

Statistische Auswertungen

Ordinalmerkmale

DIN
53 804
Teil 3

Statistical evaluations; ordinal characteristics

Inhalt

	Seite		Seite
1 Zweck und Anwendungsbereich	1	7 Tabellen	9
2 Begriffe	2	7.1 Grenzen des Vertrauensbereichs für den Median der Gesamtheit	9
3 Ermittlung von Kennwerten einer Stichprobe bei Ordinalmerkmalen	3	7.2 Tabellenwerte $U_{n_1, n_2; 0,025}$ für den Test von Mann-Whitney-Wilcoxon zum Signifikanzniveau $\alpha \leq 0,05$ bei zweiseitigem Test und zum Signifikanzniveau $\alpha \leq 0,025$ bei einseitigem Test	12
4 Vertrauensbereich für den Median der Gesamtheit und notwendiger Stichprobenumfang	3	7.3 Tabellenwerte $U_{n_1, n_2; 0,005}$ für den Test von Mann-Whitney-Wilcoxon zum Signifikanzniveau $\alpha \leq 0,01$ bei zweiseitigem Test und zum Signifikanzniveau $\alpha \leq 0,005$ bei einseitigem Test	12
5 Vergleich von Gesamtheiten	4	7.4 Tabellenwerte $\chi^2_{f; 1-\alpha}$ der χ^2 -Verteilung zum Freiheitsgrad f und zum Signifikanzniveau α	13
5.1 Vergleich einer Gesamtheit mit einem vorgegebenen Wert	4	Anhang: Beispiele aus der Textiltechnik	14
5.2 Vergleich zweier Gesamtheiten bei unabhängigen Stichproben	5	Vergleich zweier Gesamtheiten: Ölabweisung für zwei ölabweisend ausgerüstete Gewebe	14
5.3 Vergleich mehrerer Gesamtheiten bei unabhängigen Stichproben	7	Vergleich mehrerer Gesamtheiten: Selbstglättungsverhalten eines Baumwollgewebes	14
6 Empfehlungen für den Aufbau einer Ordinalskala, die Beurteilung eines Ordinalmerkmals und den Umfang einer Stichprobe	8	Übersicht über die benutzten Formelzeichen	15
6.1 Empfehlungen für den Aufbau einer Ordinalskala	8	Zitierte Normen und Unterlagen	16
6.2 Empfehlungen für die Beurteilung eines Ordinalmerkmals und den Umfang einer Stichprobe	8	Weitere Normen und Unterlagen	17
		Erläuterungen	17

1 Zweck und Anwendungsbereich

Die Eigenschaften von Produkten oder Tätigkeiten werden durch Merkmale erfaßt. Den Merkmalsausprägungen werden Werte einer jeweils geeigneten Skala zugeordnet. Die Skalenwerte sind

- bei meßbaren (kontinuierlichen) Merkmalen beliebige reelle Zahlen (als Zahlenwerte von physikalischen Größen)
- bei zählbaren (diskreten) Merkmalen ganze Zahlen (Zählwerte)
- bei Ordinalmerkmalen Eigenschaftskategorien, die einer Rangordnung folgen (z.B. glatt, etwas verknittert, stark verknittert)
- bei Attributmerkmalen Attribute (z.B. vorhanden/nicht vorhanden oder rot/gelb/blau).

Meßbare und zählbare Merkmale werden als quantitative, Ordinalmerkmale und Attributmerkmale als qualitative (beurteilbare) Merkmale bezeichnet. Diese Merkmalsarten entsprechen den Grundbegriffen der Meßtechnik: Messen, Zählen, Sortieren und Klassieren (siehe DIN 1319 Teil 1).

Da es im allgemeinen nicht sinnvoll ist, Merkmalswerte an allen Einheiten einer Gesamtheit zu bestimmen, werden Stichproben gezogen und Kennwerte der Stichproben ermittelt.

Fortsetzung Seite 2 bis 17

Normenausschuß Materialprüfung (NMP) im DIN Deutsches Institut für Normung e. V.
Ausschuß Qualitätssicherung und angewandte Statistik (AQS) im DIN

Nachdruck, auch auszugsweise, nur mit Genehmigung des DIN Deutsches Institut für Normung e. V., Berlin, gestattet.

Parameter der Wahrscheinlichkeitsverteilung, die das Verhalten des Merkmals in der Gesamtheit beschreiben, werden mit Hilfe von Kennwerten der Stichprobe geschätzt. Die Schätzung ist mit einer angebbaren Unsicherheit behaftet. Hypothesen über die durch eine Stichprobe untersuchte Gesamtheit können mittels statistischer Tests geprüft werden.

