	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
	--	0	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
6	--	--	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
	--	0	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
7	--	--	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
	--	0	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
8	--	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
	--	1	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
9	--	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
	--	1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
10	--	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
	--	1	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
11	--	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
	--	1	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
12	--	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
	--	2	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
13	--	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
	--	2	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84
14	--	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
	--	2	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92
15	--	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
	--	3	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
16	--	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
	--	3	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
17	--	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
	--	3	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
18	--	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
	--	4	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
19	--	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
	0	4	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
20	--	2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127
	0	4	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

A.4 Tabelle 4: Kritische Werte w_α zum *Wilcoxon-Test*

n	α: 0.005	0.01	0.025	0.05	0.10	$1-\alpha$: 0.90	0.95	0.975	0.99	0.995
4	0	0	0	0	0	10	10	10	10	10
5	0	0	0	0	2	13	15	15	15	15
6	0	0	0	2	3	19	19	21	21	21
7	0	0	2	3	5	23	25	25	28	28
8	0	1	3	5	8	29	31	33	35	36
9	1	3	5	8	10	35	37	40	42	44
10	3	5	8	10	14	41	45	47	50	52
11	5	7	10	13	17	49	53	56	59	61
12	7	9	13	17	21	57	61	65	69	71
13	9	12	17	21	26	65	70	74	79	82
14	12	15	21	25	31	74	80	84	90	93
15	15	19	25	30	36	84	90	95	101	105
16	19	23	29	35	42	96	101	107	113	117
17	23	27	34	41	48	105	112	119	125	130
18	27	32	40	47	55	116	124	131	139	144
19	32	37	46	53	62	128	137	144	153	158
20	37	43	52	60	69	141	150	153	167	173