Diese Norm beschreibt statistische Verfahren, mit denen Merkmalswerte aufbereitet und Parameter der zugrunde liegenden Wahrscheinlichkeitsverteilung geschätzt oder getestet werden können.

Die statistischen Verfahren richten sich nach den benutzten Skalenarten. Die Normenreihe erscheint deshalb in vier Teilen. DIN 53 804 Teil 1 befaßt sich mit meßbaren (kontinuierlichen) Merkmalen. Zählbare Merkmale und Attributmerkmale werden in weiteren Teilen *) behandelt.

In der vorliegenden DIN 53 804 Teil 3 werden Ordinalmerkmale behandelt, deren Werte auf einer diskontinuierlichen Skala liegen. Diese Skala heißt Ordinalskala; die Skalenwerte heißen Noten. Die Ordinalskala hat zwar eine eindeutig festgelegte Ordnung (z. B. die Ordnung steigender Ausprägungen), jedoch keine definierten Abstände zwischen den Noten. Da der Begriff „Abstand“ im Sinne einer numerischen Differenz zwischen Noten sinnlos ist, sind Zwischenwerte zwischen zwei Noten nicht zulässig. Die Ordinalskala hat außerdem keinen Nullpunkt (siehe Abschnitt 6.1).

Damit unterscheidet sich die Ordinalskala ¹⁾ wesentlich von den Skalen für zählbare und für meßbare Merkmale. Hierauf ist bei der statistischen Auswertung von Ordinalmerkmalen zu achten.

Die Noten einer Ordinalskala können mit Wörtern, Buchstaben, Ziffern oder anderen Symbolen bezeichnet werden, die diese Ordnung widerspiegeln. Bei Buchstaben wird die alphabetische Reihenfolge (z. B. *D, E, F, G, H, ...*) gewählt, bei Zahlen die Reihenfolge ihrer Größe (z. B. 10, 9, 8, ..., 4, 3; oder 8, 13, 20, 21, ...). Bei der Verwendung von Zahlen sind unbedingt die Eigenschaften der Ordinalskala zu beachten, die die meisten der üblicherweise mit Zahlen durchführbaren Operationen nicht zuläßt [9].

Empfehlungen für den Aufbau von Ordinalskalen werden in Abschnitt 6.1 gegeben.

Anmerkung 1: Beispiele für benutzte Ordinalskalen sind:

- Der Blaumaßstab für die Beurteilung der Lichtechtheit nach DIN 54 003 und DIN 54 004 mit den Blautypen 8 bis 1
- Die Graumaßstäbe für die Änderung der Farbe nach DIN 54 001 mit den Noten 5 bis 1, das Anbluten nach DIN 54 002 mit den Noten 5 bis 1
- Der Maßstab für die Beurteilung des Wasser-Abperleffekts nach DIN 53 888 mit den Noten 5 bis 1
- Der Maßstab für die Beurteilung des Aussehens von Baumwollgarnen nach ASTM D 2255 mit den Noten *A* bis *D*
- Der Maßstab für die Beurteilung der Ölabweisung (3M-Test) für ölabweisend ausgerüstete Textilien mit den Noten *A* bis *D*
- Die Maßstäbe für die Beurteilung des Pillverhaltens, der Reutlinger Standard mit den Noten 1 bis 8, der amerikanische Standard mit den Noten 5 bis 1 nach ASTM D 3511
- Der Maßstab für die Beurteilung des Selbstglättungsverhaltens von Textilien nach dem Waschen und Trocknen, nach DIN 53 895 mit den Noten 10, 8, 6, 4, 2, 1, der dem AATCC ²⁾-Standard ³⁾ mit den Stufen 5, 4, 3.5, 3, 2, 1 entspricht (Three-Dimensional Durable Press Replicas for Use with AATCC Test Method No. 124 – 1973)
- Der Schulnoten-Maßstab mit den Noten sehr gut, gut, befriedigend, ausreichend, mangelhaft, ungenügend. In dieser Aufstellung sind die Noten stets in der Ordnung von gut nach schlecht angegeben worden.

Anmerkung 2: Die Beurteilung von Ordinalmerkmalen an Hand einer Ordinalskala darf **nicht** verwechselt werden mit dem Rangieren von Einzelwerten nach aufsteigender Größe und der Zuordnung von Rangzahlen im Sinne von DIN 53 804 Teil 1. Für eine Stichprobe vom Umfang n laufen die Rangzahlen von 1 bis n , sie sind also vom Stichprobenumfang abhängig.

Die Noten dagegen sind Namen für die Stufen der Ordinalskala, die mit Symbolen (z. B. Buchstaben oder ganzen Zahlen) belegt sind; die Anzahl der Noten einer Ordinalskala ist vom Stichprobenumfang unabhängig.

Zu einer Rangzahl gehört stets genau ein Merkmalswert, eine Note dagegen kann mehrfach als Beurteilungswert vergeben werden.

2 Begriffe

Die in der vorliegenden Norm benutzten statistischen Begriffe sind den Normen DIN 13 303 Teil 1**) und Teil 2**) und DIN 55 350 Teil 12, Teil 21**), Teil 22**), Teil 23**), Teil 24**) sowie DIN 55 303 Teil 2**) zu entnehmen. Der allgemeine Begriff „Einzelwert“ wird durch den speziellen Begriff „Beurteilungswert“ ersetzt.

*) Z. Z. in Vorbereitung

**) Z. Z. Entwurf

1) Weitere Eigenschaften der Ordinalskala sowie ihre Abgrenzung gegen andere Skalen siehe Schrifttum, z. B. [3], [4], [6].

2) AATCC = American Association of Textile Chemists and Colourists

3) Lieferant: AATCC, POB 12215, Research Triangle Park, N. C. 27709, USA

3 Ermittlung von Kennwerten einer Stichprobe bei Ordinalmerkmalen

Ein Beurteilungswert (im Sinne dieser Norm) ist eine bei einer einzelnen Beurteilung (Beobachtung, Benotung, Einstufung) gefundene Note.

Als Kennwert der Lage läßt sich der Median aus den geordneten Beurteilungswerten ohne Rechenaufwand entnehmen. Bei ungerader Anzahl n von Beurteilungswerten ist der Median $\tilde{x}$ der mittelste der – entsprechend der Ordinalskala – geordneten Beurteilungswerte $x_{(1)}, x_{(2)}, \dots, x_{(n)}$:

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)}; n \text{ ungerade} \quad (1)$$

Bei einer geraden Anzahl von Beurteilungswerten ist der Median nicht eindeutig definiert; je nach Problem oder nach Vereinbarung wird der kleinere oder der größere der beiden mittelsten Beurteilungswerte der – entsprechend der Ordinalskala – geordneten Beurteilungswerte $x_{(1)}, x_{(2)}, \dots, x_{(n)}$:

$$\tilde{x} = x_{\left(\frac{n}{2}\right)} \text{ oder } \tilde{x} = x_{\left(\frac{n}{2} + 1\right)}; n \text{ gerade} \quad (2)$$

oder eine der möglicherweise zwischen diesen beiden Beurteilungswerten liegende unbesetzte Note als Median gewählt. Damit der Median eindeutig bestimmt werden kann, sind Stichproben mit ungeradzahligem Umfang vorzuziehen.

Anmerkung: Die Gleichung (6) aus DIN 53 804 Teil 1 zur Berechnung des Medians als arithmetischer Mittelwert aus zwei Einzelwerten eines meßbaren Merkmals in der Mitte der Rangfolge kann bei Ordinalmerkmalen nicht verwendet werden, da bei Noten die Berechnung des Mittelwertes sinnlos ist.

Beispiel 1:

Eine Stichprobe vom Umfang $n = 15$ wurde an Hand einer sechsstufigen Ordinalskala mit den Noten 1 bis 6 beurteilt. Dabei wurden als Beurteilungswerte folgende Noten vergeben:

4, 3, 3, 2, 4, 3, 4, 2, 3, 3, 3, 2, 5, 3, 4

Die geordneten Beurteilungswerte sind:

2, 2, 2,

3, 3, 3, 3, 3, 3, 3,

4, 4, 4, 4,

5.

Da n ungeradzahlig ist, ist der Median:

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)} = x_{(8)} = 3$$

In diesem Falle ist der an achter Stelle der geordneten Beurteilungswerte stehende Wert der Median.

Beispiel 2:

Eine Stichprobe vom Umfang $n = 12$ wurde an Hand einer sechsstufigen Ordinalskala mit den Noten A bis F beurteilt. Als Beurteilungswerte wurden folgende Noten vergeben:

E, C, E, E, C, B, C, E, C, E, E, C

Die geordneten Beurteilungswerte sind:

B, C, C, C, C, C, E, E, E, E, E, E

Wurde vor Beginn der Untersuchung vereinbart, den kleineren der beiden mittelsten Beurteilungswerte als Median zu wählen, ist

$$\tilde{x} = x_{(6)} = C$$

Wurde die Wahl des größeren der beiden mittelsten Beurteilungswerte vereinbart, ist

$$\tilde{x} = x_{(7)} = E$$

Wurde die Wahl einer gegebenenfalls unbesetzten Note als Median zugelassen, ist

$$\tilde{x} = D$$

Nicht nur der arithmetische Mittelwert als Kennwert der Lage ist nicht angebbbar, sondern auch Kennwerte der Streuung wie Standardabweichung, Varianz oder Spannweite sind bei Ordinalmerkmalen nicht angebbbar, weil diese Kennwerte von numerischen Differenzen ausgehen, die bei Noten sinnlos sind.

4 Vertrauensbereich für den Median der Gesamtheit und notwendiger Stichprobenumfang

Der Median $\tilde{x}$ der Stichprobe ist Schätzwert für den Median ζ der Gesamtheit.

Für den Median der Gesamtheit läßt sich ein Vertrauensbereich angeben. Die Grenzen $x_{(r)}$ und $x_{(s)}$ des Vertrauensbereichs für den Median der Gesamtheit stimmen überein mit zwei Beurteilungswerten der geordneten Stichprobe; sie gehören mit zum Vertrauensbereich.

Die Tabelle im Abschnitt 7.1 enthält für Stichproben mit dem Umfang $n = 4$ bis $n = 100$ die Rangzahlen r und s der Beurteilungswerte aus der geordneten Stichprobe, die für das festgelegte Vertrauensniveau bei einseitiger oder zweiseitiger Abgrenzung die Grenzen des Vertrauensbereichs sind. Die Tabelle beginnt mit dem Stichprobenumfang $n = 4$, weil bei kleineren Stichproben ein Vertrauensbereich zu einem Vertrauensniveau $1 - \alpha \geq 0,90$ nicht abgrenzbar ist.

Anmerkung: Von den Grenzen $x_{(r)}$ und $x_{(s)}$ des Vertrauensbereichs für den Median ζ der Gesamtheit können die eine oder die andere oder beide mit dem Median $\tilde{x}$ der Stichprobe übereinstimmen, wenn in der Umgebung des Medians die Beurteilungswerte gehäuft liegen. Die Grenzen des Vertrauensbereichs liegen oft unsymmetrisch zum Median $\tilde{x}$ der Stichprobe.

Beispiel 3:

Für die Beurteilungswerte aus Beispiel 1 wird der zum Vertrauensniveau $1 - \alpha \geq 0,95$ zweiseitig abgegrenzte Vertrauensbereich nach der Tabelle im Abschnitt 7.1 durch die Beurteilungswerte $x_{(r)} = x_{(4)}$ und $x_{(s)} = x_{(12)}$ der geordneten Stichprobe begrenzt. Damit ist der Vertrauensbereich für den Median der Gesamtheit zum Vertrauensniveau $1 - \alpha \geq 0,95$:

$$x_{(4)} \leq \zeta \leq x_{(12)}; \quad 3 \leq \zeta \leq 4$$

Der Median der Stichprobe $\tilde{x} = 3$ stimmt in diesem Falle mit der unteren Grenze des Vertrauensbereichs überein. Die drei kleinsten und die drei größten Beurteilungswerte der geordneten Stichprobe haben hier keinen Einfluß auf den Vertrauensbereich.

Der mindestens notwendige Stichprobenumfang wird direkt aus der Tabelle im Abschnitt 7.1 abgelesen.

Beispiel 4:

Soll der Vertrauensbereich zum Vertrauensniveau $1 - \alpha \geq 0,975$ einseitig abgegrenzt werden, und können alle Beurteilungswerte einer Stichprobe zur Abgrenzung des Vertrauensbereichs des Medians herangezogen werden, sind also keine Ausreißerwerte zu befürchten, ist nach der Tabelle im Abschnitt 7.1 der notwendige Stichprobenumfang $n \geq 6$.

Beispiel 5:

Soll der Vertrauensbereich des Medians der Gesamtheit zum Vertrauensniveau $1 - \alpha \geq 0,95$ zweiseitig abgegrenzt werden, und sollen bei einer Stichprobe etwa 20 % aller Beurteilungswerte am oberen sowie etwa 20 % am unteren Ende der geordneten Stichprobe zur Bestimmung des Vertrauensbereichs nicht herangezogen werden, ist nach der Tabelle im Abschnitt 7.1 der dazu notwendige Stichprobenumfang $n = 15$ oder $n = 16$. In diesem Falle haben die drei kleinsten Beurteilungswerte $x_{(1)}$, $x_{(2)}$ und $x_{(3)}$ und die drei größten Beurteilungswerte $x_{(13)}$, $x_{(14)}$ und $x_{(15)}$ bei $n = 15$ oder $x_{(14)}$, $x_{(15)}$ und $x_{(16)}$ bei $n = 16$ keinen Einfluß auf den Vertrauensbereich des Medians. Bei $n = 15$ sind die drei kleinsten und die drei größten Beurteilungswerte jeweils 20 % aller 15 Beurteilungswerte, bei $n = 16$ sind die drei kleinsten und die drei größten Beurteilungswerte jeweils 18,8 % aller Beurteilungswerte.

Ausreißerverdächtige Beurteilungswerte müssen in der Stichprobe verbleiben. Andernfalls würde die Verkleinerung des Stichprobenumfangs zu einer Vergrößerung des Vertrauensniveaus führen, wodurch die Aussage verfälscht würde.

Beispiel 6:

Nach der Tabelle im Abschnitt 7.1 grenzen die Beurteilungswerte $x_{(4)}$ und $x_{(12)}$ der geordneten Stichprobe vom Umfang $n = 15$ einen Vertrauensbereich zum Vertrauensniveau $1 - \alpha \geq 0,95$ zweiseitig ab; der genaue Wert ist $1 - \alpha = 0,965$. Läßt man aus dieser Stichprobe jedoch entgegen der Festlegung den größten und den kleinsten Beurteilungswert weg, wobei der Umfang auf $n = 13$ sinkt, und grenzt man jetzt den Vertrauensbereich durch $x_{(3)}$ und $x_{(11)}$ ab, dann gehört zu diesem Vertrauensbereich das Vertrauensniveau $1 - \alpha = 0,978$ (genauer Wert). Dieses (falsche) Vertrauensniveau ist größer als das Vertrauensniveau 0,965, obwohl der Vertrauensbereich in beiden Fällen durch genau dieselben Beurteilungswerte abgegrenzt wird.

5 Vergleich von Gesamtheiten

Eine Vermutung über Mediane läßt sich mit Hilfe eines statistischen Tests prüfen. Dabei wird in der Regel das Gegenteil der Vermutung als Nullhypothese H_0 aufgestellt, der die Vermutung selbst als Alternativhypothese H_1 gegenübergestellt wird. Werden beispielsweise bei zwei Materialien hinsichtlich eines bestimmten Merkmals unterschiedliche Mediane vermutet, werden folgende Hypothesen formuliert:

Nullhypothese H_0 : Der Median des Merkmals ist bei beiden Materialien gleich.

Alternativhypothese H_1 : Der Median des Merkmals ist bei beiden Materialien verschieden.

Mit Hilfe der Mediane als Kennwerte der Stichproben wird entschieden, ob die Nullhypothese auf einem vorgegebenen Signifikanzniveau α zugunsten der Alternativhypothese zu verwerfen ist (womit die Alternativhypothese zur Arbeitshypothese wird) oder nicht (womit die Nullhypothese die Arbeitshypothese bleibt).

Soll andererseits das Erfüllen einer vorgegebenen Forderung geprüft werden, wird das Erfüllen der Forderung als Nullhypothese H_0 , das Nichterfüllen als Alternativhypothese H_1 aufgestellt.

Das Signifikanzniveau ist nach technischen und wirtschaftlichen Gesichtspunkten **vor Beginn der Untersuchung** festzulegen; gebräuchlich sind die Werte $\alpha = 0,05$ und $\alpha = 0,01$. Es gibt die maximale Wahrscheinlichkeit für den Fehler erster Art an [1], [2], d. h. für das Verwerfen der Nullhypothese, obwohl diese richtig ist.

5.1 Vergleich einer Gesamtheit mit einem vorgegebenen Wert

Ist die Frage zu entscheiden, ob der Median ζ der Gesamtheit von einem vorgegebenen Wert ζ_0 verschieden ist oder nicht, wird entsprechend der Fragestellung ein einseitiger oder der zweiseitige Test zum Signifikanzniveau α durchgeführt. Dazu bildet man mit Hilfe der Tabelle im Abschnitt 8.1 den Vertrauensbereich für den Median der Gesamtheit zum Vertrauensniveau $1 - \alpha$.

Für den **zweiseitigen** Test mit der Nullhypothese $H_0: \zeta = \zeta_0$ und der Alternativhypothese $H_1: \zeta \neq \zeta_0$ wird der zweiseitig abgegrenzte Vertrauensbereich des Medians zum Vertrauensniveau $1 - \alpha$ ermittelt, $x_{(r)} \leq \zeta \leq x_{(s)}$. Liegt der vorgegebene Wert ζ_0 innerhalb dieses Vertrauensbereichs einschließlich seiner Grenzen, wird die Nullhypothese nicht verworfen; der Median ζ wird auf dem Signifikanzniveau α als nicht verschieden vom vorgegebenen Wert ζ_0 angesehen. – Liegt der vorgegebene Wert ζ_0 außerhalb dieses Vertrauensbereichs, wird die Nullhypothese zugunsten der Alternativhypothese verworfen; in diesem Falle wird der Median ζ auf dem Signifikanzniveau α als verschieden vom vorgegebenen Wert ζ_0 angesehen.

Beispiel 7:

Bei den Beurteilungswerten des Beispiels 1 ist gemäß Beispiel 3 der zweiseitig abgegrenzte Vertrauensbereich zum Vertrauensniveau $1 - \alpha \geq 0,95$:

$$x_{(4)} \leq \zeta \leq x_{(12)}; \quad 3 \leq \zeta \leq 4$$

Mit dem vorgegebenen Wert $\zeta_0 = 3$ ist ein zweiseitiger Test durchzuführen. Die Nullhypothese $H_0: \zeta = \zeta_0$ wird nicht verworfen; der Median ζ wird auf dem Signifikanzniveau $\alpha \leq 0,05$ als nicht verschieden vom vorgegebenen Wert $\zeta_0 = 3$ angesehen.

Beim **einseitigen** Test sind zwei Fälle zu unterscheiden:

Ist die Nullhypothese $H_0: \zeta \leq \zeta_0$ und die Alternativhypothese $H_1: \zeta > \zeta_0$, wird der einseitig nach unten abgegrenzte Vertrauensbereich $x_{(r)} \leq \zeta$ des Medians zum Vertrauensniveau $1 - \alpha$ ermittelt. Liegt der vorgegebene Wert ζ_0 nicht unterhalb der einseitigen unteren Vertrauensgrenze $x_{(r)}$, wird die Nullhypothese nicht verworfen, und der Median ζ wird auf dem Signifikanzniveau α als kleiner als der oder gleich dem vorgegebenen Wert ζ_0 angesehen. – Liegt der vorgegebene Wert ζ_0 unterhalb der einseitigen unteren Vertrauensgrenze $x_{(r)}$, wird die Nullhypothese zugunsten der Alternativhypothese verworfen, und der Median ζ wird auf dem Signifikanzniveau α als größer als der vorgegebene Wert ζ_0 angesehen.

Ist die Nullhypothese $H_0: \zeta \geq \zeta_0$ und die Alternativhypothese $H_1: \zeta < \zeta_0$, wird der einseitig nach oben abgegrenzte Vertrauensbereich $\zeta \leq x_{(s)}$ des Medians zum Vertrauensniveau $1 - \alpha$ ermittelt. Liegt der vorgegebene Wert ζ_0 nicht oberhalb der einseitigen oberen Vertrauensgrenze $x_{(s)}$, wird die Nullhypothese nicht verworfen, und der Median ζ wird auf dem Signifikanzniveau α als größer als der oder gleich dem vorgegebenen Wert ζ_0 angesehen. – Liegt der vorgegebene Wert ζ_0 oberhalb der einseitigen oberen Vertrauensgrenze $x_{(s)}$, wird die Nullhypothese zugunsten der Alternativhypothese verworfen, und der Median ζ wird auf dem Signifikanzniveau α als kleiner als der vorgegebene Wert ζ_0 angesehen.

Beispiel 8:

Bei den Beurteilungswerten des Beispiels 1 ist analog zu Beispiel 3 der zum Vertrauensniveau $1 - \alpha \geq 0,975$ einseitig nach oben abgegrenzte Vertrauensbereich:

$$\zeta \leq x_{(12)}; \quad \zeta \leq 4$$

Mit dem vorgegebenen Wert $\zeta_0 = 5$ ist ein einseitiger Test zum Signifikanzniveau $\alpha = 0,025$ bei der Alternativhypothese $H_1: \zeta < \zeta_0$ durchzuführen. Die Nullhypothese $H_0: \zeta \geq \zeta_0$ wird verworfen zugunsten der Alternativhypothese, d. h. auf dem Signifikanzniveau $\alpha \leq 0,025$ wird der Median ζ als kleiner als der vorgegebene Wert $\zeta_0 = 5$ angesehen.

Faßt man die vorstehenden Ausführungen zusammen, erhält man folgende Übersicht:

Art des Tests	Nullhypothese H_0	Alternativhypothese H_1	H_0 wird nicht verworfen, wenn	H_0 wird verworfen, wenn
zweiseitig	$\zeta = \zeta_0$	$\zeta \neq \zeta_0$	$x_{(r)} \leq \zeta_0 \leq x_{(s)}$	oder $\zeta_0 < x_{(r)}$ $x_{(s)} < \zeta_0$
einseitig	$\zeta \leq \zeta_0$	$\zeta > \zeta_0$	$x_{(r)} \leq \zeta_0$	$\zeta_0 < x_{(r)}$
einseitig	$\zeta \geq \zeta_0$	$\zeta < \zeta_0$	$\zeta_0 \leq x_{(s)}$	$x_{(s)} < \zeta_0$

5.2 Vergleich zweier Gesamtheiten bei unabhängigen Stichproben

Liegen zwei voneinander unabhängige Stichproben 1 und 2 mit den Stichprobenumfängen n_1 und n_2 vor, läßt sich durch den statistischen Test feststellen, ob zwischen den beiden Gesamtheiten ein Unterschied besteht oder nicht. Mit dem Test läßt sich der Unterschied zwischen zwei Prüfstellen oder der Unterschied zwischen zwei Materialien prüfen.

Der hier dargestellte Test von Mann-Whitney-Wilcoxon geht von den Rangzahlen der Noten aus, während der Median-Test nur die Anzahl der Noten berücksichtigt. Er berücksichtigt also mehr an Informationen, die in den Noten enthalten sind, als der Median-Test, und ist deshalb dem (mit weniger Aufwand verbundenen) Median-Test vorzuziehen. Zum Median-Test siehe Schrifttum, z. B. [1], [3], [4].

Der Test von Mann-Whitney-Wilcoxon spricht auch auf Unterschiede in der Form der Verteilungen an, d. h. auch wenn die Mediane der Stichproben miteinander übereinstimmen, kann er zum Verwerfen der Nullhypothese führen.

Hinweis: Im folgenden werden neben den Noten die Rangzahlen benutzt, für die die Rangzahlsumme gebildet werden kann. Die Summenbildung hat bei Noten keinen Sinn.

Für den Test werden die $(n_1 + n_2)$ Beurteilungswerte ohne Rücksicht auf ihre Zugehörigkeit zur Stichprobe 1 oder 2 in eine gemeinsame Rangfolge mit den Rangzahlen 1 bis $(n_1 + n_2)$ gebracht. Bei mehreren gleichen Beurteilungswerten wird für

jede dieser Wertgruppen das arithmetische Mittel der auf sie entfallenden Rangzahlen gebildet und jedem Beurteilungswert der betreffenden Gruppe zugeordnet.

Anschließend werden die Rangzahlen, deren zugehörige Beurteilungswerte aus der Stichprobe 1 stammen, und die Rangzahlen, deren zugehörige Beurteilungswerte aus der Stichprobe 2 stammen, getrennt addiert zu den Rangzahlsummen R_1 und R_2 . Man berechnet die Prüfwerte

$$U_1 = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - R_1 \quad \text{und} \quad U_2 = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - R_2 \quad (3)$$

wobei als Rechenkontrolle die Beziehung $U_1 + U_2 = n_1 \cdot n_2$ dient.

Für den **zweiseitigen** Test ist die Nullhypothese $H_0: \zeta_1 = \zeta_2$ und die Alternativhypothese $H_1: \zeta_1 \neq \zeta_2$. In diesem Falle ist der kleinere der beiden Werte U_1 und U_2 der Prüfwert $U_{\min}$, der mit dem von n_1 und n_2 und dem vorgeschriebenen Signifikanzniveau α abhängigen Tabellenwert $U_{n_1, n_2; \alpha/2}$ verglichen wird. Diese Tabellenwerte sind in der Tabelle im Abschnitt 7.2 und der Tabelle im Abschnitt 7.3 angegeben. Weitere Tabellenwerte sind z. B. in [5] zu finden.

Die Nullhypothese wird für $U_{\min} > U_{n_1, n_2; \alpha/2}$ nicht verworfen, in diesem Falle ist zwischen den Medianen ζ_1 und ζ_2 auf dem Signifikanzniveau α ein Unterschied nicht nachweisbar. Für $U_{\min} \leq U_{n_1, n_2; \alpha/2}$ wird die Nullhypothese zugunsten der Alternativhypothese verworfen, und die Mediane ζ_1 und ζ_2 werden auf dem Signifikanzniveau α als voneinander verschieden angesehen.

Beim **einseitigen** Test sind zwei Fälle zu unterscheiden:

Ist die Nullhypothese $H_0: \zeta_1 \leq \zeta_2$ und die Alternativhypothese $H_1: \zeta_1 > \zeta_2$, wird U_1 als Prüfwert benutzt. Ist $U_1 > U_{n_1, n_2; \alpha}$, wird die Nullhypothese nicht verworfen, und der Median ζ_1 wird auf dem vorgeschriebenen Signifikanzniveau α als kleiner als der oder gleich dem Median ζ_2 angesehen. – Ist $U_1 \leq U_{n_1, n_2; \alpha}$, wird die Nullhypothese zugunsten der Alternativhypothese verworfen, und der Median ζ_1 wird auf dem vorgeschriebenen Signifikanzniveau α als größer als der Median ζ_2 angesehen.

Ist die Nullhypothese $H_0: \zeta_1 \geq \zeta_2$ und die Alternativhypothese $H_1: \zeta_1 < \zeta_2$, wird U_2 als Prüfwert benutzt. Ist $U_2 > U_{n_1, n_2; \alpha}$, wird die Nullhypothese nicht verworfen, und der Median ζ_1 wird auf dem vorgeschriebenen Signifikanzniveau α als größer als der oder gleich dem Median ζ_2 angesehen. – Ist $U_2 \leq U_{n_1, n_2; \alpha}$, wird die Nullhypothese zugunsten der Alternativhypothese verworfen, und der Median ζ_1 wird auf dem vorgeschriebenen Signifikanzniveau α als kleiner als der Median ζ_2 angesehen.

Faßt man die vorstehenden Ausführungen zusammen, erhält man folgende Übersicht:

Art des Tests	Nullhypothese H_0	Alternativhypothese H_1	H_0 wird nicht verworfen, wenn	H_0 wird verworfen, wenn
zweiseitig	$\zeta_1 = \zeta_2$	$\zeta_1 \neq \zeta_2$	$\min(U_1, U_2) > U_{n_1, n_2; \alpha/2}$	$\min(U_1, U_2) \leq U_{n_1, n_2; \alpha/2}$
einseitig	$\zeta_1 \leq \zeta_2$	$\zeta_1 > \zeta_2$	$U_1 > U_{n_1, n_2; \alpha}$	$U_1 \leq U_{n_1, n_2; \alpha}$
einseitig	$\zeta_1 \geq \zeta_2$	$\zeta_1 < \zeta_2$	$U_2 > U_{n_1, n_2; \alpha}$	$U_2 \leq U_{n_1, n_2; \alpha}$

Beispiel 9:

Zwei Stichproben 1 und 2 vom Umfang $n_1 = 11$ und $n_2 = 9$ wurden unabhängig voneinander gezogen und an Hand einer zehnstufigen Ordinalskala mit den Noten A bis K beurteilt. Die Beurteilungswerte sind:

	Beurteilungswerte											Median
Stichprobe 1	F	F	G	H	E	G	F	F	H	H	F	$\tilde{x}_1 = F$
Stichprobe 2	F	D	E	F	F	G	D	F	E			$\tilde{x}_2 = F$

Die beiden Gesamtheiten sind mit dem zweiseitigen Test von Mann-Whitney-Wilcoxon auf dem Signifikanzniveau $\alpha \leq 0,05$ zu vergleichen, und es ist zu prüfen, ob ein Unterschied zwischen den beiden Gesamtheiten nachweisbar ist oder nicht.

Die gemeinsam geordneten Beurteilungswerte werden wie folgt ausgewertet:

geordnete Beurteilungswerte	D	D	E	E	E	F	F	F	F	F	F	F	F	F	G	G	G	H	H	H
zugehörige Stichprobe	2	2	2	2		1	1	1	1	1			2	2	2	2	2	1	1	1
Rangzahlen	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Rangzahlen nach Berücksichtigung der Häufigkeit gleicher Beurteilungswerte	1,5	1,5	4	4		4	10	10	10	10	10				16	16	19	19	19	